Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 176 (2020) 561–571

Procedia Computer Science

www.elsevier.com/locate/procedia

24th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

# Hyperspectral Feature Extraction by Tensor Modeling and Intrinsic Decomposition

Asma FEJJARI[a*], Karim SAHEB ETTABAA[b], Ouajdi KORBAA[a]

*[a]MARS Research Laboratory, ISITCom, 4011, Hammam Sousse, University of Sousse, Tunisia.*
*[b]IMT Atlantique, Iti Department, Telecom Bretagne, 655 Street of Technopôle 29200 Plouzané, France.*

**Abstract**

Recently, hyperspectral tensor modeling has been assessed by its capacity of determining more compact as well as by its useful intrinsic data representation. In this paper, to enhance the hyperspectral tensor data representation and to eliminate non-relevant spatial datum, we integrated the intrinsic decomposition (ID) as a pre-processing step. The suggested approach acts in agreement with the joint use of spectral and spatial features provided in hyperspectral scenes, and it incorporates more usefully with the spatial data in the dimensional reduction step. The suggested framework consists of three steps: firstly, the intrinsic decomposition is employed to remove useless spatial data from hyperspectral image (HSI). Secondly, after modelling ID results as a tensor structure, the tensor principal component analysis is used to reduce tensorial data redundancy. Finally, we evaluated the proposed approach during classification tasks using real hyperspectral data sets. Compared to other methods, experiment results have proved that our approach can pave the way for the best classification accuracy.



***Keywords:*** Tensor modelling; intrinsic decomposition; dimensional reduction; classification; hyperspectral data.

* Corresponding author. *E-mail address:* asmafejjari@gmail.com

1877-0509 
10.1016/j.procs.2020.08.058

## 1. Introduction

Nowadays, with the fast progress of remote sensing technologies, the new generation of hyperspectral satellite sensors can produce a set of images of the same land surface according to several, fine and contiguous spectral bands[1]-[3], [33]. Due to its discrimination ability and its performance in terms of spatial and spectral resolution, hyperspectral imaging systems are used worldwide for various applications [4]-[6], [38]. Despite its motivating potentialities, the processing and exploitation of hyperspectral classification is a complicated and expensive procedure. In fact, the exceptional nature of the hyperspectral images (HSIs) produces also a huge amount of irrelevant or redundant data. Consequently, it poses several troubles during data rectifying and storing, particularly a significant computing time increase. To address the high HSI dimensionality issue, various dimensionality reduction methods have been suggested during the last decade [7], [34]-[35]. In the early phase of HSI dimensionality reduction, the majority of techniques have focused on using spectral features only. However, the reached classification results are often unsatisfactory, since spatial properties are not taken in consideration. Lately, spatial contextual information has been incorporated, in addition to spectral features, into dimensionality reduction process. Hence, more and more spectral-spatial features extraction methods have been developed [8]-[10].

Tensor decomposition has become, recently, a growing trend in data analysis and processing. A considerable progress has been made with this method in many remote sensing tasks, e.g. image classification [11], target detection [12]-[13] and feature extraction [14]. On account to the 3-D form: two spatial dimensions and one spectral mode, HSI data can be modeled as 3-order tensor. Hence, tensor learning methods, known also by multilinear tensor algebra based methods, can be applied on 3-dimensional HSI, where spectral and spatial properties are used jointly [11]. Nevertheless, HSI data are often subjected to spatial variations such like environmental, atmospheric and illumination conditions; which can affect accuracy and generate non relevant spatial data. Therefore, to overcome this issue, many features have been adopted, lately, to get rid of useless spatial data [8], [15].

In this work, an intrinsic decomposition based tensor learning approach was our proposition to remove non relevant spatial information from HSI data set. The new suggested method exploits the energy optimisation [16] concept to define the tensor organisation scheme. It includes three steps: to preserve the useful spatial content, an intrinsic decomposition is adopted at the first step. Thereafter, the tensor principal component analysis (TPCA) [17] is employed to lessen spatial-spectral data redundancy. Finally, we adopted the support vector machine (SVM) classifier [18] to resolve multiclass image classification. The present work contains mainly three contributions:

- A novel feature extraction approach for HSI classification tasks is suggested in this work. The proposed method integrates the intrinsic decomposition concept with the tensor learning model, for the first time.
- The new suggested method is able to extract more useful spatial and spatial features, independently from spatial variability.
- A perceivable experimental study, on real HSIs, was deeply presented and discussed in this work, to prove the efficacy of the proposed approach.

The remainder of this manuscript is organised in this way: the second section is consecrated to the proposed approach. Section 3 conducts on experimental analysis and results. Finally, conclusions and suggestions are presented in section 4.

## 2. Proposed approach

In this proposed approach, as an attempt to remove useless spatial information in HSI and to achieve a more compact spatial feature representation, we seek to enhance the spatial feature representation in the classical tensor structure, through an additive intrinsic decomposition.

### *2.1. Intrinsic data decomposition*

The intrinsic photo graph decomposition was introduced by Barrow and Tenenbaum [19] to separate the different physical properties of a scene into reflectance components (spectral properties) and its illumination (shading or lighting). In this light, several methods have been proposed to estimate the reflectance and illumination from a single image such as Retinex methods [20], optimization approaches [16], and constraints-based methods [21]. In

this paper, we adopted an optimization decomposition approach based on energy minimization [22] for the following causes: first of all, complex mathematical models are no longer required. Secondly, it has proved its efficiency to recover intrinsic components (reflectance and Shading) for image processing. Finally, even though training simples are relatively low, nevertheless it offers very high precision. The proposed method tends to eliminate useless spatial information by preserving only the reflectance components of the image and eliminating shadow-dependent properties (light and shape). The intrinsic decomposition of an image $\boldsymbol{I}$ to a reflectance layer $\boldsymbol{R}$ and a shading layer $\boldsymbol{S}$ is based on the following model [16]:

$$\boldsymbol{I_i} = \boldsymbol{R_i}.\boldsymbol{S_i} \tag{1}$$

The reflectance $\boldsymbol{R_i}$ represents the material intrinsic color for a pixel $i$ and it is invariant to illumination and shading conditions. $\boldsymbol{S_i}$ registers the quantity of reflected light on pixel $i$, it summarizes the proportional alteration of shading values next to color channel values (red, green and blue).Therefore, the intrinsic decomposition of an image requires the determination of the two unknowns $\boldsymbol{S_i}$ and $\boldsymbol{R_i}$ at each pixel $i$ of the input image. In a local window of an image, changes in pixel values are usually caused by changes in reflectance. In other words, pixels those have the same intensities, can achieve the same reflectance values, so the value of a pixel reflectance may be defined by the weighted sum of its neighborhood pixel values as follows:

$$\boldsymbol{R_i} = \sum_{\boldsymbol{j}\in\boldsymbol{\varphi(i)}} \boldsymbol{\omega_{ij}}\boldsymbol{R_j} \tag{2}$$

