Article

Morphological Principal Component Analysis for Hyperspectral Image Analysis †

Gianni Franchi *,‡ and Jesús Angulo ‡

MINES ParisTech, PSL-Research University, CMM-Centre de Morphologie Mathématique,
35 rue Saint-Honoré , 77305 Fontainebleau, France; jesus.angulo@mines-paristech.fr

* Correspondence: gianni.franchi@mines-paristech.fr; Tel.: +331-6469-4706

† This paper is an extended version of our paper published in G. Franchi, J. Angulo, Comparative Study on Morphological Principal Component Analysis of Hyperspectral Images. In Proceedings of the 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS'14), Lausannne, Switzerland, 24–27 June 2014.

‡ These authors contributed equally to this work.

Abstract: This article deals with the issue of reducing the spectral dimension of a hyperspectral image using principal component analysis (PCA). To perform this dimensionality reduction, we propose the addition of spatial information in order to improve the features that are extracted. Several approaches proposed to add spatial information are discussed in this article. They are based on mathematical morphology operators. These morphological operators are the area opening/closing, granulometries and grey-scale distance function. We name the proposed family of techniques the Morphological Principal Component Analysis (MorphPCA). Present approaches provide new feature spaces able to jointly handle the spatial and spectral information of hyperspectral images. They are computationally simple since the key element is the computation of an empirical covariance matrix which integrates simultaneously both spatial and spectral information. The performance of the different feature spaces is assessed for different tasks in order to prove their practical interest.

Keywords: spatial machine learning; hyperspectral images; dimensionality reduction; mathematical morphology

1. Introduction

Hyperspectral images allow us to reconstruct the spectral profiles of objects imaged by the acquisition of several tens or hundreds of narrow spectral bands. Conventionally, in many applications hyperspectral images are reduced in the spectral dimension before any processing. Most hyperspectral image reduction methods are linear and are not concerned with the multiple sources of nonlinearity present in this kind of image [1]. Nonlinear reduction techniques are nowadays widely used on data reduction, and some of them have been used for hyperspectral images [2]. Nevertheless, most of these techniques present some disadvantages [3] in comparison to the canonical linear principal component analysis (PCA). That is the rationale behind our choice of PCA as starting point. In particular, one major drawback of those nonlinear techniques is that they are computationally too complex in comparison to PCA. Hence most of the time, they cannot be applied on real full resolution images. Another common disadvantage of both classical linear and nonlinear dimensionality reduction techniques is that they consider a hyperspectral image as a set of vectors. They are appropriate when the data do not present useful spatial information, and therefore they are not totally adapted to images.

As mentioned above, dimensionality reduction in hyperspectral images is usually considered as a preprocessing step for supervised pixel classification as well as for other hyperspectral image tasks such as unmixing, target detection, *etc*. Hence, our goal is to incorporate spatial information into the dimensionality reduction (DR).

The contribution of our approach can be summarized as follows. We propose to add spatial information on the estimation of the covariance matrix used for PCA computation. This is done by means of morphological image representations, which involve a nonlinear embedding of the original hyperspectral image into a morphological feature space.

Many previous works have considered how to introduce spatial information into hyperspectral dimensionality image reduction. We can divide these techniques into different fields. The first family of techniques is close to our paradigm since they are based on mathematical morphology [4–8]. Other approaches are founded on Markov random field image representation such as [9,10]. Another family of techniques uses kernel methods, where kernels for spatial information and kernels for spectral information are combined together [11–14]. Techniques based on tensor representation of hyperspectral images [15,16] have been successfully considered. Finally, wavelet representation and image extracted features have also been used to add spatial information [17].

The rest of the paper is organized as follows. Section 2 provides a remind on the mathematical morphology multi-scale representation tools used in our approach. Section 3 introduces in detail our approach named morphological principal component analysis (MorphPCA). In order to justify our framework, a summary of the classical theory underlying the standard PCA is provided as well as the notion of Pearson image correlation. Then, the four variants of MorphPCA are discussed, including an analysis of their corresponding covariance matrix meaning. The application of MorphPCA to hyperspectral dimensionality image reduction is considered in Section 4. That involves an assessment of the different variants according to different criteria. For some of the criteria, new techniques to evaluate the quality of dimensionality reduction techniques on image processing are introduced. Techniques arising from manifold learning are also considered in the comparaison. Finally, Section 5 closes the paper with the conclusions.

We note that this paper is an extended and improved version of the conference contribution [18]. Besides a more detailed discussion on the techniques, a larger set of hyperspectral images is used in the assessment of the different approaches. The first one which was acquired over the city of Pavia (Italy), is a hyperspectral image of spatial size : 610×340 pixels, with 103 spectral bands. The second image, which represents the University of Houston, is a hyperspectral image of spatial size 349×1905 pixels and with 144 spectral bands [19]. The last one called Indian Pines is a hyperspectral image of spatial size 145×145 pixels, and with 224 spectral bands. Moreover, on this extended version new measures of comparison have been introduced.

2. Basics on Morphological Image Representation

The goal of this section is to introduce a short background on morphological operators and transforms used in the sequel. Notation considered in the rest of this paper is also stated.

2.1. Notation

Let E be a subset of the discrete space $\mathbb{Z}^2$, which represents the support space of a 2D image and $F \subseteq \mathbb{R}^D$ be a set of pixel values in dimension D. Hence, it is assumed in our case that the value of a pixel $x \in E$ is represented by a vector $v \in F$ of dimension D, where discrete space E has a size of $n_1 \times n_2$ pixels. This vector v represents the spectrum at position x. Additionally, we will write higher order tensors by calligraphic upper-case letters ($\mathcal{I}, \mathcal{S}, \ldots$). The order of tensor $\mathcal{I} \in \mathbb{R}^{n_1 \times n_2 \times \ldots \times n_J}$ is J. Moreover, if $\mathcal{I} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, for all $i \in [1, n_3]$ $\mathcal{I}_{:,:,i}$ represents a matrix of size $n_1 \times n_2$ where the third component is equal to i. In our case we can also associate a tensor to the hyperspectral image $\mathcal{F} \in \mathbb{R}^{n_1 \times n_2 \times D}$.

2.2. Nonlinear Scale-Spaces and Morphological Decomposition

Mathematical morphology is a well known nonlinear image processing methodology based on the application of complete lattice theory to spatial structures. Let $f : E \to \mathbb{Z}$ be a grey-scale image. Area openings $\gamma_{s_l}^a(f)$ (resp. area closings $\varphi_{s_l}^a(f)$) are morphological filters that remove from the image f the bright (resp. dark) connected components having a surface area smaller than the parameter $s_l \in \mathbb{N}$ [20]:

$$\gamma_{s_l}^a(f) = \bigvee_i \{\gamma_{B_i}(f) | B_i \text{ is connected and } \mathrm{card}(B_i) = s_l\} \tag{1}$$

$$\varphi_{s_l}^a(f) = \bigwedge_i \{\varphi_{B_i}(f) | B_i \text{ is connected and } \mathrm{card}(B_i) = s_l\} \tag{2}$$

where $\gamma_B(f)$ and $\varphi_B(f)$ represent respectively the morphological flat opening and closing according to structuring element B [21]. We note that these connected filters can be implemented as binary filters on the stack decomposition of f into upper level sets. Figure 1 illustrates how area opening and area closing modify a simple image f. The image f in this toy example is composed of one black triangle of area equal to 30, 2 diamonds, one black and one white of area equal to 15. Finally the last connected components are 4 white circles and 5 black ones of area equal to 5. When an area opening is used (respectively closing) of threshold $s_l = 7$, just the white (respectively black) circles are removed.
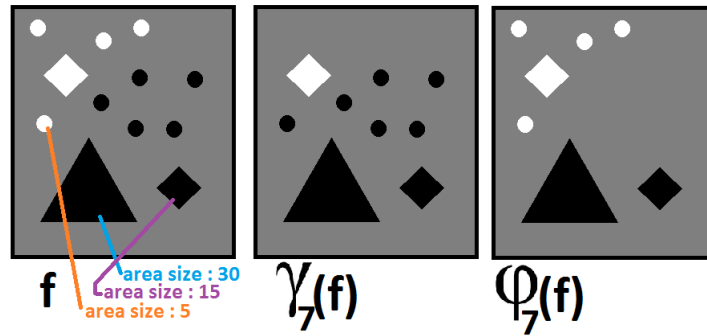

Figure 1. Illustration of an area opening $\gamma_{s_l}^a$ and an area closing $\varphi_{s_l}^a$ of image f, with $s_l = 7$ pixels. We can see that the connected components removed by the opening operator are the white circles since their area is 5, so below 7, and similarly for the black circles in the closed image.

Area opening and area closing are very relevant to simplify images, without deforming the contours of the objects remaining. In addition, area opening and closing can be used to produce a multi-scale decomposition of an image. The notion of morphological decomposition is related to the granulometry axiomatic [21]. Let us consider $\{\gamma_{s_l}^a\}$, $1 \le l \le S$ and $\{\varphi_{s_l}^a\}$, $1 \le l \le S$, two indexed families of area openings and closings respectively. Typically, the index l is associated to scale, or more precisely to the surface area. Namely, we have on the one hand:

$$f = \sum_{l=1}^{S} (\gamma_{s_{l-1}}^a(f) - \gamma_{s_l}^a(f)) + \gamma_{s_S}^a(f) \tag{3}$$

$$f = \varphi_{s_S}^a(f) - \sum_{l=1}^{S} (\varphi_{s_l}^a(f) - \varphi_{s_{l-1}}^a(f)) \tag{4}$$

On the other hand, we can rewrite the decomposition [15]:

$$f = 1/2 \left((\gamma_{s_S}(f) + \varphi_{s_S}(f)) + \sum_{l=1}^{S} (\gamma_{s_{l-1}}^a(f) - \gamma_{s_l}^a(f)) - \sum_{l=1}^{S} (\varphi_{s_l}^a(f) - \varphi_{s_{l-1}}^a(f)) \right)$$

Therefore we have an additive decomposition of the initial image f into S scales, together with the average largest area opening and closing. We remark that the residue $(\gamma^a_{s_{l-1}}(f) - \gamma^a_{s_l}(f))$ represents bright details between levels s_l and s_{l-1}. Similarly, $(\varphi^a_{s_l}(f) - \varphi^a_{s_{l-1}}(f))$ stands for dark details between levels s_l and s_{l-1}. At this point, some issues must be taken into account. First, after decomposing an image into S scales, we have now to deal with an image representation of higher dimensionality. Second, the decomposition may not be optimal since it depends on the discretization of S scales, *i.e.*, size of each scale. In order to illustrate that issue, we have represented in Figure 2a a channel of Pavia hyperspectral image and in Figure 2b its morphological decomposition by area openings that we have over-estimated. As it may be noticed from Figure 2b, the choice of the scales is fundamental in order to avoid a redundant decomposition.
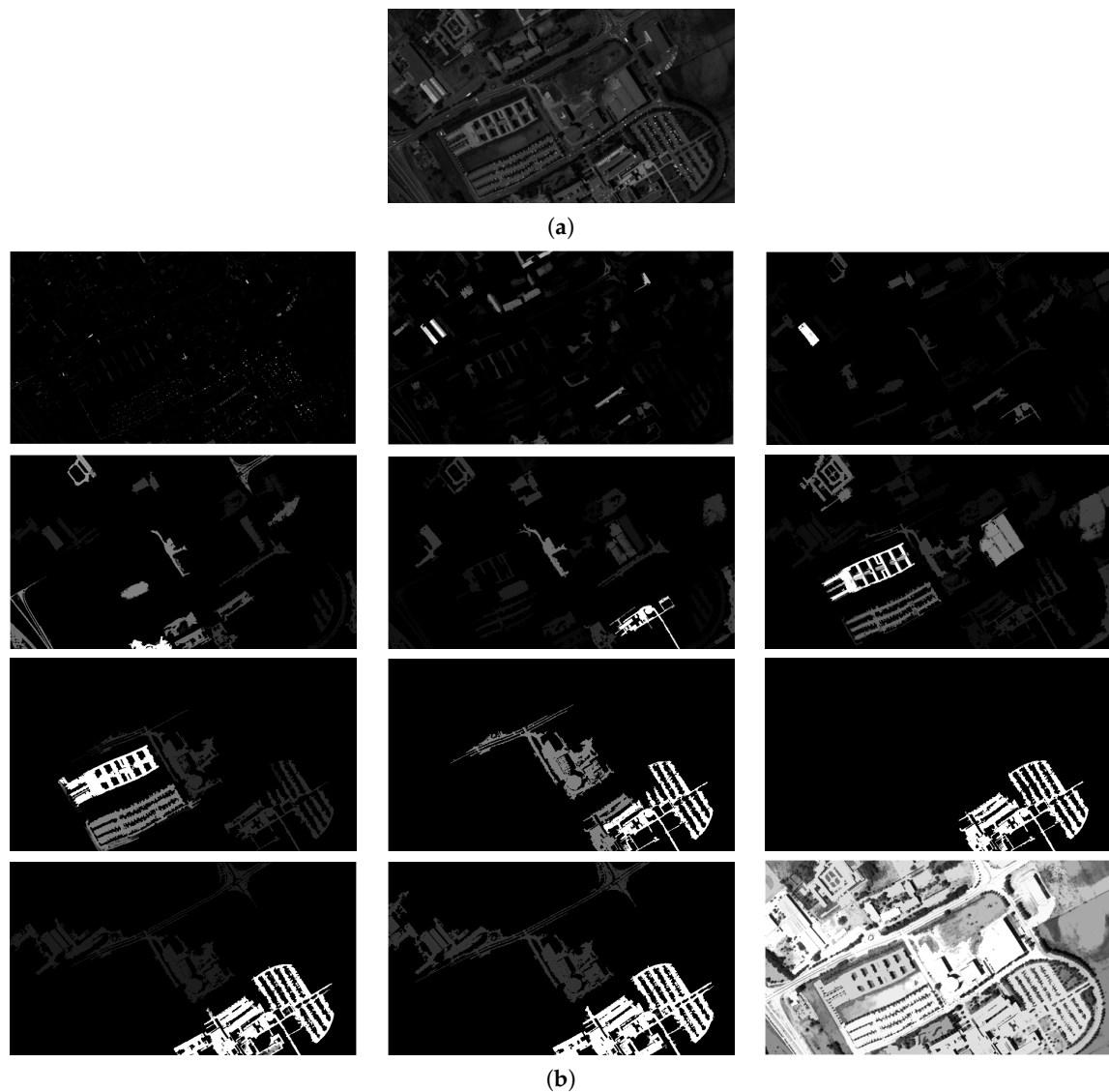

