Improvement of Classification Based on Noise and Spectral Dimensionality Reduction for Hyperspectral Image

Salah Bourennane, Caroline Fossati and Josselin Juan

Aix Marseille Univ, CNRS, Centrale Marseille, Institut Fresnel, Marseille, France

Abstract: Hyperspectral image (HSI) classification requires spectral dimensionality reduction and spatial filtering. While common dimensionality reduction and denoising methods use linear algebra, we propose a tensorial method to jointly achieve denoising and dimensionality reduction.

Firstly, we propose a new method for pre-whitening the noise (PW) in HSI. Then we propose a method based on quadtree decomposition adapted to tensor data in order to take into account the local image characteristics in the multi-way Wiener filter (LMWF) which performs both noise and spectral dimensionality reduction, referred to as PW-LMWF$_{dr}$-(K_1, K_2, P_3). Classification algorithm SVM is applied to the output of dimensionality and noise reduction methods to compare their efficiency : The proposed PW-LMWF$_{dr}$-(K_1, K_2, P_3), PW-MWF$_{dr}$-(K_1, K_2, P_3), PCA$_{dr}$, MNF$_{dr}$ associated with Wiener filtering.

1. Introduction

A hyperspectral image (HSI) is a multidimensional array also named as a tensor and it normally consists of hundreds of spectral bands. So, HSI data has two spatial dimensions and one spectral dimension. Hyperspectral imaging sensors provide a huge number of spectral bands, typically up to several hundreds. This unreasonably large dimension of HSI not only increases computational complexity but also degrades classification accuracy [1]. Reduction of spectral dimensionality has proven necessary to apply classification algorithms. Due to its simplicity and ease of use, the most popular dimensionality reduction (DR) approaches are principal component analysis (PCA), independent component analysis (ICA), maximum noise fraction (MNF) and discrete wavelet transform (DWT) [2]. But those DR methods require a preliminary data arrangement. Indeed, when dealing with tensor data a first step consists in vectorizing all images yielding matrix data, permitting the use of signal processing but neglecting spatial rearrangement. To overcome it, [3] proposes a multichannel mathematical morphology operator-based DR method which incorporates the image representation. In this paper, a tensor-based DR method is proposed to extract spectral principal components by taking into account spatial information. Moreover acquired images are unavoidably distorted by additive noise [4, 5, 6, 7], which impairs the useful information and can degrade classification results [8]. As HSIs are normally produced by a series of sensors, the noise mainly comes from two aspects: circuity noise and photonic noise [9]. Although the photonic noise has become as dominant as the circuity noise in HSI data collected by new-generation hyperspectral sensors due to the improved sensitivity in the electronic components [9], the additive circuity noise is still an important part of noise. Since the denoising methods for those two types of noise are not same, we mainly focus on the reduction of the additive circuity noise in this paper and the term noise in the following will only refer to the additive circuity noise. To reduce the noise, a HSI is commonly split into vectors or matrices so any 2D filtering method could be applied, but this splitting way does not consider the related information between different bands [10, 11]. So, some approaches, such as tensor decomposition methods [12], have been used to remove the noise and have shown some prospects in this field [13].

A multi-way Wiener filter (MWF) [14] has been proposed to process a HSI as a whole entity based on TUCKER3 decomposition. In MWF, the filter in each mode is computed as a function of the filters in other modes, which reflects its capability in integrally exploiting the information in each mode of the multidimensional

data. This model has been successfully applied in the reduction of white noise. In practice, HSIs are always distorted by non-white noise [15], but the MWF method can not deal with the cases of colored noise.

In this paper, a 3-dimensional pre-whitening method (PW) for HSIs to change the colored noise into a white one is proposed. After that MWF can be used to filter the whitened HSI(PW-MWF). Although MWF or PW-MWF preserve the data structure of HSI, they also have some negative side effects, in practice, the MWF provides, generally, blurry restored tensor. It does not consider local details. In order to preserve edges, it is necessary to apply the filtering (PW-LMWF) on the HSI's homogeneous parts. Whereas a fixed size window may not cover homogeneous parts of the image, an adapted quadtree decomposition to HSI permits to process successively homogeneous blocks.

Since reducing spectral dimensionality is an important issue in the HSI processing field [16, 17] for the classification improvement, we propose in this paper a multilinear-algebra based DR method by integrating spectral DR in the PW-LMWF(K_1, K_2, K_3). This new tool is referred to as PW-LMWF$_{\mathrm{dr}}$-(K_1, K_2, P_3), where P_3 represents the number of spectral principal components. PW-LMWF$_{\mathrm{dr}}$-(K_1, K_2, P_3) is proposed to reduce simultaneously non-white noise and spectral dimensionality and to preserve the local image characteristics and hence improve the classification performance. The experiments of simulated and real-world images are given to present the performance of classification after denoising by PW-LMWF$_{\mathrm{dr}}$-(K_1, K_2, P_3) and compared to the most popular DR approaches, i.e., the principal component analysis, referred to as PCA$_{\mathrm{dr}}$, minimum noise fraction MNF$_{\mathrm{dr}}$ associated with Wiener filtering and PW-MWF$_{\mathrm{dr}}$ method.

