Classification of hyperspectral images by tensor modeling and additive morphological decomposition

Santiago Velasco-Forero, Jesus Angulo

Classification of hyperspectral images by tensor modeling and additive morphological decomposition

Santiago Velasco-Forero, Jesus Angulo

CMM-Centre de Morphologie Mathématique, Mathématiques et Systèmes, MINES Paristech, France

Abstract

Pixel-wise classification in high-dimensional multivariate images is investigated. The proposed method deals with the joint use of spectral and spatial information provided in hyperspectral images. Additive morphological decomposition (AMD) based on morphological operators is proposed. AMD defines a scale-space decomposition for multivariate images without any loss of information. AMD is modeled as a tensor structure and tensor principal components analysis is compared as dimensional reduction algorithm versus classic approach. Experimental comparison shows that the proposed algorithm can provide better performance for pixel classification of hyperspectral image than many other well-known techniques.

Keywords: Hyperspectral images; Mathematical Morphology; Pixelwise Classification; Tensor modeling.

Classification of hyperspectral images by tensor modeling and additive morphological decomposition

Santiago Velasco-Forero, Jesus Angulo

CMM-Centre de Morphologie Mathématique, Mathématiques et Systèmes, MINES Paristech, France

1. Introduction

Hyperspectral imaging (HSI) is a remote sensing technique that acquires two dimension spatial images in typically hundreds of contiguous bands of high spectral resolution covering the visible, near-infrared, and shortwave infrared bands. This technique has been applied in several applications, for instance, face detection [1], planetary exploitation [2], and biology [3]. That technology produces a signature for each pixel in the image in many highly correlated bands presenting considerable amounts of spectral redundancy. On the one hand, *dimension reduction* of multivariate images is one of the main subject of interest for the hyperspectral community. Target detection, image segmentation, pixel classification and spectra unmixing in HSI have the additional difficulty that pixels are located in a high dimension space increasing computational complexity and degrading accuracy [4, 5]. On the other hand, identification of relatively small objects incorporates issues because spatial resolution is necessary for accurate classification. Accordingly, if the spatial contents of the image is not used, the resulting thematic map sometimes looks noisy (salt and pepper classification noise). In the particular

case of supervised classification, that topic is called *spatial/spectral classification*. The aim is to assign each image pixel to one class using a feature vector based on its own spectral value (the spectral information) and information extracted from its neighborhood (referred to as the spatial information). The pioneer work in introducing spatial context into a multivariate image classification is ECHO (Extraction and Classification of Homogeneous Objects) classifier [6]. Since then, many studies have been led to propose new algorithms to perform spectral-spatial classification. Recent works in HSI have seen a surge of research toward developing approaches that exploit various features specific to the spatial/spectral classification. The approaches due to ([7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17]) show some degree of success. Pixelwise classification incorporating spatial information in HSI can be roughly divided according to their mathematical formulation as follows.

- Smoothing by partial differential equation [9, 13]: Anisotropic diffusion from classic grey-scale image processing [18] is extended to multivariate scenarios, by using a general definition of vector gradient.

- Markov random field, which takes into account the spatial dependence between the pixels based on the observed intensity field [14].

- Mathematical Morphology [16, 19, 17]: Results of morphological operators over features calculated by some dimensionality reduction technique are incorporate into the classification.

- Classifiers with spatial information [7]: Pairwise classification based on kernel formulation where the spatial information is incorporated as an operation among spatial, spectral and spatial-spectral kernels.

- Segmentation and post-processing [10, 20, 15]: Approaches start with a preliminary spatial/spectral clustering/segmentation followed by a fusion-area stage based on supervised criterium.

- Tensor modeling [21, 12]: Three dimension array or third-order tensor preserves the usual image representation and band continuity is represented as the third tensor dimension. Spatial information is included as row-column correspondence in the mathematical structure.

- Context-based classification [11, 8] attempts to identify relevant models to a test sample through context estimation in the feature space, by using random set framework [11] or hierarchical segmentation [8].

To the best of our knowledge, there has been no previous work on modeling multivariate images by using additive morphological decompositions as tensor structures, which is the subject of this paper. Our approach is motivated by the desire to discover "interesting" low-dimensional linear projections of high-dimensional images where the spatial information plays an important role. In this paper, we present an additive scale-space decomposition which incorporates spatial information into the dimensionality reduction stage for multivariate images. In summary, the main contributions of this paper are as follows.

- A new image decomposition based on mathematical morphology which is more compact and performs better in supervised classification.

- Tensor-PCA based on morphological decomposition producing a workflow where the spatial information is included in the dimensionality reduction step instead of in the classification stage.
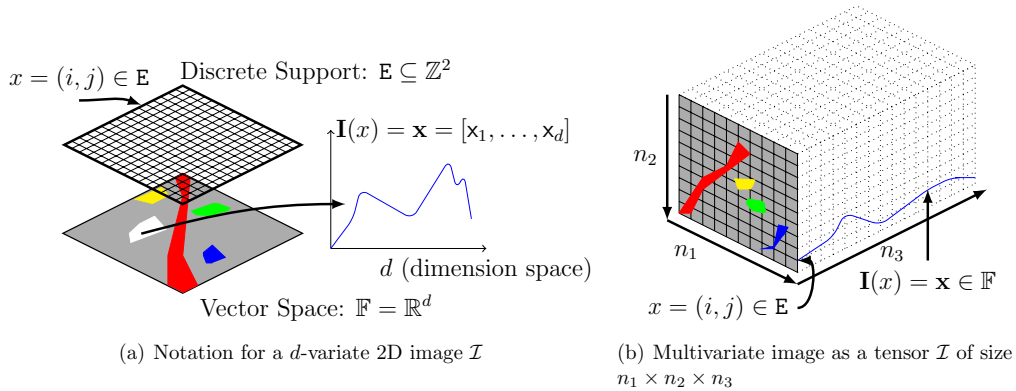
(a) Notation for a d-variate 2D image $\mathcal{I}$

(b) Multivariate image as a tensor $\mathcal{I}$ of size $n_1 \times n_2 \times n_3$

Figure 1: Mathematical notation for a 2D multivariate image, $\mathbf{I} : \mathtt{E} \to \mathbb{F}$

- We show in practical examples that our workflow allows to include the spatial information in the dimensionality reduction stage.

- State of the art for classification of HSI in remote sensing based on morphological decomposition.

The paper is organized as follows. Section 2 presents the additive scale-space decomposition with morphological transformations. Section 3 introduces tensor modeling of morphological decomposed multivariate images. Section 4 proposes a formulation of classification for reduced tensors using support vector machines (SVM) and shows the effectiveness of the modified approach via practical examples with a comparison versus classical approaches. Section 5 concludes the paper.

2. Additive Morphological Decomposition

In this section we focus on mathematical morphology (MM) as a nonlinear image processing methodology composed of a larger family of operators based

on the set theory and defined on an abstract structure known as complete lattice of spatial structures [22].

2.1. Notation

Let us precise the terms and notation to be used in the rest of the paper. Let E be a subset of the discrete space $\mathbb{Z}^2$, considered as the support space of the 2D image, and $\mathbb{F} \subseteq \mathbb{R}^d$ be a set of pixels values in dimension d, corresponding to the vector space of values of the image of size $n_1 \times n_2$ and d channels. A *vector-valued image* is represented by the mapping,

$$
\mathbf{I} : \begin{cases} E & \rightarrow & \mathbb{F} \\ x = (i,j) & \rightarrow & \mathbf{x} \end{cases} \tag{1}
$$

i.e., $\mathbf{I} \in \mathcal{F}(E, \mathbb{F})$ is the set of maps from a point x at the discrete spatial coordinates $(i,j) \in E$ into a vector value $\mathbf{x} \in \mathbb{F} \subseteq \mathbb{R}^d$. Let us assume that the pixel x is represented by a d-dimensional vector $\mathbf{x}(i,j) = [\mathsf{x}_1(i,j), \mathsf{x}_2(i,j), \ldots, \mathsf{x}_d(i,j)] \in \mathbb{R}^d$, where $\mathbb{R}$ denotes the set of real numbers in which the pixels spectral response $\mathsf{x}_l(i,j)$ at sensor channels $l = 1, \ldots, d$. Figure 1 shows the notation in two graphical schemes. Additionally, let $\mathbf{X}$ be an $n \times d$ matrix representing d spectral bands for each n pixels in the vector-value image $\mathbf{I}$. We use the following notations to facilitate presentation: Scalars are denoted by lower case letters $(a, b, \ldots)$, vectors by bold lower case letters $(\mathbf{a}, \mathbf{b}, \ldots)$, matrices or images by bold upper-case letters $(\mathbf{X}, \mathbf{Y}, \ldots)$, and higher-order tensors by calligraphic upper-case letters $(\mathcal{I}, \mathcal{S}, \ldots)$. The order of tensor $\mathcal{I} \in \mathbb{R}^{n_1 \times n_2 \ldots \times n_J}$ is J. We use subscripts to illustrate the tensor order, for example $\mathcal{I}_{ijkl}$ is a tensor of order four.
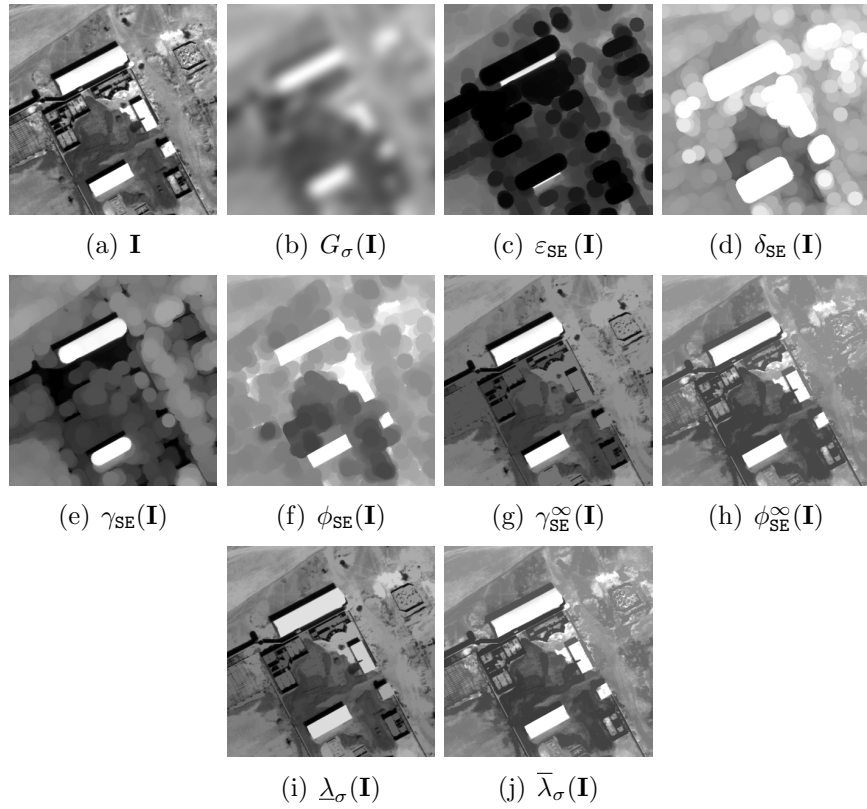
6

Figure 2: Morphological transformations of a scalar (grey level) image. Original image (a) is a 342×342 pixels in 70-cm-resolution satellite image from the panchromatic band of Quickbird.