(**a**)

(**b**)

Figure 2. (**a**) Channel number 50 of Pavia hyperspectral image and (**b**) its morphological decomposition by area openings $\gamma^a_{s_l}$, $s_l = \{0.5 \times 10^2, 1 \times 10^2, 5 \times 10^2, 7 \times 10^2, 1 \times 10^3, 2 \times 10^3, 5 \times 10^3, 7 \times 10^3, 1 \times 10^4, 1.2 \times 10^4, 1.5 \times 10^4, 2.5 \times 10^4\}$. Last image in (b) corresponds to $\gamma^a_{s_S}$, $s_S = 2.5 \times 10^4$; the other images in (b) are $(\gamma^a_{s_{l-1}}(f) - \gamma^a_{s_l}(f))$. Note that the contrast of images has been enhanced to improve visualization.

In order to deal with the problem of scale discretization, we propose to use the pattern spectrum that provides information about the image component size distribution. We can also notice another technique to find the optimal discretization [22].

2.3. Pattern Spectrum

The notion of pattern spectrum (PS) [23] corresponds to the probability density function (pdf) underlying a granulometric decomposition by morphological openings and closings [21,24]. The area-based PS of f at size s_l is given by

$$PS^a(f,l) = \left[\mathrm{Mes}(\gamma^a_{s_l}(f) - \gamma^a_{s_{l+1}}(f))\right] / \mathrm{Mes}(f) \tag{5}$$

$$PS^a(f,-l) = \left[\mathrm{Mes}(\varphi^a_{s_{l+1}}(f) - \varphi^a_{s_l}(f))\right] / \mathrm{Mes}(f) \tag{6}$$

where Mes represents here the integral of the grey-scale image. Two images having the same pattern spectrum have the same morphological distribution according to the choice of the family of openings/closings. Since our goal is to have a non-redundant multi-scale representation with the same morphological representation than the original image, then by sampling the PS and choosing the scales of the distribution which keep it as similar as possible to the image PS, we can expect to find the appropriate discretization of scales. However, one can see in Figure 3, the PS is not a smooth function, and consequently, sampling it with a limited number of scales would not lead to a good result.
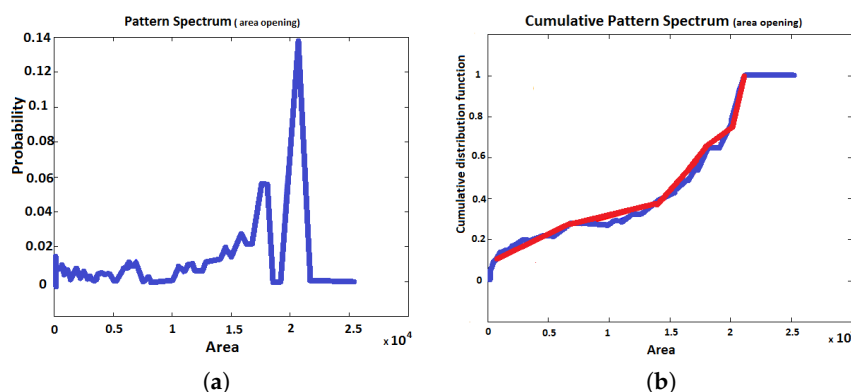

Figure 3. The pattern spectrum (PS) by area openings of a grey-scale image using 100 scales in (**a**); In (**b**), in blue, its corresponding cumulative pattern spectrum (CPS); in red, its approximation with $S = 8$ scales.

Based on the analogy between the PS and probability density function, we can compute its corresponding cumulative pattern spectrum (CPS) for both sides $l \geq 0$ and $l \leq 0$. Naturally, this function is smoother than the PS. In order to select the appropriate scales, the CPS for openings and closings are sampled, where the number of samples is fixed and is equal to S, under the constraint that the sampled function must be as similar as possible to the original function.

An example of such sampling is given in Figure 3, where the approximation of the CPS is depicted in red and the CPS of the original image in blue. It is well known in probability that two distributions that have the same cumulative distribution function have the same probability distribution function. Based on this property, we can expect that the discretization from the CPS approximates the original PS of the image and consequently, the selected scales properly represent the size distribution of the image.

2.4. Grey-Scale Distance Function

Let X be the closed set associated to a binary image. The distance function corresponding to set X gives at each point $x \in X$ a positive number that depends on the position of x with respect to X and is given by [25]:

$$\text{dist}(X)(x) \quad = \quad \min\{d(x,y) : y \in X^c\} \tag{7}$$

where $d(x,y)$ is the Euclidean distance between points x and y, and where X^c is the complement of set X. This well known transform is very useful in image processing [25].

Distance function of binary images can be extended to grey-scale images f by considering its representation into upper level sets $\{X_h(f)\}_{a \leq h \leq b}$, where

$$X_h(f) = \{x \in E \,:\, f(x) \geq h\}$$

such that $a = \min\{f(x), x \in E\}$, and $b = \max\{f(x), x \in E\}$. Then, the so-called the grey-scale distance transform of f is defined as [26]:

$$\text{dist}(f)(x) \quad = \quad (b-a)^{-1} \sum_{h=a}^{b} \text{dist}\left(X_h(f)\right)(x) \tag{8}$$

That is, the grey-scale distance transform of f is equal to the sum of the distance functions from its upper level sets.

3. Morphological Principal Component Analysis

We introduce in this section the notion of Morphological Principal Component Analysis (MorphPCA) and its variants. Before that, a mathematical background on PCA and covariance/correlation matrix is provided in order to state the rationale behind MorphPCA.

3.1. Remind on Classical PCA

Principal Component Analysis (PCA), also known as Karhunen-Loève transform, Hotelling transform, SVD transform, *etc.*, is without any doubt the most useful technique for data dimensionality reduction.

Let us start with a set of vectors $\{v_i\} \in \mathbb{R}^D$, $1 \leq i \leq n$, where n represents the number of vectors; in our case it corresponds to the number of image pixels, *i.e.*, $n = n_1 n_2$ since n_1 and n_2 are the two spatial dimensions. The goal of the PCA is to reduce the dimension of this vector space thanks to a projection on the principal component space, namely

$$F = \{v_i\}_{i=1}^{n} \longrightarrow F' = \{v_i'\}_{i=1}^{n} \tag{9}$$

with $v_i' \in \mathbb{R}^d$, where $d \ll D$. In our case, dataset $F \in M_{n,D}(\mathbb{R})$ represents the hyperspectral image $\mathcal{F}$, where each column $F_k \in \mathbb{R}^n$, $1 \leq k \leq D$ corresponds to a vectorized spectral band. PCA should find a projective space such that the mean squared distance between the original vectors and their projections is as small as possible. As we just show, this is equivalent to find the projection that maximizes the variance.

Let us call $w_j \in \mathbb{R}^D$, where j are the principal components. The aim of PCA is to find the set of vectors $\{w_j, 1 \leq j \leq D\}$ such as

$$\underset{w_j}{\arg\min} \left[n^{-1} \sum_{i=1}^{n} \|v_i - <v_i, w_j> w_j\|^2 \right], \quad \forall 1 \leq j \leq D \tag{10}$$

Developing now the distance we have:

$$\|v_i - <v_i, w_j> w_j\| = \|v_i\|^2 - 2 <v_i, w_j>^2 + <v_i, w_j>^2 \cdot \|w_j\|^2$$

then, by adding the additional constraint that $\|w_j\|^2 = 1$, replacing in Equation (10), and keeping only terms that depend on w_j, we obtain the following new objective function:

$$\underset{w_j, \|w_j\|^2=1}{\arg\max} \; n^{-1} \sum_{i=1}^{n} <v_i, w_j>^2, \;\; \forall 1 \leq j \leq D \tag{11}$$

Since

$$\mathrm{var}(<v_i, w_j>) = n^{-1} \sum_{i=1}^{n}(<v_i, w_j>)^2 - (n^{-1}\sum_{i}^{n}(<v_i, w_j>))^2$$

if we consider that the dataset F has been column-centered, which means that $\sum_{i=1}^{n} v_i = 0$, then $\mathrm{var}(<v_i, w_j>) = n^{-1}\sum_{i=1}^{n}(<v_i, w_j>)^2$. Thus we can see that the PCA aims at finding principal components that maximize the variance. The problem can be rewritten in a matrix way using the development:

$$
\begin{aligned}
n^{-1} \sum_{i=1}^{n} <v_i, w_j>^2 &= n^{-1}(Fw_j)^T(Fw_j) \\
&= w_j^T(n^{-1}(F^TF))w_j = w_j^T V w_j
\end{aligned}
$$

where $V = n^{-1}(F^TF)$, $V \in M_{D,D}(\mathbb{R})$, is the covariance matrix of F. Hence the problem to be optimized is written as

$$\underset{w_j, \|w_j\|^2=1}{\arg\max} \; w_j^T V w_j, \;\; \forall 1 \leq j \leq D \tag{12}$$

Thanks to Lagrange multiplier theorem, we can rewrite the objective function Equation (12) as:

$$\mathbf{L}(w_j, \lambda) = w_j^T V w_j - \lambda(w_j^T w_j - 1) \tag{13}$$

where $\lambda \in \mathbb{R}$. Since one should maximize this function, we had to derive it and to make it equal to zero, *i.e.*,

$$\frac{\partial \mathbf{L}}{\partial w_j}(w_j, \lambda) = 2V w_j - 2\lambda w_j = 0$$

Finally, we obtain the solution:

$$V w_j = \lambda w_j \tag{14}$$

Thus, the principal component w_j that satisfies the objective function is an eigenvector of the covariance matrix V, and the one maximizing $\mathbf{L}(w_j, \lambda)$ is the one with the largest eigenvalue. Then we can have all the w_j by simply computing the SVD of V.

There are different approaches to choose the reduced dimension d, that is the number of the principal component to be kept. The underlying assumption is the following: if the intrinsic dimension of the data is d, then the remaining $d - D$ eigenvalues, corresponding to the eigenvectors that are discarded, should be significantly small. This principle is expressed using Prop $= \sum_{j=1}^{d} \lambda_j / \sum_{j=1}^{D} \lambda_j$, which is equal to the proportion of the original variance kept. Typically, in all our examples we fix Prop $= 0.9$.

