

Copyright notice

Copyright 2016 Society of Photo-Optical Instrumentation Engineers (SPIE). One print or electronic copy may be made for personal use only. Systematic reproduction and distribution, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper are prohibited

Citation

Pablo Quesada-Barriuso, Dora B. Heras, and Francisco Argüello "Exploring the impact of wavelet-based denoising in the classification of remote sensing hyperspectral images", Proc. SPIE 10004, Image and Signal Processing for Remote Sensing XXII, 100040R (18 October 2016); <https://doi.org/10.1117/12.2240854>

DOI

<https://doi.org/10.1117/12.2240854>

Exploring the Impact of Wavelet-based Denoising in the Classification of Remote Sensing Hyperspectral Images

Pablo Quesada-Barriuso^{*a}, Dora B. Heras^a, and Francisco Argüello^b

^aCentro Singular de Investigación en Tecnoloxías da Información (CiTIUS),

^bDepartamento de Electrónica e Computación, Universidade de Santiago de Compostela, Spain

ABSTRACT

The classification of remote sensing hyperspectral images for land cover applications is a very intensive topic. In the case of supervised classification, Support Vector Machines (SVMs) play a dominant role. Recently, the Extreme Learning Machine algorithm (ELM) has been extensively used. The classification scheme previously published by the authors, and called WT-EMP, introduces spatial information in the classification process by means of an Extended Morphological Profile (EMP) that is created from features extracted by wavelets. In addition, the hyperspectral image is denoised in the 2-D spatial domain, also using wavelets and it is joined to the EMP via a stacked vector. In this paper, the scheme is improved achieving two goals. The first one is to reduce the classification time while preserving the accuracy of the classification by using ELM instead of SVM. The second one is to improve the accuracy results by performing not only a 2-D denoising for every spectral band, but also a previous additional 1-D spectral signature denoising applied to each pixel vector of the image. For each denoising the image is transformed by applying a 1-D or 2-D wavelet transform, and then a NeighShrink thresholding is applied. Improvements in terms of classification accuracy are obtained, especially for images with close regions in the classification reference map, because in these cases the accuracy of the classification in the edges between classes is more relevant.

Keywords: Remote sensing, Land cover classification, Hyperspectral analysis, Wavelet transform, Feature extraction, Morphological profiles, Denoising, Spectral-Spatial processing.

1. INTRODUCTION

Hyperspectral sensors have the ability to sense electromagnetic signals in a very detailed spectral resolution. The availability of hundreds of spectral bands at different wavelength channels offers possibilities for extracting information that can be useful for different applications. In the case of remote sensing, the problems that are typically addressed for such hyperspectral data include anomaly detection, target recognition, and background characterization, including land-cover classification,¹ which is the aim of this work.

Although many methods have been applied to different problems associated with the classification of multi-spectral and hyperspectral data,²⁻⁵ Support Vector Machines² (SVMs) are generally recognized as the supervised classification method that offers the best results in terms of accuracy of the classification.⁶ Also considering supervised classification, in the field of neural networks, Extreme Learning Machines (ELMs) have been recently used in remote sensing⁷⁻⁹ and slightly improve SVM results. ELM is the name for a class of single-hidden layer feedforward neural networks with random weights.¹⁰ It is important to note that the computational requirements of the classification of hyperspectral datasets are high, so the computational efficiency in terms of execution time of the methods must be considered when they are designed. In this sense ELM and ELM-based classification techniques for remote sensing hyperspectral images have been compared to similar SVM-based techniques¹¹ showing much lower execution time.

The accuracy of the results obtained with a pixel-wise classifier such as SVM or ELM when only spectral information is considered is limited.⁴ In general, the accuracy of the classification of hyperspectral data is improved when information corresponding to the spatial structures present in the image is also introduced.⁵ The

^{*}*Corresponding author.* Send correspondence to CiTIUS, Rúa de Jenaro de la Fuente Domínguez, 15782, Santiago de Compostela, Spain. E-mail addresses: {pablo.quesada, dora.blanco, francisco.arguello}@usc.es.

methods for extracting spatial information include those based on segmentation,^{12,13} on partitional clustering techniques,¹⁴ on minimum spanning forest,¹⁵ or on local filtering¹⁶ among others.