The remainder of the paper is organized as follows: Section 2 introduces some basic knowledge about the multilinear algebra. Section 3 introduces the signal model. Section 4 presents the proposed method and its formulation of the classical noise reducing problem. Section 5 introduces the proposed spectral dimensionality reduction method, the PW-LMWF$_{\mathrm{dr}}$-(K_1, K_2, P_3). Section 6 contains some comparative results concerning the performance of support vector machine (SVM) classifier [18] when it is applied after either denoising and/or dimensionality reduction of HSI. Section 7 concludes the paper.

Within the scope of this paper, scalar is denoted by x, vector by $\mathbf{x}$, matrix by $\mathbf{X}$ and tensor by $\mathcal{X}$.

2. Multilinear algebra tools

In the following, some basic multilinear algebra tools used in tensor decompositions are introduced.

2.1. n-mode unfolding

$\mathbf{X}_n \in \mathbb{R}^{I_n \times M_n}$ $(n = 1, 2, 3)$ denotes the n-mode unfolding matrix of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, with size $I_n \times M_n$ where $M_n = I_p \times I_q$ with $p, q \neq n$ $(p, q = 1, 2, 3)$. The columns of $\mathbf{X}_n$ are the I_n-dimensional vectors obtained from $\mathcal{X}$ by varying index i_n while keeping the other indices fixed [14].

2.2. n-mode product

An n-mode product is defined as the product between a data tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ and a matrix $\mathbf{B} \in \mathbb{R}^{J \times I_n}$ in mode n and is used to extend matrix singular value decomposition. It is of size $I_1 \times \cdots \times I_{n-1} \times J \times I_{n+1} \times \cdots \times I_N$ and denoted by $\mathcal{X} \times_n \mathbf{B}$. Elementwise, it is $(\mathcal{X} \times_n \mathbf{B})_{i_1, \cdots i_{n-1}, j, i_{n+1}, \cdots i_N} = \sum_{i_n=1}^{I_n} x_{i_1, i_2, \cdots i_N} b_{j, i_n}$ where b_{j, i_n} denotes the (j, i_n) element of matrix $\mathbf{B}$ and $j = 1, \cdots J$.

3. Data model

A noisy HSI is modeled as a tensor $\mathcal{R} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, which means $I_1 \times I_2$ pixels and I_3 spectral bands, resulting from a pure HSI $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ impaired by an additive tensor noise $\mathcal{N} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$. The tensor $\mathcal{R}$ can be expressed as

$$\mathcal{R} = \mathcal{X} + \mathcal{N} \tag{1}$$

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 $M_n = I_p I_q$, with $p, q \neq n$. $\mathbf{R}_n$ columns are I_n-dimensional vectors obtained from $\mathcal{R}$ by varying the index i_n and keeping the other indices fixed. These vectors are called n-mode vectors. According to (1) the n-mode flattened matrix $\mathbf{R}_n$ can be expressed as

$$\mathbf{R}_n = \mathbf{X}_n + \mathbf{N}_n \tag{2}$$

There are several approaches to filter multidimensional data. A common one is to consider the modes of the tensor data as separable to enable classical 1D or 2D methods. However, that could lead to a loss of inter-dimension relationships. An interesting approach uses a hybrid filtering relying on the decorrelation of channels.

In this paper, we propose a tensor method, Tucker3 model, which permits to process the tensor data as a whole entity.

4. Proposed denoising method

4.1. Pre-whitening procedure

If the noise in HSI is colored, the noise covariance matrix of the n-mode unfolding matrix

$$\mathbf{C}_{\mathrm{N}}^{(n)} = E\left[\mathbf{N}_n\mathbf{N}_n^{\mathrm{T}}\right] \neq \sigma^2\mathbf{I} \tag{3}$$

where σ^2 is the variance of the corresponding white noise, $\mathbf{I}$ is an identity matrix and the superscript T denotes the transpose, consequently, MWF cannot effectively remove this type of noise and estimate the expected signal $\mathcal{X}$. In this paper, we propose a method to modify the colored/non-white noise in $\mathcal{R}$ into a white one, then MWF can be used effectively to denoise the whitened data tensor $\mathcal{R}_{\mathrm{w}}$. A whitening matrix $\mathbf{P}_n$ can be applied to $\mathcal{R}$. The matrix $\mathbf{P}_n$ is given by

$$\mathbf{P}_n = \mathbf{\Lambda}_n^{-1/2}\mathbf{V}_n^{\mathrm{T}} \tag{4}$$

where $\mathbf{V}_n$ is the orthonormal n-mode matrix holding the eigenvectors and $\mathbf{\Lambda}_n$ is the matrix of the corresponding eigenvalues of the matrix $\mathbf{C}_{\mathrm{N}}^{(n)}$, for instance

$$\mathbf{C}_{\mathrm{N}}^{(n)} = \mathbf{V}_n\mathbf{\Lambda}_n\mathbf{V}_n^{\mathrm{T}}. \tag{5}$$

