Wavelet-Based Multicomponent Denoising Profile for the Classification of Hyperspectral Images

Pedro G. Bascoy, Pablo Quesada-Barriuso, Dora B. Heras and Fancisco Argüello

Version: accepted article

How to cite:

Copyright information:

Wavelet-based Multicomponent Denoising Profile for the Classification of Hyperspectral Images

Pedro G. Bascoy, Pablo Quesada-Barriuso, Dora B. Heras, *Member, IEEE*, and Francisco Argüello

Abstract—The high resolution of the hyperspectral remote sensing images available allows the detailed analysis of even small spatial structures. As a consequence, the study of techniques to efficiently extract spatial information is a very active realm. In this paper we propose a novel denoising wavelet-based profile for the extraction of spatial information that does not require parameters fixed by the user. Over each band obtained by a wavelet-based feature extraction technique, a denoising profile (DP) is built through the recursive application of discrete wavelet transforms (DWT) followed by a thresholding process. Each component of the DP consists of features reconstructed by recursively applying inverse wavelet transforms (IWT) to the thresholded coefficients. Several thresholding methods are explored. In order to show the effectiveness of the extended DP (EDP), we propose a classification scheme based on the computation of the EDP and supervised classification by extreme learning machine (ELM). The obtained results are compared to other state-of-the-art methods based on profiles in the literature. An additional study of behavior in the presence of added noise is also performed showing the high reliability of the EDP proposed.

Index Terms—Remote Sensing, Profile, Classification, Wavelet Transform, Denoising.

I. INTRODUCTION

CLASSIFICATION is currently an area of great interest in the field of remote sensing. The high dimensionality of hyperspectral imagery allows the development of techniques that can improve the operation of applications which target panchromatic, color and multispectral images [1], [2]. However, due to the high number of bands and the strong spectral correlation [3], feature extraction (FE) methods must be introduced to perform a dimensionality reduction previous to any further processing. Essentially, FE methods modify the original data (e.g. selecting bands or projecting data into a low dimensional space) whilst preserving as much relevant spectral information as possible. One example that retains physical data of the extracted features is band selection [4], [5], which unlike feature reduction methods, extracts information without modifying the original space of the data. On the other hand, principal component analysis (PCA) constitutes a quite popular method for feature extraction [6]–[10]. PCA estimates projections of the original data so that most of the variance is concentrated in a few components. The first principal components (PCs) thus consist of the most relevant data from the whole original features and the remaining ones are expected to contain irrelevant and/or noisy information. Independent component analysis (ICA) [10]–[12] builds components from original features as independent as possible keeping the variance among them constant, while Kernel PCA (KPCA) [13] improves PCA using a non-linear function

enabling the extraction of higher order statistics with positive results compared to the linear version. Unfortunately, all three PCA, ICA-based methods and KPCA are computationally expensive and do not take advantage of spatial information [14]. Signal analysis tools such as wavelets [15] are widely used in remote sensing. Among its recent uses, are included: feature reduction [16], [17], textural classification [18], [19], region-of-interest coding [20], data compression [21], data fusion [22], [23], noise removal [17], [24], spectral-spatial classification [25], [26], vegetation and classification analysis [27], [28], and change detection [29], [30]. Regarding FE, [31], [32] show that the classification results obtained extracting features by either PCA or wavelets are comparable, the former being significantly more efficient from a computational point of view.

Besides the spectral features, spatial information (such as shapes and sizes of structures) can be included in spectral-based classifiers in order to improve the results (spectral-spatial classification). Two groups of techniques are commonly used for the extraction of spatial information: those based on a fixed neighborhood and those based on an adaptive neighborhood. Profiles, such as the morphological profile (MP) [6], are based on adaptive neighborhoods. MPs build granulometries through the opening and closing transformations with any arbitrary structuring element over the bands extracted by any FE method. Either bright or dark structures of different sizes are highlighted depending on the size and shape of the structuring element and the estimated morphological transformation [33]. The extended morphological profile (EMP) is formed thereby fusing the MPs carried out for each band into a single dataset. The differential area profile (DAP) [34] is a feature descriptor which, unlike MPs, computes just the difference in the area metric as an attribute instead of using a structuring element, which makes it independent from its shape and size. Attribute profiles (APs), popularly considered as a generalization of MPs, have been proposed to overcome their problems [35]. APs are based on the application of morphological attribute filters (AFs) over the connected components of the image. In [36] the authors considered area and standard deviation as attributes in a completely automatic method. For further techniques based on APs, a comprehensive survey can be found in [37]. Recently, a novel approach that uses extinction filters (EFs) has been proposed [38], which unlike AFs, uses the regional extrema of attributes (maximal size of AFs) to build the so-called extinction profiles.

Although the use of wavelets with denoising purposes after an EMP computation has already been considered in [39], it should be noted that the concept of EDPs is presented for

the first time in this paper as well as its application to the classification of remote sensing hyperspectral imagery.

The aim of this paper is to contribute with a reliable, fast and accurate new approach for the extraction of spatial information in a completely automatic fashion. More precisely:

1) An approach for extracting spatial features of hyperspectral images based on creating denoising profiles (DPs) is proposed. The extended denoising profile (EDP) is built by stacking the DPs carried out for each feature-reduced band of the input image.
2) The selection of the thresholding method and the optimal maximal wavelet decomposition level is discussed.
3) The proposed profile size is small, which makes the method suitable for fast classification.

It is worth noting that the proposed approach is fully automatic, tuning its internal parameters based on the spatial dimensions of the hyperspectral image. Furthermore, a noise tolerance study shows that the approach is highly reliable when datasets are contaminated with white noise. The rest of the article is organized as follows: the mathematical formulation is presented within Section II. In Section III the approach for the classification of hyperspectral images is analyzed. Section IV discusses the thresholding method selection and compares the results to other profiles in the literature. The conclusions, final remarks and future work are included in Section V.

II. Extended Denoising Profile

In this section, the methodology for building the EDP is presented. First the application of wavelets for FE is explained as well as the denoising process based on wavelets and the thresholding process. Finally, the EDP construction is detailed.

A. Feature Extraction Using Wavelets

Wavelets are tools for signal processing analysis. A generic signal $x(n)$ can be decomposed as $a(n)$ and $d(n)$, the approximation and detail signals, through the convolution operation of a low-pass and a high-pass filter, respectively. The iterative application of these filters over $a(n)$ leads to the Mallat's decomposition tree [40].

At each tree level, the number of approximation features is reduced by the applied filters, as shown Fig. 1, where three pixel-vectors are reduced with up to three levels of wavelet transform. The approximation coefficients shown in the figure contain the most relevant information that represents the original signal. The detail coefficients however keep the information related to high frequency features of the original image and can be discarded without losing most of the variability of the pixel. The proposed scheme uses a Mallat's tree decomposition in the spectral domain to reduce the dimensionality of the images before building the EDP, retaining b' bands from the original image.