2.2. Basic Morphological Transformation

The morphological image transformation Φ is an *image to image transformation*, i.e., $\Phi : \mathcal{F}(\mathrm{E}, \mathbb{F}) \rightarrow \mathcal{F}(\mathrm{E}, \mathbb{F})$. Additionally, it is a *neighborhood-image transform* [22], i.e., the output value at a given pixel x is a function of the values of the pixels falling in the neighborhood induced by the structuring element SE and centered at the considered pixel x. The shape of SE plays the role of the a priori knowledge about the geometry of the interesting and uninteresting spatial structures in the image. In general a transformation $\Phi \in \mathcal{F}(\mathrm{E}, \mathbb{F}) \rightarrow \mathcal{F}(\mathrm{E}, \mathbb{F})$ is called:

- *extensive* if $\mathbf{I}(x) \leq \Phi(\mathbf{I}(x))$,

- *anti-extensive* if $\Phi(\mathbf{I}(x)) \leq \mathbf{I}(x)$,

- *idempotent* if $\Phi(\Phi(\mathbf{I}(x))) = \Phi(\mathbf{I}(x))$,

for all $x \in \mathrm{E}$ and $\mathbf{I} \in \mathcal{F}(\mathrm{E}, \mathbb{F})$. There are two basic operators in MM named erosion and dilation. The *erosion* of an image $\mathbf{I}$ at pixel $x \in \mathrm{E}$ by the structuring element $\mathrm{SE} \subset \mathrm{E}$ is the transformation given by

$$\varepsilon_{\mathrm{SE}}(\mathbf{I})(x) = \{\mathbf{I}(y) : \mathbf{I}(y) = \bigwedge_{z \in \mathrm{SE}(x)} \mathbf{I}(z)\} \tag{2}$$

where $\bigwedge$ is the infimum according to a total ordering in $\mathbb{F}$ and $\mathrm{SE}(x)$ is the structuring element centered at the considered pixel x. The dual operator called *dilation* is the transformation given by

$$\delta_{\mathrm{SE}}(\mathbf{I})(x) = \{\mathbf{I}(y) : \mathbf{I}(y) = \bigvee_{z \in \mathrm{SE}(x)} \mathbf{I}(z)\} \tag{3}$$

For binary or grey-scale images, they are simple in the sense that they usually have an intuitive interpretation. Erosion $\varepsilon_{\mathrm{SE}}(\mathbf{I})$ shrinks bright objects,

Table 1: Key notations used in the paper formulation. $\mathbf{I}$ is the original image and $\mathbf{M}$ a marker image. $\mathtt{B}$ is the unitary isotropic structuring element useful in the geodesic operators. $\mathtt{SE}$ is the structuring element. Idempotent means that it applies the operator twice with the same set the parameter yields the same result. Transformations are illustrated in Fig. 2 in a practical example.

Notation	Name	Definition	Idempotent
$G_\sigma(\cdot)$	Gaussian Filter	$G_\sigma(\mathbf{I}) = \mathbf{I} * \mathcal{N}(0,\sigma)$	No
$\varepsilon_{\mathtt{SE}}(\cdot)$	Erosion	$\varepsilon_{\mathtt{SE}}(\mathbf{I})(x) = \{\mathbf{I}(z) : \bigwedge_{y \in \mathtt{SE}(x)} \mathbf{I}(y)\}$	No
$\delta_{\mathtt{SE}}(\cdot)$	Dilation	$\delta_{\mathtt{SE}}(\mathbf{I})(x) = \{\mathbf{I}(z) : \bigvee_{y \in \mathtt{SE}(x)} \mathbf{I}(y)\}$	No
$\gamma_{\mathtt{SE}}(\cdot)$	Opening	$\gamma_{\mathtt{SE}}(\mathbf{I}) = \delta_{\mathtt{SE}}(\varepsilon_{\mathtt{SE}}(\mathbf{I}))$	Yes
$\phi_{\mathtt{SE}}(\cdot)$	Closing	$\phi_{\mathtt{SE}}(\mathbf{I}) = \varepsilon_{\mathtt{SE}}(\delta_{\mathtt{SE}}(\mathbf{I}))$	Yes
$\delta_{\mathtt{B}}^i(\cdot,\cdot)$	Geodesic dilation of size i	$\delta_{\mathtt{B}}^i(\mathbf{M},\mathbf{I}) = \delta_{\mathtt{B}}^1(\delta_{\mathtt{B}}^{i-1}(\mathbf{M},\mathbf{I}),\mathbf{I})$, with $\delta_{\mathtt{B}}^1(\mathbf{M},\mathbf{I}) = \delta_{\mathtt{SE}}(\mathbf{M}) \wedge \mathbf{I}$	No
$\varepsilon_{\mathtt{B}}^i(\cdot,\cdot)$	Geodesic erosion of size i	$\varepsilon_{\mathtt{B}}^i(\mathbf{M},\mathbf{I}) = \varepsilon_{\mathtt{B}}^1(\varepsilon_{\mathtt{B}}^{i-1}(\mathbf{M},\mathbf{I}),\mathbf{I})$, with $\varepsilon_{\mathtt{B}}^1(\mathbf{M},\mathbf{I}) = \varepsilon_{\mathtt{SE}}(\mathbf{M}) \vee \mathbf{I}$	No
$\delta_{\mathtt{B}}^\infty(\cdot,\cdot)$	Reconstruction by Dilation	$\delta_{\mathtt{B}}^\infty(\mathbf{M},\mathbf{I}) = \{\delta_{\mathtt{B}}^i(\mathbf{M},\mathbf{I}) \mid \delta_{\mathtt{B}}^{i+1}(\mathbf{M},\mathbf{I}) = \delta_{\mathtt{B}}^i(\mathbf{M},\mathbf{I})\}$	Yes
$\varepsilon_{\mathtt{B}}^\infty(\cdot,\cdot)$	Reconstruction by Erosion	$\varepsilon_{\mathtt{B}}^\infty(\mathbf{M},\mathbf{I}) = \{\varepsilon_{\mathtt{B}}^i(\mathbf{M},\mathbf{I}) \mid \varepsilon_{\mathtt{B}}^{i+1}(\mathbf{M},\mathbf{I}) = \varepsilon_{\mathtt{B}}^i(\mathbf{M},\mathbf{I})\}$	Yes
$\gamma_{\mathtt{SE}}^\infty(\cdot)$	Opening by reconstruction	$\gamma_{\mathtt{SE}}^\infty(\mathbf{I}) = \delta_{\mathtt{B}}^\infty(\gamma_{\mathtt{SE}}(\mathbf{I}),\mathbf{I})$	Yes
$\phi_{\mathtt{SE}}^\infty(\cdot)$	Closing by reconstruction	$\phi_{\mathtt{SE}}^\infty(\mathbf{I}) = \varepsilon_{\mathtt{B}}^\infty(\phi_{\mathtt{SE}}(\mathbf{I}),\mathbf{I})$	Yes
$\overline{\lambda}_\sigma(\cdot)$	Gaussian Upper-Leveling[23]	$\overline{\lambda}_\sigma(\mathbf{I}) = \delta_{\mathtt{B}}^\infty(G_\sigma(\mathbf{I}) \wedge \mathbf{I},\mathbf{I})$	No
$\underline{\lambda}_\sigma(\cdot)$	Gaussian Lower-Leveling [23]	$\underline{\lambda}_\sigma(\mathbf{I}) = \varepsilon_{\mathtt{B}}^\infty(G_\sigma(\mathbf{I}) \vee \mathbf{I},\mathbf{I})$	No

whereas dilation $\delta_{\mathtt{SE}}(\mathbf{I})$ expands bright objects at the boundary. The size effect is controlled by the structuring element $\mathtt{SE}$. They are not inverses, owing to the non-linear character of the operators, however, they constitute an algebraic-adjunction [24]. The morphological *opening* $\gamma_{\mathtt{SE}}(\cdot)$ is an idempotent transformation defined by composition of erosion and dilation, i.e. $\gamma_{\mathtt{SE}}(\mathbf{I}) = \delta_{\mathtt{SE}}(\varepsilon_{\mathtt{SE}}(\mathbf{I}))$. Duality, the morphological *closing* $\phi_{\mathtt{SE}}(\cdot)$ is defined as the composition of dilation and erosion i.e. $\gamma_{\mathtt{SE}}(\mathbf{I}) = \varepsilon_{\mathtt{SE}}(\delta_{\mathtt{SE}}(\mathbf{I}))$. Their effect are also intuitive: Closing removes "holes" and thin cavities, and opening removes small object protuberances. Additionally, one of the most interesting properties for $(\gamma_{\mathtt{SE}}(\cdot), \phi_{\mathtt{SE}}(\cdot))$ is that they form a *Matheron semi-group* and they obey the *absorption law*[25], i.e.

$$\gamma_{\mathtt{SE}_1}(\gamma_{\mathtt{SE}_2}(\mathbf{I})) = \gamma_{\mathtt{SE}_1}(\mathbf{I}) \text{ and } \phi_{\mathtt{SE}_1}(\phi_{\mathtt{SE}_2}(\mathbf{I})) = \phi_{\mathtt{SE}_1}(\mathbf{I}) \tag{4}$$

if $\text{SE}_2 \subseteq \text{SE}_1$ in a family of scaled structuring elements. For the case of the family of concentric discrete disks, see [22] p.325. Fig. 2 shows the basic morphological transformations in a high resolution panchromatic image.