Another common approach in order to extract the spatial information is to consider Mathematical Morphology¹⁷ (MM). Using the morphological operators known as erosion and dilation a Morphological Profile (MP) can be built¹⁸ from hyperspectral data. These profiles contain information on the spatial structures on the image at different resolutions. The MPs could be computed over each band of the hyperspectral image but this would generate a huge amount of data with redundancy. For this reason different feature extraction (FE) techniques are applied before calculating the MP. For example, Principal Component Analysis (PCA) is used for building MPs^{19,20} and, if several components are retained, then a so-called Extended Morphological Profile^{21,22} (EMP) is obtained. Other FE methods can be used. For example, independent component analysis as in Ref. 18, Decision Boundary Feature Extraction (DBFE) and Non-parametric Weighted Feature Extraction (NWFE) as in Refs. 19 and 21, or wavelets as proposed in Ref. 23, have been used for building extended morphological profiles. If attribute operators related, for example, to the size or geometry of the spatial structures in the image are considered instead of morphological operations, then the profile is called Attribute Profile²⁴ (AP). The profile is called EMAP²⁵ if multiple attributes and several components are used for extracting the spatial information.

In order to combine the spatial and the spectral information different methods have been used being majority voting^{12,15} and regularization²⁶ two common approaches. Marker selection is also used in combination with SVM in different techniques.⁵ In other cases a simpler method consisting in concatenating the spectral and spatial features via stacked vectors²⁷ is used. The data fusion process can also be performed via kernel methods.^{28,29} Data fusion can also be implicitly performed by concatenating different methods. For example, Empirical Mode Decomposition (EMD) has been proposed to spatially decompose hyperspectral bands,³⁰ and in this scheme 1-D discrete wavelet transforms (DWTs) are used to enhance spectral information from the EMD intrinsic functions.³¹ So, data fusion is implicitly performed by applying the spectral wavelet transform along the data found by the 2-D EMD.

Hyperspectral data suffers from atmospheric haze or instrument noise.³² Noise in remote sensing is identified as additive, multiplicative or impulsive noise³³ (or a mixture of them). Gaussian additive noise in hyperspectral images is considered in this paper. Different noise reduction techniques have been presented of which wavelets are commonly used as a common pre-processing stage. For example, a hyperspectral image band by band denoising^{23,34} improves the classification accuracy as compared to the original noisy image. BM4D³⁵ is a denoising technique based on block-matching filtering applied to volumetric datasets. The idea is to extract similar patches from 3-D data and stack them into a volume that is filtered using a discrete cosine transform. A discrete Fourier transform and 2-D discrete wavelet transform (2-D DWT) are proposed in Ref. 36 for spectral and spatial denoising, respectively. In general, if the spectral and the spatial information are considered by the denoising method, better results in terms of peak signal-to-noise ratio (PSNR) and classification accuracy are expected.

A 3-D denoising method (SarUWT)³⁷ based on sparse regularization with 3-D undecimated wavelet transform outperformed in the literature other 2-D and 3-D wavelet-based denoising schemes, such as BiShri2D/3D.^{38,39} A wavelet-based denoising method on components extracted by PCA (PCABiShr)⁴⁰ applies 2-D bivariate shrinkage (BiShri2D) in the spatial domain and 1-D dual-tree complex wavelet denoising⁴¹ in the spectral domain for each principal component. First order roughness penalty based on enforcing smoothness on spectra has been also used for denoising (FORPDN)⁴² hyperspectral data. The proposed cost function is formulated in the wavelet domain to exploit the multiresolution property of wavelets. Whatever the wavelet-based denoising method used, the noisy coefficients in the wavelet domain are shrunk based on a threshold, while the noisy free wavelet coefficients remain unchanged (hard-thresholding) or are attenuated⁴³ (soft-thresholding). Several algorithms can be used to estimate threshold values, such as the universal threshold or VisuShrink,⁴³ the BayesShrink,⁴⁴ or Stein's Unbiased Risk Estimator (SURE) shrinkage method.⁴⁵ In addition, wavelet coefficients at different scales can be used in the thresholding formula, such as the BiShri2D scheme. Cai and Silverman⁴⁶ proposed a wavelet shrinkage method that incorporates neighbour wavelet coefficients into the thresholding formula. As the wavelet transform produces correlated coefficients, large wavelet coefficients will probably have large wavelet coefficients at their neighbours. This approach was extended to the 2-D case in Ref. 47 using a two dimensional window and the method was named *NeighShrink*.