B. Wavelet Denoising and EDP

The 1D-DWT described in the previous section can be extended to be used on 2D images by applying it first by rows and then by columns (2D-DWT). A denoising approximation
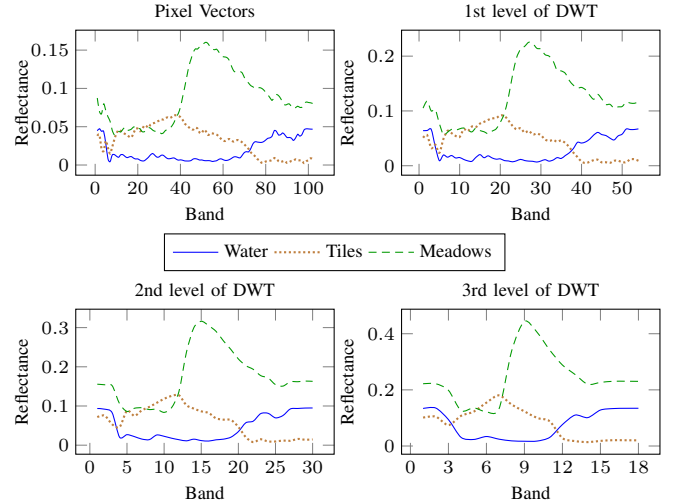

Fig. 1. Representation of three spectral signatures varying the Mallat decomposition depth level.

based on 2D-DWT applies a thresholding process at each level of decomposition. This task is commonly known as spatial wavelet denoising or wavelet thresholding. It is divided into two stages: the first one iteratively applies 2D-DWT transforms and thresholds the detail coefficients at each stage using a thresholding function T and a thresholding value λ. The second one iteratively performs a band reconstruction through the application of 2D inverse wavelet transforms (2D-IWT) to the shrinked coefficients. Depending on the T function and the λ value used in the denoising stage, the denoised image will highlight spatial structures at a completely different scale. The proposed profiles consists of the reconstructed bands obtained by varying the Mallat's tree level of decomposition in the range $l = 1, \ldots, M$ (i.e. as many times as possible) for each feature-reduced band W_i in W. M is estimated as $\log_2(\min\{h, w\})$, where h and w are the height and width of the original image. Therefore, for each feature-reduced band W_i, its DP (denoted as $\mathrm{DP}(W_i)$) will be a set containing the feature-reduced band W_i plus the denoised bands $\theta^{(l)}, \forall l$:

$$\mathrm{DP}(W_i) = \{W_i, \theta^{(1)}, \ldots, \theta^{(M)}\} . \tag{1}$$

The extended denoising profile is estimated as the set of all the DPs for all the bands W_i in W:

$$\mathrm{EDP} = \{\mathrm{DP}(W_1), \mathrm{DP}(W_2), \ldots, \mathrm{DP}(W_{b'})\} , \tag{2}$$

where b' is the number of bands of W, the feature-reduced image by wavelets.

C. Thresholding

The manipulation of the detail coefficients in the thresholding stage is carried out by a thresholding function T and a threshold value λ, transforming the k-th wavelet detail at level j, $d_j(k)$ into the denoised value $n_\lambda(d_j(k))$ according to the formula,

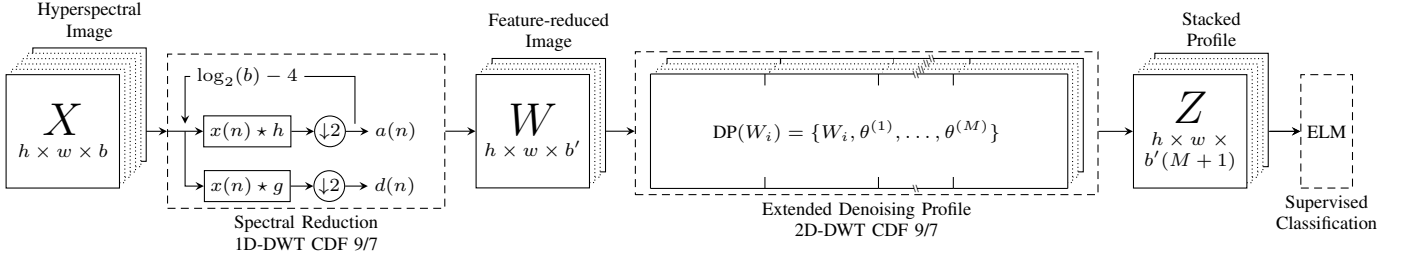$$n_\lambda(d_j(k)) = T(d_j(k), \lambda) . \tag{3}$$

Fig. 2. Complete classification flowchart based on a feature extraction using wavelets and the novel extended denoising profile (EDP).

Two different techniques have been considered to estimate the threshold value: universal and BayesShrink thresholds. Regarding the former, in the literature an universal threshold λ_U [41] was proposed for situations where the noise level is unknown. Summarizing, it is estimated as:

$$\lambda_\text{U} = \sigma \sqrt{2 \log N} \ , \tag{4}$$

where N corresponds to the length of the signal and σ to the standard deviation estimated by:

$$\sigma = \frac{\text{median}(|hh(n)|)}{0.6745} \ , \tag{5}$$

where $hh(n)$ is the signal of the first wavelet decomposition level. The latter, the BayesShrink threshold λ_B [42], [43] is estimated as:

$$\lambda_\text{B} = \frac{\sigma^2}{\sigma_x^2} \ , \tag{6}$$

where σ^2 and σ_x^2 are the noise and signal standard deviations respectively. The noise variance is calculated as the square of the median of the absolute deviation of the $hh(n)$ signal:

$$\sigma^2 = \left[\frac{\text{median}(|hh(n)|)}{0.6745} \right]^2 \ . \tag{7}$$

The estimation of the standard deviation signal corresponds to:

$$\sigma_x = \sqrt{\max\{\sigma_y^2 - \sigma^2, 0\}} \ , \tag{8}$$

where σ_y^2 is the standard deviation of the whole image.

As shown in Equation 3, any estimated threshold λ is suitable for being used in a thresholding function T. The following functions will be discussed below: hard thresholding, soft thresholding, neighboring thresholding and the removal of all the details. The first one, hard thresholding (T_H), is defined as:

$$T_\text{H}(d_j(k), \lambda) = \begin{cases} d_j(k) & \text{if } |d_j(k)| > \lambda \ , \\ 0 & \text{otherwise} \ , \end{cases} \tag{9}$$

in which $d_j(k)$ is removed if its value is lower than the threshold. Hard thresholding usually keeps edges from structures sharp at the cost of introducing salt-and-pepper-like noise in the denoised bands.

Soft thresholding (T_S) provides smoother edges in comparison to hard thresholding. This is achieved by modifying the original signal so that the removed coefficients and the remaining data present closer values:

$$T_\text{S}(d_j(k), \lambda) = \begin{cases} \text{sign}(d_j(k))|d_j(k) - \lambda| & \text{if } |d_j(k)| > \lambda \ , \\ 0 & \text{otherwise} \ . \end{cases} \tag{10}$$

The difference between hard and soft thresholding can be clearly noticed in Fig. 3. As shown, hard thresholding transforms the original continuous function in a piecewise-defined function, however, soft thresholding produces a continuous function due to the modification of the original signal.
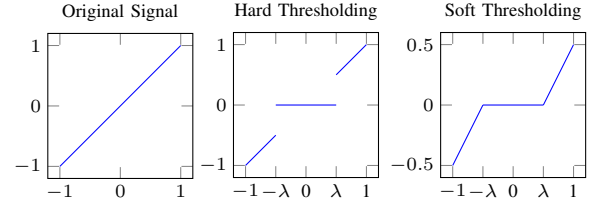

Fig. 3. Comparison between a hard thresholded and a soft thresholded signal.

The third function considered is neighboring thresholding (T_N). It is a 2D extension [44] of a wavelet shrinkage method that incorporates neighbor wavelet coefficients [45]. The formula to shrink the noisy coefficients corresponds to:

$$T_\text{N}(d_j(k), \lambda) = \begin{cases} (d_j(k)) \dfrac{1 - \lambda^2}{S_j^2(k)} & \text{if } |d_j(k)| > \lambda \ , \\ 0 & \text{otherwise} \ , \end{cases} \tag{11}$$

where λ is the universal threshold shown in Equation (4). In our implementation, a two dimensional 3×3 window is used to modify the noisy coefficients:

$$S_j^2(k) = \sum_{x=-1}^{1} \sum_{y=1}^{-1} d_j(k + x + yw)^2 \ , \tag{12}$$

where w is the number of columns of the image and k corresponds to the k-th detail of the wavelet to be shrunk.