3.2. Covariance Matrix and Pearson Correlation Matrix

The covariance between two channels (or spectral bands) of an hyperspectral image $\mathcal{F}$ is computed as

$$\text{Covar}\left(\mathcal{F}_{:,:,k}, \mathcal{F}_{:,:,k'}\right) = \frac{1}{n}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\left[\mathcal{F}_{i,j,k} - \mathbb{E}(\mathcal{F})\right]\left[\mathcal{F}_{i,j,k'} - \mathbb{E}(\mathcal{F})\right] \tag{15}$$

where $\mathbb{E}(\mathcal{F})$ is the mean of the hyperspectral image. The covariance is very meaningful; however, this is not a similarity measure [27], in the sense of a metric, since it is not range limited. In order to fulfill this requirement, a solution consists in normalizing the covariance, which leads to the notion of Pearson correlation:

$$\text{Corr}\left(\mathcal{F}_{:,:,k}, \mathcal{F}_{:,:,k'}\right) = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\left[\frac{\mathcal{F}_{i,j,k} - \mathbb{E}(\mathcal{F})}{\sigma_k}\right]\left[\frac{\mathcal{F}_{i,j,k'} - \mathbb{E}(\mathcal{F})}{\sigma_k'}\right] \tag{16}$$

where $\sigma_k = \left[\frac{1}{n}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\left(\mathcal{F}_{i,j,k} - \mathbb{E}(\mathcal{F})\right)^2\right]^{1/2}$ The correlation coefficient varies between $+1$ and -1, such that $\text{Corr}\left(\mathcal{F}_{:,:,k}, \mathcal{F}_{:,:,k'}\right) = 1$ involves that $\mathcal{F}_{:,:,k}$ and $\mathcal{F}_{:,:,k'}$ perfectly coincide. It has been proved that the best fitting case corresponds to [28]:

$$\mathcal{F}_{i,j,k} = \text{Corr}\left(\mathcal{F}_{:,:,k}, \mathcal{F}_{:,:,k'}\right)\frac{\sigma_k}{\sigma_k'}(\mathcal{F}_{i,j,k'} - \mathbb{E}(\mathcal{F})) + \mathbb{E}(\mathcal{F}) \tag{17}$$

Therefore, from Equation (17), we can see that the correlation is a linear coefficient between $\mathcal{F}_{i,j,k}$ and $\mathcal{F}_{i,j,k'}$. This means that Pearson correlation is a similarity criterion which depends on the intensities of the images and their linear relations.

3.3. MorphPCA and Its Variants

The fundamental idea of MorphPCA consists in replacing the covariance matrix V of PCA, which represents the statistical interaction of spectral bands, by a covariance matrix V_{Morpho} computed from a morphological representation of the bands. Therefore, mathematical morphology is fully integrated in the dimensionality reduction problem by standard SVD computation to solve

$$V_{\text{Morpho}}w_j = \lambda_j w_j$$

The corresponding principal components w_j provide the projection space for the hyperspectral image $\mathcal{F}$. This principle is illustrated in the diagram of Figure 4.
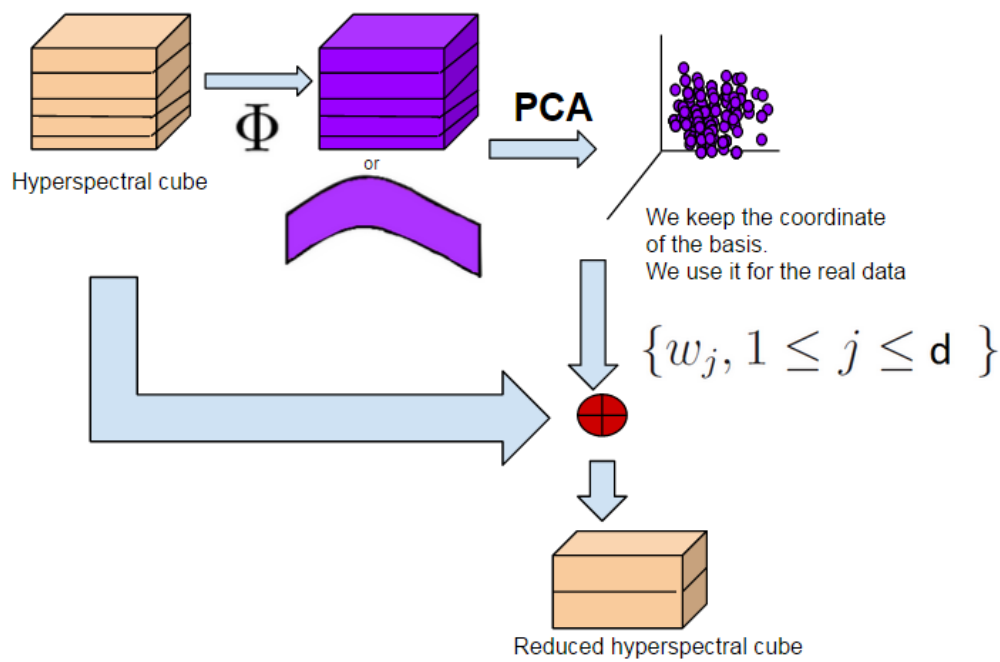
Figure 4. Global process of MorphPCA.

We propose three variants of MorphPCA which are summarized in the flowchart of Figures 5–7. An example of three different bands embedded in the space produced by these MorphPCA techniques is depicted in Figure 8.
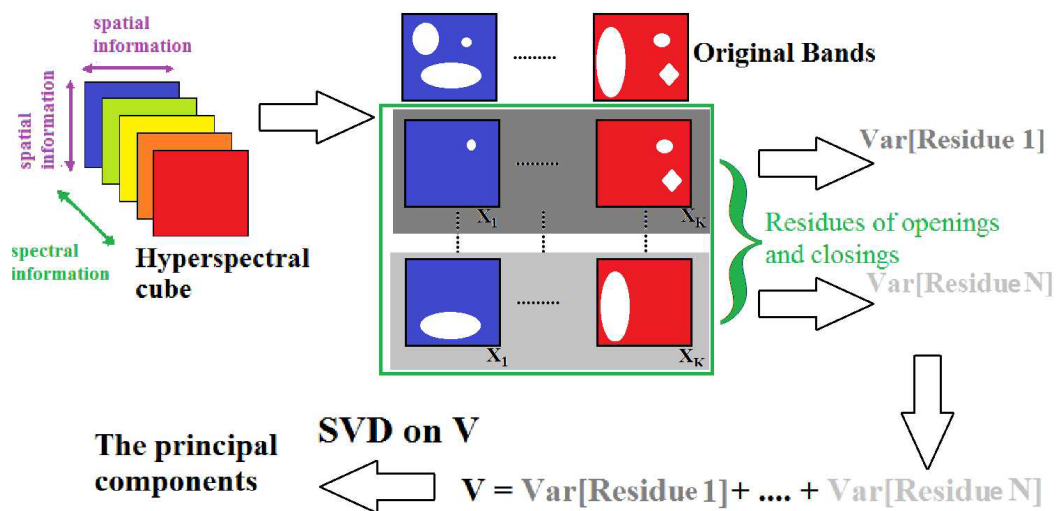

Figure 5. Process of scale-space decomposition Morphological Principal Component Analysis (MorphPCA).
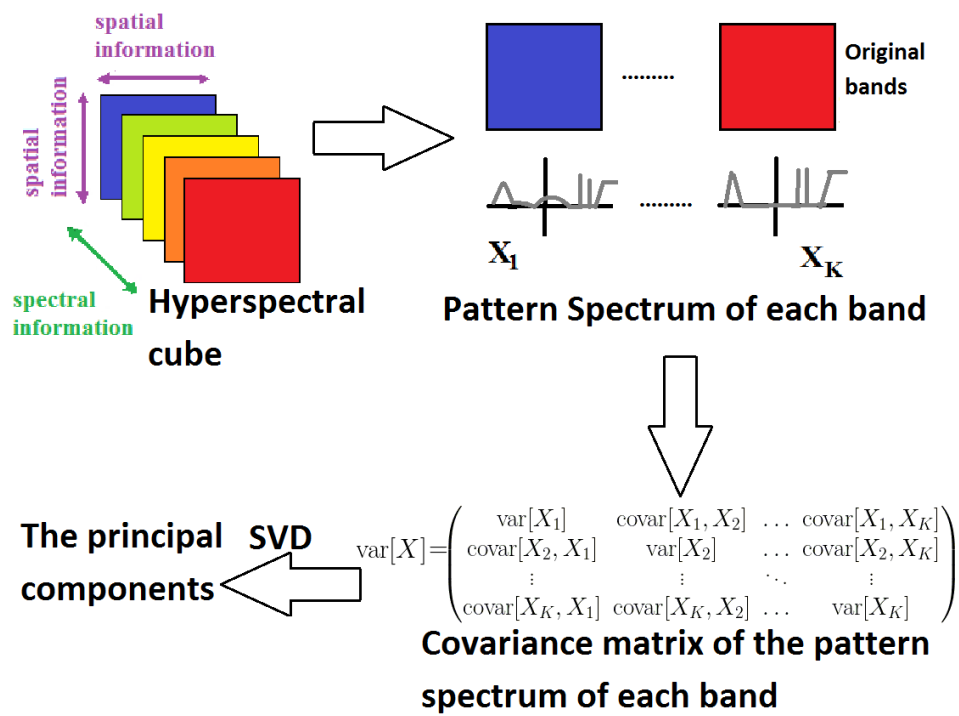
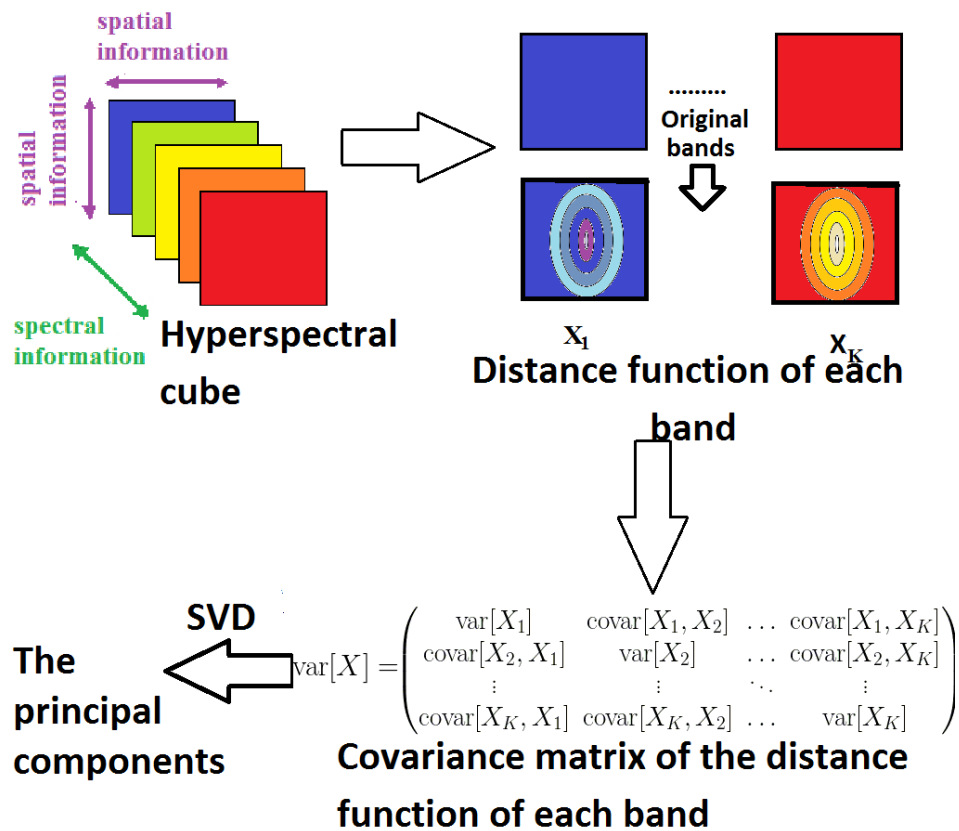**Figure 6.** Process of pattern spectrum MorphPCA.
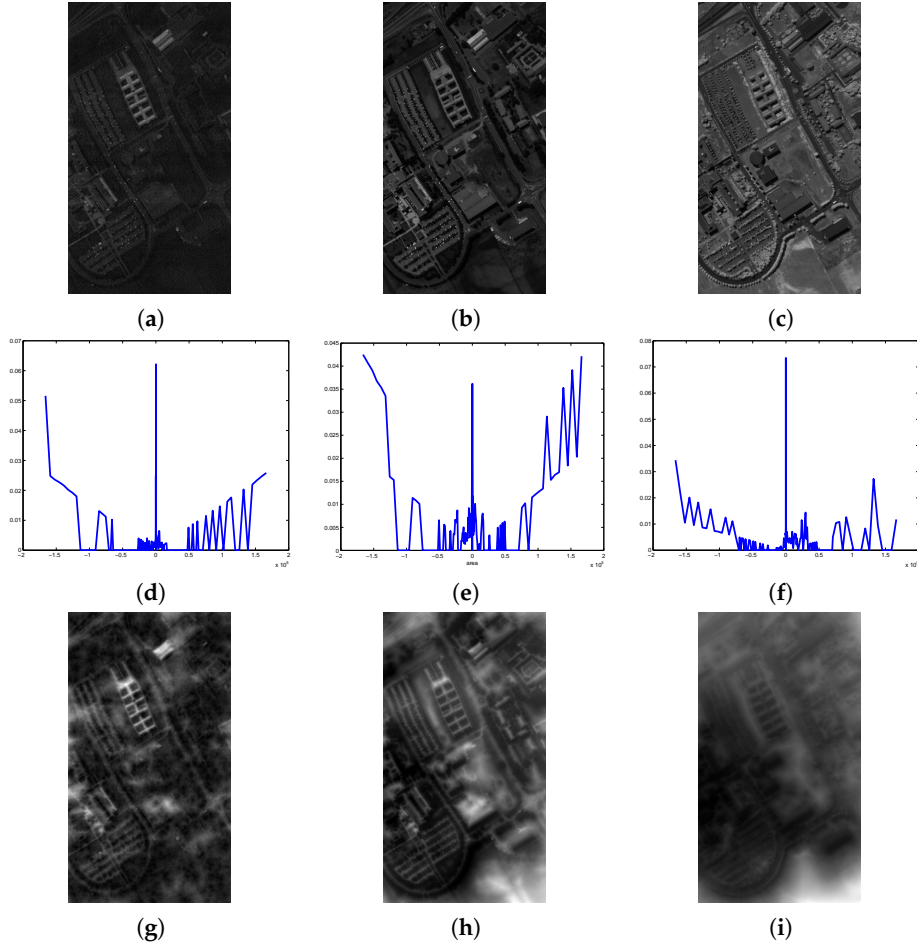

Figure 7. Process of distance function MorphPCA.