2.3. Morphological Reconstruction

It is often desirable to remove small objects from the image, while keeping larger objects totally intact. A morphological approach to answer this is the *morphological reconstruction*. For example, in the case of reconstruction by dilation, an image $\mathbf{M}$ is dilated in the usual way, but constrained so as to never grow outside the "control" image $\mathbf{M}$, called marker image. This operator is iterated until convergence is reached. Similarly, the reconstruction by erosion uses standard erosion and the dual constrain. We use the notation $(\varepsilon_{\text{B}}^{\infty}(\mathbf{M}, \mathbf{I}), \delta_{\text{B}}^{\infty}(\mathbf{M}, \mathbf{I}))$ for the couple erosion and dilation by reconstruction [22]. Table 1 gives the key notations used and the detailed definitions of morphological transformations required in the following formulation. Additionally, transformations by reconstruction are shown in Fig. 2 in a practical example.

2.4. Additive Morphological Decomposition

Let $\{\underline{\Phi}^i\}, i = 1 \ldots, m$ be a set of m anti-extensive transformations indexed for its scale i, such that:

$$\underline{\Phi}^m(\underline{\Phi}^{m-1}(\mathbf{I})) \leq \ldots \leq \underline{\Phi}^2(\underline{\Phi}^1(\mathbf{I})) \leq \underline{\Phi}^1(\mathbf{I}) \leq \mathbf{I} \tag{5}$$

Similarly, let $\{\overline{\Phi}^i\}$ be a set of m extensive transformations, where i is associated with the parameter of scale, such that:

$$\mathbf{I} \leq \overline{\Phi}(\mathbf{I}) \leq \overline{\Phi}^2(\overline{\Phi}^1(\mathbf{I})) \leq \ldots \leq \overline{\Phi}^m(\overline{\Phi}^{m-1}(\mathbf{I})). \tag{6}$$

Let us define the consecutive residuals from (5) and (6), as follows

$$\mathbf{R}_i^+ = \overline{\Phi}^i(\overline{\Phi}^{i-1}(\mathbf{I})) - \overline{\Phi}^{i-1}(\overline{\Phi}^{i-2}(\mathbf{I})) \geq 0 \tag{7}$$

$$\mathbf{R}_i^- = \underline{\Phi}^{i-1}(\underline{\Phi}^{i-2}(\mathbf{I})) - \underline{\Phi}^i(\underline{\Phi}^{i-1}(\mathbf{I})) \geq 0 \tag{8}$$

with $\overline{\Phi}^0 = \underline{\Phi}^0 = \mathbf{Id}$, the identity transform. From (5) and (6) we obtain,

$$\mathbf{I} = \overline{\Phi}^m(\overline{\Phi}^{m-1}(\mathbf{I})) - \sum_{i=1}^{m} \mathbf{R}_i^+ \tag{9}$$

and likewise,

$$\mathbf{I} = \underline{\Phi}^m(\underline{\Phi}^{m-1}(\mathbf{I})) + \sum_{i=1}^{m} \mathbf{R}_i^- \tag{10}$$

combining (9) and (10) provides us with an additive decomposition of the original image as follows

$$
\begin{aligned}
\mathbf{I} &= \underbrace{\frac{\overline{\Phi}^m(\overline{\Phi}^{m-1}(\mathbf{I})) + \underline{\Phi}^m(\underline{\Phi}^{m-1}(\mathbf{I}))}{2}}_{\mathbf{S}} + \sum_{i=1}^{m} \underbrace{\frac{(\mathbf{R}_i^- - \mathbf{R}_i^+)}{2}}_{\mathbf{R}_i} \\
&= \mathbf{S} + \sum_{i=1}^{m} \mathbf{R}_i = \mathbf{S} + \mathbf{R}.
\end{aligned}
$$

We now need to determine what kind of transformations $(\underline{\Phi}^i, \overline{\Phi}^i)$ should use to have interesting additive decomposition. Firstly, we consider the case of a family of morphological operators by reconstruction indexed by the size of the structuring element, i.e., $(\underline{\Phi}^i, \overline{\Phi}^i) = (\gamma_{\mathsf{SE}_i}^\infty, \phi_{\mathsf{SE}_i}^\infty)$ such that $\mathsf{SE}_i \subseteq \mathsf{SE}_j$ for all $i < j$. In this *additive morphological decomposition* (AMD), the couple $(\mathbf{R}_i^-, \mathbf{R}_i^+)$ is essentially composed by image structures associated with bright and dark objects in the image at different scales. The results for a spectral band of a hyperspectral image are shown in Fig.3(a)-(d). In this case, thanks to idempotence and absorption laws of the openings [22], i.e.

11

$\phi_{\text{SE}_i}^\infty(\phi_{\text{SE}_j}^\infty) = \phi_{\text{SE}_j}^\infty$ if $\text{SE}_i \subseteq \text{SE}_j$, the implementation of AMD does not require the composition of transformations associated with different scales. We remark that AMD has the same residues produced by the *differential morphological profile* (DMP)[26] but its representation has dimension $(m+1)d$ instead than $2md$, and in addition the AMD includes the term $\mathbf{S}$ associated with the image structure.

Secondly, the additive decomposition (11) can be applied even if the transformations do not satisfy the absorption law as in the previous case. That is the case of $(\underline{\Phi}^i, \overline{\Phi}^i) = (\underline{\lambda}_\sigma(\cdot), \overline{\lambda}_\sigma(\cdot))$ the pair of *Upper/Lower Leveling* where marker $\mathbf{M}$ is the minimum (maximum) between $G_\sigma(\mathbf{I})$ (the convolution of the original image with a Gaussian kernel with variance σ^2) and the original $\mathbf{I}$ [23]. See Table 1 to get details of the definition. We use the acronym ADL for *additive decomposition by leveling* to refer to this approach. In ADL, the multiscale effect is controlled by the value σ in the gaussian kernel associated to the marker. Results for ADL are shown in Fig.3(f)-(h) for a practical example. In remote sensing applications, the leveling transformation was advocated in [27]. Other kind of function can be considered instead of a gaussian convolution, for instance, subtraction of the original image and a constant as in [28]. The decomposition step extracts the most relevant parts from the original image $\mathbf{I}$, resulting to a cartoon-like image $\mathbf{S}$, formed by homogeneous regions with sharp boundaries. Expressed in a different way, image $\mathbf{S}$ retains all contrast and boundary information but loses all small scale pattern details. The correspondent residue $\mathbf{R} = \{\mathbf{R}_1, \ldots, \mathbf{R}_m\}$, constitutes a hierarchy of multiscale texture components. It should be remarked that the scope of this decomposition is not to find the optimum additive
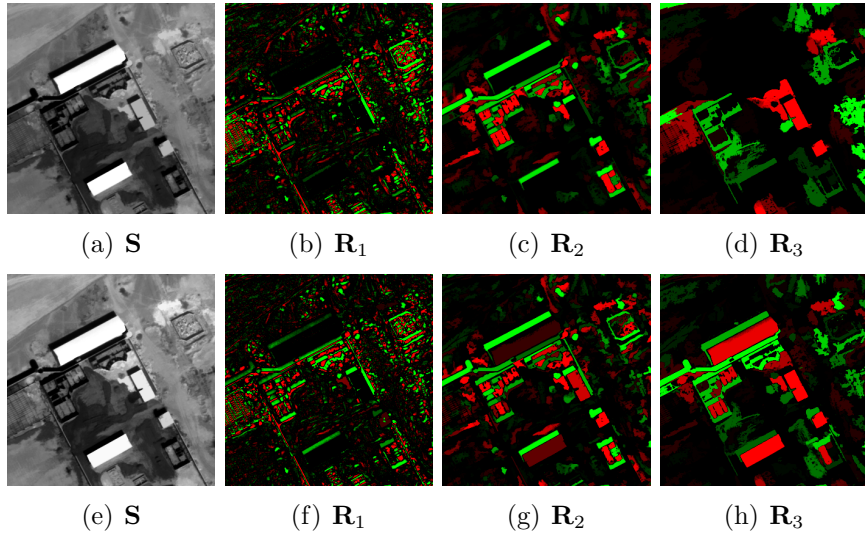
Figure 3: (a)-(d) Additive Morphological Decomposition (AMD) with SE's disks of radius 3,7,11. (e)-(h) Additive Decomposition by Leveling (ADL) with σ's 3,7,11. Note that residuals can be negatives(red) or positives(green).

decomposition as done in [29, 30], but it is a simple decomposition scheme where the spatial size of the texture can be interpreted. In the particular case of remote sensing imagery, several morphological decompositions have been proposed as summarized in Table 2. In section 4, we include experiments of performance of AMD, ADL and DMP in well-known hyperspectral images. Recently, difference of morphological attributes filters have been also introduced by [31, 16, 32]. The extension of additive morphological decomposition using these filters is straightforward but it is out of the scope of this paper. At this point, we have introduced an additive decomposition for a multivariate image. The next challenge is to find a way to handle the increase of the dimensionality.

Table 2: Different morphological decompositions for an image $\mathbf{I}$ of size $n_1 \times n_2 \times d$ in m levels.