Lastly, the removal method (T_R) consists in filling signals $lh(n), hl(n)$ and $hh(n)$ of all levels with zeros [31]:

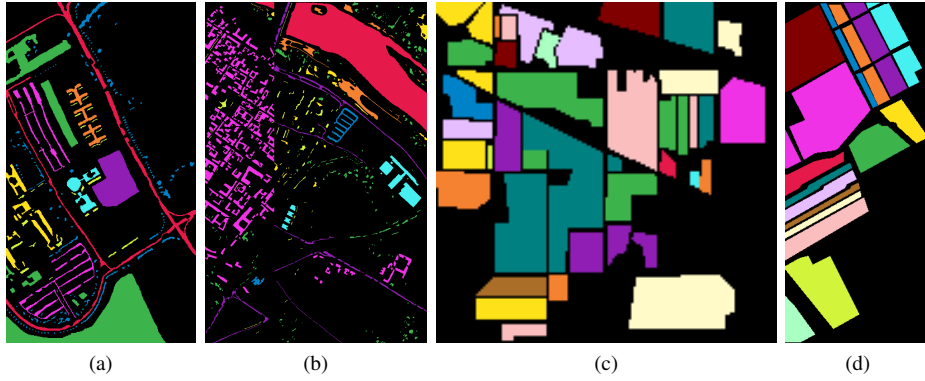$$T_\text{R}(d_j(k)) = 0 \ . \tag{13}$$

Fig. 4. Reference data for (a) Pavia University, (b) Pavia Centre, (c) Indian Pines and (d) Salinas Valley.

TABLE I
NUMBER OF TRAINING AND TEST SAMPLES FOR THE DATASETS USED IN THE EXPERIMENTAL STAGE.

Pavia University					Indian Pines									
#	Classes	Train	Test	Color	#	Classes	Train	Test	Color	#	Classes	Train	Test	Color
1.	Asphalt	548	6083		1.	Alfalfa	15	31		10.	Soybean-notill	50	915	
2.	Meadows	540	18109		2.	Corn-notill	50	1365		11.	Soybean-mintill	50	2386	
3.	Gravel	392	1707		3.	Corn-mintill	50	767		12.	Soybean-clean	50	525	
4.	Trees	524	2540		4.	Corn	50	177		13.	Wheat	50	158	
5.	Metal	265	1080		5.	Grass/pasture	50	426		14.	Woods	50	1209	
6.	Bare soil	532	4497		6.	Grass-trees	50	673		15.	Bld-Grass-Trees	50	321	
7.	Bitumen	375	955		7.	Grass-pasture-mowed	15	11		16.	Stone-Steel	50	41	
8.	Bricks	514	3168		8.	Hay-windrowed	50	430						
9.	Shadows	231	716		9.	Oats	15	5						

Pavia Centre					Salinas Valley									
#	Classes	Train	Test	Color	#	Classes	Train	Test	Color	#	Classes	Train	Test	Color
1.	Water	824	65147		1.	Brocoli gr. weeds 1	40	1969		10.	Corn gr. weeds	65	3213	
2.	Trees	820	6778		2.	Brocoli gr. weeds 2	74	3652		11.	Lettuce rom 4 weeks	21	1047	
3.	Meadows	824	2266		3.	Fallow	39	1937		12.	Lettuce rom 5 weeks	38	1889	
4.	Bricks	808	1877		4.	Fallow rough plow	27	1367		13.	Lettuce rom 6 weeks	18	898	
5.	Soil	820	5764		5.	Fallow smooth	53	2625		14.	Lettuce rom 7 weeks	21	1049	
6.	Asphalt	816	8432		6.	Stubble	79	3880		15.	Vineyard untrained	145	7123	
7.	Bitumen	808	6479		7.	Celery	71	3508		16.	Vineyard ver. trellis	36	1771	
8.	Tiles	1260	41566		8.	Grapes untrained	225	11046						
9.	Shadows	476	2387		9.	Soil vineyard dev.	124	6079						

D. EDP construction

A summary of the EDP construction is depicted in Algorithm 1. It starts off by estimating the depth of the Mallat's tree L in order to perform the FE by using 1D-DWT transforms on each pixel-vector of the input hyperspectral image. Once the FE is carried out, the estimation of the size of each DP is calculated based on the spatial dimensions of the image. Afterwards, an iterative process begins to build the EDP from the DPs built over the feature-extracted bands. The estimation of the threshold is performed using data from the first 2D-DWT decomposition level.

III. EDP-BASED SCHEME FOR THE CLASSIFICATION OF HYPERSPECTRAL IMAGES

The developed classification scheme combines the novel extended denoising profile (EDP), which includes a spectral feature reduction, with a well-known supervised classifier, Extreme Learning Machine (ELM) [46]. Fig. 2 illustrates the proposed scheme flowchart. The process begins with a feature reduction using 1D-DWT transforms over the pixel-vectors of the input hyperspectral image X, whose dimensions are h (height), w (width) and b (number of bands). The parameter that sets the number of stages of the 1D-DWT transform is calculated as:

$$L = \lceil \log_2(b) - 4 \rceil , \qquad (14)$$

where b is the number of bands of the hyperspectral image.

Filters used both in the 1D-DWT and the 2D-DWT stages correspond to the CDF 9/7 biorthogonal wavelet transform proposed by Cohen, Daubechies, and Feauveau [47] used in the JPEG 2000 compression algorithm. The next step consists in building the EDP by creating the DPs over the bands of
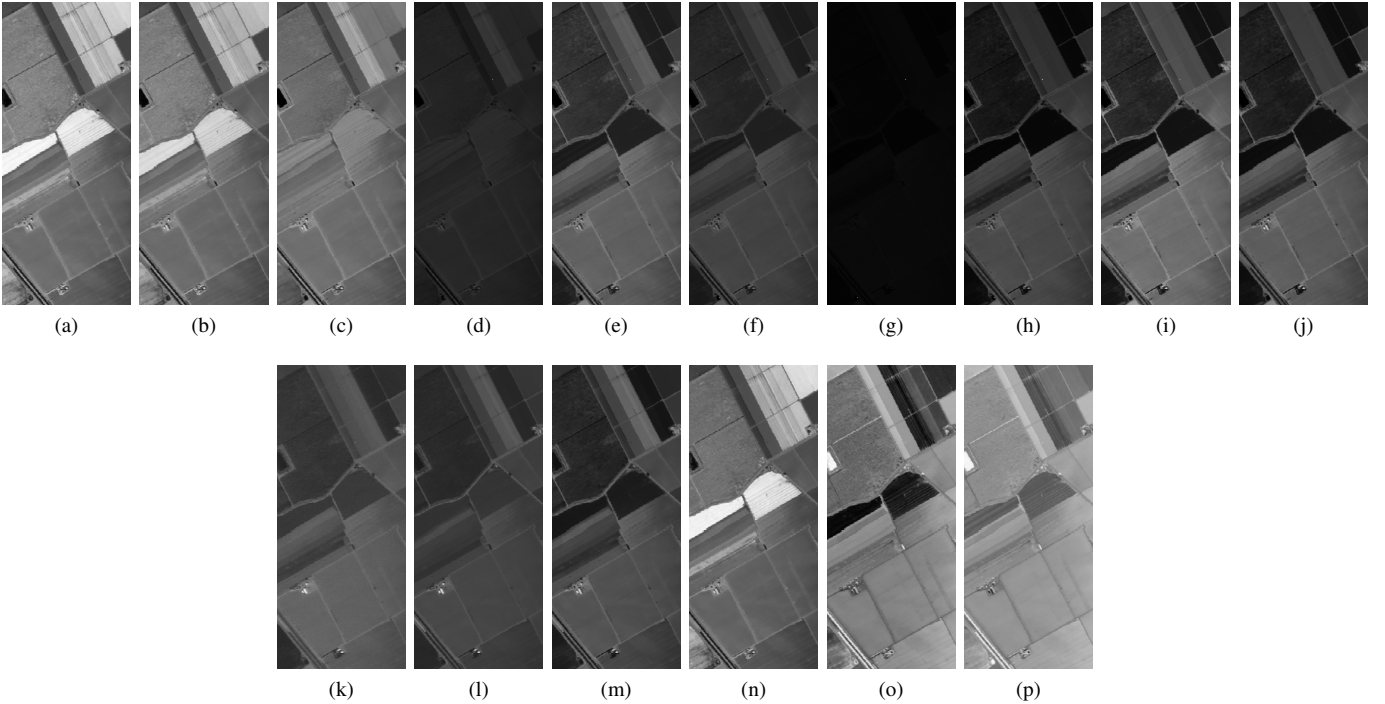
Fig. 5. Salinas dataset. Greyscale representation of the feature-reduced image W obtained by recursively applying a three level 1D-DWT *Cohen-Daubechies-Feauveau* 9/7 (CDF 9/7) transform to every pixel vector in X. (a) corresponds to the first feature-extracted band (W_1), (b) to the second (W_2) and so on.