The local neighborhood window of the pixel $i$ is described by $\varphi(i)$, while $\omega_{ij} = e^{-(\boldsymbol{Y}_i-\boldsymbol{Y}_j)^2/(2\sigma_i^2)}$ measures the reflectance similarity between two pixels $i$ and $j$. $\boldsymbol{Y}_i$ and $\boldsymbol{Y}_j$ represent the intensity values of pixel $i$ and $j$ respectively. $\sigma_i$ defines the intensity variance in local window. The weight function can be improved as following:

$$\omega_{ij} = e^{[(\boldsymbol{Y}_i-\boldsymbol{Y}_j)^2/(\sigma_{iY}^2)+T(\boldsymbol{I}_i,\boldsymbol{I}_j)^2/(\sigma_{iT}^2)]} \tag{3}$$

Where $(\boldsymbol{I}_i,\boldsymbol{I}_j)$ = arccos $(\boldsymbol{I}_i * \boldsymbol{I}_j)$ = arccos $(\boldsymbol{I}_{ir}\boldsymbol{I}_{jr} + \boldsymbol{I}_{iv}\boldsymbol{I}_{jv} + \boldsymbol{I}_{ib}\boldsymbol{I}_{jb})$.$(\boldsymbol{I}_i,\boldsymbol{I}_j)$ denotes the angle between vector $\boldsymbol{I}_i$and $\boldsymbol{I}_j$, $\sigma_{i\boldsymbol{T}}$ and $\sigma_{i\boldsymbol{Y}}$ represent angle variance and pixel values in the local window around $i$. As long as changes in shading values are due to proportional modifications of its color channel values (red, green, and blue), the shading element could be come off from the input image. The reflectance $\boldsymbol{R}$ components can be obtained by optimizing the energy function $\boldsymbol{E(R, S)}$:

$$\text{argmin}_{\boldsymbol{R},\boldsymbol{S}}\boldsymbol{E}\,(\mathbf{R}, \boldsymbol{S}) = \sum_{i\in p}\ (\boldsymbol{R}_i - \sum_{j\in\varphi(i)} \omega_{ij}\boldsymbol{R}_i)^2 - \sum_{i\in p}(\boldsymbol{I}_i/\boldsymbol{S}_i - \boldsymbol{R}_i)^2 \tag{4}$$

*2.2. Tensor modeling*

***1) Multilinear algebra and tensor properties:*** A tensor [23] is a multidimensional array of elements characterized by a linear transformation; it can be seen as a higher order of a vector and matrix generation. An $N^{\text{th}}$ order tensor is denoted with multilinear algebra as: $\boldsymbol{X} \in \mathbb{R}^{L_1\times L_2\times\ldots\times L_N}$, where $N$, called way or mode, is the order of the tensor. The $N^{th}$ tensor order is of size $L_i$. An arbitrary element of $\boldsymbol{X}$ is a scalar noted by $x_{l_1,l_2,\ldots,l_N}$ where $1\leq l_i \leq L_i$and $1 \leq i \leq N$. $l_i$designates the location of this element in mode $i$. Hyperspectral data are considered as 3-order tensor $\boldsymbol{X}\in \mathbb{R}^{L_1\times L_2\times L_3}$ where both modes 1 and 2 represent the spatial modes and mode 3 denotes the spectral ones. Some definitions related to the basic tensor algebra are given as follows:

***Mode-d vectors:*** The mode-$d$ vectors of $\boldsymbol{X}$, defined as $l_n$-dimensional vectors, are obtained from $\boldsymbol{X}$ by varying its index $l_n$while keeping the others fixed.

***Mode-d matricing:*** It consists of unfolding an $N$-order tensor $\boldsymbol{X} \in R^{L_1\times L_2\times\ldots\times L_N}$ to a matrix $mat_d(\boldsymbol{X}) \in \mathbb{R}^{L_d\times\overline{L_d}}$, in which $\overline{L_d} = \prod_{i=1,i\neq d}^{N} L_i$, by keeping the index$l_d$fixed and varying the other index. The column vectors of the resulting matrix $mat_d(\boldsymbol{A})$ is mode-$d$ set vectors.

***Mode-d product (dU):*** The mode-$d$ product of a tensor $\boldsymbol{X}\in \mathbb{R}^{L_1\times L_2\times\ldots L_N}$by a matrix $\boldsymbol{U}\in \mathbb{R}^{L_d\times L_{d'}}$, designated as $\boldsymbol{X} \times_d \boldsymbol{U}$, is a tensor with the following entries:

$$(\boldsymbol{X}\times_d \boldsymbol{U^T})_{\boldsymbol{l_1},\boldsymbol{l_2},\ldots,\boldsymbol{l_{d-1}},\boldsymbol{l'_d},\boldsymbol{l_{d+1}},\ldots,\boldsymbol{l_N}} = \sum_{l_d}^{L_d}(\boldsymbol{X}_{\boldsymbol{l_1},\boldsymbol{l_2},\ldots,\boldsymbol{l_d},\ldots,\boldsymbol{l_N}}\boldsymbol{U}_{\boldsymbol{l'_d},\boldsymbol{l_d}}) \tag{5}$$

***Frobenus norm:*** The Frobenus norm of a tensor $\boldsymbol{X} \in \boldsymbol{R}^{L_1 \times L_2 \times \ldots L_N}$ is defined as:

$$||\boldsymbol{X}|| = \sqrt{< \boldsymbol{X}, \boldsymbol{X} >} = \sqrt{\sum_{l_1=1}^{L_1} \sum_{l_2=1}^{L_2} \ldots \sum_{l_N=1}^{L_N} a_{l_1,l_2,\ldots,l_N}^2} \tag{6}$$

***2) Tensor modeling and dimension reduction:*** Since they are able to pack up the multi aspect organisation of high dimensional data, tensor decompositions are considered as an important tool for feature extraction and dimension reduction processes. The High-order singular value decomposition (HOSVD) [24], knows as higher-order SVD, is among the effective means for tensor feature extraction and dimension reduction. According to HOSVD algorithm, an $N$-order tensor $\boldsymbol{R} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times \ldots \times I_N}$ can by composed into a core tensor $\boldsymbol{C}$ multiplied by an ensemble of component matrices $\boldsymbol{U}_n$. $\boldsymbol{C}$ describes the interaction between the matrices $\boldsymbol{U}_n$, for $n = 1,\ldots,N$ and $\boldsymbol{U}_n$ contains the orthonormal vectors including the column space of the matrix $\boldsymbol{R}_n$, produced from the mode-$n$ flattening of $\boldsymbol{R}$. Hence, $\boldsymbol{R}$ can be written as a mode $d$-product of $N$-orthogonal spaces:

$$\boldsymbol{R} = \boldsymbol{C} \times_1 \boldsymbol{U_1} \times_2 \boldsymbol{U_2} \times_3 \ldots \times_N \boldsymbol{U_N} \tag{7}$$