Acronym	Additive	Trans.$(\underline{\Phi}/\overline{\Phi})$	Dimension
Differential Morphological Profile (DMP) [26]	No	$\gamma_{\mathrm{SE}}^{\infty}(\cdot)/\phi_{\mathrm{SE}}^{\infty}(\cdot)$	$2 \times m \times d$
Morphological Profile by Leveling (MPL) [33]	Yes	$\underline{\lambda}_\sigma(\cdot)/\overline{\lambda}_\sigma(\cdot)$	$2 \times m \times d$
Additive Morphological Decomposition (AMD)	Yes	$\gamma_{\mathrm{SE}}^{\infty}(\cdot)/\phi_{\mathrm{SE}}^{\infty}(\cdot)$	$(m+1) \times d$
Additive Decomposition by Leveling (ADL)	Yes	$\underline{\lambda}_\sigma(\cdot)/\overline{\lambda}_\sigma(\cdot)$	$(m+1) \times d$

3. Tensor Modeling

3.1. Introduction

The most popular dimensional reduction approach in HSI is PCA. However, PCA requires a preliminary data arrangement, i.e., the original hyperspectral image $\mathbf{I}$ of size $n_1 \times n_2 \times n_3$ is firstly vectorized into a matrix $\mathbf{X}$ of size $(n_1 n_2) \times n_3$ permitting the use of classic linear algebra approaches, but neglecting spatial rearrangement. The main shortcoming of this method is the assumption of separability between spatial processing and spectral processing. The dimensional reduction approach based on tensor decomposition considers the multivariate image $\mathbf{I}$ as a third order tensor $\mathcal{I}$ [34]. This kind of model based on tensor signal processing had been previously applied in HSI [35] [12]. Let us introduce the notation commonly used within tensor analysis literature, followed by the core of dimensional reduction problem and its solution. Let the tensor $\mathcal{I} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be an $n_1 \times n_2 \times n_3$ array containing the original information of image $\mathbf{I}$. Note that $n_3 = d$. Each index in the tensor is called *mode*: the first two are spatial and the third is spectral. Our

14

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.

3.2. Tensor Decomposition

A matrix $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}$ is a two-mode mathematical object that has two associated vector spaces, a row space and a column space. Singular Value Decomposition (SVD) orthogonalizes these two spaces and decomposes the matrix as $\mathbf{X} = \mathbf{U}_1 \boldsymbol{\Sigma} \mathbf{U}_2^T$, where $\mathbf{U}_1$ and $\mathbf{U}_2^T$ represent orthogonal column space, and $\boldsymbol{\Sigma}$ is a diagonal singular value matrix. In terms of the i-mode products, this SVD decomposition can be rewritten as $\mathbf{X} = \boldsymbol{\Sigma} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2$, where $\times_i$ is the i-mode product [36] [34]. Extension to a J-order tensor $\mathcal{I} \in \mathbb{R}^{n_1 \times n_2 \times n_3 \times \ldots \times n_J}$ was presented by [36] orthogonalizing J spaces and expressing the tensor as the J-mode product of J-orthogonal spaces

$$\mathcal{I} = \mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \ldots \times_J \mathbf{U}_J \tag{11}$$

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 has no a diagonal structure; rather, $\mathcal{C}$ is in general a full tensor. The core tensor governs the interaction between the mode matrices $\mathbf{U}_i$, for $i = 1, \ldots, J$. Mode matrix $\mathbf{U}_i$ contains the orthonormal vectors spanning the column space of the matrix $\mathbf{X}_i$ that results from the i-mode flattening of $\mathcal{I}$. Flattening, also known as matricization or unfolding, is the process of reordering the elements of an i-mode into a matrix [34]. This method is know as higher-order SVD (HOSVD) from the work of De Lathauwer, De Moor, and Vandewalle [36], who showed that the HOSVD is a convincing

15

generalization of the matrix SVD and discussed ways to efficiently compute the leading left singular vectors of $\mathbf{X}_i$. An excellent compendium about tensor decomposition is presented in [34]. The HOSVD is usually performed using Alternative Least Square algorithm used to jointly find i-mode matrices $\mathbf{U}_i$, but recently other approaches have been introduced [37]. In the case of three mode tensors $\mathcal{I}$, the objective of HOSVD is to select subspaces $\mathbf{U}_1, \mathbf{U}_2$ and $\mathbf{U}_3$ and the core tensor $\mathcal{C}$ such that the L_2-norm reconstruction error is minimized [34],

$$\min_{\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \mathcal{C}} E_1 = ||\mathcal{I} - \mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3||^2 \tag{12}$$

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 $\mathcal{E} = \mathcal{I} \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \times_3 \mathbf{U}_3^T$, and (12) can be written as:

$$\min_{\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3} E_1 = ||\mathcal{I}||^2 - ||\mathcal{E}||^2$$
$$\Leftrightarrow \max_{\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3} E_2 = ||\mathcal{E}||^2 \tag{13}$$

As it was presented by Huang [38], the equation (13) is equivalent to maximize:

$$\max_{\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3} E_2 = \text{Trace}(\mathbf{U}_1^T\mathbf{F}\mathbf{U}_1) =$$
$$\text{Trace}(\mathbf{U}_2^T\mathbf{G}\mathbf{U}_2) = \text{Trace}(\mathbf{U}_3^T\mathbf{H}\mathbf{U}_3) \tag{14}$$

where:

$$\mathbf{F}_{ii'} = \sum_{ll'}(\mathbf{X}_{(l)}\mathbf{U}_2\mathbf{U}_2^T\mathbf{X}_{(l')}^T)_{ii'}(\mathbf{U}_3\mathbf{U}_3^T)_{ll'}$$

$$\mathbf{G}_{jj'} = \sum_{ll'}(\mathbf{X}_{(l)}\mathbf{U}_1\mathbf{U}_1^T\mathbf{X}_{(l')}^T)_{jj'}(\mathbf{U}_3\mathbf{U}_3^T)_{ll'}$$

$$\mathbf{H}_{ll'} = \sum_{ii'jj'}\mathcal{I}_{ijl}\mathcal{I}_{i'j'l'}(\mathbf{U}_1\mathbf{U}_1^T)_{ii'}(\mathbf{U}_2\mathbf{U}_2^T)_{jj'}$$

16

Since $\mathbf{F}, \mathbf{G}, \mathbf{H}$ are semipositive definite, $||\mathcal{E}^2||$ is monotonically increasing, therefore HOSVD algorithm converges to a local optimal. Thus theoretically, the solutions HOSVD are not unique. That issue was already pointed in [34] and studied in detail for [39] in real life databases concluding that the convergence depends on the eigenvalue distribution for the matrix $\mathbf{F}$, $\mathbf{G}$ and $\mathbf{H}$. However, convergence problem in real HSIs has shown that cumulative values in the eigenvalues of $\mathbf{F}$ and $\mathbf{G}$ is a better criterion [40].

3.3. Tensor Principal Component Analysis (TPCA)

In high-dimensional images as HSI, it is of great interest to reduce the spectral dimension in order to exceed problems as "Curse of Dimensionality" in distance-based analysis or nonparametric analysis and "Hughes phenomenon" in linear classifiers. Commonly a pre-processing step consists in performing a PCA to reduce feature space. We present a tensor version for PCA based on [21]. It is a lower rank approximation, where classical PCA is a particular case, if no subspace reduction is performed in the modes associated with rows and columns. We assume that the hyperspectral image $\mathcal{I}$ is a zero-mean tensor in the flattening matrix related to the J-mode, i.e., $\overline{\mathbf{X}}_J = \mathbf{0}$. That is equivalent to subtracting the empirical mean vector from each column of the data matrix $\mathbf{X}$ as in PCA. In addition, the best lower rank tensor approximation of $\mathcal{I}$ [36], denoted by $\tilde{\mathcal{I}}$ is:

$$\tilde{\mathcal{I}} = \mathcal{I} \times_1 \mathbf{P}_1 \times_2 \mathbf{P}_2 \times_3 \ldots \times_J \mathbf{P}_J \qquad (15)$$

where $\mathbf{P}_i = \mathbf{U}_i \mathbf{U}_i^T$, and $\mathbf{U}_i$ is found by using expression (11). This representation allows to include noise filtering in the sense of SVD filtering [41] if only the largest eigenvectors are considered per mode. Thus, the tensor-PCA

(TPCA) approximation of the image $\mathcal{I}$ with parameters $(s_1, s_2, \ldots, s_{J-1}, k)$, $1 \le s_i \le n_i, \forall i = 1, \ldots, J-1$, is defined as follows:

$$\widetilde{\mathcal{I}} = \mathcal{I} \times_1 \widetilde{\mathbf{U}}_{s_1} \widetilde{\mathbf{U}}_{s_1}^T \times_2 \widetilde{\mathbf{U}}_{s_2} \widetilde{\mathbf{U}}_{s_2}^T \times_3 \ldots \times_J \widetilde{\mathbf{U}}_k^T \tag{16}$$

where k denotes the dimension in the J-mode, i.e., the number of components in the dimensional reduction. Additionally, s_i is the number of eigenvectors included in the filtering with respect to the i-mode and $\widetilde{\mathbf{U}}_{s_i}$ contains the s_i eigenvectors associated with the s_i largest eigenvalues holding of the unfolding matrix $\mathbf{X}_i$. We define the first k *tensor principal components* with parameters $(s_1, \ldots, s_{j-1})$ of $\mathcal{I}$ as the first k column of the matrix $\widetilde{\mathbf{U}}_k^T$ from (16). Summarizing, for a HSI $\mathbf{I}$, the tensor principal component analysis with parameters (s_1, s_2, k) is a transformation $\mathcal{F}(\mathbf{E}, \mathbb{R}^d) \to \mathcal{F}(\mathbf{E}, \mathbb{R}^k)$. The equivalence to the principal component analysis is presented in the next section.