W. Finally, the EDP features are stacked to constitute Z, the input to the supervised classification.

IV. EXPERIMENTAL RESULTS

The aim of this section is to analyze how the parameters of the proposed scheme influence the results of the classification. The results have been measured in terms of the following accuracies: the overall accuracy (OA), a percentage which grades how many pixels have been successfully classified, average accuracy (AA), the mean of the class-specific OA, and the κ statistic [48], which is the percentage of agreement corrected by the amount of it that could be expected due to chance.

A. Hyperspectral images and experimental setup

The datasets used in the experiments correspond to four images commonly used in the remote sensing literature: Pavia Centre (PaviaC), Pavia University (PaviaU), Salinas Valley (Salinas), and Indian Pines (IndianP). The first two images were obtained by the ROSIS-03 sensor over the city of Pavia (Italy) with a spatial resolution of 2.6 meters/pixel and covering the spectral range from 430 to 860 nm. In the case of PaviaC, the dimensions are 1096×715 and 102 spectral bands while in the case of PaviaU the dimensions are 610×340 pixels and 103 bands. The last two images were obtained by the AVIRIS sensor from NASA with a spectral range from 400 to 2500 nm. The main properties of these images are a resolution of 3.7 meters/pixel and dimensions of 512×217 and 224 spectral bands for Salinas, and a resolution of 20 meters/pixel with dimensions of 145×145 and 220 bands for

IndianP. As far as the training and test features are concerned, two disjoint sets were set up for each dataset, Train and Test, as presented in Table I (as in [49] and in [2]). The ground truths are shown in Fig. 4 for each dataset. Pixels with the same color represent the same land-cover class.

The number of hidden neurons of the ELM corresponds to 1000, 500, 385 and 350 respectively for PaviaU, PaviaC, IndianP and Salinas. The experiments have been carried out using MATLAB in a workstation consisting of 16 GB of RAM, a first-generation Intel Core i7 860 CPU at 2.80 GHz and a 64-bit Ubuntu 18.04 provided with gcc 7.3.0.
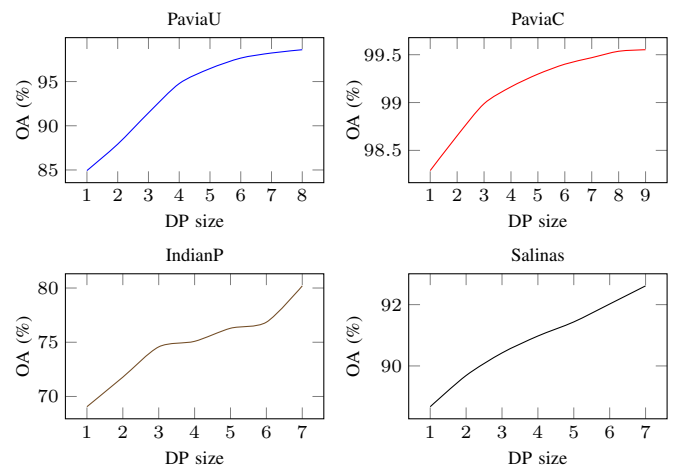

Fig. 6. OA accuracy obtained by increasing the decomposition level (l) of the 2D-DWT using soft thresholding and the universal threshold.
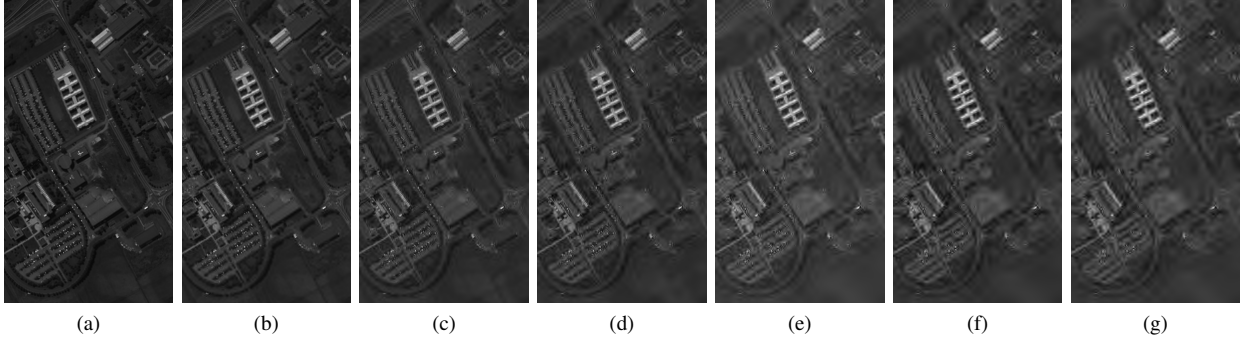
Fig. 7. PaviaU dataset. Greyscale representation of the DP elements corresponding to W_1 using hard thresholding and $\lambda = 2$. (a) corresponds to $\theta^{(1)}$, (b) corresponds to $\theta^{(2)}$ and so on.

Algorithm 1 EDP-Construction

Input:
    X: hyperspectral image
    h :   image height
    w :   image width
    b :   number of bands
Output:
    EDP: extended denoising profile

Feature Extraction:
1: $L \leftarrow \lceil \log_2(b) \rceil - 4$
2: **for each** pixel-vector in X **do**
3:     Compute a L-stage 1D-DWT
4:     Discard detail coefficients
5: **end for**

Profile Construction:
6: $M \leftarrow \lfloor \log_2(\min\{h, w\}) + 1 \rfloor$
7: EDP $\leftarrow \emptyset$
8: **for each** band in the feature-reduced image **do**
9:     DP $\leftarrow \emptyset$
10:    Add the extracted band to the DP set.
11:    **for** $l \leftarrow 1$ **to** $M - 1$ **do**
12:       **if** l is 1 **then**
13:         Estimate the threshold value λ
14:       **end if**
15:       Compute a l-stage 2D-DWT
16:       Threshold detail coefficients using λ and T.
17:       Apply a l-stage 2D-IWT transform to the thresholded coefficients.
18:       Add reconstructed band to the DP set.
19:    **end for**
20:    Add current DP to the EDP set.
21: **end for**

B. Wavelet decomposition level

Considering the numbers of bands the images (102, 103, 220 and 224) and using Eq. 14, the resulting feature-reduced image W consists of $b' = 16$ for all the images. A bandwise greyscale representation of W for the Salinas dataset is shown in Fig. 5. An experiment in order to see how M (the parameter which varies the DP size) influences the classification results was carried out. It is expected that the accuracies will improve up to a certain level when the number of denoised bands included in the DPs grows due to the addition of new granulometry levels (wavelet decomposition levels).

In Fig. 6, the OA values for each dataset varying the number of included $\theta^{(l)}, \forall l = 1, \ldots, M$ are depicted using the universal threshold and the soft thresholding function. As can be seen, the general behavior of all datasets is pretty similar: the higher the number of denoised bands included the better the accuracies achieved due to the addition of new granulometry levels. These results show thus that the features extracted by wavelets provide relevant information which the spectral classifier can take advantage of. However, as a trade-off decision between accuracy of the classification and computational time spent on it, the DP size M was set to seven, keeping the profile size small (128 features per pixel).

Fig. 7 illustrates how the number of levels of wavelet decomposition affects a band when the hard thresholding function with $\lambda = 2$ is used over the band W_1. The removal of coarse textures as a consequence of the iterative loss of irrelevant details leads to the production of flat zones, i.e., the characterization of structures within images.

C. Comparison of thresholding approaches

Different approaches to either delete or shrink detail coefficients can be used to improve classification accuracies. We presented in Section II-C four different functions and two thresholds. The aim of this section is to discuss which combination of method and threshold is best. The classification results might be inspected by looking at Table II and Fig. 8. Regarding the chart, the x-axis label is divided in a 2-level hierarchy which indicates for each dataset how the threshold is estimated ("Threshold") and the thresholding function used ("Function").