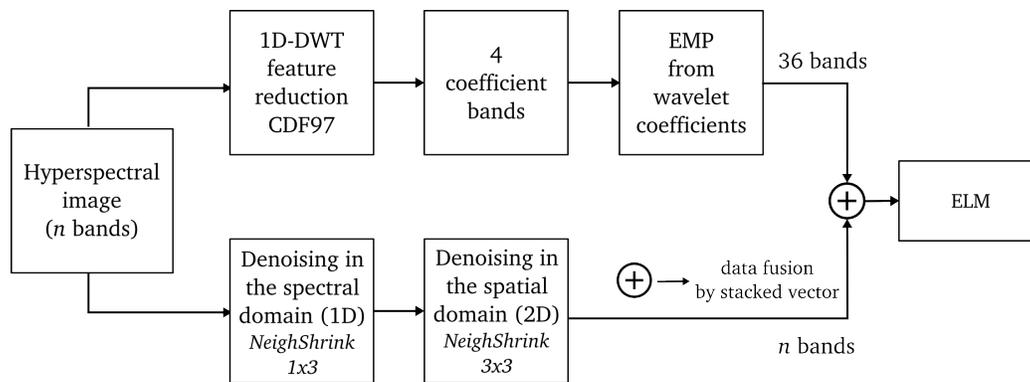


Figure 1. Flow-chart of the WTSS-EMP spectral-spatial classification scheme based on feature extraction and denoising by wavelets, and mathematical morphology.

In this paper the spectral-spatial classification scheme called WT-EMP presented in Ref. 23 is extended mainly by changing the classification method and modifying its denoising stage. The original scheme is based on EMPs and classification by SVM. The EMP is built from features extracted by wavelets and it is combined with the denoised image via stacked vectors. The denoising is performed band by band using 2-D wavelet-based denoising, with the objective of removing undesirable artifacts introduced in the acquisition of the data collected by the hyperspectral sensor. In this paper two goals are achieved reducing the computational cost of the method and improving its accuracy. In order to reduce the computational cost of the method an ELM classifier is introduced. In order to improve the accuracy a new denoising stage operating also on the spectral dimension is applied producing a new scheme called WTSS-EMP. Neighbour wavelet coefficients are used in 1-D and 2-D with an universal threshold for denoising the hyperspectral image.

The rest of the paper is organized as follows. Section 2 describes the spectral-spatial classification scheme based on wavelet and morphological profiles and presents the improvements proposed in this work. Section 3 discusses the results over different hyperspectral datasets and compares the classification to other classification schemes. Finally, Section 4 presents the conclusions and future works.

2. SPECTRAL-SPATIAL CLASSIFICATION SCHEME BASED ON WAVELETS AND EXTENDED MORPHOLOGICAL PROFILES

In this section we present the proposed classification scheme, named WTSS-EMP and based on WT-EMP. The main changes regarding WT-EMP are first that ELM is used instead of SVM and that a new denoising stage is applied. Figure 1 shows the proposed classification scheme. The upper branch in Fig. 1 illustrates the spatial processing. A 1-D DWT is applied in the spectral domain to reduce the dimensionality of the hyperspectral image and the EMP is built from the remaining coefficient bands. The wavelet-based denoising stage is illustrated in the lower branch in Fig. 1. A 1-D spectral denoising to each pixel vector and a 2-D spatial denoising to each hyperspectral band are applied using neighbour coefficients for wavelet shrinkage. The two branches are combined via stacked vectors prior to the supervised classification by ELM. The following sections explain in detail the different steps of this scheme.

2.1 Wavelet-based denoising

Denoising by wavelets plays a key role in the proposed scheme. Wavelet-based denoising applied band by band, as in the WT-EMP scheme, improves the accuracy on hyperspectral image classification. In this work, this stage is improved by a spectral denoising to each pixel vector on the hyperspectral image. By exploiting spectral and spatial denoising better results in terms of classification accuracy are expected, as it will be shown in Section 3.

Wavelets are mathematical tools for signal processing analysis at different scales.⁴⁸ The discrete wavelet transform (DWT) of a signal x can be computed as the convolution of x with low-pass and high-pass filters. The low-pass filters generate an approximation of the signal (a), whereas the high-pass filters highlight its details (d).