In the non-white noise case, we consider the unfolding matrix

$$\mathbf{R}_n = \mathbf{X}_n + \mathbf{N}_n \tag{6}$$

and substitute $\mathbf{R}_n$ to

$$\mathbf{R}_{\mathrm{w}} = \mathbf{P}_n\mathbf{R}_n, \tag{7}$$

then

$$\mathbf{R}_{\mathrm{w}} = \mathbf{P}_n\mathbf{X}_n + \mathbf{P}_n\mathbf{N}_n \tag{8}$$

with the assumption that the signal is independent of the non-white noise. So, the covariance matrix

$$E\left[\mathbf{R}_{\mathrm{w}}\mathbf{R}_{\mathrm{w}}^{\mathrm{T}}\right] = \mathbf{P}_n E\left[\mathbf{X}_n\mathbf{X}_n^{\mathrm{T}}\right]\mathbf{P}_n^{\mathrm{T}} + \mathbf{I} \tag{9}$$

that is to say that the non-white/colored noise has been whitened. Thus the MWF algorithm can be applied to the whitened unfolding data matrix $\mathbf{P}_n\mathbf{R}_n$. To get the estimated signal $\hat{\mathcal{X}}$, an inverse process of whitening is necessary after we get the denoised image.

4.2. Multi-way Wiener filtering with preserving local image characteristics

The noisy HSI tensor can be expressed as :

$$\mathcal{R} = \mathcal{X} + \mathcal{N}. \tag{10}$$

MWF filter aims at estimating the desired signal $\mathcal{X}$ from data tensor $\mathcal{R}$ using multilinear algebra tools [19] :

$$\hat{\mathcal{X}} = \mathcal{R} \times_1 \mathbf{H}^{(1)} \times_2 \mathbf{H}^{(2)} \times_3 \mathbf{H}^{(3)}, \tag{11}$$

where $\times_n$ is the n-mode product, that is, the matrix product between n-mode flattening matrix $\mathbf{R}_n$ and matrix $\mathbf{H}^{(n)}$, $n = 1$ to 3.

Equation (11) represents the n-mode filtering of data tensor $\mathcal{R}$ by n-mode filters $\mathbf{H}^{(n)}$, $n = 1$ to 3.

The optimal n-mode filter $\mathbf{H}^{(n)}$ is computed by the minimization of the mean squared error

$$e(\mathbf{H}^{(1)}, \mathbf{H}^{(2)}, \mathbf{H}^{(3)}) = \mathrm{E}\left[\left\|\mathcal{X} - \hat{\mathcal{X}}\right\|^2\right]. \tag{12}$$

The n-mode filters $\mathbf{H}^{(n)}$ are obtained using an Alternating Least Squares (ALS) algorithm.

Thanks to this procedure any filter along a given mode depends on the filters along all other modes. In this iterative algorithm, the n-mode filters are initialized to corresponding identity matrices. Every m-mode filter $\mathbf{H}^{(m)}$ fixed, $m \neq n$, the expression of the optimal n-mode filter $\mathbf{H}^{(n)}$ is [19] :

$$\mathbf{H}^{(n)} = \gamma_{\mathbf{XR}}^{(n)} \mathbf{\Gamma}_{\mathbf{RR}}^{(n)}{}^{-1}, \tag{13}$$

where

$$\gamma_{\mathbf{XR}}^{(n)} = \mathrm{E}\left[\mathbf{X}_n \mathbf{q}^{(n)} \mathbf{R}_n^T\right] \tag{14}$$

$$\mathbf{\Gamma}_{\mathbf{RR}}^{(n)} = \mathrm{E}\left[\mathbf{R}_n \mathbf{Q}^{(n)} \mathbf{R}_n^T\right] \tag{15}$$

with $\mathbf{q}^{(n)} = \mathbf{H}^{(m)} \otimes \mathbf{H}^{(p)}$, $\mathbf{Q}^{(n)} = \mathbf{H}^{(m)T}\mathbf{H}^{(m)} \otimes \mathbf{H}^{(p)T}\mathbf{H}^{(p)}$ where $m \neq n$, $p \neq n$ and $\otimes$ defines the Kronecker product. By assuming that $\mathbf{X}_n$ can be expressed as a linear weighted combination of K_n vectors associated with the largest eigenvalues of $\mathrm{E}[\mathbf{R}_n\mathbf{R}_n^T]$, the optimal n-mode filter $\mathbf{H}^{(n)}$ is expressed as [19] :

$$\mathbf{H}^{(n)} = \mathbf{V}_s^{(n)} \Lambda_s^{(n)} \mathbf{V}_s^{(n)T} \tag{16}$$

where,

$$\Lambda_s^{(n)} = diag\left\{ \frac{\lambda_1^\gamma - \sigma_\gamma^{(n)2}}{\lambda_1^\Gamma}, ..., \frac{\lambda_{K_n}^\gamma - \sigma_\gamma^{(n)2}}{\lambda_{K_n}^\Gamma} \right\} \tag{17}$$

in which $\{\lambda_i^\gamma, i = 1, \ldots, K_n\}$ and $\{\lambda_i^\Gamma, i = 1, \ldots, K_n\}$ are the K_n largest eigenvalues of matrices $\mathrm{E}\left[\mathbf{R}_n \mathbf{q}^{(n)} \mathbf{R}_n^T\right]$ and $\mathrm{E}\left[\mathbf{R}_n \mathbf{Q}^{(n)} \mathbf{R}_n^T\right]$ respectively. The noise power $\sigma_\gamma^{(n)2}$ is estimated by computing the average of the $I_n - K_n$ smallest eigenvalues of $\mathrm{E}\left[\mathbf{R}_n \mathbf{q}^{(n)} \mathbf{R}_n^T\right]$:

$$\hat{\sigma}_\gamma^{(n)2} = \frac{1}{I_n - K_n} \sum_{i=K_n+1}^{I_n} \lambda_i^\gamma \tag{18}$$

The estimate of $\mathbf{H}^{(n)}$ can be computed out of the tensor data $\mathcal{R}$.