The HOSVD, implemented adopting the Alternative Least Square (ALS) [25] algorithm, is used to get mode-$d$ matrices $\boldsymbol{U}_n$. In the case of three mode tensor $\boldsymbol{R} = [\boldsymbol{R}_1, \boldsymbol{R}_2, \ldots, \boldsymbol{R}_{I_3}]$, the objective of HOSVD is to define the core tensor $\boldsymbol{C}$ and the basic matrices $\boldsymbol{U}_1$, $\boldsymbol{U}_2$ and $\boldsymbol{U}_3$ in a way that the $L_2$-norm reconstruction error is minimized:

$$min_{\boldsymbol{U}_1,\boldsymbol{U}_2,\boldsymbol{U}_3,\boldsymbol{C}} \boldsymbol{E_1} = ||\boldsymbol{R} - \boldsymbol{C} \times_1 \boldsymbol{U_1} \times_2 \boldsymbol{U_2} \times_3 \boldsymbol{U_3}||^2 \tag{8}$$

$\boldsymbol{U}_1$, $\boldsymbol{U}_2$, $\boldsymbol{U}_3$ should to be orthogonal, i.e., $\boldsymbol{U}_1^T\boldsymbol{U}_1 = \boldsymbol{U}_2^T\boldsymbol{U}_2 = \boldsymbol{U}_3^T\boldsymbol{U}_3$. Due to the orthonormality hypothesis, we can get:

$$\boldsymbol{W} = \boldsymbol{R} \times_1 \boldsymbol{U}_1^T \times_2 \boldsymbol{U}_2^T \times_3 \boldsymbol{U}_3^T \tag{9}$$

Hence (8) can be written as follows:

$$min_{\boldsymbol{U}_1,\boldsymbol{U}_2,\boldsymbol{U}_3} E_1 = ||\boldsymbol{R}||^2 - ||\boldsymbol{W}||^2 \tag{10}$$

$$= max_{\boldsymbol{U}_1,\boldsymbol{U}_2,\boldsymbol{U}_3} E_2 = ||\boldsymbol{W}||_F^2 \tag{11}$$

The equation (11) is equivalent to maximize:

$$\max_{\boldsymbol{U}_1,\boldsymbol{U}_2,\boldsymbol{U}_3} E_2 = \text{Trace}\,(\boldsymbol{U}_1^T \boldsymbol{O}\, \boldsymbol{U}_1) = \text{Trace}\,(\boldsymbol{U}_2^T\, \boldsymbol{Z}\, \boldsymbol{U}_2) = \text{Trace}\,(\boldsymbol{U}_3^T\, \boldsymbol{V}\, \boldsymbol{U}_3) \tag{12}$$

Where:

$$\boldsymbol{O}_{ii'} = \sum_{ll'} (\boldsymbol{R}_{(l)} \boldsymbol{U}_2 \boldsymbol{U}_2^T \boldsymbol{R}_{(l')}^T)_{ii'} (\boldsymbol{U}_3 \boldsymbol{U}_3^T)_{ll'} \tag{13}$$

$$\boldsymbol{Z}_{jj'} = \sum_{ll'} (\boldsymbol{R}_{(l)} \boldsymbol{U}_1 \boldsymbol{U}_1^T \boldsymbol{R}_{(l')}^T)_{jj'} (\boldsymbol{U}_3 \boldsymbol{U}_3^T)_{ll'} \tag{14}$$

$$\boldsymbol{V}_{ii'jj'} = \sum_{ll'} (\boldsymbol{R}_{(ijl)} \boldsymbol{R}_{i'j'l'} \boldsymbol{U}_1 \boldsymbol{U}_1^T)_{ii'} (\boldsymbol{U}_2 \boldsymbol{U}_2^T)_{jj'} \tag{15}$$

Since HOSVD algorithm converges to a local optimal, the solutions HOSVD are not unique. To avoid the convergence problem, O, Z and V have been set as suggested in [26], [27]. Since high dimensional feature space of HSIs, known as "Hughes phenomenon", poses several challenges during data processing and storage, a pre-processing step which performs the Tensor PCA (TPCA) [17] approach was adopted to reduce the dimensional space. TPCA is an extension of the PCA [28] technique where the input can be matrices or higher-order tensors. The proposed feature extraction allows capturing most of the original tensorial input variation through multilinear projections. In the case of hyperspectral image, the TPCA [26] approximation of an image $\boldsymbol{R}$ of size $n_1$, $n_2$, $n_3$ is defined as follows:

$$\widehat{\boldsymbol{R}} = \boldsymbol{R} \times_1 \widehat{\boldsymbol{U}_{J_1}} \widehat{\boldsymbol{U}}_{J_1}^T \times_2 \widehat{\boldsymbol{U}_{J_2}} \widehat{\boldsymbol{U}}_{J_2}^T \times_3 \widehat{\boldsymbol{U}}_k^T \tag{16}$$

$\widehat{\boldsymbol{U}_{J_1}}$ and $\widehat{\boldsymbol{U}_{J_2}}$ have the $J_1$ and $J_2$ largest eigenvectors associated of the unfolding matrix $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ respectively, and $\widehat{\boldsymbol{U}}_k^T$ includs the $k$-largest eigenvectors associated of the unfolding matrix $\boldsymbol{R}_3$. If $J_1 = n_1$ and $J_2 = n_2$, $\widehat{\boldsymbol{U}_{J_1}} \widehat{\boldsymbol{U}}_{J_1}^T = \boldsymbol{I}_{n_1 \times n_1}$ and $\widehat{\boldsymbol{U}_{J_2}} \widehat{\boldsymbol{U}}_{J_2}^T = \boldsymbol{I}_{n_2 \times n_2}$ in that case, (16) is defined as follows:

$$\widehat{\boldsymbol{R}} = \boldsymbol{R} \times_3 \widehat{\boldsymbol{U}}_k^T \tag{17}$$

The influence of parameters $J_1$ and $J_2$ will be studied according to classification accuracy in section 3.