This wavelet decomposition can be recursively applied to the approximation signal that is decomposed into a new pair (a, d) of approximation and detail coefficients. Both wavelet coefficients, a and d , are down-sampled to produce half the coefficients at each new level of decomposition j . This multi-resolution analysis using filters is known as the Mallat-tree decomposition.⁴⁹

Signal denoising by wavelets is a common task known as wavelet thresholding or wavelet shrinkage. The smallest high frequency coefficients are usually considered as noise and might be reduced without substantially affecting the original signal. A general wavelet-based denoising scheme can be summarized as follows:⁴³

1. Transform the image x into an orthogonal domain by a discrete wavelet transform.
2. Apply a thresholding formula to the detail wavelet coefficients d using a threshold λ .
3. Performs an inverse discrete wavelet transform to reconstruct the original image with lower amount of noise.

In addition, the thresholding of step 2 can be applied remaining unchanged the noisy free wavelet coefficients or attenuating them by hard-thresholding or soft-thresholding, respectively. The following nonlinear transform is used for soft-thresholding:⁴³

$$\eta_\lambda(d_{j,k}) = \text{sgn}(d_{j,k})(|d_{j,k}| - \lambda)_+, \quad (1)$$

where λ is the threshold and $d_{j,k}$ the k th detail wavelet coefficient at level j to be denoised. The subscript $+$ in Eq. (1) indicates that the coefficients $|d_{j,k}| > \lambda$ are attenuated, whereas the rest are set to zero.

Donoho⁴³ proposed an universal threshold

$$\lambda = \sigma\sqrt{2 \log N} \quad (2)$$

for situations where the noise level is unknown. In Eq. (2), N is the length of the signal and σ is the noise standard deviation estimated by $\sigma = \frac{\text{median}(|\mathbf{d}_1|)}{0.6745}$, with \mathbf{d}_1 the detail wavelet coefficients obtained at the first level of decomposition. Other approaches can be used to estimate threshold values, such as the BayesShrink⁴⁴ or the SURE shrinkage method.⁴⁵ In this work, soft-thresholding is applied and the universal threshold λ of Eq. (2) is used for 1-D and 2-D wavelet denoising.

A 1-D wavelet shrinkage method that incorporates neighbour wavelet coefficients was proposed in Ref. 46 and it was extended to the 2-D case in the work presented in Ref. 47. As mentioned above, the wavelet transform produces correlated coefficients. So, by considering the neighbourhood of each detail coefficient centered on $d_{j,k}$, each coefficient is influenced by its neighbour to be (or not) shrunk by soft-thresholding. In the WTSS-EMP scheme, a 1-D spectral and a 2-D spatial denoising using neighbour coefficients are applied to the hyperspectral image.

For 1-D denoising, the hyperspectral image is first transformed into an orthogonal domain by a 1-D DWT. If \mathbf{d}_j is the set of detail wavelet coefficients of the transformed image at level j ,

$$S_{j,k}^2 = d_{j,k-1}^2 + d_{j,k}^2 + d_{j,k+1}^2 \quad (3)$$

defines a 1×3 neighbourhood window for each coefficient k , with the magnitude of the square sum of all the wavelet coefficients centered on $d_{j,k}$. The following thresholding formula proposed in Ref. 46 is applied to shrink the noisy coefficients:

$$d_{j,k} = d_{j,k}(1 - \lambda^2/S_{j,k}^2)_+ \quad (4)$$

where λ is the universal threshold and k indicates the k th detail wavelet coefficient at level j . The subscript $+$ in Eq. (4) indicates to keep the value if positive or set it to zero otherwise. After thresholding by Eq. (4), the inverse wavelet transform is applied to reconstruct the hyperspectral image.

For 2-D denoising, a separable 2-D DWT is applied to each hyperspectral band. The high frequency subbands created by this wavelet decomposition are thresholded by Eq. (4) but using a two dimensional window:

$$S_{j,k}^2 = \sum_{x=-1, y=-1}^{x=1, y=1} d_{x,y}^2 \quad (5)$$

with $d_{x,y}$ the set of neighbour wavelet coefficients centered on the 3×3 neighbourhood window defined by Eq. (5). Finally, each hyperspectral band is reconstructed by an inverse 2-D wavelet transform.