The computation of n-mode filters $\mathbf{H}^{(n)}$, $n = 1$ to 3 involves the n-mode rank values K_1, K_2 and K_3. These values are estimated using the criterion [20].

In practice the MWF provides, generally, blurry restored tensor. It does not consider local details. In order to preserve edges, it is necessary to apply the filtering on the HSI's homogeneous parts. Whereas a fixed size window may not cover homogeneous parts of the image, an adapted quadtree decomposition to HSI permits to process successively homogeneous blocks. Their sizes are linked with the local characteristics of the image. Quadtree decomposition has often been used to represent the underlying structure of digital data [21]. A quadtree decomposition is based on the recursive regular decomposition of space into blocks whose sides are of size power of two. The quadtree decomposition starts from a $T \times T$ block where T is a power of two and it divides the array into quadrants if the image is not homogeneous. Each sub-block is then recursively processed like providing a decomposition in which every block is homogeneous. We adapt, in this paper, the quadtree decomposition to improve the restoration of details after noise removal. The approach consists in filtering separately homogeneous regions to preserve local characteristics. The decision function associated with the split homogeneity test relies on the variance of each block image W to measure its homogeneity [21]:

$$\sigma_W^2 = \frac{1}{N_{pix}} \sum_{k=1}^{N_{pix}} (p_k - m_W)^2 \tag{19}$$

where σ_W^2 denotes the variance of N_{pix} pixel intensities p_k in the block W with mean value m_W. The comparison of this variance to an experimental *a priori* fixed threshold, permits to decide whether to split or not a block into four sub-blocks. The Local Multidimensional Wiener Filtering (LMWF) method can be summarized as follows

1. For each mode: Decompose the tensor into homogeneous sub-blocks using a variance based quadtree method.
2. Flatten each sub-block along the mode.
3. Compute LMWF using Equation (11)
4. Compute the average filtered HSI.

In the next section we show how PW-LMWF can be used to improve the classification. It simultaneously reduces the non-white noise and spectral dimensionality which leads to improvement of classification algorithms.

5. Spectral dimensionality reduction

In HSI context, we are interested in reducing the number of spectral bands by selecting more significant spectral features in order to improve classification. The principles of PCA are the following: I_3 images of full size $I_1 \cdot I_2$ are considered. Each image is transformed into a vector by row concatenation. Data tensor $\mathcal{R} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ composed of all I_3 images as slice matrices becomes a matrix $\mathbf{R_3} \in \mathbb{R}^{I_3 \times p}$ where $p = I_1 \times I_2$. The aim of DR is to extract a small number $P_3 < I_3$ of features, called principal components (PCs). Therefore the P_3 PCs generate a reducing matrix $\mathbf{Z} \in \mathbb{R}^{P_3 \times p}$,

$$\mathbf{Z} = \mathbf{\Lambda}_s^{(3)^{-1/2}} \mathbf{V}_s^{(3)^T} \times \mathbf{R}_3. \tag{20}$$

Where $\mathbf{V}_s^{(3)}$ is a matrix holding P_3 selected eigenvectors, $\mathbf{\Lambda}_s^{(3)}$ is the diagonal eigenvalue matrix holding the P_3 largest eigenvalues. The data can be reshaped as a tensor $\mathcal{Z} \in \mathbb{R}^{I_1 \times I_2 \times P_3}$. In tensor formulation [19], the previously obtained matrix $\mathbf{Z}$ is equivalent to the 3-mode flattened matrix of $\mathcal{R}$ noted $\mathbf{R}_3$. Then $\mathbf{Z}$ can be written

$$\mathcal{Z} = \mathcal{R} \times_3 \mathbf{\Lambda}^{(3)^{-1/2}} \mathbf{V}_s^{(3)^T}. \tag{21}$$