The aim of this work is to improve the accuracy results obtained by the classical tensor modeling by integrating an intrinsic decomposition as tensor structures. Before the spatial-spectral dimension reduction step, the intrinsic decomposition is used to eliminate useless spatial information. Tensor modeling, based on intrinsic decomposition, generates a workflow where the spatial-spectral data is involved in the dimensionality reduction stage. Accordingly, this integration provides a good accuracy compared to the results offered by the tensor modeling only. The processing procedure of the proposed algorithm is as follows:

- We extract reflectance components from input hyperspectral data set using (4).
- The reflectance components are modeled as a tensor, and then, the TPCA algorithm is used for spatial-spectral dimensionality reduction.
- The SVM classifier is employed to classify the test set and generate classification accuracy and maps.

## 3. Experiments and discussions

### *3.1. Test data and classification process*

The proposed approach was analysed and evaluated using three real hyperspectral data sets:

***a. Indian Pines data set:*** Indian Pines image was acquired by the Visible/Infrared Aerial Imaging Spectrometry Sensor (AVIRIS) in June of 1992 and in the north-west of Indiana, US. The scene includes 145×145 pixels and 224 spectral bands in the 0.4-2.5μm range. By eliminating water absorption bands (104-108, 150-163, 220), the number of bands is reduced to 204 and a total of 16 classes have been identified.

***b. Pavia data set:*** Pavia scene was recorded by the ROSIS sensor (Imaging Spectrometer of the Reflective Optics System) capturing the university of Pavia in Italy. Pavia data set holds 115 bands of the size 610×340 in the spectral range of 0.43 to 0.86 μm. The ground truth contains 9 classes. In our experience, we did not manipulate the whole scene, we just contented on 300×200 pixels part size of the original image.

***c. Salinas data set:*** Salinas image was gathered by the AVIRIS sensor in October of 1998 over the Salinas valley in California, USA. It contains initially 224 bands; each band consists of 512×217 pixels, with spatial resolution equal to 3.7 m/pixel. After rejecting 20 bands correspond to water absorption (108-112, 154-167, 224), only 204 bands can be exploited. Salinas data set includes 16 object classes. We only used a part of 300×200 pixels and 14 classes.

The data sets and their ground truth maps, shown in Fig. 4, were downloaded from [29].We used confusion matrix results to calculate classification rates: overall accuracy (OA), average accuracy (AA) and Kappa coefficient. Computing time and visual outputs (classification maps) were also adopted to evaluate the proposed feature extraction method efficiency.

$$\mathrm{OA} = \frac{\text{The number of pixels properly classified}}{\text{The entire number of pixels to classify}} \times 100 \tag{18}$$

$$\mathrm{AA} = \frac{\text{The number of correctly labeled pixels for each class}}{\text{Total number of object classes}} \times 100 \tag{19}$$

$$\mathrm{Kappa} = \frac{\text{Pixels properly classified.}}{\text{Pixels properly classified + Number of confusion}} \times 100 \tag{20}$$

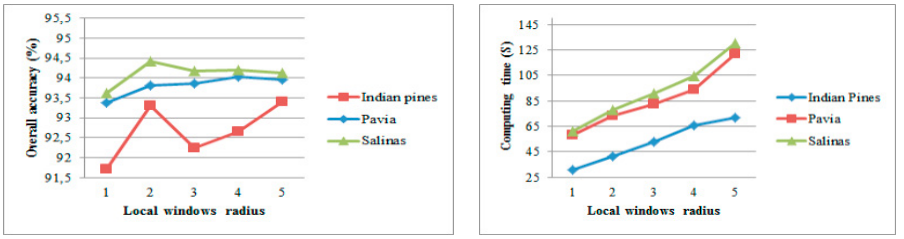
In order to prove the effectiveness of the proposed method, we compared it with six other feature extraction techniques which are: t-Distributed Stochastic Neighbor embedding (t-SNE) [36], Linear Discriminant Analysis (LDA) [37], 3D Wavelet filter [30], 3D Gabor filter [31], morphological decomposition (MD) [26] and the original TPCA approach [17].

### *3.2. Parameter analysis*

Choosing the appropriate parameters for the proposed feature extraction approach is important to evaluate accuracy and efficiency. Experiment sets were adopted for each data set to determine the most suitable settings. The implementation of the proposed approach includes essentially the evaluation of three parameters which are: 1) The local window radius for intrinsic decomposition, 2) The parameters $J_1$ and $J_2$ used for spatial dimension reduction

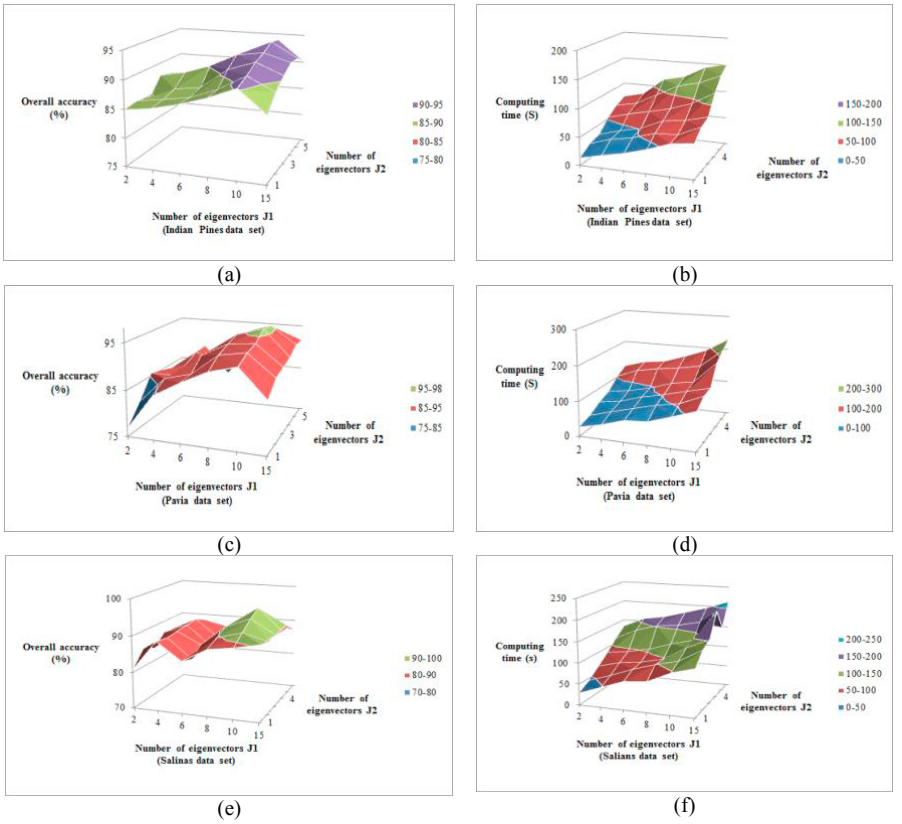
and 3) The number of bands to be reduced by the TPCA technique.