2.2 Supervised classification by ELM

As we have mentioned in the introduction, ELM is a classification technique that has been introduced in the classification scheme for classification of hyperspectral images^{9,50} offering good results (as it will be shown in the results section). It is a supervised learning technique for a class of Single-hidden Layer Feedforward Neural Networks (SLFN) with random weights.¹⁰ Figure 2 shows the structure of an SLFN. Given an input \mathbf{x} , the output function of an SLFN with L hidden nodes can be written as:

$$f(\mathbf{x}) = \sum_{i=1}^L \beta_i G(\mathbf{a}_i, b_i, \mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^d, \beta_i \in \mathbb{R}^m, \quad (6)$$

where $G(\mathbf{a}_i, b_i, \mathbf{x})$ denotes the output function of the i th hidden node and β_i , \mathbf{a}_i , and b_i are the weights and biases which must be generated in the training phase. The output function G of the hidden node in Eq. (6) can be expressed with an activation function g (usually a sigmoidal function) as

$$G(\mathbf{a}_i, b_i, \mathbf{x}) = g(\mathbf{a}_i \cdot \mathbf{x} + b_i), \quad \mathbf{a}_i \in \mathbb{R}^d, b_i \in \mathbb{R}. \quad (7)$$

An SLFN with L hidden nodes can approximate N arbitrary distinct samples and targets $(\mathbf{x}_i, \mathbf{t}_i) \in \mathbb{R}^d \times \mathbb{R}^m$, if the following equation system can be solved:

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{T}, \quad (8)$$

where \mathbf{H} is called the hidden layer output matrix of the neural network

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}(x_1) \\ \vdots \\ \mathbf{h}(x_N) \end{bmatrix} = \begin{bmatrix} G(\mathbf{a}_1, b_1, \mathbf{x}_1) & \cdots & G(\mathbf{a}_L, b_L, \mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ G(\mathbf{a}_1, b_1, \mathbf{x}_N) & \cdots & G(\mathbf{a}_L, b_L, \mathbf{x}_N) \end{bmatrix}_{N \times L} \quad (9)$$

and $\boldsymbol{\beta} = [\beta_1^T \dots \beta_N^T]^T$, and $\mathbf{T} = [\mathbf{t}_1^T \dots \mathbf{t}_N^T]^T$ the training data target matrix. The hidden node parameters (\mathbf{a}_i, b_i) remain fixed after being randomly generated and training an SLFN is equivalent to finding a least-squares solution $\hat{\boldsymbol{\beta}}$ of the linear system $\mathbf{H}\boldsymbol{\beta} = \mathbf{T}$, i.e.,

$$\hat{\boldsymbol{\beta}} = \mathbf{H}^\dagger \mathbf{T}, \quad (10)$$

where \mathbf{H}^\dagger is the *Moore-Penrose generalized inverse*⁵¹ of matrix \mathbf{H} . For further details on ELM, we refer the reader to Refs. 6 and 10.

The parameter configuration of the ELM requires only setting the number of nodes in the hidden layer. The accuracy of the ELM classifiers, as in any other supervised classification method, depends on the number of the training samples.

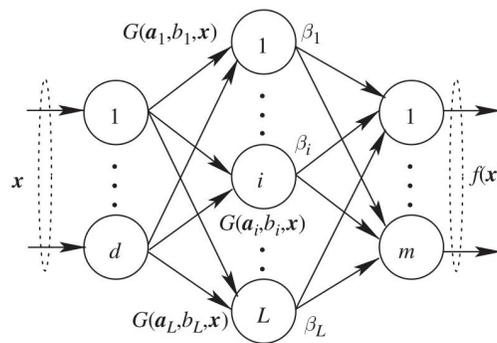


Figure 2. A single-hidden layer feedforward network (SLFN).

2.3 Spatial processing

In the proposed WTSS-EMP scheme, the spatial information is introduced by the EMP that is built from the main characteristics of the spectral signature. As in WT-EMP introduced in Ref. 23, the main features are retained by a 1-D wavelet transform applied to the spectral dimension of the hyperspectral data.

Wavelet transform methods have been proposed for dimensionality reduction in the spectral domain. As described in Section 2.1, a multi-resolution wavelet analysis recursively decomposes a signal in approximation and detail coefficients. Figure 3 shows an example of a 2-level decomposition of a pixel vector of 103 spectral bands. Only the approximation coefficients are shown in the figure. As can be observed in Fig. 3, increasing the level of decomposition implies halving the number of coefficient bands and smoothing the original pixel vector. The feature reduction technique of the WTSS-EMP scheme uses a 1-D DWT applied to the hyperspectral image m times, being $m = \log_2(N) - 2$, with N the number of spectral bands, reducing the dimensionality in the spectral domain up to 4 coefficient bands.