On the one hand, regarding the threshold value, Fig. 8 shows that a general trend cannot be observed. In relation to the threshold function, on the other hand, the experiments have shown that the most dramatic approximation (i.e. removing all the details in the coefficients, T_R in Table II) achieves the best accuracies with a significant difference, especially for the PaviaU, IndianP and Salinas datasets. In second position, both neighboring and soft thresholding achieve similar results
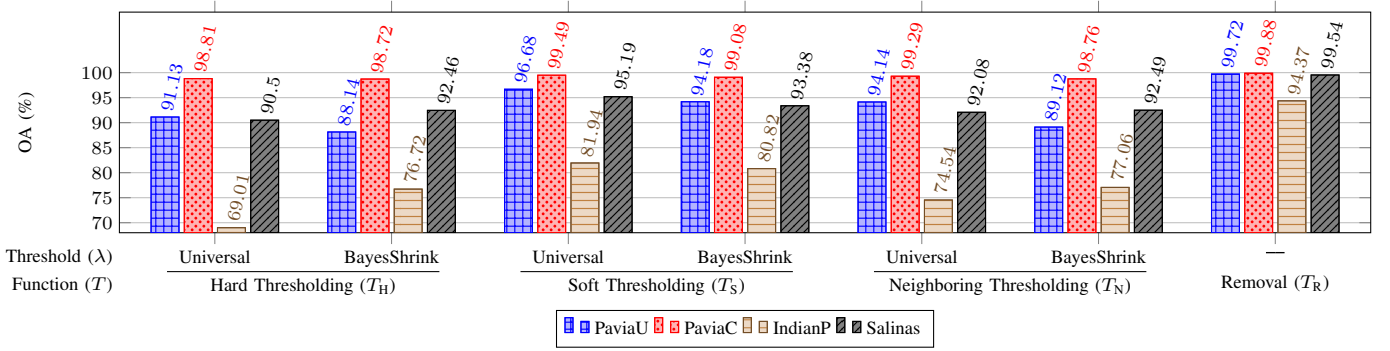
Fig. 8. OA accuracies obtained for every combination of the proposed thresholding methods and values.

TABLE II
ACCURACIES OBTAINED BY EDP FOR EVERY COMBINATION OF THE THRESHOLDING FUNCTIONS AND VALUES

| λ | Universal (λ_U) | | | BayesShrink (λ_B) | | | Universal (λ_U) | | | BayesShrink (λ_B) | | | Universal (λ_U) | | | BayesShrink (λ_B) | | | – | | |
T	Hard (T_H)			Hard (T_H)			Soft (T_S)			Soft (T_S)			Neighboring (T_N)			Neighboring (T_N)			Removal (T_R)		
	OA	AA	κ	OA	AA	κ	OA	AA	κ	OA	AA	κ	OA	AA	κ	OA	AA	κ	OA	AA	κ
PaviaU	91.13	88.79	87.99	88.14	87.83	84.09	96.68	94.23	95.47	94.18	92.20	92.09	94.14	91.65	92.01	89.12	88.21	85.36	**99.72**	**99.46**	**99.62**
PaviaC	98.81	97.24	98.27	98.72	96.81	98.15	99.49	98.84	99.26	99.08	97.89	98.66	99.29	98.35	98.97	98.76	96.97	98.19	**99.88**	**99.78**	**99.83**
IndianP	69.01	77.06	64.97	76.72	83.80	73.64	81.94	88.63	79.48	80.82	87.67	78.24	74.54	82.43	71.20	77.06	83.81	74.00	**94.37**	**96.85**	**93.55**
Salinas	90.50	93.87	89.40	92.46	95.47	91.59	95.19	96.64	94.64	93.38	95.90	92.62	92.08	95.14	91.17	92.49	95.45	91.63	**99.54**	**99.41**	**99.48**
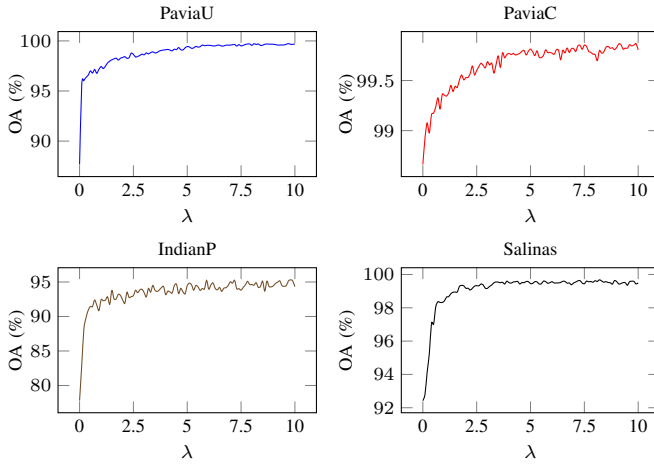
Fig. 9. Obtained OA Accuracy for each dataset varying using the hard thresholding function and varying λ in the range $[0, 0.1, \dots, 10]$.

since they reduce the impact of the removed details. Hard thresholding lies in the third position probably due to the harsh boundaries created in the denoised bands. Notice that the removal function is a special case of hard thresholding whose threshold value is ∞. Charts showing the OA values for the hard thresholding function when the threshold value is manually increased with values $[0, 0.1, \dots, 10]$ are shown in Fig. 9 for each dataset. As can be seen, in all the four cases almost monotonically increasing functions are depicted. When threshold values are high, in the tens, the accuracies tend to be constant in a range of values fairly close to the results obtained by removing all the details. The classification map obtained by the proposed classification method (EDP) when the removal threshold is applied can be seen in Fig. 10, where misclassified pixels are wrapped by white bounding circles and boxes.

D. Comparison to other classification methods

In this section a comparison between the proposed EDP-based classification scheme and relevant methods in the literature is presented. Table III summarizes the results.

TABLE III
COMPARISON BETWEEN THE PROPOSED PROFILE (EDP) AND OTHER STATE-OF-THE-ART PROFILES [2].

Dataset	Method	OA	AA	κ
PaviaU	EDP	**99.70**	**99.72**	**99.59**
	RF	71.51	82.15	64.98
	EMP	91.82	93.54	89.12
	EAP$_a$	90.33	93.47	87.71
	EEP$_a$	94.82	96.17	93.32
	EMAP	93.52	94.82	91.65
	EMEP	95.46	96.57	94.07
IndianP	EDP	**94.02**	**96.15**	**93.14**
	RF	69.36	76.55	65.41
	EMP	91.99	95.04	90.85
	EAP$_a$	91.38	93.54	90.15
	EEP$_a$	92.99	95.58	91.99
	EMAP	91.65	95.15	90.46
	EMEP	93.70	96.00	92.79
Houston	EDP	**84.53**	**86.29**	**83.21**
	RF	77.47	80.34	75.63
	EMP	80.01	82.78	78.34
	EAP$_a$	79.50	82.47	77.70
	EEP$_a$	80.32	83.36	78.66
	EMAP	78.92	82.23	77.21
	EMEP	80.83	83.64	79.20

The approaches selected in this comparison represent the state-of-the-art regarding spectral-spatial classification meth-
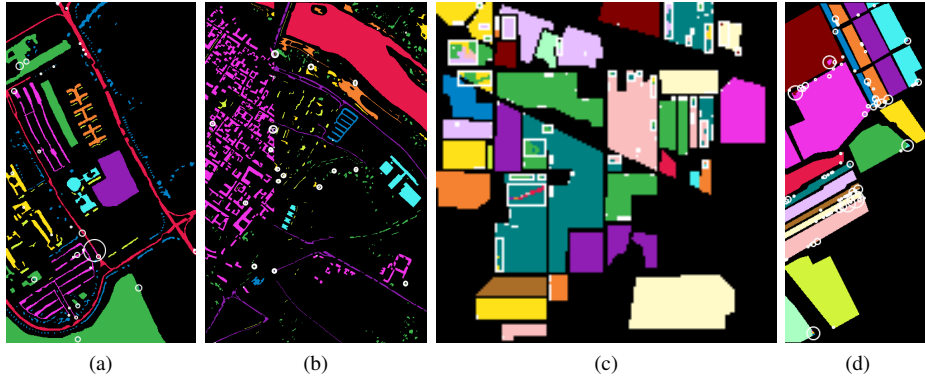
Fig. 10. Classification map of our novel EDP-based classification scheme. (a), (b), (c) and (d) correspond, respectively, to PaviaU, PaviaC, IndianP and Salinas. Areas whose pixels are misclassified have been enhanced by white bounding circles (in (a), (b) and (d)) and with white bounding boxes in (c).

ods which build profiles [2]. In order to present a fairer comparison, the setup has been slightly modified for this table according to [2], leading to the following changes:

1) The ELM classifier was replaced by a random forest (RF) with 200 trees.
2) A new dataset capturing the area over the University of Houston is considered. The image dimensions are 349×1905 with 2.5 meters of spatial resolution. The number of available bands are 144 ranging from 0.38 to 1.5 μm and the reference data consists of 15 different classes [2].
3) The 2D wavelet corresponds to the Cohen–Daubechies–Feauveau 5.