***a. Local window radius φ:*** $\varphi$ was used during the intrinsic decomposition and was analyzed by evaluating classification accuracy (OA) and computing time using the SVM library [32]. Intrinsic Image decomposition was constructed using [16]. Empirically, a large windows radius can complicate the process with the possibility that more of one class can be included in the windows. On the other hand, large windows required a lot of time for processing. Fig. 1 shows the classification rate (OA) and the computing time during the intrinsic decomposition of the three datasets with different values of $\varphi$. From this figure, it can be seen that a high classification precision can



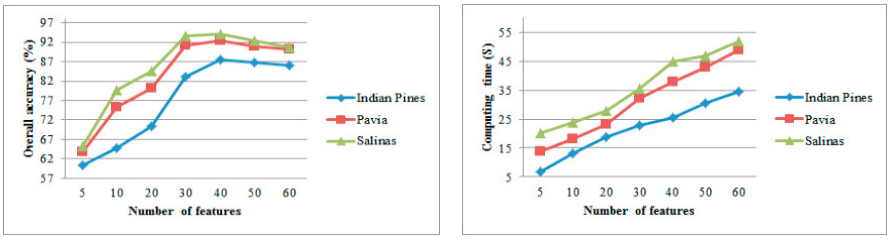

(a) (b)

Fig. 1. Intrinsic decomposition results with different values of $\boldsymbol{\varphi}$: (a) Overall classification accuracy and (b) computing time.




(a) (b) (c) (d) (e) (f)

Fig. 2. Overall accuracy and computing time, with respect to $\mathbf{J_1}$ and $\mathbf{J_2}$: (a) and (b) Indian Pines, (c) and (d) Pavia and (e) and (f) Salina.




(a) (b)

Fig. 3. The reduced dimensionality in TPCA: (a) Overall classification accuracy and (b) computing time.

be achieved when the radius of the local window $\varphi$ is fixed at 2 pixels for the two data sets Indian Pines and Salinas the three datasets with different values of $\varphi$. From this figure, it can be seen that a high classification precision can be achieved when the radius of the local window $\varphi$ is fixed at 2 pixels for the two data sets Indian Pines and Salinas as well as $\varphi = 4$ pixels for Pavia data set. To accomplish a compromise between classification accuracy and computing efficiency, $\varphi = 2$ is chosen to be the default parameter used for the intrinsic decomposition of the three data sets for the rest of the experiments.

***b. Parameters $J_1$ and $J_2$:*** In this part, we assessed the effect of the parameters $J_1$ and $J_2$ used by TPCA. We studied the impact of $J_1$ and $J_2$, according to overall accuracy (OA) and computing time. The corresponding classification OA and computing time with respect to variations of $J_1$ and $J_2$, are shown in Fig. 2. This last exhibited that the computing time of the proposed approach increases when the number of selected eigenvectors is rising. Consequently, using a small number of $J1$ and $J2$ leads to a trade-off between classification proficiency and computing efficiency. Experimentally, it was proved that the optimal values of $J_1$ and $J_2$ are 10 and 4 respectively, for Pavia and Salinas data set, while $J_1 = 10$ and $J_2 = 2$ are the optimal sizes for Indian Pines data set. These parameters will be used later for the rest of the experiments

***c. Spectral reduced dimensionality for TPCA:*** Here, we analyzed the influence of spectral dimension reduction for the TPCA technique: we implemented the TPCA algorithm described in [17]. Fig. 3 shows the classification accuracy (OA) and the computing time of the proposed technique with different number of features. It is shown that the increasing of spectral feature number rises the computing time. When the reduced dimension is set at 40 for Indian Pines data set and at 30 for Pavia and Salinas data sets, a compromise between high classification rate and fast implementation can be noticed.

### *3.3. Results and discussions*

The aim of this work is to well represent the spatial-spectral input features in the reduction process. To show the utility of incorporating intrinsic decomposition and tensor modeling, a quantitative comparison was made with the three HSIs. Three experiments were carried out respectively on Indian Pines, Pavia and Salinas scenes, in which different training samples were randomly chosen to represent the learning samples. Each classification scenario is repeated 10 times, and the average of the classification results is reported in the evaluation of the classification performance.

***a. Indian Pines data set:*** 10% of the reference data was randomly selected as learning samples for this data set and the rest was for testing. Classification accuracies (OA, AA and Kappa) and computation time results are described in table 1. For this data set, we can deduce that the proposed approach has the best classification rates with twelve classes: Alfalfa (99,98%), Corn-M (95,38%), Corn (99,35%), Grass-P (99,38%), Grass-T (98,13%), Grass-PM (99.85%), Hay(100%), Soybean–N (96,39%), Soybean–M (91,67%), Soybean–C (96.34%), Buildings (97,46%) and Stone (99,93%), whereas the worst classification rates were obtained with Corn-N (92,84%) and Soybean-M (91,67) classes. On the other side, we noted that the proposed approach is the most expensive in term of computing time, it needs two times the time required for the TPCA technique and three times the time intended for the 3D wavelet method. Fig. 4 (First row) shows the thematic classification maps for t-SNE, LDA, 3D Wavelet filter, 3D Gabor filter, the morphological decomposition (MD) method, TPCA technique and the proposed approach respectively. From this figure, we can observe that the proposed method outperforms the other tested techniques in term of visualization effectiveness.

***b. Pavia data set:*** In the second experiment, the training samples were chosen at random to represent 4% of reference data while 96% of pixels were dedicated to testing. Table 2 describes the classification accuracies and computing time for Pavia image. From this table, it can be seen that the proposed approach is characterized by the highest classification accuracies for six classes: Asphalt (95,85%), Meadows (96,68%), Trees (97,52%), Soil (98,68%), Bitumen (99,85%) and Bricks (98,43%). Empirically, our proposed method exceeds the other tested algorithms for the OA (96,62%), AA(97,55%) and Kappa (94,35%) indices. The adaptation of the intrinsic decomposition can increase in a great way the precision of the TPCA technique; it enhanced the accuracy (OA) by about 10%. Taking into consideration the computing time, the proposed approach proves again, to be the most costly for the Pavia data set, it needs for almost 3 minutes.