3.4. Equivalence with PCA

In the case of a typical hyperspectral image represented as a tensor, $\mathcal{I}$ of size $n_1 \times n_2 \times n_3$, the expression (16) is particularized as

$$\widetilde{\mathcal{I}} = \mathcal{I} \times_1 \widetilde{\mathbf{U}}_{s_1} \widetilde{\mathbf{U}}_{s_1}^T \times_2 \widetilde{\mathbf{U}}_{s_2} \widetilde{\mathbf{U}}_{s_2}^T \times_3 \widetilde{\mathbf{U}}_k^T, \tag{17}$$

where $\widetilde{\mathbf{U}}_1$ and $\widetilde{\mathbf{U}}_2$ has the s_1 and s_2 largest eigenvectors associated of the unfolding matrix $\mathbf{X}_1$ and $\mathbf{X}_2$, respectively. Firstly, it is important to remark that if $s_1 = n_1$ and $s_2 = n_2$, $\widetilde{\mathbf{U}}_1 \widetilde{\mathbf{U}}_1^T = \mathbf{I}_{n_1 \times n_1}$ and $\widetilde{\mathbf{U}}_2 \widetilde{\mathbf{U}}_2^T = \mathbf{I}_{n_2 \times n_2}$ in that case, expression (17) becomes:

$$\widetilde{\mathcal{I}} = \mathcal{I} \times_3 \widetilde{\mathbf{U}}_k^T,$$

where $\widetilde{\mathbf{U}}_k$ contains the k-largest eigenvectors associated of the unfolding matrix $\mathbf{X}_3$, i.e., the matrix $\mathbf{X}$ of $(n_1 \times n_2)$ rows and n_3 columns that is the
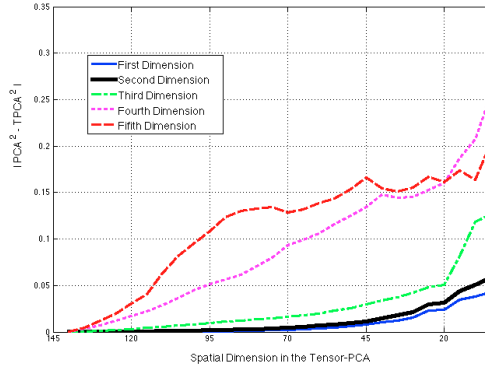
Figure 4: Experiment shows clearly that TPCA is equivalent to PCA. Experiments are presented with $s_1 = s_2$ in the interval $[n_1 = n_2, \ldots, 1]$, for Indian Pines HSI used in the experiments.

traditional unfolding of $\mathcal{I}$. Using the assumption that $\mathcal{I}$ is a zero-mean tensor in the third order, the eigenvectors associated to $\mathbf{X}$ are the same as the expression $(\mathbf{X} - \mu)^T (\mathbf{X} - \mu)$ which are the projections calculated by PCA. To illustrate this results in a practical example, we calculate the first five components in both PCA and TPCA.The absolute value of the differences between the squares of projections calculated by PCA and TPCA are illustrated in Fig. 4 for a real HSI (Indian Pine). It is easy to see that when the components number in the spatial dimension (s_1, s_2) for TPCA are equal to the image original dimension, the projections calculated by TPCA and PCA become similar. The differences become larger as soon as the spatial dimension reduces. Consequently, by this spatial dimension reduction with $s_1 < n_1$ and $s_2 < n_2$ we obtain through TPCA a spatial smoothing separately in rows and columns of the image, which is not produced in PCA.
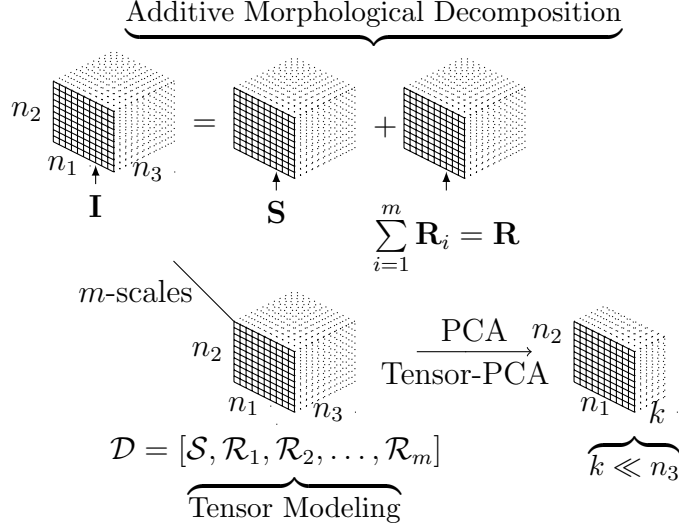
Figure 5: Illustration of dimensional reduction stage using additive morphological decomposition.

3.5. Modeling additive morphological decomposition with TPCA

The basic idea of our approach is summarized in Fig. 5. For a hyperspectral image $\mathcal{I}$ we find the additive decomposition in m levels, as it was introduced in Section 2, i.e. $\mathcal{I} = \mathcal{S} + \mathcal{R}_1 + \ldots + \mathcal{R}_m$. We regroup the whole decomposition in a four-order tensor $\mathcal{D} = [\mathcal{S}, \mathcal{R}_1, \ldots, \mathcal{R}_m]$. We apply the TPCA with parameters s_1, s_2, k_1, k_2 with $k = k_1 \times k_2$,

$$\widetilde{\mathcal{D}} = \mathcal{D} \times_1 \widetilde{\mathbf{U}}_{s_1} \widetilde{\mathbf{U}}_{s_1}^T \times_2 \widetilde{\mathbf{U}}_{s_2} \widetilde{\mathbf{U}}_{s_2}^T \times_3 \widetilde{\mathbf{U}}_{k_1}^T \times_4 \widetilde{\mathbf{U}}_{k_2}^T \tag{18}$$

where $\widetilde{\mathcal{D}}$ is a tensor of size $n_1 \times n_2 \times k_1 \times k_2$. The parameters s_1 and s_2 are associated with the spatial filtering in the sense of SVD filtering in the rows and columns space, k_1 is the reduction in the spectral space and k_2 is corresponding to the scale decomposition. The connection to precedent subsection is established for TPCA in four-order tensors, to traditional PCA in the case

of $k_1 = n_1$, $k_2 = n_2$ and no scale decomposition. In summary, the proposed workflow yields a reduced feature space as PCA where the spatial information included in the morphological decomposition is relevant. Additionally, spatial filtering can be included through the tensor decomposition.

4. Experiments

In this section, we present the experimental results obtained in our analysis. Firstly, we overview the characteristics of the data used in the experimental setup. After that, several experiments are presented in order to compare the effectiveness of the proposed additive decompositions and tensor dimensional reduction. The application of the introduced approach requires a morphological transformation for vector images. Unfortunately, the generalization to multivariate image of mathematical morphology operators is still an open problem [42, 43]. We present the results of our approach applying the transformations marginally, i.e. for each channel $i = \{1, 2, \ldots, d\}$ independently. For instance, the dilation of the d-variate image $\mathbf{I}$ is given by $\delta_{\text{SE}}(\mathbf{I})(\mathbf{x}) = [\delta_{\text{SE}}(\mathsf{x}_1), \delta_{\text{SE}}(\mathsf{x}_2), \ldots, \delta_{\text{SE}}(\mathsf{x}_d)]$, where $\mathbf{x} = [\mathsf{x}_1, \mathsf{x}_2, \ldots, \mathsf{x}_d]$. And similarly, for all the other operators summarized in Table 1.

4.1. Data Description and Experimental Setup

In order to further evaluate and compare the proposed algorithm with other state-of-the-art approaches for spatial-spectral classification, we use two real hyperspectral images.

1. Airborne Visible/Infrared imaging spectrometer hyperspectral image (AVIRIS) Indian Pines Scene. The AVIRIS sensor generates 220 bands across the spectral range from 0.2 to 2.4 μm. In the experiments, the number of bands is
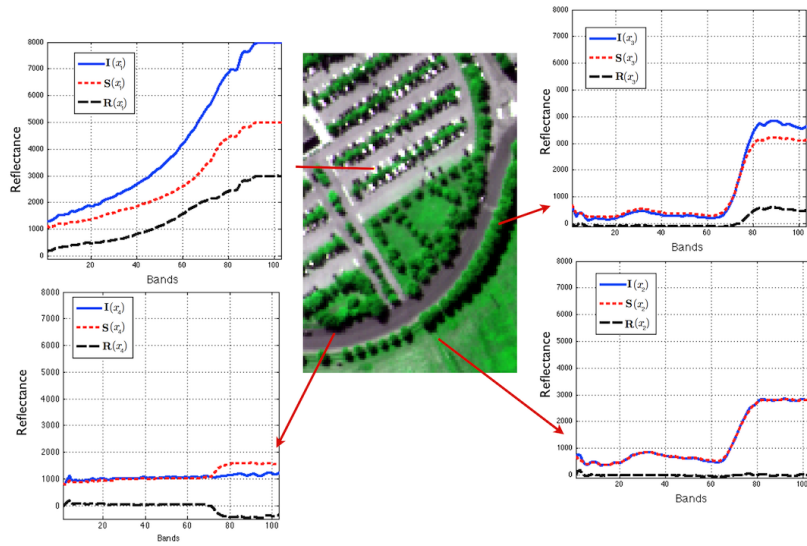
21

Figure 6: AMD using $\{\mathtt{SE}_1, \mathtt{SE}_2\}$ a disk of diameter 2 and 4 is shown for four pixels in the ROSIS Pavia University HSI. First pixel (Left-Up) is a one-pixel set object. Its residual component ($\mathbf{R}$) is large in comparison to the structure one $\mathbf{S}$, i.e., the pixel is very different from its neighbors. Second pixel (Right-Up) is a tree. The residual component is important only in the spectrum associated with vegetation. Third pixel (Left-Down) is a shadow-tree. Residual is negative and significative only in the vegetation section of the spectrum. Fourth pixel (Right-Down) is a pixel in a homogeneous zone, i.e., it has no texture component.

reduced to 200 by removing 20 water absorption bands. The image has a spatial resolution of 20 meters per pixel and a spatial dimension of 145×145 pixels.This image is a classical benchmark to validate the accuracy of HSI classification algorithms and constitutes a challenging problem due to the significant presence of mixed pixels in all available classes and also because of the unbalanced number of available labeled pixels per class. We follow the experiment proposed in [44] to analyze HSI classification in a very difficult situation. From the 16 different land-cover classes available in the original ground-truth, seven were discarded due to an insufficient number of training samples. The finally selected classes were: Corn-no till (1434), Corn-min till (834), Grass/Pasture (497), Grass/Trees (747), Hay-wind-rowed (489), Soybean-no till (968), Soybean-min till (2468), Soybean-clean till (614), and Woods (1294). Summarizing, the ground-truth contains nine classes, as seen in Fig. 8(a). In the experiment, we test the introduced method in different ill-posed scenarios where only five pixels are used as training samples per class. Our results are compared with those reported by [44, 45].