In the same way, we can turn the PW-LMWF-(K_1, K_2, K_3) into a spectral dimensionality reduction tool. This tool is referred to PW-LMWF$_{\mathrm{dr}}$-(K_1, K_2, P_3) in this paper, where P_3 represents the number of spectral principal components. PW-LMWF$_{\mathrm{dr}}$-(K_1, K_2, P_3) extracts P_3 spectral PCs in order to obtain the three-way array $\mathcal{Z} \in \mathbb{R}^{I_1 \times I_2 \times P_3}$. The challenge is carried out thanks to the LMWF$_{\mathrm{dr}}$-(K_1, K_2, P_3) is to jointly reduce the dimensionality of the spectral mode and to project the information along the spatial modes onto lower (K_1, K_2)-dimensional subspaces. The latter processing permits to compress and to spatially denoise the data. The PW-LMWF$_{\mathrm{dr}}$-(K_1, K_2, P_3) model reads

$$\mathcal{Z} = \mathcal{R} \times_1 \mathbf{H}^{(1)} \times_2 \mathbf{H}^{(2)} \times_3 \mathbf{\Lambda}_s^{(3)^{-1/2}} \mathbf{V}_s^{(3)^T} \tag{22}$$

where

$\mathbf{H}^{(n)}$ is the filter for the n-mode, defined in Equation (13). First, LMWF$_{\mathrm{dr}}$-(K_1, K_2, P_3) joint uses spatial and spectral information to extract the spectral principal components. Secondly, PW-LMWF$_{\mathrm{dr}}$-(K_1, K_2, P_3) denoises the extracted spectral principal components thanks to the estimated spatial projectors, $\mathbf{H}^{(n)}$ $n = 1, 2$ and the pre-whitening procedure.

6. Experiments

In this section, we focus on the classification results obtained after denoising by PW-LMWF$_{\mathrm{dr}}$-(K_1, K_2, P_3) and other methods. Two real-world images are considered for this investigation. The first one, referred to as HYDICE HSI, was acquired by HYperspectral Digital Imagery Collection Experiment (HYDICE) and has 148 spectral bands (from 435 to 2326 nm), 310 rows, and 220 columns. The scene, is shown in figure 1 (a). This HSI is modeled as a tensor $\mathcal{R} \in \mathbb{R}^{310 \times 220 \times 148}$ and its ground truth is shown in figure 1 (b). According to the ground truth, there are 7 land cover classes in HYDICE HSI: field, trees, road, shadow and three different targets. The second one, referred to as AVIRIS HSI was collected by the airborne visible/infrared imaging spectrometer (AVIRIS)from a mixed forest/agricultural site at the Indian Pine test site in north-west Indiana. This image was taken by the National Aeronautics and Space Administration (NASA)/Jet Propulsion Laboratory. The raw image size is $145 \times 145 \times 220$ ($I_1 = 145$, $I_2 = 145$, $I_3 = 220$) (figure 1 (c)). This HSI can be represented as a tensor $\mathcal{R} \in \mathbb{R}^{145 \times 145 \times 220}$ and its ground truth is shown in figure 1 (d). According to the ground truth, there are 16 land cover classes in AVIRIS HSI: Corn-min, hay-windrowed, stone-steel towers, woods, wheat, soybean-clean, Oats, soybean-notill, corn, bldg-grass-tree-drives, alfalfa, corn-notill, grass/trees, grass/pasture, grass/pasture-mowed, soybeans-min.

To compare the classification results quantitatively, the overall accuracy (OA) in percentage is defined as: for P classes C_i ($i = 1, \cdots, P$), if $a_{i,j}$ is the number of test samples that actually belong to class C_i and is classified into C_j ($j = 1, \cdots, P$), then OA is $OA = \frac{1}{M} \sum_{i=1}^{P} a_{i,i}$ where M is the total number of samples, P is the number of classes C_i and $a_{i,i} = a_{i,j}$ for $i = j$. And to evaluate quantitatively the denoising results, the signal to noise ratio of image after denoising also named as SNR output is defined as: $\mathrm{SNR}_{\mathrm{out}} = 10 \log_{10} \|\hat{\mathcal{X}}\|^2 / \|\mathcal{X} - \hat{\mathcal{X}}\|^2$ (dB) where $\hat{\mathcal{X}}$ means the estimated signal after denoising. Correspondingly, SNR input is defined as: $\mathrm{SNR}_{\mathrm{in}} = 10 \log_{10} \|\mathcal{X}\|^2 / \|\mathcal{N}\|^2$ (dB).

The experiment shows the ability of PW-LMWF$_{\mathrm{dr}}$-(K_1, K_2, P_3) as a DR method; and its robustness in the presence of non-white noise.
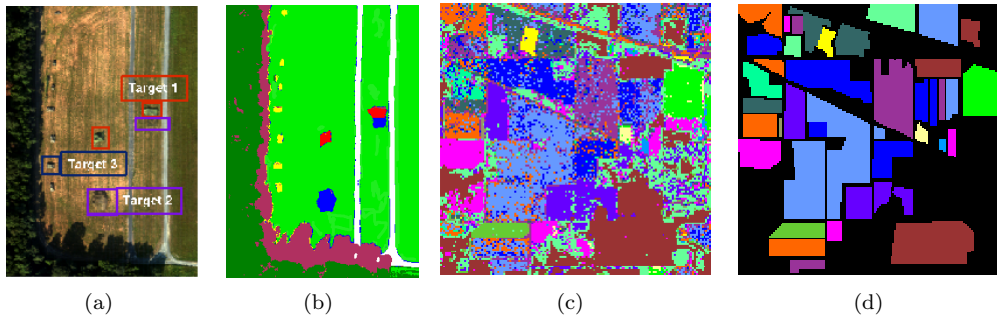
Figure 1: HSI images: (a) Classes in HYDICE HSI, (b) Ground truth, (c) Raw AVIRIS HSI, (d) Ground truth of the area with 16 classes

.