In Table III, RF stands for the classification of the raw hyperspectral features through random forest. On the other hand, the extended morphological profile (EMP) is built by opening and closing transformations using a circular structuring element of size 2, 4, 6, 8, 10, 12 and 14. Both the extended attribute profile (EAP_a) and the extended extinction profile (EEP_a) only consider area, while the extended multi-attribute profile (EMAP) and the extended multi-extinction profile (EMEP) consider the following attributes: area, height, standard deviation, volume and diagonal of the bounding box [2].

Regarding the results, including the spatial information extracted by EDP and classifying with RF overtake the accuracies obtained by not only the methods that exploit the area attribute (EMP, EAP_a and EEP_a) but the multi-attribute ones as well. This can be explained taking into account that the image simplification achieved within the features of the proposed profile is bigger and more relevant than the one achieved by EEP or EAP. Besides, unlike the methods in comparison, all the information generated by EDP is accomplished without the tuning of any inner parameter, which is a time-consuming weakness of the above mentioned methods.

E. Spatial properties of EDPs

As shown in the previous section, EDP achieves better classification results than other profiles, such as EMP. In this section we propose a comparison analyzing the structural similarity and the image simplification for both profiles built

by EDP and EMP, the latter being one of the most referenced in the literature. For the estimation of the simplification performance we used the flat-zones simplification rate [38]. It is calculated as the mean of the number of the flat zones present for each $\theta^{(l)}$ of each DP divided by the number of flat zones of the original band W_i in W. The rate varies from zero to one and the higher the value the higher the similarity between the bands and vice versa. The image similarity is shown by the structural similarity (SSIM) estimator [50] between each $\theta^{(l)}$ and the original band. It ranges from zero to one, one meaning a perfect similarity matching.
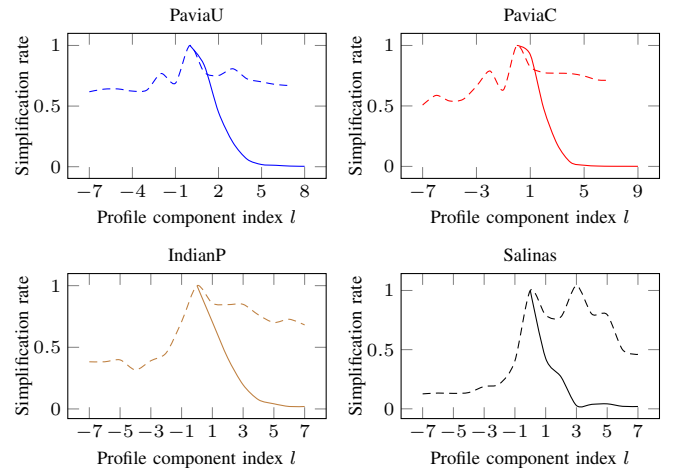

Fig. 11. Simplification rate (%) obtained for each component of the profile. EDP and EMP correspond respectively to solid and dashed lines.

In Fig. 11 we illustrate the flat-zones simplification rate for the EDP and EMP for the four datasets. Average values for the b' bands are represented. The $l = 0$ in the x-axis corresponds to the rates obtained with the feature reduced band W_i over which both EDP and EMP build the profile. In the case of EMP, the negative and positive sides of the axis corresponds to the opening and closing by reconstruction morphological operations, respectively. As can be seen, EMP and EDP present a different behavior: the former produces a higher number of flat-zones at the cost of displaying small variations throughout the components of the profile. In contrast, the latter simplifies the image in a quicker manner, keeping the smallest
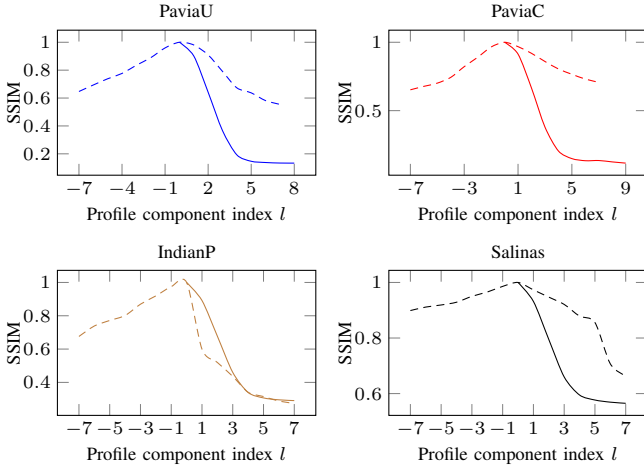
Fig. 12. Similarity rate (SSIM) (%) obtained for each component of the profile. EDP and EMP correspond respectively to solid and dashed lines.

details only in the very first components.

In relation to the SSIM index (Fig. 12), the values are lower for the EDP showing that the EMP features are more similar to the original image. In this sense, both simplification indexes show that the features within EDP summarize the spatial properties in a more rapid manner. Besides, the fixed shape of the structuring element to produce EMPs imposes a constraint to model spatial structures within a scene, making them less flexible when fitting the variability in size of the objects present in images than EDP.

On the other hand, EDP characterises structures within images by approximations at different resolution scales using wavelets, what is know as a granulometry. An additional advantage of its usage is that they have beneficial mathematical properties for spatial structure enhancement, allowing the capture of both frequency and allocation information since the scaling can be interpreted as discrete filtering. In addition, wavelets aim to extract information considering the regularity and frequency content present in the image.

It is worth noting that as for the profiles considered in the comparison, the image details are always preserved in the original feature-reduced bands within each DP, making the combination of detail preservation and big image simplification very powerful as shown in the classification results.

F. Results in the presence of noise

In this section, the results of the classification using images corrupted with Additive White Gaussian Noise (AWGN) is presented. All four datasets were contaminated by a signal-to-noise ratio (SNR) varying in the range $[5, \ldots, 30]$ dB. A peak signal-to-noise ratio (PSNR) was also estimated from the resulting corrupted images. The results can be seen in Table IV, where the best results are highlighted in bold font. Datasets corrupted by a huge amount of noise, e.g. SNR of 5 dB, improve the OA from 17.60% produced by ELM to 85.18% (in the most extreme of the cases, looking at IndianP data) using our profile method. Regarding the remaining datasets, our scheme significantly outperforms the results with any amount of noise.

TABLE IV
OA, AA VALUES (%) AND κ COEFFICIENT OF AGREEMENT OBTAINED FOR EACH DATASET IN PRESENCE OF AWGN.