2. University of Pavia, is an urban image acquired by Reflective Optics System Imaging Spectrometer (ROSIS). The ROSIS sensor generates 115 spectral bands ranging from 0.43 to 0.86 μm with a band of 4nm and has a spatial resolution of 1.3-meter per pixel. The image consists of 610×340 pixels, each having 103 bands with 12 most noisy bands removed. There are nine ground-truth classes of interest, as shown in Fig. 10(a). Nine thematic land-cover classes were identified in the university campus: Trees, Asphalt, Bitumen, Gravel, Metal sheets, Shadows, Self-blocking Bricks, Meadows, and Bare soil. For this data set, a total of 3921 and 42776 pixels were available as
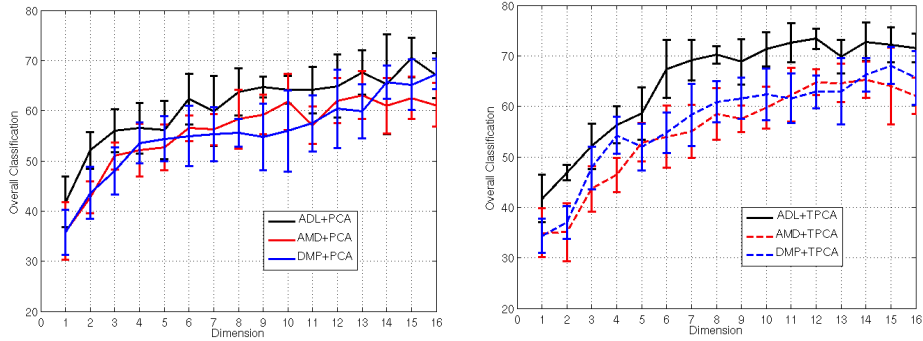
(a) Overall accuracy of morphological decompositions as a function of the number of components in PCA.

(b) Overall accuracy of morphological decompositions as a function of the number of component in TPCA.

Figure 7: Indian Pine first scenario. Only five pixels per class are selected for the training set. The results show the average and standard deviation in 25 repetitions.

training and test sets, respectively, as seen in Fig. 10(b). Proposed approach results are compared with those obtained from [6, 46, 14, 15, 17].

4.2. Classification

Support Vector Machine (SVM) have shown promising results in terms of prediction accuracy in HSI [48]. A pixelwise classification was performed using the multiclass "one versus one" SVM classifier in the correspondent dimension produced by PCA and TPCA in the morphological decompositions considered in Section 2. The reduced space is scaled to the range of $[0, 1]$ and SVM is trained with Gaussian kernel, and parameters tuned in the range $\{-1, \ldots, 3\}$ for the regularization parameter and $\{-4, \ldots, 1\}$ for the Gaussian kernel parameter by using cross-validation. The following measures of accuracy were used: Overall accuracy (OA) is the percentage of correctly classified pixels, average accuracy (AA) is the mean of class-specific accuracies, i.e., the percentage of correctly classified pixels for each class, and

24

(a) Groundtruth   (b) PCA+SVM   (c) DMP+PCA+SVM   (d) DMP+TPCA+SVM

(e) AMD+PCA+SVM   (f) ADL+PCA+SVM   (g) AMD+TPCA+SVM   (h) ADL+TPCA+SVM
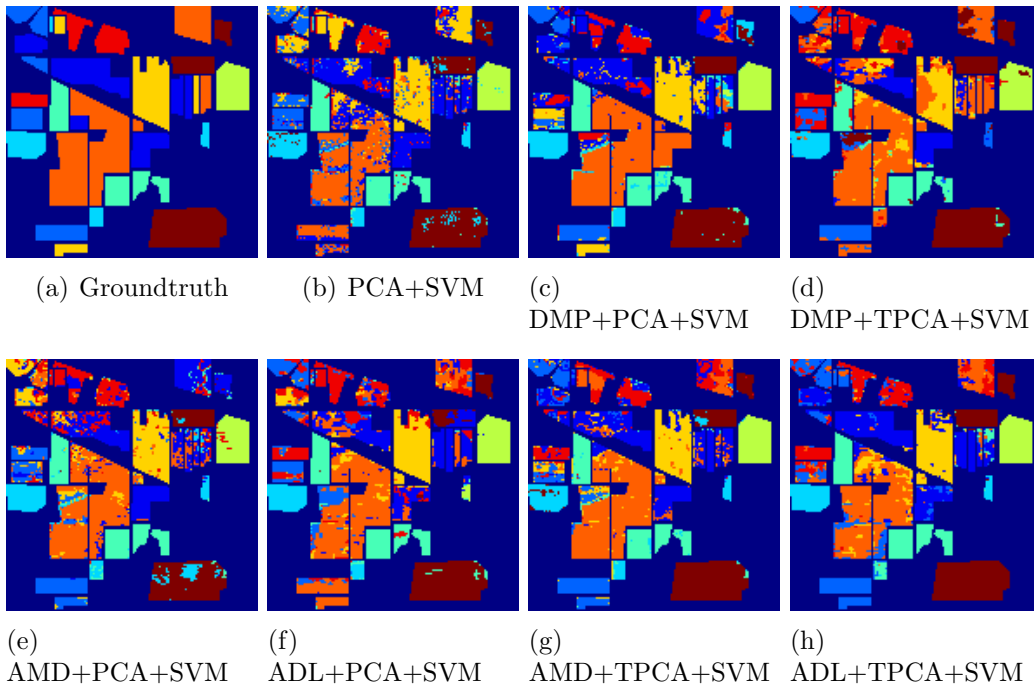
Figure 8: Classification maps for the Indian Pines HSI using different approaches. Only five training pixels in nine classes are considered. The classification map is the best result in 25 random repetitions.

Table 3: Classification accuracy for Indian Pines HSI. Only five samples per class are included in the training set. For proposed methods in parentheses the number of components.

| Method	OA %	κ
Spatial Kernel [44]		
Spectral +SVM	45.79	0.43
Spectral +Graph	48.96	0.46
Summation + SVM	48.88	0.46
Summation + Graph	52.27	0.49
Cross +SVM	61.75	0.60
Cross+Graph	66.04	0.64
Random Field+Multinomial Logistic Regression [45]		
SS	72.62	N/A
LORSAL	58.10	N/A
Differential Morphological Profile		
DMP + PCA(16)+SVM	67.33	0.62
DMP + TPCA(15)+SVM	68.10	0.63
Additive Morphological Decomposition		
ADL + PCA(15)+SVM	70.57	0.66
ADL + TPCA(12)+SVM	73.39	0.69
AMD + PCA(13)+SVM	63.13	0.57
AMD + TPCA(14)+SVM	65.31	0.60
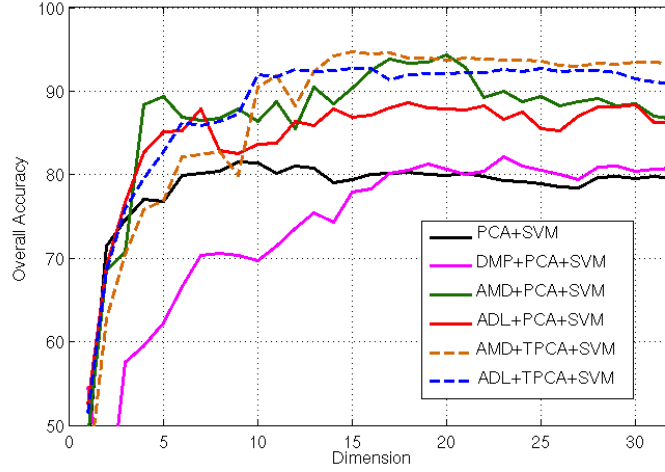
Figure 9: Behavior of the overall accuracy in the Pavia University dataset for different morphological decompositions and dimensional reduction approach.

26

Table 4: Overall and average classification accuracies and κ statistics obtained after comparing the proposed framework with other spatial-spectral classifiers for Pavia University data set. The number of components is reported in parentheses.

Method	Overall Accuracy	Average Accuracy	κ
ECHO [6]	87.58	92.16	.8390
SVMMSF+MV [46]	91.08	94.76	.8830
SSK [17]	86.11	91.98	.8235
LORSAL-MLL [14]	85.57	92.54	.8180
MLRsubMLL [15]	94.10	93.45	.9224
PCA+SVM(9)	81.57	87.39	.7662
DMP+PCA+SVM(23)	82.12	83.54	.7675
AMD+PCA+SVM(20)	94.32	94.64	.9253
ADL+PCA+SVM (18)	88.67	91.39	.8508
AMD+TPCA+SVM(15)	94.70	94.51	.9301
ADL+TPCA+SVM(15)	92.72	91.77	.9043

Table 5: Overall and average classification accuracies and κ statistics obtained after comparing the proposed framework with other spatial-spectral classifiers for Pavia University data set. Classification task is performed by SVM. For proposed methods in parentheses the number of components.