6.1. Classification of real-world HSI

The PW-LMWF$_{dr}$-(K_1, K_2, P_3) is tested on real-world HSIs which are impaired by an additive noise because of some properties of imaging system. This experiment evaluates the necessity of denoising real-world data [22] to improve the classification results. For this experiment, $(K_1, K_2, K_3) = (39, 39, 85)$ are used to apply PW-LMWF-(K_1, K_2, K_3).

Table 1 shows SVM classification results and the SNR$_{out}$ of the denoised images by PW-LMWF$_{dr}$-(K_1, K_2, P_3), PW-MWF$_{dr}$, PCA$_{dr}$, PCA$_{dr}$-Wiener, MNF$_{dr}$ and MNF$_{dr}$-Wiener denoising methods. The OA of SVM classification obtained from the raw HYDICE HSI is 93.49% and 81.69% from AVIRIS HSI. Therefore, both the OA values of classification and SNR$_{out}$ results show that denoising and dimensionality reduction are necessary preprocessing steps. This table shows the DR usefulness. In fact, the DR permits to increase OA for each filtering method and for this HSI the PW-LMWF$_{dr}$-(K_1, K_2, P_3) outperforms better PW-LMWF and other methods in denoising and preserving local characteristics of objects in HSI.

Table 1: OA(%) of SVM results of denoised HYDICE HSI and AVIRIS HSI

Methods	HYDICE HSI		AVIRIS HSI	
	OA (%)	SNR$_{out}$ (dB)	OA (%)	SNR$_{out}$ (dB)
PCA-(40)	94.57	36.30	82.30	30.10
PCA-(40)-Wiener	95.60	37.31	84.37	31.62
MNF-(38)	94.71	36.51	84.87	31.38
MNF-(38)-Wiener	95.66	37.83	87.78	32.73
PW-MWF	96.20	38.36	92.41	35.50
PW-MWF$_{dr}$-(40,40,80)	96.82	38.79	93.18	35.8
PW-LMWF-(39,39,85)	97.03	39.01	95.37	38.12
PW-LMWF$_{dr}$-(39,39,85)	99.88	40.20	99.51	38.91

We notice from the OA values that denoising is a necessary preprocessing step, but as the noise in this HYDICE HSI is not obvious in the next experiment we test the noise robustness of the proposed method for reducing more non-white noise in HSIs than in previous images.

6.2. Classification in noisy environment

In this experiment, we test the noise robustness of the PW-LMWF$_{dr}$-(K_1, K_2, P_3) method, and the results highlight the advantage of applying this method on HSIs before classification. For this issue, non-white noise [15] is added to real-world HSI data with SNR$_{in}$ varies from 15 to 40 dB into the real-world HSI. We compare the classification results after applying our proposed method and other denoising methods.

Figure 2 shows the OA values obtained from the denoised HSI and shows that the PW-LMWF$_{dr}$-(K_1, K_2, P_3) method permits to reduce jointly the spectral dimension and noise which is of great interest for SVM classifier. The comparison of the OA values calculated for each preprocessing of DR and denoising, shows that multilinear algebra-based DR method PW-LMWF$_{dr}$-(K_1, K_2, P_3) leads to better classification results than PCA$_{dr}$-Wiener,

PW-MWF$_{\text{dr}}$ and MNF$_{\text{dr}}$-Wiener considered in this experiment.

Figure 3 (a) presents the SVM classification results obtained from a raw AVIRIS HSI, figures 3 (b)-(e) are the
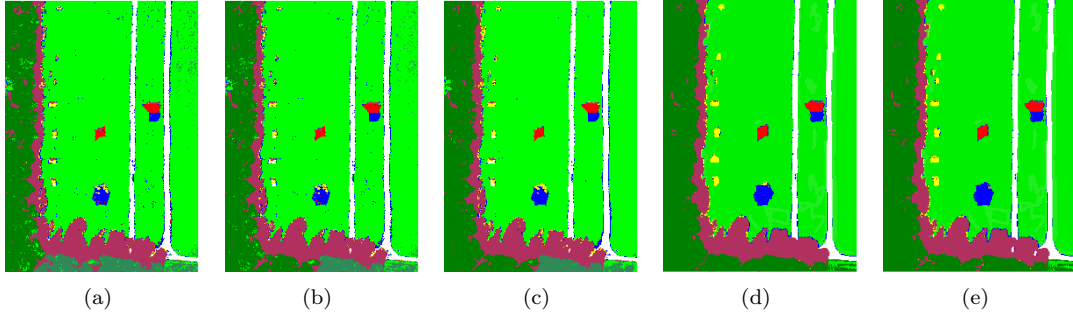

| (a) | (b) | (c) | (d) | (e) |

Figure 2: Spectral dimensionality reduction outcome for classification: (a) Raw HYDICE data OA = 93.49, (b) PCA$_{\text{dr}}$-Wiener OA = 95.60, (c) MNF$_{\text{dr}}$-Wiener OA = 95.66, (d)PW-MWF$_{\text{dr}}$ OA = 96.82, (e) PW-LMWF$_{\text{dr}}$ OA = 99.88

.