Dataset	SNR (dB)	PSNR (dB)	EDP OA	EDP AA	EDP κ	ELM OA	ELM AA	ELM κ
PaviaU	5	24.40	**99.67**	**99.56**	**99.55**	58.73	64.93	48.27
	10	29.16	**99.78**	**99.69**	**99.70**	65.10	70.36	55.55
	15	34.05	**99.79**	**99.68**	**99.71**	72.64	77.15	64.55
	20	39.02	**99.78**	**99.61**	**99.70**	79.68	83.23	73.33
	25	44.01	**99.77**	**99.57**	**99.68**	84.64	87.33	79.64
	30	49.00	**99.77**	**99.54**	**99.69**	87.41	89.04	83.19
PaviaC	5	24.92	**99.69**	**99.45**	**99.54**	91.10	82.48	87.26
	10	29.62	**99.78**	**99.63**	**99.68**	95.33	89.20	93.25
	15	34.43	**99.84**	**99.70**	**99.77**	96.55	91.74	95.01
	20	39.33	**99.86**	**99.70**	**99.79**	97.30	93.49	96.09
	25	44.29	**99.87**	**99.73**	**99.81**	97.89	94.75	96.94
	30	49.27	**99.88**	**99.75**	**99.83**	98.32	95.90	97.57
IndianP	5	20.89	**85.18**	**92.45**	**83.17**	17.60	18.51	9.69
	10	25.65	**87.48**	**94.28**	**85.77**	22.85	24.11	15.15
	15	30.42	**89.15**	**95.28**	**87.67**	29.00	30.01	21.44
	20	35.20	**90.94**	**96.03**	**89.68**	35.11	38.63	28.02
	25	39.99	**91.92**	**96.37**	**90.79**	42.80	46.37	35.85
	30	44.81	**92.24**	**96.55**	**91.15**	49.19	55.85	43.11
Salinas	5	24.53	**99.05**	**99.06**	**98.95**	49.21	37.31	42.17
	10	29.21	**99.34**	**99.26**	**99.27**	58.76	49.81	53.34
	15	33.99	**99.46**	**99.41**	**99.40**	68.42	63.28	64.41
	20	38.85	**99.58**	**99.48**	**99.53**	74.16	71.54	71.00
	25	43.74	**99.57**	**99.49**	**99.52**	79.22	78.89	76.74
	30	48.65	**99.57**	**99.52**	**99.52**	83.35	86.59	81.38

Once the noise reaches 15 dB SNR, the accuracies obtained over the images are close to those results shown in Table II, indicating the high performance of the filters used for denoising.

G. Analysis of computational cost

A disadvantage of the methods in comparison shown in Section IV-D is the computational cost. Although it is often claimed that they can provide results rapidly, the amount of time for the construction of the max tree required by APs and EPs is large as compared to the computation of 2D wavelets required by EDP.

Focusing in the proposed profile, the computational cost of building the EDP is the sum of the cost of applying the wavelet filters along the spectral-spatial domain and the cost of the corresponding thresholding function, as shown in Fig. 2. Table V summarizes the complexity of each stage within the EDP construction technique. To start off, the cost of computing 1D-DWT transforms is completely linear with the image dimensionality, therefore, the cost of carrying out the FE corresponds to the cost of applying 1D-DWT wavelets in the spectral domain, i.e., the cost of the wavelet transform for every pixel-vector in the image (having h rows and w columns) and for every level of wavelet decomposition L. On the one hand, the cost of the 2D-DWT stage is the cost of applying wavelets by rows and columns for every feature-reduced band

(being b' the number of feature reduced bands) until the last decomposition level M. On the other hand, the 2D-IWT cost is the same as the cost of 2D-DWT. To sum up, the cost of building the EDP (named EDP in Table V) corresponds to the sum of the FE, 2D-DWT and 2D-IWT costs.

TABLE V
COMPLEXITY ANALYSIS OF THE DIFFERENT STAGES IN THE EDP CONSTRUCTION METHOD.

Operation	Complexity
FE by 1D-DWT	$\mathcal{O}(hwLb)$
Wavelet thresholding by 2D-DWT	$\mathcal{O}(hwMb')$
Band reconstruction by 2D-IWT	$\mathcal{O}(hwMb')$
EDP	$\mathcal{O}(hw(Lb + Mb'))$

Table VI shows a comparative of the execution time required by optimized single-core C codes for both the FE and the construction of the EDP and the EMP. The table also shows the number of features extracted (sixteen in both cases) and the size of the profiles that are built from the features. To keep the EMP size close to EDP's, the EMP is built by opening and closing transformations using a circular structuring element of size 2, 4 and 6.

Regarding the computational time, even though the size of the EMP is slightly smaller considering the same number of features extracted, the time required for building it is much higher as compared to the building of EDP profile for all the datasets, showing the great advantage of the use of wavelets.

TABLE VI
PROCESSING TIME SPENT ON COMPUTING EDP AND PCA+EMP IN SECONDS.

		Size	PaviaU	PaviaC	IndianP	Salinas
EDP	FE (1D-DWT)	16	**1.32**	**5.27**	**0.29**	**1.52**
	Extended Profile	128	**5.40**	**20.96**	**0.59**	**2.23**
	Total	–	**6.72**	**26.22**	**0.89**	**3.75**
PCA+EMP	FE (PCA)	16	4.47	20.16	0.79	4.96
	EMP	112	18.24	75.75	3.17	12.07
	Total	–	22.71	95.91	3.96	17.03

V. CONCLUSIONS AND FUTURE WORK

In this paper we presented a novel denoising profile based on wavelets and on a thresholding process for the extraction of spatial features from hyperspectral imagery. It is built over the bands extracted by discrete wavelet transforms, which reduces the spectral size of the dataset. Over them, the extended denoising profile (EDP) is created by recursively performing 2D discrete wavelet transforms, thresholding the detail coefficients at each stage and, finally, applying inverse wavelet transforms to them. The whole process is carried out in a fully automatic fashion without requiring parameter tuning by the user.

A number of techniques for thresholding the detail coefficients have been explored using four well-known datasets: Pavia University, Pavia Centre, Indian Pines and Salinas Valley. In terms of overall accuracy (OA), our experiments have shown that the best thresholding approach consists in completely removing the detail coefficients. In particular, 99.72%, 99.88%, 94.37% and 99.54% of OA were achieved for the above-mentioned datasets, respectively. Furthermore, another aim of this paper was to study the impact of the size of the profile in the results, finding that every component yields important information which the classifier can take advantage of. However, the profile size is reduced as a trade-off between accuracy and computational time.

The proposed profile has been compared to state-of-the-art spectral-spatial classification methods based on profiles, obtaining better classification accuracies than EMAP and EMEP for three well-known hyperspectral datasets. Regarding the presence of noise, our method presents a high reliability when datasets are contaminated by up to 5 dB of additive white Gaussian noise.

Provided that the EDP offers a competitive technique in order to effectively extract spatial information from hyperspectral imagery as it has been shown throughout this paper, its application to different processing tasks should be carried out. Its automatic tuning process and short execution time makes it especially appealing. The first promising area to apply it is the object-based change detection including a classification stage [51].

REFERENCES

[1] T. Lillesand, R. W. Kiefer, and J. Chipman, *Remote sensing and image interpretation*. John Wiley & Sons, 2014.

[2] P. Ghamisi, E. Maggiori, S. Li, R. Souza, Y. Tarabalka, G. Moser, A. De Giorgi, L. Fang, Y. Chen, M. Chi, S. Serpico, and J. Benediktsson, "Frontiers in spectral-spatial classification of hyperspectral images," *IEEE Geoscience and Remote Sensing Magazine*, Jul. 2018.

[3] F. Y. Kuo and I. H. Sloan, "Lifting the curse of dimensionality," *Notices of the AMS*, vol. 52, pp. 1320–1329, 2005.

[4] P. Su, D. Liu, X. Li, and Z. Liu, "A saliency-based band selection approach for hyperspectral imagery inspired by scale selection," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 4, pp. 572–576, April 2018.

[5] S. Jia, G. Tang, J. Zhu, and Q. Li, "A novel ranking-based clustering approach for hyperspectral band selection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 1, pp. 88–102, 2016.

[6] J. A. Benediktsson and P. Ghamisi, *Spectral-spatial classification of hyperspectral remote sensing images*. Artech House, 2015.

[7] X. Tong, H. Xie, and Q. Weng, "Urban land cover classification with airborne hyperspectral data: What features to use?" *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 10, pp. 3998–4009, Oct 2014.