Class	SVM	EMP[47]	SSK [17]	PCA(9)	DMP + PCA(23)	AMD + PCA(20)	ADL + PCA(18)	AMD + TPCA(15)	ADL+ TPCA(15)
Asphalt	80.64	93.33	84.36	83.52	88.30	**96.56**	92.75	93.45	93.30
Meadow	68.47	73.40	78.52	74.89	84.06	94.01	89.72	**95.77**	94.07
Gravel	73.80	52.45	**84.80**	70.32	55.03	84.52	88.71	82.80	65.41
Tree	97.49	**99.31**	96.87	98.07	84.30	98.56	97.91	98.86	98.56
Metal sheet	99.49	99.48	99.88	99.48	99.78	99.48	**100**	99.48	99.55
Bare soil	94.83	61.90	**95.61**	82.86	57.49	88.86	59.14	89.72	87.69
Bitumen	91.50	97.67	95.56	90.30	99.02	**99.17**	96.84	99.02	97.44
Brick	91.88	95.17	95.44	88.78	93.78	98.23	**99.13**	98.86	98.56
Shadow	97.04	92.29	97.78	98.21	90.07	92.40	**98.31**	92.61	91.34
Overall	80.13	79.83	86.11	81.57	82.12	94.32	88.67	**94.70**	92.72
Average	88.33	85.00	91.98	87.39	83.54	**94.64**	91.39	94.51	91.77
κ	.7519	.7415	.8235	.7662	.7675	.9253	.8508	**.9301**	.9043

(a) Training Set　　(b) Test Set　　(c) PCA (81.57%)　　(d)　　DMP+PCA (82.12%)

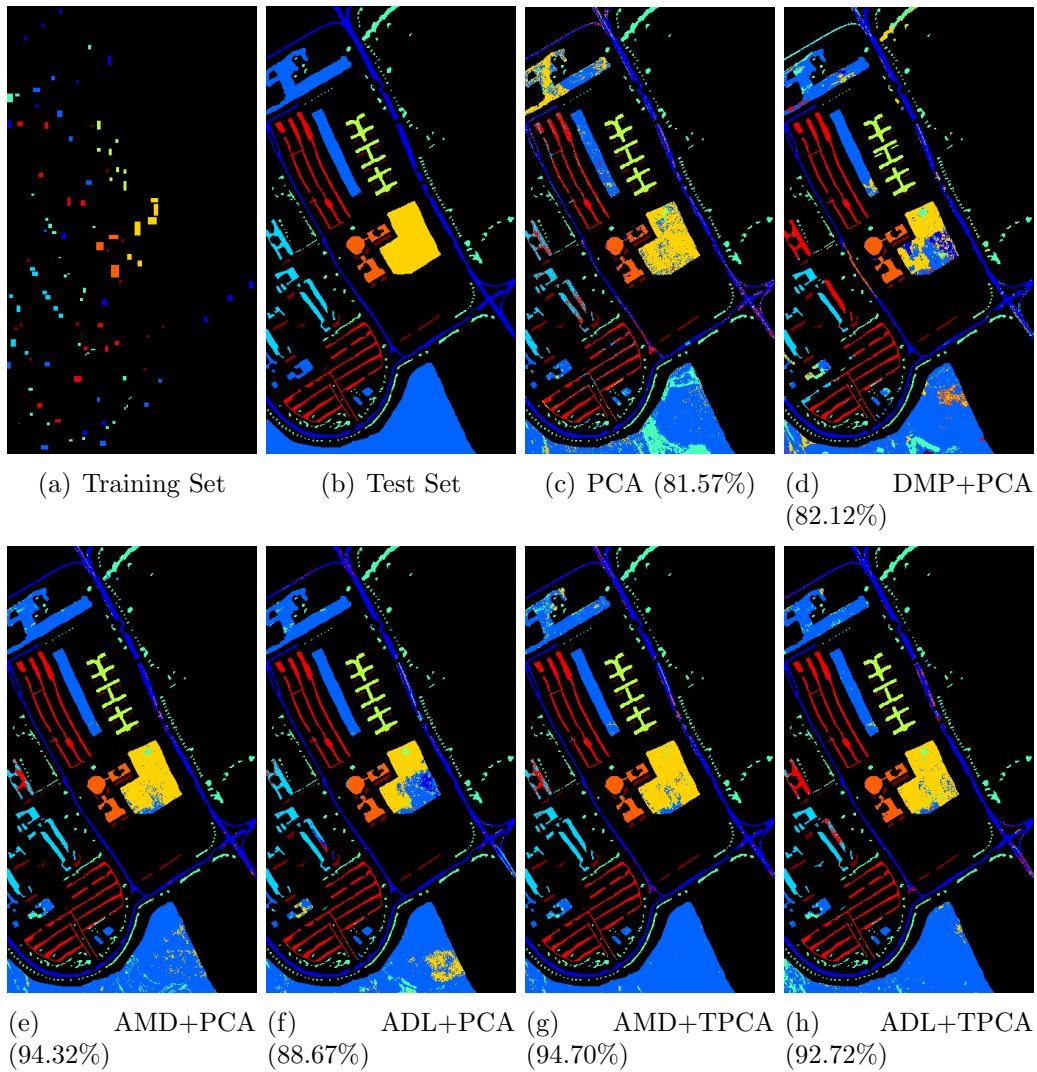(e)　　AMD+PCA (94.32%)　　(f)　　ADL+PCA (88.67%)　　(g)　　AMD+TPCA (94.70%)　　(h)　　ADL+TPCA (92.72%)

Figure 10: Classification maps obtained by the different tested methods for Pavia University data set (Overall accuracies are reported in parentheses)

kappa coefficient (κ) is the percentage of agreement, i.e., correctly classified pixels, corrected by the number of agreements that would be expected purely by chance. In order to compare the performance of the proposed technique to include the spatial information into a classification task, we have also included results of the previously proposed methods: ECHO spatial classifier [6], Spatial kernels [44], Markov Random Field [45], Bayesian approach to active learning [14], subspace multinomial logistic regression [15] and classification followed by post-processing [46]. Morphological approaches to analysis HSI are also included [17] and [47]. Additionally, the objective in the experiment is to complete the comparative analysis in feature spaces of different dimension size produced by the proposed additive morphological decomposition and extracted by PCA and TPCA. The parameters in the TPCA (s_1, s_2) have been set to avoid the convergence problem in the tensor decomposition, as it has been suggested in [40].

4.3. Results and discussion

Firstly, to illustrate the motivation behind this work and to clarify the concept of "additive decomposition" for multivariate images, Fig. 6 visualizes the result of ADL for four types of pixels for a HSI. Spectra are decomposed accordingly to their relationship in the spatial neighborhood. Simple interpretations can be done regarding the structure+texture decomposition in the spectrum range (See caption in Fig. 6). Secondly, a quantitive comparison is carried out in two real HSI. The Indian Pines experiment shows the importance to incorporate spatial information previously to do feature reduction. We point out that the experiment was set in a extremely difficult case (only 5 samples per class as training set) to illustrate the effect of

29

spatial information, additive decomposition and tensor-PCA in classification of high dimensional images. In other words, in this scenario the influence of the classifier is limited and allow us to compare the different techniques of dimensionality reduction. To reliably evaluate the performance of the proposed method, the results were averaged over 25 different randomly selected training (of size five) for a number of feature yield by PCA and TPCA in the range of $[1, \ldots, 16]$. For a best understanding of this comparison, the experiment considered in Fig. 7 illustrates the performance of ADL, AMD and DMP in both dimensional reduction approaches (PCA and TPCA). In the broader range of results, ADL exhibits a higher classification rate for the dimensionality size considered in this example. ADL led to the best classification results, as it can also be seen from Table 3. On the other hand, this experiment confirms our intuition that the inclusion of a spatial prior can significantly improve the classification results provided by using only spectral information. Fig. 8 shows the thematic classification maps for the pixel wise SVM and the spectral-spatial classification by morphological decomposition after the dimensional reduction step. Our approach involving morphological information is clearly better that its spectral equivalent. Additionally, ADL and tensor reduction has the best performance with more than 73% in overall classification in this very difficult scenario.

Additionally, to get fair comparison with others approaches, we analyze the Pavia University HSI in Fig. 9. This is the standard scenario to do comparisons in HSI. We can observe how additive decomposition have a better performance than classical approaches. However, ADL and AMP produce equivalent results. We note that in this experiments the training set is fixed,

so the result depend of this selection, i.e., we can not argument that in the statistical sense than the difference are significative or not. Comparison with state-of-the-art approaches are presented in Table 3 (summarizing Fig. 9), where we can see: a) The performance of tensor-PCA is in general better (in most of the dimensions) than PCA. b) Classification from AMD are better than original image or DMP in all the considered cases. c) AMD+TPCA gives the best result of the reported approaches in this supervised scenario. In turn, it can also be seen in Fig. 9 that the inclusion of the tensor structure provides much higher classification accuracies than those reported for PCA. From Table 4, it can be observed that the proposed additive decompositions (AMD and ADL) obtain good result when compared with other methods. Tensor structure improves the classification accuracy and yields a representation with better separability in lower dimension, for instance, from 20 features for ADL+PCA+SVM (OA-88.67%) to 15 features for ADL+TPCA+SVM (OA-92.72%). Table 5 gives the class-specific accuracies of the best pixel wise classification in the projected space induced by the correspondent dimensional reduction algorithm. The performances of the proposed additive decomposition are compared with those obtained by other morphological based approaches [47, 17]. The AMD+TPCA yields the best OA and kappa coefficient. Most of the best class-specific accuracies are obtained by AMD with PCA or TPCA. This approach significantly outperforms other classification approaches. For illustrative purposes, the effectiveness of the proposed framework with Pavia University HSI is further shown in Fig.10 and the classification maps obtained are displayed along with their associated OA scores. Fig. 10 shows the classification maps of PCA, DMP+PCA,

and additive decompositions reduced by PCA and TPCA. As it can be seen, the ADM map contains much more homogeneous spatial structures when compared with the PCA map.