Figure 8. Top, three examples of spectral bands of Pavia image: (**a**) ♯1; (**b**) ♯50; (**c**) ♯100; middle; (**d**–**f**) pattern spectrum (PS) of corresponding spectral bands; (**g**–**i**) Molchanov distance functions of corresponding spectral bands.

3.3.1. Scale-Space Decomposition MorphPCA

In the first variant, we just use the area-based nonlinear scale-space discussed in the previous section. So the grey-scale image of each spectral band $\mathcal{F}_{:,:,k}$ is decomposed into residues of area openings and area closings according to the discretization into S scales for each operator, *i.e.*, $r_l(\mathcal{F}_{:,:,k}) = \gamma^a_{s_{l-1}}(\mathcal{F}_{:,:,k}) - \gamma^a_{s_l}(\mathcal{F}_{:,:,k})$ and $r_{-l}(\mathcal{F}_{:,:,k}) = \varphi^a_{s_l}(\mathcal{F}_{:,:,k}) - \varphi^a_{s_{l-1}}(\mathcal{F}_{:,:,k})$, $1 \leq l \leq S$. Thus we have increased the dimensionality of the initial dataset from a tensor (n_1, n_2, D) to a tensor $(n_1, n_2, D, 2S + 1)$. As discussed in [15], this tensor can be reduced using high order-SVD techniques. We propose here to simply compute a covariance matrix as the sum of the covariance matrices from the various scales. More precisely, we introduce $V_{\text{Morpho-1}} \in M_{D,D}(\mathbb{R})$ with :

$$V_{\text{Morpho-1}} \quad = \quad \sum_{l=1}^{S}(V(l)) + \sum_{l=1}^{S}(V(-l)) \tag{18}$$

where the covariance matrices at each scale l is obtained as

$$V(l)_{k,k'} = \text{Covar}\left(r_l(\mathcal{F}_{:,:,k}), r_l(\mathcal{F}_{:,:,k'})\right), \; 1 \leq k, k' \leq D$$

We note that it involves an assumption of independence of the various scales. We remark also that this technique is different from the classical approaches of differential profiles as [5] where the

morphological decomposition is applied after computing the spectral PCA (*i.e.*, morphology plays a role in spatial/spectral classification but not for spatial/spectral dimensionality reduction as in our case).

3.3.2. Pattern Spectrum MorphPCA

In the second variant, we can consider a very compact representation of the morphological information associated to the area-based nonlinear scale-space of each spectral band. It simply involves considering the area-based PS of each spectral band as the variable to be used to find statistical redundancy on the data. In other words, the corresponding covariance matrix $V_{\text{Morpho-2}} \in M_{D,D}(\mathbb{R})$ is defined as :

$$V_{\text{Morpho-2}\ k,k'} = \text{Covar}\left(PS^a(\mathcal{F}_{:,:,k}, l), PS^a(\mathcal{F}_{:,:,k'}, l)\right) \tag{19}$$

with $1 \leq k, k' \leq D$ and where $PS^a(\mathcal{F}_{:,:,k}, l)$, $-S \leq l \leq S$, is the area-based pattern spectrum obtained by area-openings and area-closings. We note that the pattern spectrum can be seen as a kind of pdf of image structures. Consequently the MorphPCA associated to it explores the intrinsic dimensionality of sets of distributions instead of sets of vectors. For illustrating the information carried out by the PS, we have provided in Figure 8 the pattern spectra computed from three different bands of a hyperspectral image.

In order to better understand the interest of $V_{\text{Morpho-2}}$, we propose an analysis based on its Pearson correlation counterpart. Once the correlation of PS distribution is calculated, we have a linear coefficient between $PS^a(\mathcal{F}_{:,:,k}, l)$ and $PS^a(\mathcal{F}_{:,:,k'}, l)$. However, since the PS is the result of nonlinear operations, the underlying extracted features are naturally nonlinear.

Let us consider the two binary images of Figure 9a, which represent two objects having exactly the same size. If the correlations are calculated, we have:

$$\text{Corr}\,(\text{image1}, \text{image2}) = 0$$
$$\text{Corr}\,(PS^a(\text{image1}), PS^a(\text{image2})) = 1$$

Image 1   Image 2

(a)

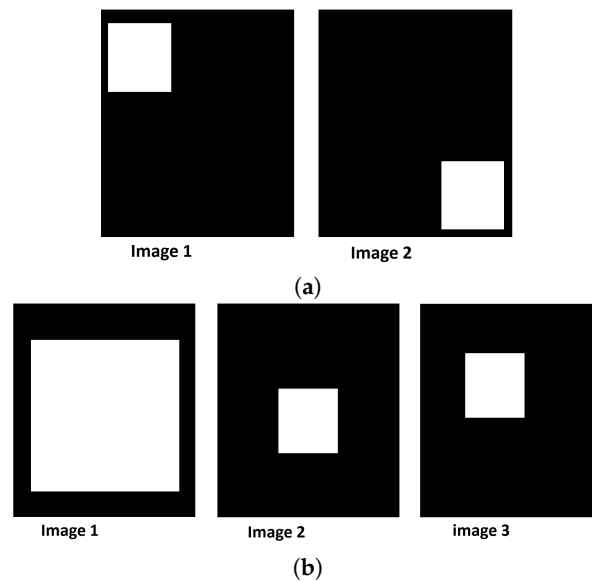Image 1   Image 2   image 3

(b)

Figure 9. (**a**) Example of a pair of binary images for pattern spectrum correlation discussion; (**b**) Example of triplet of binary images for distance function correlation discussion.

Hence, we can see that the morphological distribution being the same, the PS correlation is maximal. In a certain way, we observe that this transform builds size-invariants from the images and consequently it is robust to some groups of transforms and deformations. For instance, it is invariant to rotation and to translation.

Classical PCA on the spectral bands and the MorphPCA based on the PS can be compared by the corresponding correlation matrices from a hyperspectral image, such as the example plotted in Figure 10a,b. From this visualization, we already observe that the bands are better discriminated between them.
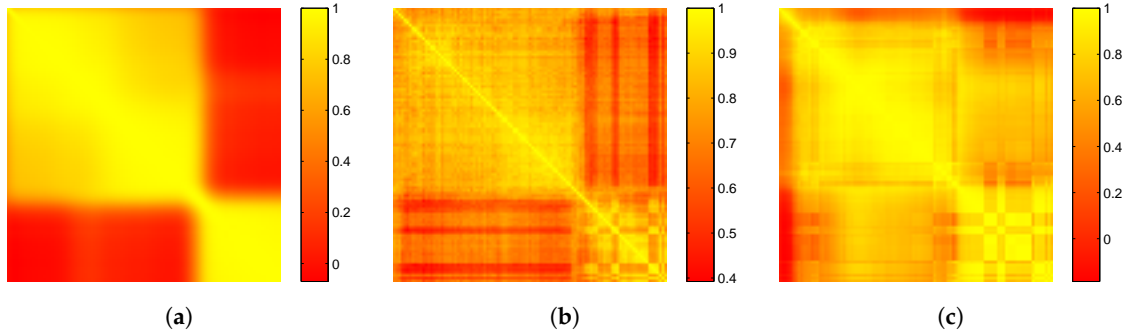

| (a) | (b) | (c) |

Figure 10. Visualization of the correlation matrix of (**a**) the spectral bands of Pavia hyperspectral image; (**b**) the PS of its spectral bands; (**c**) the distance function of its spectral bands.

3.3.3. Distance Function MorphPCA

Classical PCA for hyperspectral images is based on exploring covariances between spectral intensities. The previous MorphPCA involves changing the covariance into a morphological scale-space representation of the images. An alternative is founded when transforming each spectral band from an intensity based map to a metric based map where at each pixel the value is associated to both the initial intensity and the spatial relationships between the image structures. This objective can be achieved using the Molchanov grey-scale distance function [26] for each spectral band $\text{dist}(\mathcal{F}_{:,:,k})$. The new covariance matrix $V_{\text{Morpho-3}} \in M_{D,D}(\mathbb{R})$ is now defined as:

$$V_{\text{Morpho-3 } k,k'} \quad = \quad \text{Covar}\left(\text{dist}(\mathcal{F}_{:,:,k}), \text{dist}(\mathcal{F}_{:,:,k'})\right) \tag{20}$$

with $1 \leq k, k' \leq D$. Figure 8 depicts the corresponding grey-scale distance function from three spectral band of a hyperspectral image. We note that this function carries out simultaneously both intensity and shape information from the image.

Let consider in detail the expression of the covariance of distance functions:

$$\text{Covar}\left(\text{dist}(\mathcal{F}_{:,:,k}), \text{dist}(\mathcal{F}_{:,:,k'})\right) =$$

$$\text{Covar}\left(\sum_{h=\min(\mathcal{F}_{:,:,k})}^{\max(\mathcal{F}_{:,:,k})} d(X_s(\mathcal{F}_{:,:,k})), \sum_{h'=\min(\mathcal{F}_{:,:,k})}^{\max(\mathcal{F}_{:,:,k})} d(X_{h'}(\mathcal{F}_{:,:,k'}))\right) =$$

$$\sum_{h=\min(\mathcal{F}_{:,:,k})}^{\max(\mathcal{F}_{:,:,k})} \sum_{h'=\min(\mathcal{F}_{:,:,k})}^{\max(\mathcal{F}_{:,:,k})} \text{Covar}\left(d(X_h(\mathcal{F}_{:,:,k})), d(X_{h'}(\mathcal{F}_{:,:,k'}))\right)$$

where $X_h(\mathcal{F}_{:,:,k})$ denotes an upper level set at threshold h. The central term is the covariance between two binary distance functions and can be developed as follows:

$$\text{Covar}\left(d(X_h(\mathcal{F}_{:,:,k})), d(X_{h'}(\mathcal{F}_{:,:,k'}))\right) =$$
$$\mathbb{E}\left(d(X_h(\mathcal{F}_{:,:,k})), d(X_{h'}(\mathcal{F}_{:,:,k'}))\right) - \mathbb{E}\left(d(X_h(\mathcal{F}_{:,:,k}))\right)\mathbb{E}\left(d(X_{h'}(\mathcal{F}_{:,:,k'}))\right) =$$
$$n^{-1} < d(X_h(\mathcal{F}_{:,:,k})), d(X_{h'}(\mathcal{F}_{:,:,k'})) >_{L^2} -\mathbb{E}\left(d(X_h(\mathcal{F}_{:,:,k}))\right)\mathbb{E}\left(d(X_{h'}(\mathcal{F}_{:,:,k'}))\right)$$

where $< \cdot, \cdot >_{L^2}$ denotes the L^2 inner product. Using the classical relationship:

$$\|A - B\|_{L^2}^2 = \|A\|_{L^2}^2 + \|B\|_{L^2}^2 - 2 < A, B >_{L^2}, \ \forall A, B \in \mathbb{R}^n$$

we finally obtain that:

$$\text{Covar}\left(d(X_h(\mathcal{F}_{:,:,k})), d(X_{h'}(\mathcal{F}_{:,:,k'}))\right) = (2n)^{-1}(\|d(X_h(\mathcal{F}_{:,:,k}))\|_{L^2}^2 + \|d(X_{h'}(\mathcal{F}_{:,:,k'}))\|_{L^2}^2)$$
$$-(2n)^{-1}\|d(X_h(\mathcal{F}_{:,:,k})) - d(X_{h'}(\mathcal{F}_{:,:,k'}))\|_{L^2}^2 - n^{-1}\|d(X_h(\mathcal{F}_{:,:,k}))\|_{L^2}^2 \|d(X_{h'}(\mathcal{F}_{:,:,k'}))\|_{L^2}^2$$

From this latter expression, the term

$$\|d(X_h(\mathcal{F}_{:,:,k})) - d(X_{h'}(\mathcal{F}_{:,:,k'}))\|_{L^2}^2$$

can be identified as the Baddeley distance [29] used in shape analysis. This distance is somehow equivalent to the most classical Hausdorff distance between the upper level sets h of spectral band k and h' of spectral band k'. Thus, the underlying similarity from this covariance compares the shape of the spectral channels, and extracts a richer description than Pearson correlation from the spectral channels themselves. We note that the use of Hausdorff distance between upper level sets of hyperspectral bands was previously used in [30].