***c. Salinas data set:*** Table 3 summarizes classification outputs of the third tested HSI. For this scene, we implemented only 2% of pixels as training samples and 98% for testing. Once again with the Salinas dataset, the proposed approach has the highest classifications for nine classes: Weeds-2 (99,89%), Fallow (99,49%), Fallow-P (99,69%), Fallow-S (99,63%), Stubble (99,18%), Grapes (94,19%), Lettuce-5wk (96,69%), Lettuce-6wk (97,14%)

TABLE 1. CLASSIFICATION ACCURACIES FOR VARIOUS REDUCTION METHODS IN INDIAN PINES DATA SET (10% OF REFERENCE DATA).

| Class | t-SNE | LDA | 3D Wavelet | 3D Gabor | MD | TPCA | Our approach |
|---|---|---|---|---|---|---|---|
| Alfalfa | 88,08 | 87,18 | 75,9 | 99,95 | 98,07 | 97,27 | **99,98** |
| Corn-N | 70,19 | 65,24 | 72,85 | 91,1 | **94,43** | 92,02 | 92,84 |
| Corn-M | 83,02 | 80,66 | 82,35 | 92,25 | 92,12 | 88,25 | **95,38** |
| Corn | 93,07 | 67,44 | 78,13 | 98,07 | 92,02 | 90,04 | **99,35** |
| Grass-P | 98,14 | 87,36 | 98,65 | 78,81 | 95,5 | 92,39 | **99,38** |
| Grass-T | 88,45 | 95,24 | 81,87 | 95,99 | 97,25 | 97,2 | **98,13** |
| Grass-PM | 96,44 | 90,12 | 92,84 | 97,85 | 95,95 | 97,98 | **99.85** |
| Hay | 90,05 | 87,54 | 85,2 | 99,76 | 94,93 | 99,94 | **100,00** |
| Oats | 77,27 | 71,7 | 81,8 | **99,97** | 98,1 | 92,76 | 99,80 |
| Soybean-N | 88,4 | 88,43 | 84,37 | 95,95 | 89,02 | 94,5 | **96,39** |
| Soybean-M | 88,05 | 82,64 | 89,67 | 87,35 | 89,27 | 85,85 | **91,67** |
| Soybean-C | 75,11 | 61,4 | 77,12 | 95,25 | 92,71 | 88,01 | **96.34** |
| Wheat | 87,26 | 83,34 | 91,06 | **99,85** | 90,97 | 90,65 | 99,76 |
| Woods | 95,44 | 96,9 | 93,42 | **99,21** | 93,31 | 97,18 | 99,01 |
| Buildings | 78,17 | 51,18 | 84,13 | 96,1 | 92,55 | 90,32 | **97,46** |
| Stone | 78,14 | 82,13 | 78,84 | 99,74 | 96,15 | 90,89 | **99,93** |
| **OA (%)** | **78,03** | **70,73** | **80,11** | **91,27** | **87,03** | **84,29** | **94,16** |
| **AA (%)** | **85,95** | **79,90** | **84,26** | **95,45** | **93,46** | **92,82** | **97,79** |
| **Kappa (%)** | **76,65** | **66,88** | **78,60** | **89,88** | **85,24** | **82,53** | **92,83** |
| **Time (S)** | **23,52** | **50,47** | **16,73** | **35,73** | **27,38** | **23,69** | **45,55** |

TABLE 2. CLASSIFICATION ACCURACIES FOR VARIOUS REDUCTION METHODS IN PAVIA DATA SET (4% OF REFERENCE DATA).

| Class | t-SNE | LDA | 3D Wavelet | 3D Gabor | MD | TPCA | Our approach |
|---|---|---|---|---|---|---|---|
| Asphalt | 71,18 | 61,13 | 75,79 | 93,73 | 95,78 | 87,96 | **95,85** |
| Meadows | 86,96 | 78,86 | 83,71 | 95,52 | 86,93 | 88,13 | **96,68** |
| Gravel | 80,65 | 85,62 | 87,62 | **99,9** | 99,65 | 99,71 | 99,85 |
| Trees | 91,02 | 84,23 | 87,57 | 96,50 | 97,38 | 96,58 | **97,52** |
| Sheets | 95,05 | 92,37 | 91,48 | 97,19 | **99,10** | 97,62 | 97,13 |
| Soil | 88,28 | 81,13 | 91,17 | 98,10 | 94,49 | 95,48 | **98,68** |
| Bitumen | 89,06 | 52,05 | 85,65 | 98,55 | 96,65 | 95,85 | **99,85** |
| Bricks | 91,3 | 84,13 | 85,97 | 97,23 | 90,79 | 92,58 | **98,43** |
| Shadows | 80,12 | 72,29 | 85,28 | 99,49 | **99,67** | 99,00 | 93,96 |
| **OA(%)** | **80,71** | **68,30** | **82,42** | **92,73** | **88,29** | **86,46** | **96,62** |
| **AA(%)** | **85,95** | **76,86** | **86,02** | **97,35** | **95,60** | **94,76** | **97,55** |
| **Kappa(%)** | **71,79** | **53,03** | **80,43** | **90,55** | **86,06** | **84,74** | **94,35** |
| **Time (S)** | **48,56** | **87,71** | **42,35** | **79,89** | **68,56** | **51,54** | **85,04** |

TABLE 3. CLASSIFICATION ACCURACIES FOR VARIOUS REDUCTION METHODS IN SALINAS DATA SET (2% OF REFERENCE DATA).

| Class | t-SNE | LDA | 3D Wavelet | 3D Gabor | MD | TPCA | Our approach |
|---|---|---|---|---|---|---|---|
| Weeds-1 | 90,32 | 87,11 | 90,97 | **99,55** | 97,87 | 98,04 | 99,52 |
| Weeds-2 | 87,29 | 93,94 | 88,65 | 98,31 | 98,17 | 95,9 | **99,89** |
| Fallow | 93,7 | 92,53 | 90,86 | 97,02 | 98,79 | 97,81 | **99,49** |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Fallow-P | 96,08 | 80,75 | 90,28 | 98,71 | 92,61 | 98,28 | **99,69** |
| Fallow-S | 96,29 | 82,37 | 85,52 | 98,65 | 99,44 | 97,29 | **99,63** |
| Stubble | 96,16 | 98,29 | 94,49 | 97,49 | 92,26 | 97,88 | **99,18** |
| Celery | 91,27 | 91,8 | 92,54 | **99,58** | 98,72 | 95,29 | 99,48 |
| Grapes | 87,23 | 53,03 | 75,12 | 90,91 | 86,65 | 82,12 | **94,19** |
| Corn | 93,31 | 65,41 | 81,84 | 99,92 | 99,96 | **100** | 99,92 |
| Lettuce-4wk | 96,37 | 72,56 | 63,03 | 96,45 | 97,6 | **97,81** | 97,48 |
| Lettuce-5wk | 91,72 | 80,39 | 87,97 | 91,26 | 94,33 | 95,21 | **96,69** |
| Lettuce-6wk | 90,15 | 82,61 | 89,37 | 95,01 | 95,87 | 96,65 | **97,14** |
| Lettuce-7wk | 86,21 | 80,81 | 80,78 | 98,73 | **98,91** | 99,04 | 98,72 |
| Vignoble | 62,64 | 43 | 66,06 | 82,06 | 82,44 | 76,12 | **87,93** |
| **OA (%)** | **82,79** | **70,69** | **81,80** | **91,30** | **88,33** | **84,07** | **95,70** |
| **AA (%)** | **89,91** | **78,90** | **84,10** | **95,97** | **95,25** | **94,81** | **97,78** |
| **Kappa (%)** | **78,547** | **66,06** | **78,82** | **88,24** | **86,37** | **81,70** | **93,54** |
| **Time (S)** | **34,56** | **61,91** | **19,75** | **36,08** | **31,53** | **27,2** | **43,67** |