To conclude, in real applications the analysis of high-dimensional images by using morphological decomposition is related to the prior knowledge of object of interest. The problem that we address in this paper is a multi-class supervised problem, where prior information about 'shape/size" of the object is not available. Accordingly, multi-scale decompositions based on ADM and DMP are obtained by using "isotropic disk" for different sizes, as in [17]. Analyzing the difficulty to select the parameters, the unique parameter required in ADL is a set of σ associated with the scale of the objects of interest, which is easier to set than a "size/shape" parameter necessary in AMD.

5. Conclusions

The paper proposed a framework integrating structural/spatial information in unsupervised dimensional exploration and feature extraction for hyperspectral images. Additive morphological decomposition is a nonlinear representation that favorably incorporate the spatial information in dimension reduction approaches. Tensor modeling 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 show how the tensor approach

32

incorporates more usefully spatial information in dimensional reduction stage in comparison with its matrix equivalent version.

[1] Z. Pan, G. Healey, M. Prasad, B. Tromberg, Face recognition in hyper-spectral images, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (12) (2003) 1552–1560.

[2] A. Gendrin, N. Mangold, J. Bibring, Y. Langevin, B. Gondet, F. Poulet, G. Bonello, C. Quantin, J. Mustard, R. Arvidson, et al., Sulfates in Martian Layered Terrains: The OMEGA/Mars Express View (2005).

[3] R. A. Schultz, T. Nielsen, J. R. Zavaleta, R. Ruch, R. Wyatt, H. R. Garner, Hyperspectral imaging: a novel approach for microscopic analysis., Cytometry 43 (4) (2001) 239–247.

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

[5] Q. Du, C.-I. Chang, A linear constrained distance-based discriminant analysis for hyperspectral image classification, Pattern Recognition 34 (3) (2001) 361–373.

[6] D. Landgrebe, Signal Theory Methods in Multispectral Remote Sensing, John Wiley and Sons, 2003.

[7] G. Camps-Valls, L. Bruzzone, Kernel-based methods for hyperspectral image classification, IEEE Transactions on Geoscience and Remote Sensing 43 (6) (2005) 1351 − 1362.

[8] L. Bruzzone, L. Carlin, A multilevel context-based system for classification of very high spatial resolution images, IEEE Transactions on Geoscience and Remote Sensing 44 (9) (2006) 2587–2600.

[9] J. Duarte-Carvajalino, G. Sapiro, M. Velez-Reyes, P. E. Castillo, Multiscale representation and segmentation of hyperspectral imagery using geometric partial differential equations and algebraic multigrid methods, IEEE Transactions on Geoscience and Remote Sensing 46 (8) (2008) 2418–2434.

[10] Y. Tarabalka, J. Chanussot, J. Benediktsson, Segmentation and classification of hyperspectral images using watershed transform, Pattern Recognition 43 (7) (2010) 2367–2379.

[11] J. Bolton, P. Gader, Random set framework for context-based classification with hyperspectral imagery, IEEE Transactions on Geoscience and Remote Sensing 47 (2009) 3810–3821.

[12] S. Bourennane, C. Fossati, A. Cailly, Improvement of classification for hyperspectral images based on tensor modeling, Geoscience and Remote Sensing Letters, IEEE 7 (4) (2010) 801 –805.

[13] Y. Wang, R. Niu, X. Yu, Anisotropic diffusion for hyperspectral imagery enhancement, Sensors Journal, IEEE 10 (3) (2010) 469 –477.

[14] J. Li, J. M. Bioucas-Dias, A. Plaza, Hyperspectral image segmentation using a new bayesian approach with active learning, IEEE Transactions on Geoscience and Remote Sensing 49 (10) (2011) 3947–3960.

[15] J. Li, J. M. Bioucas-Dias, A. Plaza, Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and markov random fields, IEEE Transactions on Geoscience and Remote Sensing 50 (3) (2011) 809–823.

[16] M. Dalla-Mura, J. Benediktsson, L. Bruzzone, Self-dual attribute profiles for the analysis of remote sensing images, in: ISMM, Vol. 6671 of LNCS, Springer Berlin / Heidelberg, 2011, pp. 320–330.

[17] M. Fauvel, J. Chanussot, J. Benediktsson, A spatial-spectral kernel-based approach for the classification of remote-sensing image, Pattern Recognition 45 (2012) 381–392.

[18] P. Perona, J. Malik, Scale-space and edge detection using anisotropic diffusion, IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (1990) 629–639.

[19] M. Pesaresi, J. Benediktsson, A new approach for the morphological segmentation of high-resolution satellite imagery, IEEE Transactions on Geoscience and Remote Sensing 39 (2) (2001) 309–320.

[20] Y. Tarabalka, J. A. Benediktsson, J. Chanussot, J. C. Tilton, Multiple spectral-spatial classification approach for hyperspectral data, IEEE Transactions on Geoscience and Remote Sensing 48 (11) (2010) 4122–4132.

[21] N. Renard, S. Bourennane, Improvement of target detection methods by multiway filtering, IEEE Transactions on Geoscience and Remote Sensing 46 (8) (2008) 2407–2417.

[22] P. Soille, Morphological Image Analysis Dos, Springer-Verlag, 2003.

[23] F. Meyer, The levelings, in: ISMM, Kluwer Academic Publishers, Norwell, MA, USA, 1998, pp. 199–206.

[24] H. J. A. M. Heijmans, Morphological Image Operators, Vol. 10, Academic Press, 1994.

[25] J. Serra, Image Analysis and Mathematical Morphology, Academic Press, Inc., Orlando, FL, USA, 1983.

[26] J. Benediktsson, M. Pesaresi, K. Arnason, Classification and feature extraction for remote sensing images from urban areas based on morphological transformations, IEEE Transactions on Geoscience and Remote Sensing 41 (9) (2003) 1940–1949.

[27] P. Soille, M. Pesaresi, Advances in mathematical morphology applied to geoscience and remote sensing, IEEE Transactions on Geoscience and Remote Sensing 40 (9) (2002) 2042–2055.

[28] F. Meyer, Leveling and flatzone morphology, in: International Conference on Pattern Recognition, 2010, pp. 1570–1574.

[29] A. Buades, T. Le, J. Morel, L. Vese, Fast cartoon + texture image filters, IEEE Transaction on Image Processing 19 (18) (2010) 1978–1986.

[30] L. Vese, S. Osher, Modeling textures with total variation minimization and oscillating patterns in image processing, Journal of Scientific Computing 19 (1) (2003) 553–572.

[31] M. Dalla-Mura, J. Benediktsson, B. Waske, L. Bruzzone, Morphological attribute profiles for the analysis of very high resolution images, IEEE Transactions on Geoscience and Remote Sensing 48 (10) (2010) 3747–3762.

[32] G. Ouzounis, M. Pesaresi, P. Soille, Differential area profiles: Decomposition properties and efficient computation, IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (8) (2012) 1533–1548.

[33] S. Velasco-Forero, J. Angulo, Morphological scale-space for hyperspectral images and dimensionality exploration using tensor modeling, in: Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, 2009. WHISPERS '09. First Workshop on, 2009, pp. 1 –4.

[34] T. Kolda, B. Bader, Tensor decompositions and applications, SIAM Review 51 (3) (2009) 455–500.

[35] D. Muti, S. Bourennane, J. Marot, Lower-rank tensor approximation and multiway filtering, SIAM J. Matrix Anal. Appl. 30 (3) (2008) 1172–1204.

[36] L. Lathauwer, B. Moor, J.Vandewalle, A multilinear singular value decomposition, SIAM Journal Matrix Anal. Appl. 21 (4) (2000) 1253–1278.

[37] L. Eldén, 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 (2009) 248–271.

[38] H. Huang, C. Ding, D. Luo, T. Li, Simultaneous tensor subspace selection and clustering: the equivalence of high order svd and k-means clustering, KDD '08: 14th ACM SIGKDD 200 (2008) 327–335.

[39] D. Luo, C. H. Q. Ding, H. Huang, Are tensor decomposition solutions unique? on the global convergence hosvd and parafac algorithms, in: PAKDD (1), 2011, pp. 148–159.

[40] S.Velasco-Forero, J. Angulo, Parameters selection of morphological scale-space decomposition for hyperspectral images using tensor modeling, in: Proceeding of the SPIE: 7695, Vol. 7695, 2010, pp. 1–12.

[41] H. Andrews, C. L. Patterson, Singular value decompositions and digital image processing, IEEE Trans. Acoust., Speech, Signal Processing 24 (1) (1976) 26–53.

[42] E. Aptoula, S. Lefèvre, A comparative study on multivariate mathematical morphology, Pattern Recognition 40 (11) (2007) 2914–2929.

[43] S. Velasco-Forero, J.Angulo, Supervised ordering in $\mathbb{R}^p$: Application to morphological processing of hyperspectral images, IEEE Transaction on Image Processing 20 (11) (2011) 3301–3308.

[44] G. Camps-Valls, T. Bandos, D. Zhou, Semi-supervised graph-based hyperspectral image classification, IEEE Transactions on Geoscience and Remote Sensing 53 (10) (2007) 3044–3054.

[45] J. Li, J. M. Bioucas-Dias, A. Plaza, Semi-supervised hyperspectral image classification based on markov random field and sparse multinomial logistic regression, Proceeding of IGARSS 1 (2009) 1–4.

[46] Y. Tarabalka, M. Fauvel, J. Chanussot, J. Benediktsson, Svm- and mrf-based method for accurate classification of hyperspectral images, IEEE Geosci. Remote Sens. Lett. 7 (4) (2010) 736–740.

[47] J.A.Benediktsson, J. Palmason, J.Sveinsson, Classification of hyperspectral data from urban areas based on extended morphological profiles, IEEE Transactions on Geoscience and Remote Sensing 43 (3) (2005) 480–491.

[48] F. Melgani, L. Bruzzone, Classification of hyperspectral remote sensing images with support vector machines, IEEE Transactions on Geoscience and Remote Sensing 42 (8) (2004) 1778–1790.