[8] W. Cao, K. Wang, G. Han, J. Yao, and A. Cichocki, "A robust PCA approach with noise structure learning and spatial–spectral low-rank modeling for hyperspectral image restoration," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, no. 99, pp. 1–17, 2018.

[9] J. Wang, C. Luo, H. Huang, H. Zhao, and S. Wang, "Transferring pre-trained deep CNNs for remote scene classification with general features learned from linear PCA network," *Remote Sensing*, vol. 9, no. 3, p. 225, 2017.

[10] P. Ghamisi, N. Yokoya, J. Li, W. Liao, S. Liu, J. Plaza, B. Rasti, and A. Plaza, "Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 37–78, Dec 2017.

[11] J. Xia, L. Bombrun, T. Adalı, Y. Berthoumieu, and C. Germain, "Spectral–spatial classification of hyperspectral images using ICA and edge-preserving filter via an ensemble strategy," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4971–4982, Aug 2016.

[12] N. Falco, J. A. Benediktsson, and L. Bruzzone, "Spectral and spatial classification of hyperspectral images based on ICA and reduced morphological attribute profiles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 11, pp. 6223–6240, Nov 2015.

[13] M. P. Uddin, M. A. Mamun, and M. A. Hossain, "Feature extraction for hyperspectral image classification," in *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, Dec 2017, pp. 379–382.

[14] X. Kang, S. Li, and J. A. Benediktsson, "Feature extraction of hyperspectral images with image fusion and recursive filtering," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 6, pp. 3742–3752, June 2014.

[15] C. K. Chui, *An introduction to wavelets*. Elsevier, 2016.

[16] S. K. Meher, B. U. Shankar, and A. Ghosh, "Wavelet-feature-based classifiers for multispectral remote-sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 6, pp. 1881–1886, 2007.

[17] P. Quesada-Barriuso, F. Argüello, and D. B. Heras, "Spectral-spatial classification of hyperspectral images using wavelets and extended morphological profiles," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 4, pp. 1177–1185, April 2014.

[18] R. A. Ansari and K. M. Buddhiraju, "Textural classification based on wavelet, curvelet and contourlet features," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2016, pp. 2753–2756.

[19] O. Regniers, L. Bombrun, V. Lafon, and C. Germain, "Supervised classification of very high resolution optical images using wavelet-based textural features," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 6, pp. 3722–3735, June 2016.

[20] L. Zhang, J. Chen, and B. Qiu, "Region-of-interest coding based on saliency detection and directional wavelet for remote sensing images," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 1, pp. 23–27, Jan 2017.

[21] N. Amrani, J. Serra-Sagristà, V. Laparra, M. W. Marcellin, and J. Malo, "Regression wavelet analysis for lossless coding of remote-sensing data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 9, pp. 5616–5627, Sept 2016.

[22] L. Zhang and J. Zhang, "A new saliency-driven fusion method based on complex wavelet transform for remote sensing images," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 12, pp. 2433–2437, Dec 2017.

[23] F. Palsson, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "Model-based fusion of multi- and hyperspectral images using PCA and wavelets," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2652–2663, May 2015.

[24] P. C. Tay and Y. Yan, "Wavelet denoising using a conjointly space and 2D frequency localized filterbank," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, Oct 2018, pp. 520–524.

[25] X. Zhou, S. Prasad, and M. M. Crawford, "Wavelet-domain multiview active learning for spatial-spectral hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 9, pp. 4047–4059, Sept 2016.

[26] L. He, Y. Li, X. Li, and W. Wu, "Spectral–spatial classification of hyperspectral images via spatial translation-invariant wavelet-based sparse representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2696–2712, May 2015.

[27] R. He, H. Li, X. Qiao, and J. Jiang, "Using wavelet analysis of hyperspectral remote-sensing data to estimate canopy chlorophyll content of winter wheat under stripe rust stress," vol. 39, pp. 4059–4076, 06 2018.

[28] D. Li, T. Cheng, K. Zhou, H. Zheng, X. Yao, Y. Tian, Y. Zhu, and W. Cao, "Wrep: A wavelet-based technique for extracting the red edge position from reflectance spectra for estimating leaf and canopy chlorophyll contents of cereal crops," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 129, pp. 103 – 117, 2017.

[29] Z. Li, W. Shi, H. Zhang, and M. Hao, "Change detection based on Gabor wavelet features for very high resolution remote sensing images," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 5, pp. 783–787, May 2017.

[30] H. Zhuang, K. Deng, Y. Yu, and H. Fan, "An approach based on discrete wavelet transform to unsupervised change detection in multispectral images," *International Journal of Remote Sensing*, vol. 38, no. 17, pp. 4914–4930, 2017.

[31] S. Kaewpijit, J. L. Moigne, and T. El-Ghazawi, "Automatic reduction of hyperspectral imagery using wavelet spectral analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 4, pp. 863–871, April 2003.

[32] M. Abdolmaleki, N. Fathianpour, and M. Tabaei, "Evaluating the performance of the wavelet transform in extracting spectral alteration features from hyperspectral images," *International Journal of Remote Sensing*, pp. 1–19, 2018.

[33] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 2, pp. 309–320, Feb 2001.

[34] G. K. Ouzounis and P. Soille, "Differential area profiles," in *2010 20th International Conference on Pattern Recognition*, Aug 2010, pp. 4085–4088.

[35] M. D. Mura, A. Villa, J. A. Benediktsson, J. Chanussot, and L. Bruzzone, "Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 3, pp. 542–546, May 2011.

[36] P. Ghamisi, J. A. Benediktsson, and J. R. Sveinsson, "Automatic spectral-spatial classification framework based on attribute profiles and supervised feature extraction." *IEEE Trans. Geoscience and Remote Sensing*, vol. 52, no. 9, pp. 5771–5782, 2014.

[37] P. Ghamisi, M. D. Mura, and J. A. Benediktsson, "A survey on spectral-spatial classification techniques based on attribute profiles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2335–2353, May 2015.

[38] P. Ghamisi, R. Souza, J. A. Benediktsson, X. X. Zhu, L. Rittner, and R. A. Lotufo, "Extinction profiles for the classification of remote sensing data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp. 5631–5645, Oct 2016.

[39] P. Quesada-Barriuso, D. B. Heras, and F. Argüello, "Exploring the impact of wavelet-based denoising in the classification of remote sensing hyperspectral images," in *SPIE Remote Sensing 2016*, F. B. Lorenzo Bruzzone, Ed., 2016.

[40] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, Jul 1989.

[41] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[42] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1532–1546, Sep 2000.

[43] D. L. Donoho and J. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[44] G. Chen, T. D. Bui, and A. Krzyzak, "Image denoising using neighbouring wavelet coefficients," *Integrated Computer-Aided Engineering*, vol. 12, no. 1, pp. 99–107, 2005.

[45] T. T. Cai and B. W. Silverman, "Incorporating information on neighbouring coefficients into wavelet estimation," *Sankhyā: The Indian Journal of Statistics, Series B*, pp. 127–148, 2001.

[46] M. Pal, "Extreme-learning-machine-based land cover classification," *International Journal of Remote Sensing*, vol. 30, no. 14, pp. 3835–3841, 2009.

[47] C. Albert, D. Ingrid, and F. J-C, "Biorthogonal bases of compactly supported wavelets," *Communications on pure and applied mathematics*, vol. 45, no. 5, pp. 485–560, 1992.

[48] A. Viera and J. Garrett, "Understanding interobserver agreement: The kappa statistic," *Family Medicine*, vol. 37, no. 5, pp. 360–363, 5 2005.

[49] P. R. Marpu, M. Pedergnana, M. D. Mura, S. Peeters, J. A. Benediktsson, and L. Bruzzone, "Classification of hyperspectral data using extended attribute profiles based on supervised and unsupervised feature extraction techniques," *International Journal of Image and Data Fusion*, vol. 3, no. 3, pp. 269–298, 2012.

[50] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.

[51] J. López-Fandiño, A. S. Garea, D. B. Heras, and F. Argüello, "Stacked autoencoders for multiclass change detection in hyperspectral images," in *International Geoscience and Remote Sensing Symposium*, 2018.