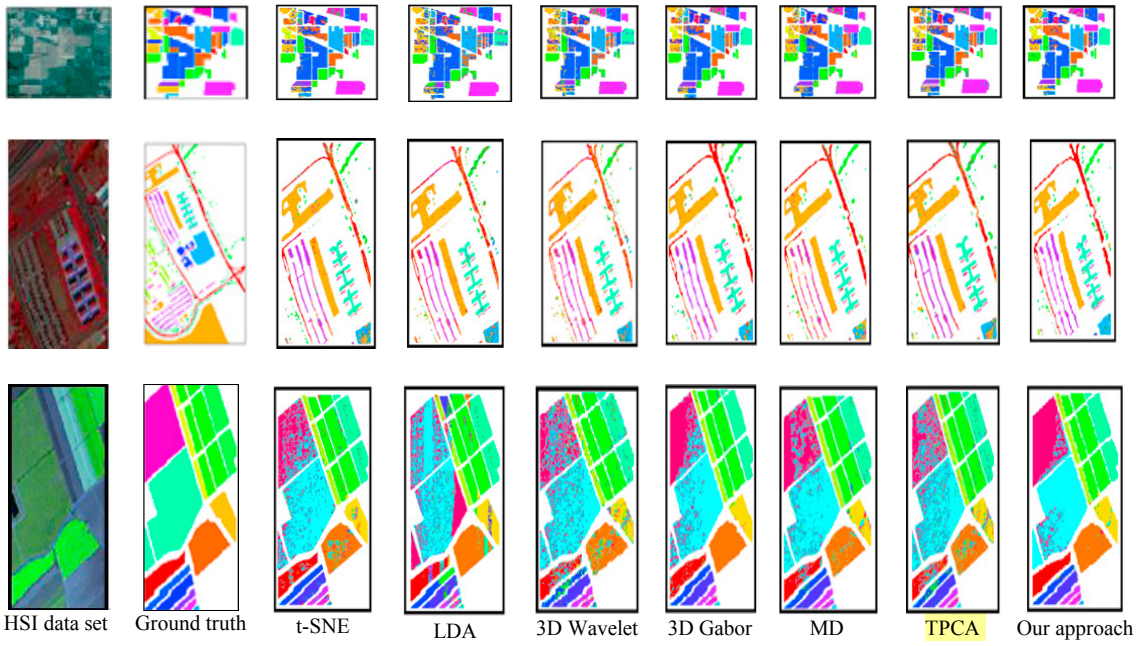



Fig. 4. Classification maps of Indian Pines (first row), Pavia (second row) and Salinas (third row) data sets, obtained by all implemented approaches.

and Vignoble (87,93%). However, the worst classification rates were obtained with both Grapes and Vignoble classes. The new suggested technique obviously exceeds the other approaches by about 4-14% of OA. Furthermore, it surpasses the other implemented methods in terms of visibility.

## 4. Conclusion and future works

In this paper, a novel feature extraction approach is proposed for HSI classification tasks. The proposed method is focused on exploring significant spectral-spatial HSI characteristics; based on intrinsic decomposition and tensor modelling concepts.The contribution of the current work is presented in two parts. On the one hand, the proposed approach is able to remove useless and non-relevant spatial data from the hyperspectral scene and integrate more useful spatial information during the dimensional reduction process. On the other hand, the followed methodology is proficient to reduce the spatial-spectral HSI dimensionality, which is seen very important to improve the classification performance. The experimental study proved the effectiveness of the proposed method compared with representative spatial-spectral feature extraction methods, such as 3D Wavelet, MD and 3D Gabor. However, the computational burden of HSI classification is not computationally efficient compared with other tested methods.

This weakness can be improved by using a fast variant of TPCA such like the framework proposed in [17].

## Acknowledgment

This work was supported and financed by the Ministry of Higher Education and Scientific Research of Tunisia.