Finally, to illustrate qualitatively the behavior of the distance function correlation, let us consider this time the three binary images depicted in Figure 9b, where image 2 and image 3 represent the same object placed at a different location on the image. One has:

$$\text{Corr (image1, image2)} \ = \ \text{Corr (image1, image3)}$$
$$\text{Corr (dist(image1), dist(image2))} \ \neq \ \text{Corr (dist(image1), dist(image3))}$$

That is, this similarity criterion related to the use of distance function is more discriminative to the relative position of the objects on the image than the classical Pearson Correlation.

From Figure 10c, one can compare now the correlation matrix using the grey-scale distance function with the usual correlation matrix Figure 10a. We note that this matrix provided also a better discrimination of bands cluster than the Pearson correlation matrix used in standard PCA.

3.3.4. Spatial/Spectral MorphPCA

As we have discussed, $V_{\text{Morpho-2}}$ represents a compact morphological representation of the image; however, the spectral intensity information is also important for dimensionality reduction. To come with a last variant of MorphPCA, we build another covariance matrix $V_{\text{Morpho-4}}$ that represents the spectral and spatial information without increasing the dimensionality by the sum of two covariance matrices:

$$V_{\text{Morpho-4 } \beta} = (1 - \beta)V + \beta V_{\text{Morpho-2}} \tag{21}$$

with $\beta \in [0, 1]$, and where obviously $V_{k,k'} = \text{Covar}\left(\mathcal{F}_{:,:,k}, \mathcal{F}_{:,:,k'}\right)$ and β stands for a regularization term that balances the spatial over the spectral information. This kind of linear combination of covariance matrices is similar to the one used in the combination of kernels, where kernels providing

different information sources are combined to have a new kernel which integrates the various contributions [12].

4. MorphPCA Applied to Hyperspectral Images

4.1. Criteria to Evaluate PCA vs. MorphPCA

We can now use PCA and the four variants of MorphPCA to achieve dimensionality reduction (DR) of hyperspectral images. In order to evaluate the interest for such a purpose, it is necessary to establish quantitative criteria that should be assessed. These criteria will evaluate both locally and globally the effectiveness of the dimension reduction techniques.

- Local criteria.

    Criterion 1 (C1)   The reconstructed hyperspectral image $\widetilde{\mathcal{F}}$ using the first d principal components should be a regularized version of $\mathcal{F}$ in order to be more spatially sparse.

    Criterion 2 (C2)   The reconstructed hyperspectral image $\widetilde{\mathcal{F}}$ using the first d principal components should preserve local homogeneity and be coherent with the original hyperspectral image $\mathcal{F}$.

    Criterion 3 (C3)   The manifold of variables (*i.e.*, intrinsic geometry) from the reconstructed hyperspectral image $\widetilde{\mathcal{F}}$ should be as similar as possible to the manifold from original hyperspectral image $\mathcal{F}$.

- Global criteria.

    Criterion 4 (C4)   The number of bands d of the reduced hyperspectral image should be reduced as much as possible. It means that a spectrally sparse image is obtained.

    Criterion 5 (C5)   The reconstructed hyperspectral image $\widetilde{\mathcal{F}}$ using the first d principal components should preserve the global similarity with the original hyperspectral image $\mathcal{F}$. Or in other words, it should be a good noise-free approximation.

    Criterion 6 (C6)   Separability of spectral classes should be improved in the dimensionality reduced space. That involves in particular a better pixel classification.

These criteria are used to analyze the effectiveness of the DR methods studying locally and globally their ability to remove redundancy and to preserve the fully richness of the spectral and spatial information.

In order to assess C1, we compute the watershed transform [25] on each channel $\mathcal{F}_k$ of the hyperspectral image. Watershed transform is a morphological image segmentation approach which in a certain way can be seen as an unsupervised classification technique. The advantage of using the watershed is that it allows us to cluster the image according to the local homogeneity; thus, an image with less details will have less spatial classes than an image with many insignificant details. Then, the number of clusters $\|N_k\|$ of $\mathcal{F}_k$ is considered as an estimation of the image complexity. To evaluate the complexity of the reconstructed hyperspectral image, the number of spatial classes is counted after having done a watershed on each band. Finally, the mean of the number of spatial classes is taken, *i.e.*,

$$\text{Error}_{\text{sparse spatially}} = (D^{-1}) \sum_{k=1}^{D} \|N_k\|$$

Assessment of C2, which involves image homogeneity, is based on a partition of the image into homogenous regions. Let us first remind the definition of a α-flat zones [31], used for such a purpose. Given a distance $d : \mathbb{R}^D \times \mathbb{R}^D \longrightarrow \mathbb{R}$, two pixels $(f(x), f(y)) \in (\mathbb{R}^D)^2$, from a vector-valued image f, belong to the same α-flat zone of f if and only if there is a path $(p_0, \ldots, p_n) \in E^n$ such as $p_0 = x$ and $p_n = y$ and $\forall i \in [1, n-1]$, $d(f(p_i), f(p_{i+1})) \leq \alpha$, with $\alpha \in \mathbb{R}^+$. Computing the α-flat zones for a given value of α produces therefore a spatial partition of the image into classes such that in each

connected class the image values are linked by paths of local bounded variation. Working on the d eigenvectors, the image partition π_α associated to the α-flat zones quantize spatially and spectrally an hyperspectral image, see example given in Figure 11. The goal of simultaneous spatial and spectral quantization of a hyperspectral image has been studied in [32], where we have studied in detail the dependency on the distance. Moreover, we have shown that in high dimensional spaces quantization results are generally not good. For the case considered here, we propose to use the Euclidean distance on the reduced space by PCA or MorphPCA. The choice of α is done in order to guarantee a number C of α-flat zones similar for all the compared approaches. We can expect that, by fixing the number of zones in the partition, the difference between a partition and another one depends exclusively on the homogeneity of the image. Now, using the partition π_α, the spectral mean value of pixels from the original image $\mathcal{F}$ in each spatial zone is computed. This quantization produces a simplified hyperspectral image, denoted $\overline{\mathcal{F}}^{\pi_\alpha}$. Finally, we assess how far pixels of the original image from each α-flat zone are from their mean; which involves computing the following error

$$\text{Error}_{\text{Homg}} = \sum_{k=1}^{D} \sum_{i,j=1}^{n_1,n_2} |\mathcal{F}_{i,j,k} - \overline{\mathcal{F}}^{\pi_\alpha}_{i,j,k}|^2$$

This criterion can consequently be seen as a way to see the trustworthiness of the DR technique, since it measures if the homogeneous partition of the reduced hyperspectral image corresponds to the homogeneous zone of the original image.
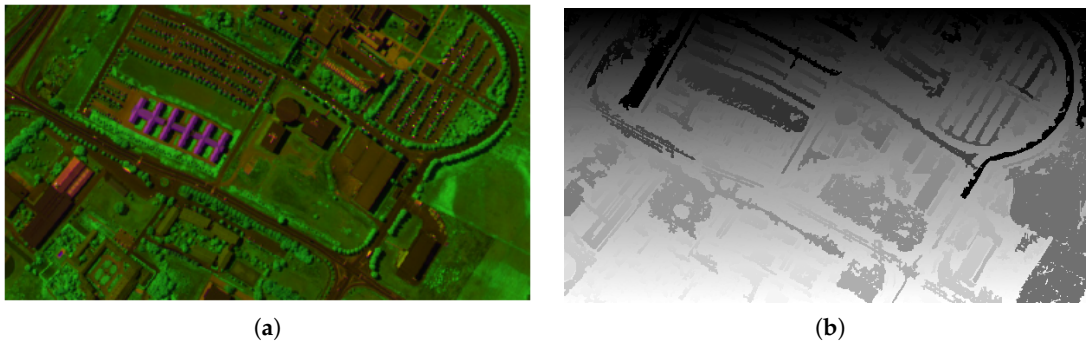

(**a**)　　　　　　　　　　　　　　　　　　　　　　　　(**b**)

Figure 11. (**a**) A 3-variate image (first three eigenimages after PCA on Pavia hyperspectral image) and (**b**) its corresponding α-flat zone partition into 84931 spatial classes using the Euclidean distance.

C3 has been evaluated by means of two manifold learning criteria called the K-intrusion and K-extrusion [33]. They are based on other criteria called continuity and trustworthiness [34]. These criteria reveal DR behavior in terms of its ability to preserve the data manifold structure. We have first sampled randomly 10 thousands spectra from our hyperspectral images, where each spectrum is a vector of dimension D. Then we have modelled the manifold by a graph where each node is a vector and each edge is the pairwise distance. We used the Euclidean distance as the pairwise distance. For the rest of the paragraph we note by x_i a point from the original manifold, v_i^K its neighborhood of size K, $\tilde{x}_i$ the same point from the manifold after a DR and $\tilde{v}_i^K$ its corresponding neighborhood of size K. A neighborhood of size K at point x_i is composed of the K closest points to x_i according to used metric. More precisely, the goal of K-extrusion is to measure how the points that were in the K-neighborhood of x_i are not preserved in the K-neighborhood of $\tilde{x}_i$ after DR. The K-intrusion evaluates if points on the K-neighborhood of $\tilde{x}_i$ on the DR manifold were in the K-neighborhood of x_i, *i.e.*,

$$M_{\text{intrusion}}(K) \quad = \quad 1 - \frac{2}{G(K)} \times \sum_{i=1}^{n} \sum_{j \in \tilde{v}_i^K \setminus v_i^K} r(i,j) - K \qquad (22)$$

$$M_{\text{extrusion}}(K) \;\; = \;\; 1 - \frac{2}{G(K)} \times \sum_{i=1}^{n} \sum_{j \in \nu_i^K \setminus \tilde{\nu}_i^K} r(\tilde{i}, j) - K \tag{23}$$

where $r(i,j)$ is the rank of the data x_j in the ordering according to the distance from x_i, and respectively $r(\tilde{i}, j)$ the rank of $\tilde{x}_j$ in the ordering according to the distance from $\tilde{x}_i$, and the term G_K scales the measure to be between zero and one, *i.e.*,

$$G(K) = \begin{cases} NK(2N - 3K - 1) \text{ if } K < N \setminus 2 \\ N(N - K)(N - K - 1) \text{ if } K \geq N \setminus 2 \end{cases} \tag{24}$$

For a better understanding of these formulae, see [35]. An important point is the dependence of these parameters on the size of the neighborhood. From Equations (22) and (23), the following parameters are computed [35]:

$$Q(K) \;\; = \;\; \frac{M_{\text{extrusion}}(K) + M_{\text{intrusion}}(K)}{2} \tag{25}$$

$$B(K) \;\; = \;\; M_{\text{intrusion}}(K) - M_{\text{extrusion}}(K) \tag{26}$$

The interest of $Q(K)$ is that it estimates in average the quality of a DR technique, whereas $B(K)$ reveals its behavior as being more intrusive or extrusive.

In order to assess C4, as classically done, the fraction of explained covariance is fixed. Then, the number of principal components needed is counted. The rationale is based on the fact that a good DR technique should reduce the number of dimensions and extract a limited number of features that would explain most of the image. However, since this criterion is linked to a sparsity criterion, we would like to add a distortion criterion, C5.

The evaluation of C5 is founded on computing a pattern spectrum of both the original hyperspectral image and the DR image. An important point is that the pattern spectrum will be computed by openings on the hyperspectral image viewed as a 3D image. By doing such assumption, the 3D openings are decomposing in a simultaneous way the spatial/spectral object of the image and the corresponding curves of the PS will represent the distribution of both the spatial and the spectral objects. Two hyperspectral images are similar if they have the same spectral/spatial size distribution. As discussed in Section 2, we prefer to use the cumulative PS in order to obtain a smoother curve. Normally we cannot deal with both spatial/spectral distortions with the reconstruction error of the two images. However we will also assess the SNR of the reconstruction error as an additional parameter.

Finally, C6 is related to supervised pixel classification of the hyperspectral image. We have considered the least square SVM algorithm [11] as a learning technique, with a linear kernel or RBF kernel, where the RBF kernel is initialized for each DR technique using cross validation. For each supervised classification run, we used for the AVIRIS Indian Pine Image 5% of the available data as a training set and the remaining 95% to validate. For the ROSIS Pavia University image we use a subset of 50 spectra (about 1% of the available data) per class as a training set and the remaining spectra to validate.

4.2. Evaluation of Algorithms

The studied DR techniques presented are listed and compared upon three mathematical and computing properties in Table 1. These properties were also considered in the excellent comparative review [3]. For comparison, we have also included in the table the Kernel-PCA (KPCA), which is a powerful generalization of PCA allowing integrating morphological and spatial features into DR.