results obtained after denoising by different methods where 16 classes are used for classification of this AVIRIS HSI. It can be seen that the 2D Wiener filtering applied on spectral components obtained by PCA permits to improve the OA value. This experiment shows that the tensor methods can do better as a preprocessing procedure than 2D methods for the classification and PW-LMWF$_{\text{dr}}$ method shows its significant advantage compared to PW-MWF and 2D methods.
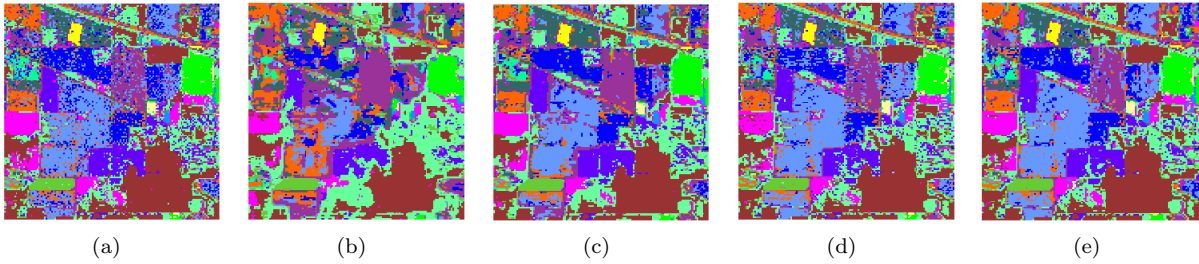

| (a) | (b) | (c) | (d) | (e) |

Figure 3: SVM classification results of raw AVIRIS.: (a) Raw AVIRIS HSI OA = 81.69, (b) PCA$_{\text{dr}}$-Wiener OA = 84.37, (c) MNF$_{\text{dr}}$-Wiener OA = 87.78, (d) PW-MWF$_{\text{dr}}$ OA = 93.18, (e) PW-LMWF$_{\text{dr}}$ OA = 99.51

.

Figure 4 presents OA values obtained from the denoised HYDICE HSI and AVIRIS HSI also shows that the PW-LMWF$_{\text{dr}}$-(K_1, K_2, P_3) performs much better than PW-MWF and other methods particularly when the SNR$_{\text{in}}$ value is low. All these results demonstrate the advantage of using the proposed method, this advantage is much more significant with AVIRIS HSI where the relevant features are localized on some regions of image. From the presence of such small local features as we can expect the proposed method provide better results compared to other methods, because PW-LMWF$_{\text{dr}}$-(K_1, K_2, P_3) permits to reduce simultaneously the spectral dimension and the dimensions of the spatial subspaces with preserving local image characteristics which is of great interest for SVM classifier.
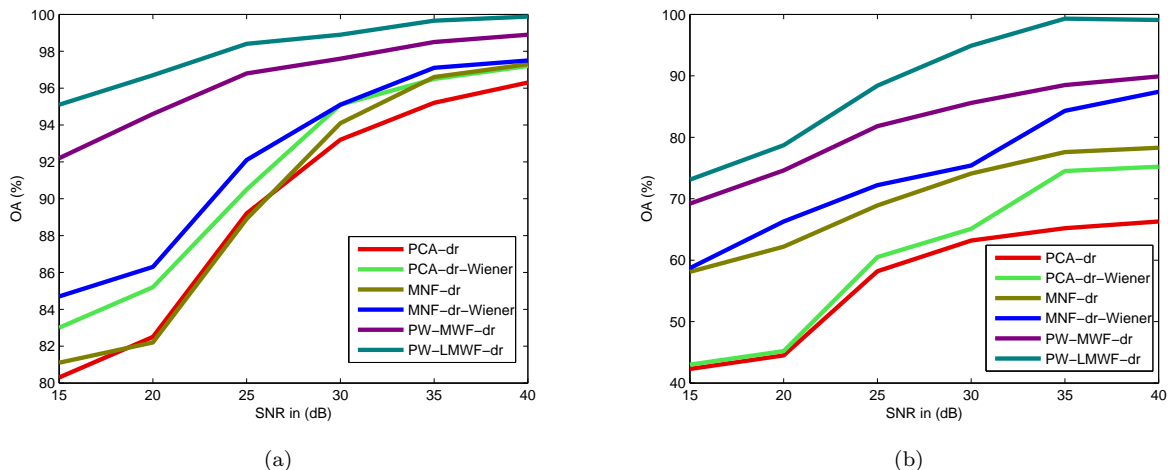
Figure 4: OA of SVM results of denoised: (a) HYDICE HSI, (b) AVIRIS HSI
.