## References

[1] Fejjari Asma, Karim Saheb Ettabaa, Ouajdi Korbaa. Fast Spatial Spectral Schroedinger Eigenmaps algorithm for hyperspectral feature extraction. *The 22nd International Conference on Knowledge-Based and Intelligent Information & Engineering System (KES)*, Belgrade, Serbia; 2018, p. 656–664.
[2] Ren Jiansi, Ruoxiang Wang, Gang Liu, Ruyi Feng, Yuanni Wang, Wei Wu. Partitioned Relief-F Method for Dimensionality Reduction of Hyperspectral Images. *Remote Sensing* 2020; 12(7).
[3] Ben Salem Manel, Karim Saheb Ettabaa, Med Salim Bouhlel. An unsupervised classification approach for hyperspectral images based adaptive spatial and spectral neighborhood selection and graph clustering. *The 22nd International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES)*, Belgrade, Serbia; 2018, p. 743–750.
[4] Fei Baowei. Hyperspectral imaging in medical applications. *Data Handling in Science and Technology* 2020; 32: 523-565.
[5] Warren Russell E., David B. Cohn. Chemical detection on surfaces by hyperspectral imaging. *J. of Applied Remote Sensing* 2017; 11(1).
[6] J. Murphy Richard, Zachary Taylor, Sven Schneider, Juan Nieto. Mapping clay minerals in an open-pit mine using hyperspectral imagery and automated feature extraction. *Vertical Geology Conference* 2014.
[7] Sun Weiwei, Qian Du. Hyperspectral band selection: A review. *IEEE Geoscience and Remote Sensing Magazine* 2019; 7: 118–139.
[8] Brajesh Kumar, Onkar Dikshit. Hyperspectral image classification based on morphological profiles and decision fusion. *International Journal of Remote Sensing* 2017; 38(20).
[9] Dey Neel, Sungmin Hong, Thomas Ach, Yiannis Koutalos, Christine A. Curcio, R. Theodore Smith, Guido Gerig. Tensor decomposition of hyperspectral images to study autofluorescence in age-related macular degeneration. *Medical Image Analysis* 2019; 56: 96-109
[10] Li Jun, José M. Bioucas-Dias, Antonio Plaza. Hyperspectral Image Segmentation Using a New Bayesian Approach with Active Learning. *IEEE Transactions on Geoscience and Remote Sensing* 2011; 49(10): 3947-3960.
[11] Sellami Akrem, Mohamed Farah, Imed Riadh Farah, Basel Solaiman. Hyperspectral Imagery Semantic Interpretation Based on Adaptive Constrained Band Selection and Knowledge Extraction Techniques. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 2018; 11(04): 1337–1347.
[12] Huang Fenghua, Ying Yu, Tinghao Feng. Hyperspectral remote sensing image change detection based on tensor and deep learning. *Journal of Visual Communication and Image Representation* 2019; 58: 233–244.
[13] Liu Yongjian, Guoming Gao, Yanfeng Gu. Tensor Matched Subspace Detector for Hyperspectral Target Detection. *IEEE Transactions on Geoscience and Remote Sensing* 2017; 55(4): 1967–1974.
[14] Yan Ronghua, Jinye Peng, Dongmei Ma, Desheng Wen. Spectral Tensor Synthesis Analysis for Hyperspectral Image Spectral–Spatial Feature Extraction. *Journal of the Indian Society of Remote Sensing* 2019; 47: 91–100.
[15] Xie Chuanqi, Yong He, Spectrum and Image Texture Features Analysis for Early Blight Disease Detection on Eggplant Leaves. *Sensors* 2016; 16 (6).
[16] Shen Jianbing, Xiaoshan Yang, Yunde Jia, Xuelong Li. Intrinsic images using optimization. *The 24th IEEE Conference on Computer Vision and Pattern Recognition CVPR*, Providence, RI, USA; 2011, p. 3481–3487.
[17] Ren Yuemei, Liang Liao, Stephen John Maybank, Yanning Zhang, Xin Liu. Hyperspectral Image Spectral-Spatial Feature Extraction via Tensor Principal Component Analysis. *IEEE Geoscience and Remote Sensing Letters* 2017; 14(9): 1431–1435.
[18] Guo Xian, Xin Huang, Lefei Zhang, Liangpei Zhang, Antonio Plaza, Jón Atli Benediktsson. Support Tensor Machines for Classification of Hyperspectral Remote Sensing Imagery. *IEEE Transactions on Geoscience and Remote Sensing* 2016; 54(6): 3248–3264.
[19] Barrow Harry G., Joan M. Tenenbaum. Recovering Intrinsic Scene Characteristics from Images. *In Computer Vision Systems,* Eds. 1978, New York: Academic, p. 3–26.
[20] Land Edwin H., John J McCann. Lightness and retinex theory. *Journal of the Optical Society of America* 1971; 61(1): 1–11.
[21] Lee Kyong Joon, Qi Zhao, Xin Tong, Minmin Gong, Shahram Izadi, Sang Uk Lee, Ping Tan, Stephen Lin. Estimation of Intrinsic Image Sequences from Image+Depth Video. *ECCV'12 Proceedings of the 12th European conference on Computer Vision* 2012; p. 327–340.
[22] Shen Jianbing, Xiaoshan Yang, Xuelong Li, Yunde Jia. Intrinsic Image Decomposition Using Optimization and User Scribbles. *IEEE Transactions on Cybernetics* 2013; 43(2): 425–36.
[23] Lu Haiping, Konstantinos N. Plataniotis, Anastasios N. Venetsanopoulos. A survey of multilinear subspace learning for tensor data. *Journal Pattern Recognition* 2011; 44(7): 1540-1551.
[24] Bergqvist G., Erik G. Larsson. The Higher-Order Singular Value Decomposition: Theory and an Application [Lecture Notes]. *IEEE Signal Processing Magazine* 2010; 27(3): 151–154.
[25] Comon Pierre, Xavier Luciani, André de Almeida. Tensor Decompositions, Alternating Least Squares and other Tales. *Journal of Chemometrics* 2009; 23: 393–405.
[26] Velasco-Forero Santiago, Jesus Angulo. Classification of hyperspectral images by tensor modeling and additive morphological decomposition. *Pattern Recognition* 2012; 46(2): 566–577.
[27] Velasco-Forero Santiago, Jesus Angulo. Parameters selection of morphological scale-space decomposition for hyperspectral images using tensor modeling. *SPIE* 2010; 7695: 1–12.
[28] Ren Jianchang, Jaime Zabalza, Stephen Marshall, Jiangbin Zheng. Effective Feature Extraction and Data Reduction in Remote Sensing Using Hyperspectral Imaging. *IEEE Signal Processing Magazine* 2014; 31(4): 149–154.
[29] Computational Intelligence search group site. *Hyperspectral Remote Sensing Scenes*. http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes. [Accessed: 05- October- 2019].
[30] Rasti Behnood, Johannes R. Sveinsson, Magnus O. Ulfarsson, Jon Atli Benediktsson. Hyperspectral Image Denoising Using 3d Wavelets. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Munich, Germany; 2012, p. 1349–1352.

[31] Bau Tien, Glenn Healey. Rotation and Scale Invariant Hyperspectral Classification Using 3D Gabor Filters. *Proceedings of SPIE - The International Society for Optical Engineering* 2009; 7334.
[32] F Chang Chih-Chung, Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2011; 2(3): 1–27.
[33] Fejjari Asma, Karim Saheb Ettabaa, Ouajdi Korbaa. Feature Extraction Techniques for Hyperspectral Images Classification. *Soft Computing Applications: Proceedings of the 8th International Workshop Soft Computing Applications (SOFA 2018)*, Arad, Romania; 2018.
[34] Fejjari Asma, Karim Saheb Ettabaa, Ouajdi Korbaa. Modified Graph-Based Algorithm for Efficient Hyperspectral Feature Extraction. *The 32nd International Symposium on Computer and Information Sciences (ISCIS),* Poznan, Poland; 2018, p. 87–95.
[35] Fejjari Asma, Karim Saheb Ettabaa, Ouajdi Korbaa. Modified Schroedinger Eigenmap Projections Algorithm for Hyperspectral Imagery Classification. *IEEE/ACS 14th International Con-ference on Computer Systems and Applications (AICCSA)*, Hamammet, Tunisia; 2017, p. 809–814.
[36] Devassy Binu Melit, Sony George. Dimensionality reduction and visualisation of hyperspectral ink data using t-SNE. *Forensic Science International* 2020, 311.
[37] Sakarya Ufuk. Hyperspectral Dimension Reduction Using Global and Local Information Based Linear Discriminant Analysis. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 2014, II-7: 61–66.
[38] Blanes Ian, Alaitz Zabala Torres, Gerard Moré, Xavier Pons, Joan Serra-Sagristà. Classification of Hyperspectral Images Compressed through 3D-JPEG2000. *The 12th International Conference on Knowledge-Based Intelligent Information and Engineering Systems (KES*), Part III, Zagreb, Croatia; 2008, p. 416–423.