The first one is the number of free parameters to be chosen. The interest of having these free parameters is that it provides more flexibility to the techniques, whereas the related inconvenient is the difficulty for properly tuning the right parameter. We notice that KPCA provides good flexibility

thanks to the choice of any possible kernel which fits the data geometry. The most simple algorithms are the PCA, and the distance function MorphPCA. Then, we have the scale-space decomposition MorphPCA, and finally the pattern spectrum MorphPCA. The second issue analyzed is the computational complexity, and the third one is the memory requirements. From a computational viewpoint, the most demanding step in the PCA is the SVD, which can be done in $O(D^3)$. PCA is the technique with the smallest computational need. On the contrary, the computational requirement of KPCA is $O(n^3)$; since $n \gg D$, this kind of algorithm seems infeasible in standard hyperspectral images. That is the reason why most of hyperspectral KPCA techniques use tricks to be able to deal with the high number of spectra [13,36,37]. All these techniques lead to a spatial distortion, which is not avoidable by the need of a sampling procedure aiming at reducing the number of spectra. Between the complexity of PCA and KPCA, we have our proposed MorphPCA algorithms. Regarding MorphPCA, the computationally demanding step is the computation of the morphological representation used in the corresponding covariance matrix. The complexity estimation has been carried out each time in the worse case; however, efficient morphological algorithms can improve this part. Distance function MorphPCA is last demanding, then the scale-space decomposition MorphPCA and finally the pattern spectrum MorphPCA. Regarding memory needs, for the PCA, the pattern spectrum MorphPCA, and the distance function MorphPCA, the steps requiring more memory is the storage of the covariance matrix, just of $O(D^2)$. The Spatial/Spectral MorphPCA needs to store 2 covariance matrices then its memory need is $O(2D^2)$; similarly the scale-space MorphPCA needs to store $2S + 1$ covariance matrices, then its required memory is $O((2S + 1)D^2)$. Note that KPCA uses a Gram matrix of size $(n \times n)$.

Table 1. Comparison of the properties of dimensionality reduction algorithms for hyperspectral images.

Technique	Parameter (1)	Computational (2)	Memory (3)
PCA	Prop	$O(D^3)$	$O(D^2)$
MorphPCA Morpho-1	Prop, S	$O(DnS_S(2S+1))$	$O(D^2(2S+1))$
MorphPCA Morpho-2	Prop	$O(DnS_S(2S+1))$	$O(D^2)$
MorphPCA Morpho- 3	Prop	$O(Dn(b-a))$	$O(D^2)$
MorphPCA Morpho-4 $_\beta$	Prop, β	$O(DnS_S(2S+1))$	$O(2D^2)$
KPCA	Prop, K	$O(n^3)$	$O(n^2)$

4.3. Evaluation on Hyperspectral Images

The assessment of the performance of PCA and MorphPCA has been carried out on three hyperspectral images. The first image was acquired over the city of Pavia (Italy) and it represents the university campus. The dimensions of the image are 610×340 pixels, with $D = 103$ spectral bands and its geometrical resolution is of 1.3 m. We also used a second hyperspectral image which represents the University of Houston campus and the neighboring urban area at the spatial resolution of 2.5 m and which dimensions are 349×1905 pixels and $D = 144$ spectral bands [19]. The third image, acquired over the region of the Indian Pines test site in North-western Indiana, is composed for two-thirds of agriculture, and one-third of forest. The dimensions of this image are 145×145 pixels, $D = 224$ spectral bands and its geometrical resolution is of 3.7 m.

We have applied classical PCA and the different variants of MorphPCA to Pavia hyperspectral image. Figure 12 shows the first three eigenimages, visualized as a RGB false color. We note that the pattern spectrum MorphPCA requires $d = 5$ to represent 92% of the variance whereas the other approaches only impose $d = 3$. An interesting aspect observed on the projection of the 103 spectral channels of Pavia hyperspectral image into the first two eigenvectors is how PCA and the scale-space decomposition MorphPCA cluster the bands linearly, see Figure 13a,b. Bands close in the projection are also near in the spectral domain, whereas the pattern spectrum MorphPCA, Figure 13c, and

distance function MorphPCA, Figure 13d, tend to cluster spectral bands which are not necessary spectrally contiguous. This can be explained thanks to Figure 10, where the MorphPCA correlation matrices are different from the classical PCA one.
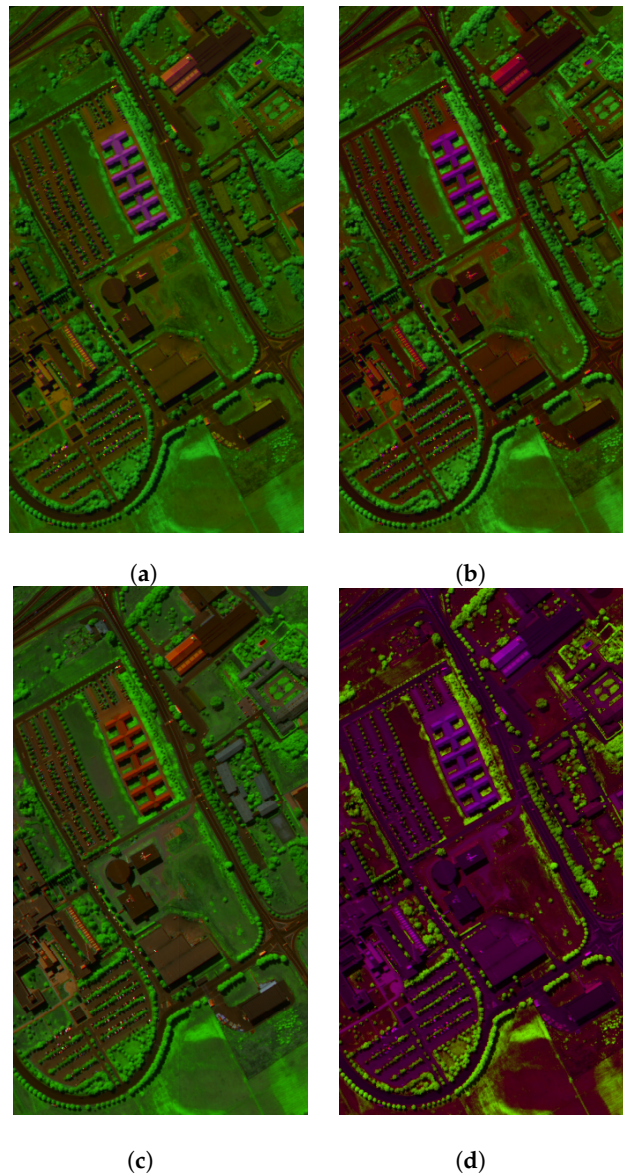

(**a**)   (**b**)

(**c**)   (**d**)

Figure 12. RGB false color visualization of first three eigenimages from Pavia hyperspectral image: (**a**) classical PCA on spectral bands; (**b**) scale-decomposition MorphPCA; (**c**) pattern spectrum MorphPCA; (**d**) distance function MorphPCA.

It can be noticed that in classical manifold learning techniques, the goal is to decrease the dimension of the data while keeping some properties on the data manifold. We work here on the manifold of the channels. This manifold is easier to use, but finding the good β in $V_{\text{Morpho-4 }\beta}$ that would maintain some properties of the manifold is not always easy, since we had to deal with a double optimization problem, *i.e.*, β and d.

From a quantitative viewpoint, one can see in Table 2 that globally MorphPCA produces a more homogenous regularization of the image than classical PCA, especially the distance function MorphPCA and Spatial/Spectral MorphPCA with an appropriate $\beta = 0.2$, which gives the lowest values of $\text{Error}_{\text{Homg}}$. We noted that $\text{Error}_{\text{sparse spatially}}$ follows a different ranking. A good method is the one with a good trade-off between both criteria, since one wants a DR to be trustworthy,

which is evaluated thanks to Error$_{\text{Homg}}$. However, if the signal is too noisy, one may prefer a sparser representation. According to these criteria, the distance function MorphPCA and the pattern spectrum MorphPCA seem to have the best result. We can also note that if we use manifold learning parameters for criterion C3, see Figure 14, the pattern spectrum MorphPCA has the best results.
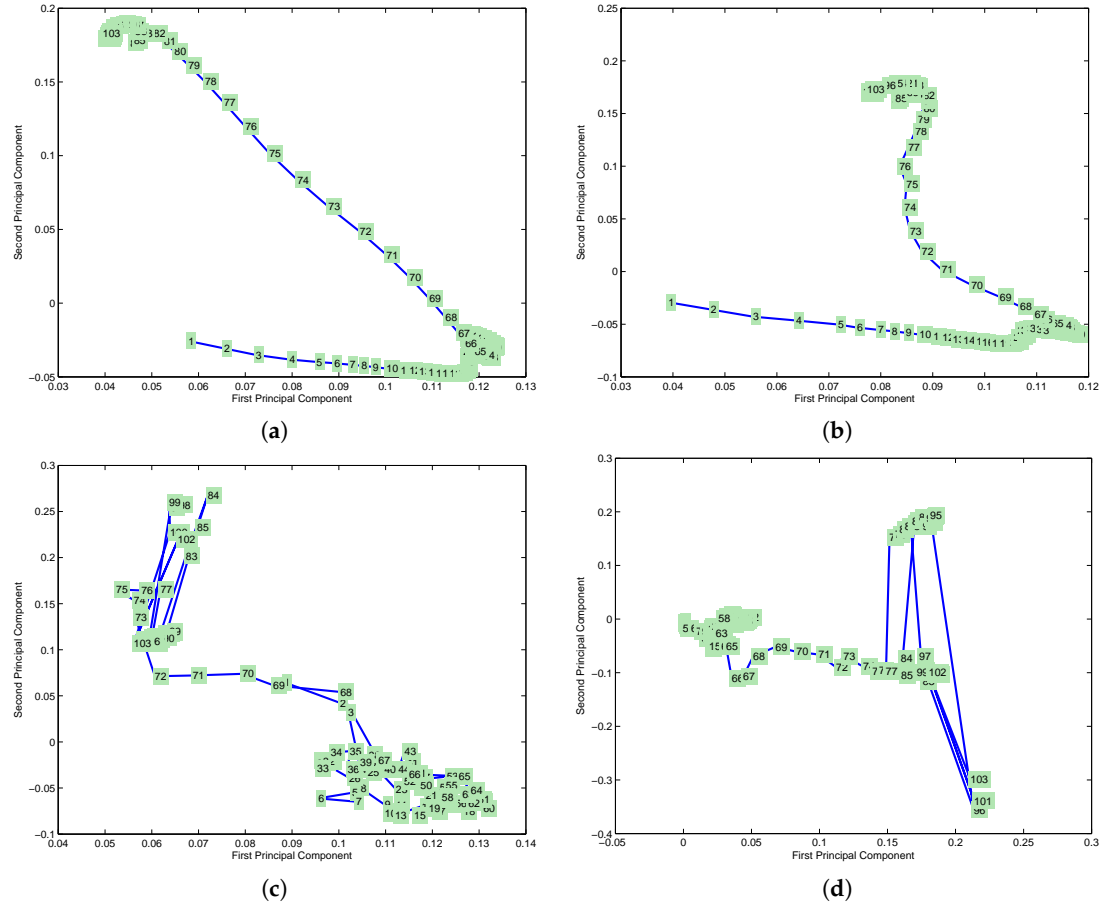

Figure 13. Hyperspectral band projection into the first two eigenvectors (*i.e.*, image manifold) from Pavia hyperspectral image: (**a**) classical PCA on spectral bands; (**b**) scale-decomposition MorphPCA; (**c**) pattern spectrum MorphPCA; (**d**) distance function MorphPCA.
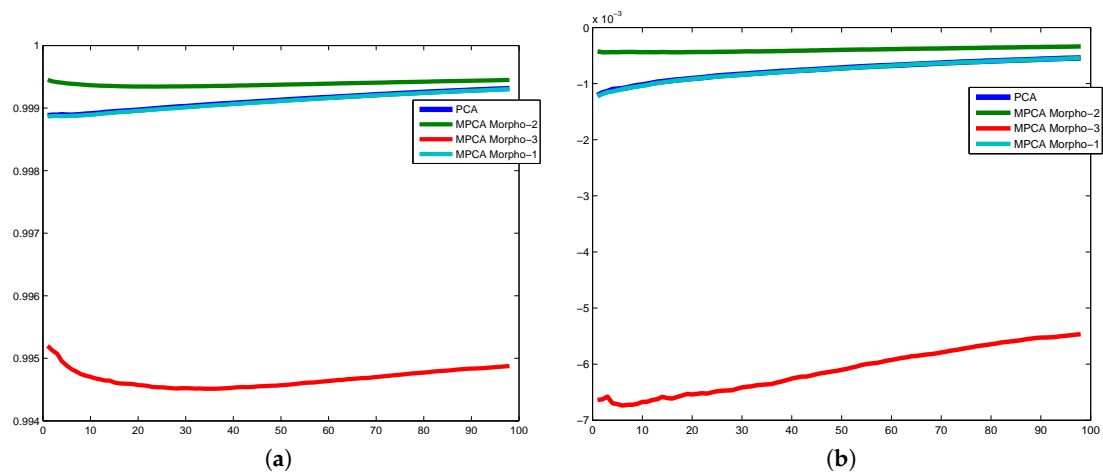

Figure 14. Intrusion/Extrusion parameters for PCA and the different variants of MorphPCA from Pavia hyperspectral image: (**a**) $Q(K)$; (**b**) $B(K)$.