7. Conclusion

With the advances of the electronic components, the reduction of photonic noise has become an important task in denoising HSI data collected by new-generation hyperspectral sensors. For this case, to reduce noise, a novel tensor-based algorithm, called PW-LMWF$_{\text{dr}}$-(K_1, K_2, P_3), is proposed for joint noise and dimensionality reduction with preserving local image characteristics. This joint spatial-spectral processing is cross-dependent thanks to the ALS algorithm. To reduce the colored/non-white noise in HSI, a pre-whitening method is proposed based on a two-stage process (PW-LMWF) composed of a noise-whitening procedure and a LMWF filter. We focused on the ability of PW-LMWF$_{\text{dr}}$-(K_1, K_2, P_3) as a preprocessing algorithm that improves SVM classification result applied to real-world AVIRIS and HYDICE data. Quantitative results based on OA criterion evaluate the impact on the spatial ranks, (K_1, K_2) values, and compare the performance with selected dimensionality reduction methods. Indeed in comparison with PCA$_{\text{dr}}$, PW-MWF$_{\text{dr}}$ and MNF$_{\text{dr}}$, the PW-LMWF$_{\text{dr}}$-(K_1, K_2, P_3) permits to extract spectral components by taking into account spatial information by simultaneously estimating spatial filters to denoise them. The comparison with selected hybrid filters, which perform 2D-spatial filtering of the retained spectral components, permits to appreciate the denoising efficiency of our method, in the application of classification in noisy data.

From the analysis and the comparative study against other similar methods in the experiments, it can be concluded that PW-LMWF$_{\text{dr}}$-(K_1, K_2, P_3) method can effectively reduce white or colored noise from HSIs. It is also necessary to take into account the colored noise when dealing with HYDICE and AVIRIS data.

These promising results encourage us to extend our experiments on the HSIs distorted by both photonic noise and spectrally correlated noise and on other hyperspectral data, for instance HSIs obtained from new generation high-resolution hyperspectral sensors.

References

[1] D. Landgrebe, *Hyperspectral image data analysis as a high dimensional signal processing problem, Special issue of the IEEE Signal Process. Magazine 1 (2002) 17–28.*

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

[3] A. Plaza, P. Martinez, J. Plaza, and R. Perez, Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations, IEEE Trans. on Geosci. and Remote Sens. 3 (2005) 466–479.

[4] H.K. Aggarwal, A. Majumdar, Exploiting spatiospectral correlation for impulse denoising in hyperspectral images. J. Electron. Imaging 24 (2015) 013027.

[5] P. Meza, J.E. Pezoa, S.N. Torres, Multidimensional Striping Noise Compensation in Hyperspectral Imaging: Exploiting Hypercubes Spatial, Spectral, and Temporal Redundancy. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 9 (2016) 4428–4441.

[6] B. Rasti, M.O. Ulfarsson, P. Ghamisi, Automatic Hyperspectral Image Restoration Using Sparse and Low-Rank Modeling. IEEE Geosci. Remote Sens. Lett. (2017) 1-5.

[7] X. Liu, S. Bourennane, C. Fossati, Reduction of Signal-Dependent Noise From Hyperspectral Images for Target Detection. IEEE Trans. Geosci. Remote Sens. 52 (2014) 5396–5411.

[8] D. Manolakis, D. Marden, G. Shaw, Hyperspectral image processing for automatic target detection applications. Linc Lab. J. 14 (2003) 79–116.

[9] M. Uss, B. Vozel, V. Lukin, K. Chehdi, Local signal-dependent noise variance estimation from hyperspectral textural images, IEEE Jounal of Selected Topics in Signal Proc. 5 (2011) 469–486.

[10] D. Letexier, S. Bourennane, J. Blanc-Talon, Nonorthogonal tensor matricization for hyperspectral image filtering. IEEE Trans. Geosci. Remote Sens. Lett. 5 (2008) 3–7.

[11] H. Song, G. Wang, K. Zhang, Hyperspectral image denoising via low-rank matrix recovery. Remote Sensing Letters 5 (2014) 872–881.

[12] T.G. Kolda, B.W. Bader, Tensor decompositions and applications. SIAM Review 51 (2009) 455–500.

[13] N. Renard, S. Bourennane, Improvement of target detection methods by multiway filtering, IEEE Trans. on Geosci. and Remote Sens. 8 (2008) 2407-2417.

[14] D. Muti, S. Bourennane, Survey on tensor signal algebraic filtering. Signal Process. 87 (2007) 237–249.

[15] R.E. Roger, Principal components transform with simple, automatic noise adjustment. International Journal of Remote Sensing 17 (1996) 2719–2727.

[16] O. Kuybeda, D. Malah, M. Barzohar, Rank estimation and redundancy reduction of high-dimensional noisy signals with preservation of rare vectors. IEEE Trans. on Signal Processing 55 (2007) 5579–5592.

[17] Z. Xiao, S. Bourennane, Constrained nonnegative matrix factorization and hyperspectral image dimensionality reduction. Remote Sensing Letters 5 (2014) 46–54.

[18] V.N. Vapnik, The nature of statistical learning Theory, New York: Springer-Verlag (1995).

[19] D. Muti, S. Bourennane, Survey on tensor signal algebraic filtering, Signal Processing 87 (2007) 237-249.

[20] C. Chang and Q. Du, Estimation of number of spectrally distinct signal sources in hyperspectral imagery, IEEE Trans. on Geosci. and Remote Sens. 3 (2004) 608–619.

[21] H. Samet, The quadtree and related hierarchical data structure, ACM Computing Surveys (1984) 188-260.

[22] R. Archibald, G. Fann, Feature selection and classification of hyperspectral images with support vector machines. IEEE Trans. Geosci. Remote Sens. 4 (2007) 674–677.