Table 2. Comparison of PCA and MorphPCA analysis using criteria C1 and C2: (**a**) for Pavia hyperspectral image; (**b**) for Houston hyperspectral image; (**c**) for Indian Pines hyperspectral image. The values have been normalized to the worst case, which gives 100.

(a)				
	V	$V_{\textbf{Morpho-1}}$	$V_{\textbf{Morpho-2}}$	$V_{\textbf{Morpho-3}}$
$Error_{\textbf{Homg}}$	100	100	95.9	79.3
$Error_{\textbf{sparse spatially}}$	99.8	99.7	100	88.3
		$V_{\textbf{Morpho-4}}\,\beta$ $\beta = 0.8$	$V_{\textbf{Morpho-4}}\,\beta$ $\beta = 0.2$	$V_{\textbf{Morpho-4}}\,\beta$ $\beta = 0.5$
$Error_{\textbf{Homg}}$		93.2	83.9	88.3
$Error_{\textbf{sparse spatially}}$		93.3	96.7	98.6
(b)				
	V	$V_{\textbf{Morpho-1}}$	$V_{\textbf{Morpho-2}}$	$V_{\textbf{Morpho-3}}$
$Error_{\textbf{Homg}}$	100	90.4	35.3	38.3
$Error_{\textbf{sparse spatially}}$	97.7	97.6	100	89
(c)				
	V	$V_{\textbf{Morpho-1}}$	$V_{\textbf{Morpho-2}}$	$V_{\textbf{Morpho-3}}$
$Error_{\textbf{Homg}}$	98.1	100	96.5	97.8
$Error_{\textbf{sparse spatially}}$	91	100	91.2	82.7

With respect to criterion C5, we have computed the 3D pattern spectrum distribution of Pavia hyperspectral image and of the different reduced images into *d* components, see Figure 15. From this result, we can see that both PCA and scale-space decomposition MorphPCA follow very well the hyperspectral image, since their spatial and spectral cumulative distributions are similar. However, if one would like to denoise the hyperspectral image thanks to a DR technique, these results are not always positive. If we compare the spatial and spectral cumulative distribution of the distance function MorphPCA and the one of the hyperspectral image, we notice that for small 3D size (*i.e.*, small spatial/spectral variations) that can be considered as noise, there are differences between these two distributions. However, when the size increases, the distribution of the the distance function MorphPCA tends to the hyperspectral image one. So it seems that the distance function MorphPCA simplifies the spectral/spatial noise, considered as the small 3D objects, but keeps the objects of interest.
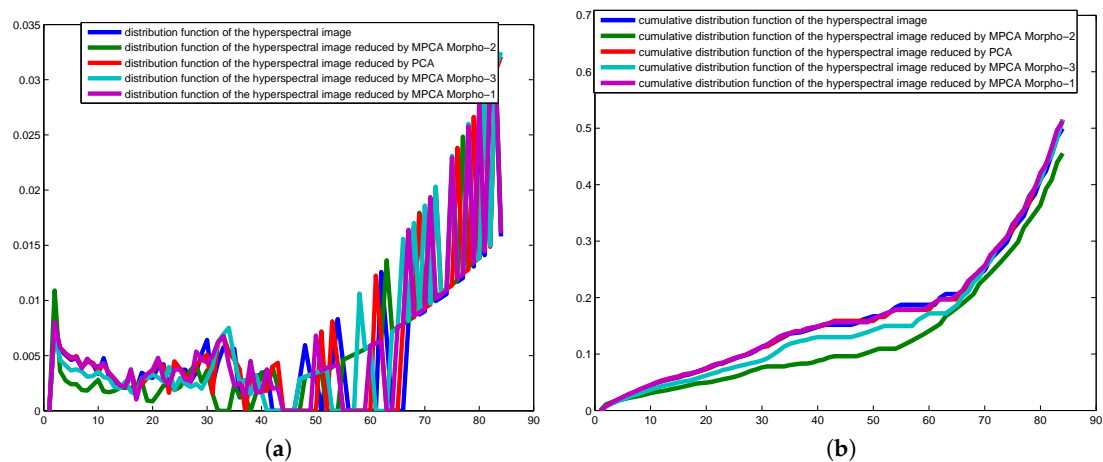

Figure 15. (**a**) 3D pattern spectrum distribution of Pavia hyperspectral image and of the different reduced images into *d* components; (**b**) Corresponding 3D cumulative pattern spectrum distributions.

Finally, Table 3 summarizes the results of supervised classification of respectively Pavia and Indian Pine hyperspectral images. We note that results for Pavia image are quite similar in all the cases even if MorphPCA seems to be better than PCA. Therefore, we have chosen to focus on the Indian Pine image, which is more challenging for supervised classification benchmark, see also Figures 16 and 17. We note that MorphPCA improves the results, especially the scale-space decomposition MorphPCA. To evaluate the classification results, first we fixed the dimension d of the reduce image. We have chosen $d = 5$. Then, we used the least square SVM, which is a multi-class classification technique, contrary to classical two-class SVM. We also used two simple kernels: the linear one, which is the simplest one, and the RBF one, which is appropriate for hyperspectral images since we can assume that these data follow Gaussian distribution. Finally, we study the influence of dimension d and of the size of the training set on the different DR techniques over the classification results. For this purpose we have depicted in Figures 18 and 19 the evolution of the kappa statistics. From the latter plot we can see that the PCA and the pattern spectrum PCA have the worst results. By combining the spectral and the spatial information, a better classification can be achieved. This is the case of the distance function MorphPCA, the scale-space decomposition MorphPCA, and the Spatial-Spectral MorphPCA for $\beta = 0.2$.
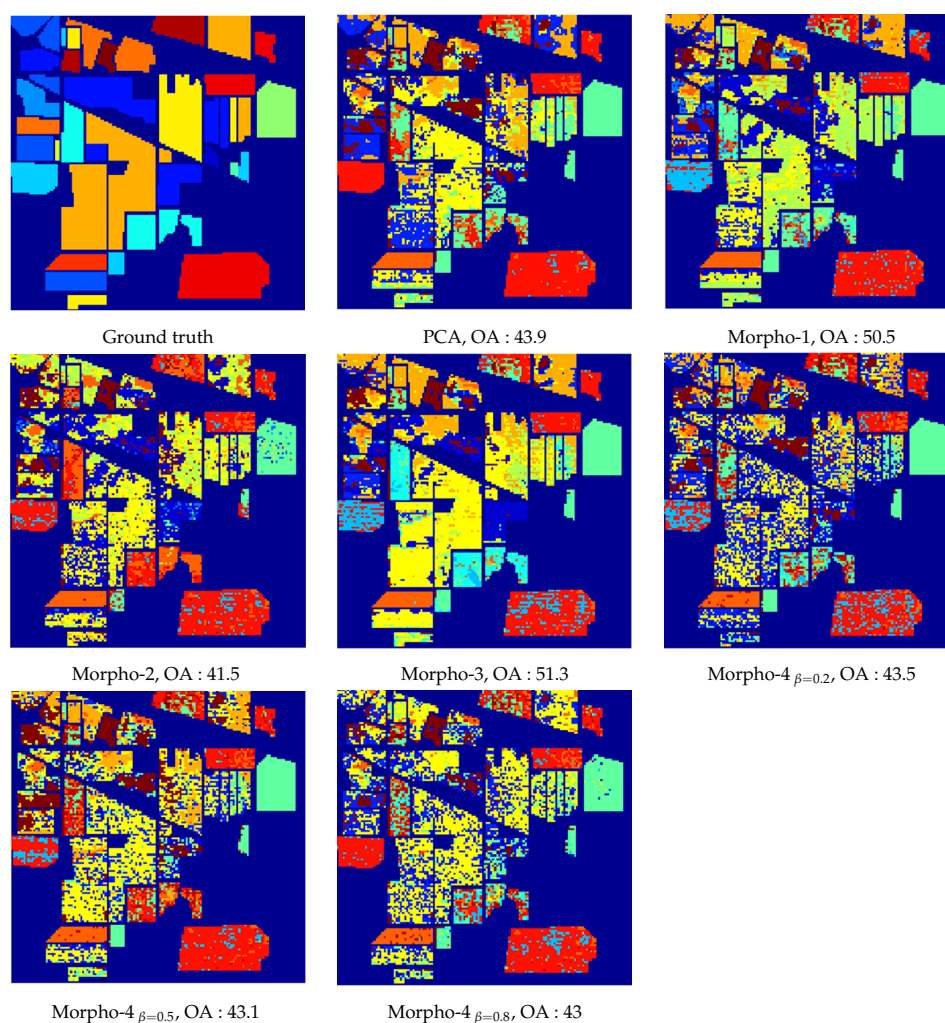

Figure 16. Results of supervised classification using least square SVM with a linear kernel on Indian Pines hyperspectral image. Note the OA is the overall accuracy.

Table 3. Comparison of hyperspectral supervised classification on PCA and MorphPCA spaces using least square SVM algorithm and different kernels: (**a**) Pavia hyperspectral image; (**b**) Indian Pines hyperspectral image.

(a) Pavia Image			
	Overall Accuracy with Linear Kernel	**Overall Accuracy with RBF Kernel**	**Kappa Statistic with RBF Kernel**
V	51.51 ± 0.9	84.9 ± 3.1	$0.84 \pm 1 \times 10^{-4}$
$V_{\textbf{Morpho-1}}$	59.6 ± 2.2	85.8 ± 2.6	$0.84 \pm 1 \times 10^{-4}$
$V_{\textbf{Morpho-2}}$	56.99 ± 1.1	85.2 ± 2.1	$0.84 \pm 1 \times 10^{-4}$
$V_{\textbf{Morpho-3}}$	59.9 ± 2.5	86.0 ± 1.9	$0.84 \pm 1 \times 10^{-4}$
$V_{\textbf{Morpho-4} \beta}, \beta = 0.2$	61.0 ± 1.73	85.2 ± 1.1	$0.83 \pm 1 \times 10^{-4}$
$V_{\textbf{Morpho-4} \beta}, \beta = 0.5$	59.9 ± 1.5	84.6 ± 1.0	$0.83 \pm 1 \times 10^{-4}$
$V_{\textbf{Morpho-4} \beta}, \beta = 0.8$	57.87 ± 3	84.7 ± 2.5	$0.83 \pm 2 \times 10^{-4}$
(b) Indian Pine image			
	Overall Accuracy with Linear Kernel	**Overall Accuracy with RBF Kernel**	**Kappa Statistic with RBF Kernel**
V	43.9 ± 3.6	75.2 ± 3.7	$0.73 \pm 4.3 \times 10^{-4}$
$V_{\textbf{Morpho-1}}$	50.5 ± 3.8	79.6 ± 3.7	$0.78 \pm 4 \times 10^{-4}$
$V_{\textbf{Morpho-2}}$	41.5 ± 3.8	66.6 ± 4.6	$0.63 \pm 4.5 \times 10^{-4}$
$V_{\textbf{Morpho-3}}$	51.3 ± 3.2	79.1 ± 3.2	$0.77 \pm 3.7 \times 10^{-4}$
$V_{\textbf{Morpho-4} \beta}, \beta = 0.2$	43.5 ± 3.3	75.1 ± 2.3	$0.72 \pm 2.6 \times 10^{-4}$
$V_{\textbf{Morpho-4} \beta}, \beta = 0.5$	43.1 ± 2.9	71.2 ± 2.6	$0.68 \pm 3 \times 10^{-4}$
$V_{\textbf{Morpho-4} \beta}, \beta = 0.8$	43.0 ± 2.2	69.7 ± 3.3	$0.67 \pm 3.9 \times 10^{-4}$

Ground truth

PCA, OA : 75.2

Morpho-1, OA : 79.6

Morpho-2, OA : 66.6

Morpho-3, OA : 79.1

Morpho-4 $_{\beta=0.2}$, OA : 75.1

Morpho-4 $_{\beta=0.5}$, OA : 71.2
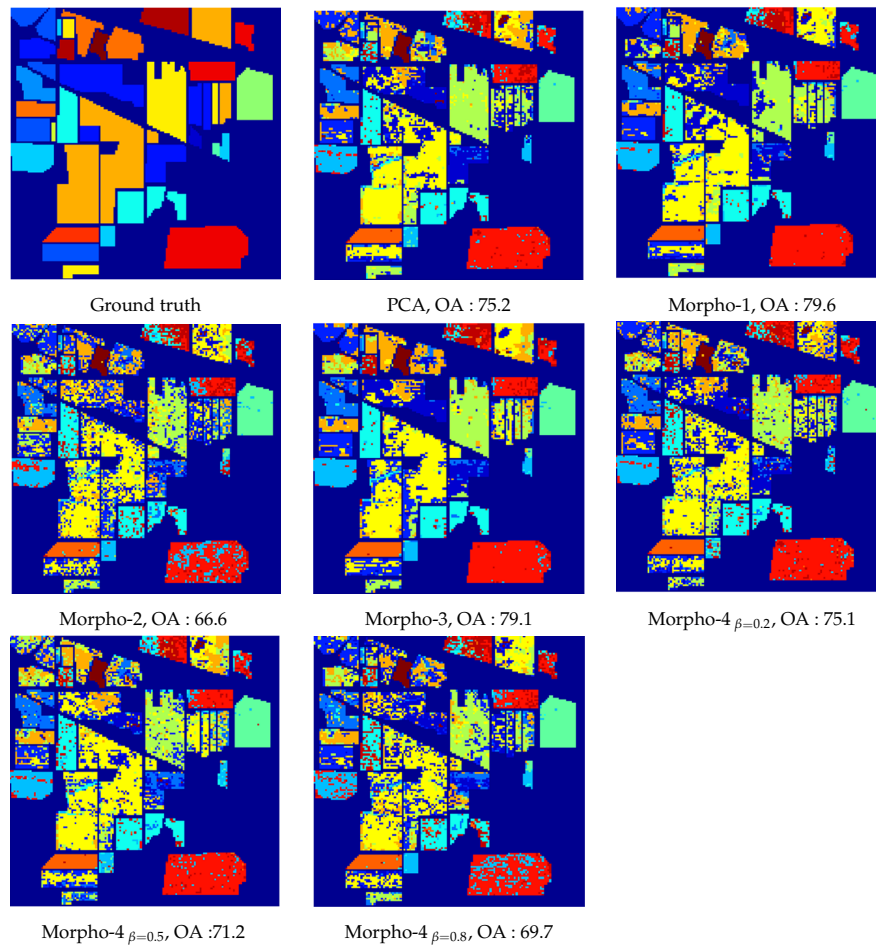

Morpho-4 $_{\beta=0.8}$, OA : 69.7

Figure 17. Results of supervised classification using least square SVM with a RBF kernel on Indian Pines hyperspectral image. Note the OA is the overall accuracy.
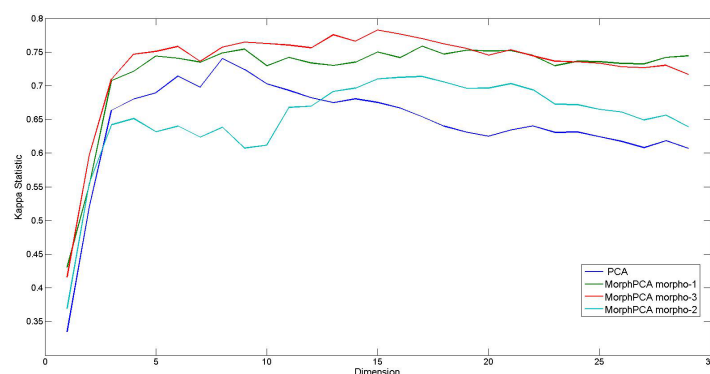
Figure 18. Results of kappa statistic for the least square SVM with a RBF kernel and different number of dimensions on Indian Pines hyperspectral image, the size of training set is equal to 5%.
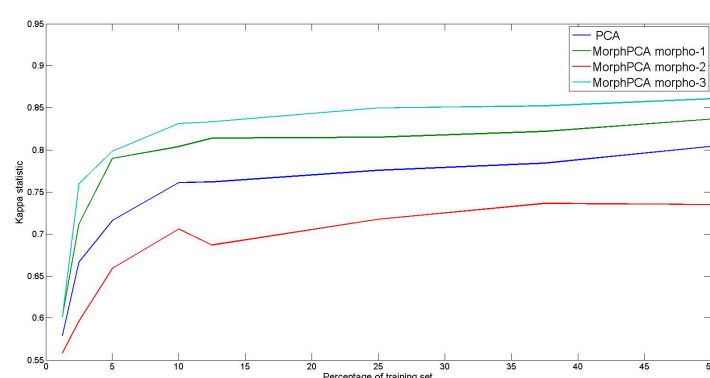

Figure 19. Results of kappa statistic for the least square SVM with a RBF kernel and different percentage of training set on Indian Pines hyperspectral image, the dimension of the reduced space is equal to 5.

5. Conclusions

We have shown in this paper how to introduce spatial information in the process of dimensionality reduction thanks to mathematical morphology operators. The representation techniques that we introduced in the paper are based on the notion the MorphPCA. As we said in the introduction, it might be possible to consider for such purpose a Kernel Principal Component Analysis (KPCA), where the kernel handles jointly spatial and spectral information. However, as discussed in Section 4.2, KPCA needs a Gram matrix of size $n \times n$, where n is the number of pixels. It is impossible to manipulate such a matrix with our test images on a standard computer. That is why there are some works trying to approximate the gram matrix in order to perform the KPCA. Some of them consider a small subset of the data which can be chosen randomly or according to some spatial information. This kind of approach has been used with hyperspectral images for spectral-spatial processing. However, these kind of works approximate the kernel, to perform dimensionality reduction on the considered Hilbert space. Here dimensionality reduction is done on the space of the data without any approximation on the morphological covariance V_{Morpho}.

That is why we proposed techniques that are simple in computation and in memory storage and that can reduce the dimension while considering the spatial information. To assess these techniques we used typical criteria of dimensionality reduction which evaluate the fact that some properties are kept on the manifold of data after the dimensionality reduction. Moreover, we also proposed some criteria to evaluate the quality of the image after the dimensionality reduction. Some of them are based on mathematical morphology, namely the 3D pattern spectrum and the α-flat zone to check that the reconstructed image preserves global and local similarity to the original hyperspectral image. Finally, we also perform a classification of the reduced data with different techniques. According to

the entire set of criteria, adding spatial information improves the dimensionality reduction. However, as we can see a good dimensionality reduction is obtained when we combine spatial and spectral feature space. A technique that seems to fulfil this optimum is the MorphPCA based on the distance function.

Finally, PCA has multiple applications on image processing; typically one can use the PCA to perform denoising. For example in the case of multiple images representing the same scene but corrupted by a Gaussian noise (like on multi-temporal images), it is possible to use the PCA to reduce the dimension of the temporal data and then to project the data back on the high dimensional space, so that we reduce the influence of the noise. This process can be done with MorphPCA. Moreover, another case would be to use MorphPCA on fusion of information techniques like pansharpening, which consists of increasing the spatial resolution of a multispectral or hyperspectral image thanks to a grey scale image at high resolution. Some techniques for pansharpening are based on PCA. In summary, MorphPCA can be an appropiate alternative to PCA in different image processing applications.

Author Contributions: Both authors contributed equally to the reported work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Guilfoyle, K.; Althouse, M.L.; Chang, C.I. Further investigations into the use of linear and nonlinear mixing models for hyperspectral image analysis. *Proc. SPIE* **2002**, doi:10.1117/12.478747.
2. Bachmann, C.; Ainsworth, T.; Fusina, R. Exploiting manifold geometry in hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 441–454.
3. Van der Maaten, L.J.; Postma, E.O.; van den Herik, H.J. Dimensionality reduction: A comparative review. *J. Mach. Learn. Res.* **2009**, *10*, 66–71.
4. Dalla Mura, M.; Benediktsson, J.; Waske, B.; Bruzzone, L. Morphological attribute profiles for the analysis of very high resolution images. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 3747–3762.
5. Fauvel, M.; Benediktsson, J.; Chanussot, J.; Sveinsson, J. Spectral and spatial classification of hyperspectral data using svms and morphological profiles. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 3804–3814.
6. Pesaresi, M.; Benediktsson, J. A new approach for the morphological segmentation of high-resolution satellite imagery. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 309–320.
7. Lefevre, S.; Chapel, L.; Merciol, F. Hyperspectral image classification from multiscale description with constrained connectivity and metric learning. In Proceedings of the 6th International Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS 2014), Lausannne, Switzerland, 24–27 June 2014.
8. Aptoula, E.; Lefevre, S. A comparative study on multivariate mathematical morphology. *Pattern Recognit.* **2007**, *40*, 2914–2929.
9. Li, J.; Bioucas-Dias, J.M.; Plaza, A. Hyperspectral image segmentation using a new bayesian approach with active learning. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 3947–3960.
10. Mercier, G.; Derrode, S.; Lennon, M. Hyperspectral image segmentation with markov chain model. In Proceedings of the 2003 IEEE International Geoscience and Remote Sensing Symposium, IGARSS '03, Toulouse, France, 21–25 July 2003; pp. 3766–3768.
11. Camps-Valls, G.; Bruzzone, L. Kernel-based methods for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 1351–1362.
12. Camps-Valls, G.; Gomez-Chova, L.; Mu noz-Marí, J.; Vila-Francés, J.; Calpe-Maravilla, J. Composite kernels for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* **2006**, *3*, 93–97.
13. Mathieu, F.; Jocelyn, C.; Jón Atli, B. Kernel principal component analysis for the classification of hyperspectral remote sensing data over urban areas. *EURASIP J. Adv. Signal Process.* **2009**, doi:10.1155/2009/783194.

14. Volpi, M.; Tuia, D. Spatially aware supervised nonlinear dimensionality reduction for hyperspectral data. In Proceedings of the 6th International Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS 2014), Lausannne, Switzerland, 24–27 June 2014.

15. Velasco-Forero, S.; Angulo, J. Classification of hyperspectral images by tensor modeling and additive morphological decomposition. *Pattern Recognit.* **2013**, *46*, 566–577.

16. Renard, N.; Bourennane, S. Dimensionality reduction based on tensor modeling for classification methods. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 1123–1131.

17. Bakshi, B.R. Multiscale pca with application to multivariate statistical process monitoring. *AIChE J.* **1998**, *44*, 1596–1610.

18. Franchi, G.; Angulo, J. Comparative study on morphological principal component analysis of hyperspectral images. In Proceedings of the 6th International Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS 2014), Lausannne, Switzerland, 24–27 June 2014.

19. I.G.D.F. Contest. Available online: http://www.grss-ieee.org/community/technical-committees/data-fusion/ (accessed on 17 December 2015).

20. Vincent, L. Morphological area opening and closing for greyscale images. In Proceedings of the NATO Shape in Picture Workshop, Driebergen, The Netherlands, May 1993.

21. Serra, J. *Image Analysis and Mathematical Morphology*; Academic Press, Inc.: Orlando, FL, USA, 1983.

22. Cavallaro, G.; Falco, N.; Dalla Mura, M.; Bruzzone, L.; Benediktsson, J.A. Automatic threshold selection for profiles of attribute filters based on granulometric characteristic functions. In *Mathematical Morphology and Its Applications to Signal and Image Processing*; Springer: Reykjavik, Islande, 2015; pp. 169–181.

23. Maragos, P. Pattern spectrum and multiscale shape representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **1989**, *11*, 701–716.

24. Matheron, G. *Random Sets and Integral Geometry*; John Wiley & Sons: New York, NY, USA, 1975.

25. Soille, P. *Morphological Image Analysis: Principles and Applications*; Springer Science & Business Media: Berlin, Germany, 2013.

26. Molchanov, I.S.; Teran, P. Distance transforms for real-valued functions. *J. Math. Anal. Appl.* **2003**, *278*, 472–484.

27. Goshtasby, A.A. Similarity and dissimilarity measures. In *Image Registration*; Springer: Berlin, Germany, 2012; pp. 7–66.

28. Huntington, E.V. Mathematics and statistics, with an elementary account of the correlation coefficient and the correlation ratio. *Am. Math. Mon.* **1919**, *26*, 421–435.

29. Baddeley, A. Errors in binary images and an lp version of the hausdorff metric. *Nieuw Archief voor Wiskunde* **1992**, *10*, 157–183.

30. Velasco-Forero, S.; Angulo, J.; Chanussot, J. Morphological image distances for hyperspectral dimensionality exploration using kernel-pca and isomap. In Proceedings of the 2009 IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2009, Cape Town, South Africa, 12–17 July 2009.

31. Meyer, F. The levelings. *Comput. Imaging Vis.* **1998**, *12*, 199–206.

32. Franchi, G.; Angulo, J. Quantization of hyperspectral image manifold using probabilistic distances. In *Geometric Science of Information*; Springer: Berlin, Germany, 2015; pp. 406–414.

33. Lee, J.A.; Verleysen, M. Quality assessment of dimensionality reduction: Rank-based criteria. *Neurocomputing* **2009**, *72*, 1431–1443.

34. Venna, J.; Kaski, S. Local multidimensional scaling. *Neural Netw.* **2006**, *19*, 889–899.

35. Venna, J.; Kaski, S. Neighborhood preservation in nonlinear projection methods: An experimental study. In *Artificial Neural Networks-ICANN 2001*; Springer: Berlin, Germany, 2001; pp. 485–491.

36. Mohan, A.; Sapiro, G.; Bosch, E. Spatially coherent nonlinear dimensionality reduction and segmentation of hyperspectral images. *IEEE Geosci. Remote Sens. Lett.* **2007**, *4*, 206–210.

37. He, J.; Zhang, L.; Wang, Q.; Li, Z. Using diffusion geometric coordinates for hyperspectral imagery representation. *IEEE Geosci. Remote Sens. Lett.* **2009**, *6*, 767–771.