Mathematical Morphology (MM) allows extracting spatial structures from images combining morphological transformations⁵² based on two basic operators: erosion and dilation. Objects that are brighter than their surrounding can be shrunk by an erosion or can be expanded by a dilation. The opening operator (erosion followed by dilation) flattens the bright objects of the image, while the closing operator (dilation followed by erosion) has the opposite effect. These operators are applied to an image I at pixel level using a structuring element (SE) of a known shape (usually a $r \times r$ square or a disk of radius r). Opening and closing do not preserve edges between objects. Therefore, opening by reconstruction and closing by reconstruction are used to completely preserve or remove the spatial structures of the image. These operators are based on geodesic reconstruction⁵³ and preserve the spatial structures of an image if the SE fits within the objects, otherwise, the structures are removed.

The EMP is defined as a set of Morphological Profiles (MPs) created with n openings by reconstruction (γ_r) and n closings by reconstruction (ϕ_r) using a SE of growing size on the different principal components extracted from the original hyperspectral image. Be W the transformed data and be W_i the set of wavelet coefficients in the i th coefficient band, the wavelet-based MP is built as:

$$MP^{(n)}(W_i) = \{\gamma_r^{(n)}(W_i), \dots, \gamma_r^{(1)}(W_i), W_i, \phi_r^{(1)}(W_i), \dots, \phi_r^{(n)}(W_i)\}.$$

The EMP is built with all the MPs as follows

$$EMP^{(n)}(W) = \{MP^{(n)}(W_1), MP^{(n)}(W_2), \dots, MP^{(n)}(W_4)\}$$

considering, for example, four coefficient bands, as is the case of this scheme.

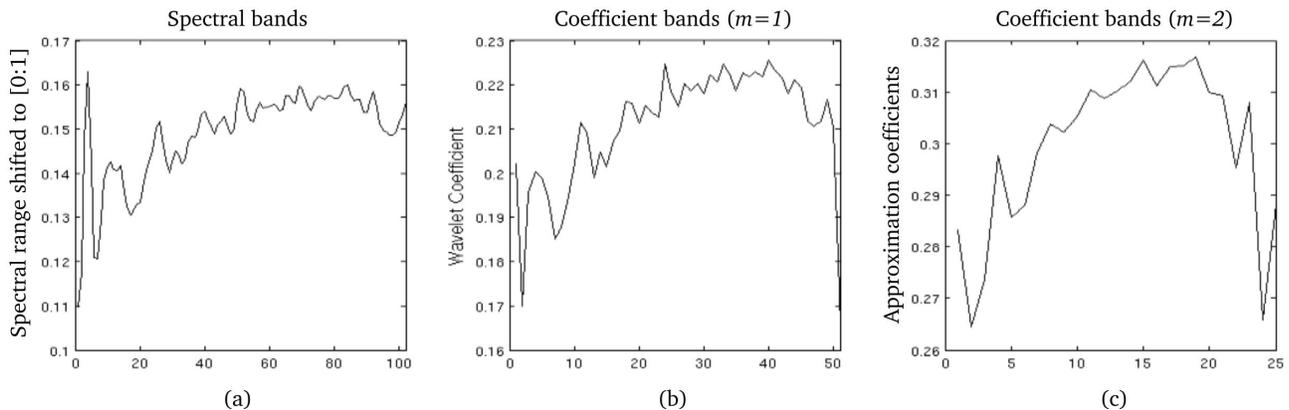


Figure 3. Example of 2 levels of 1-D wavelet decomposition of a signal representing a pixel vector of 103 spectral bands (a) and the remaining $m = 1$ (b) and $m = 2$ (c) coefficient bands.

3. RESULTS

This section presents the results in terms of classification accuracy and execution times. The accuracy of the classification scheme proposed in this work has been evaluated in terms of Overall Accuracy (OA), Average Accuracy (AA), Kappa coefficient of agreement⁵⁴ (κ) and McNemar's test⁵⁵. The OA is the percentage of correctly classified pixels in the whole image and the AA is the mean of the class-specific accuracy for all the classes. κ is the percentage of agreement corrected by the amount of agreement that could be expected due to chance alone. The McNemar's test is used to analyze whether the new proposal produces statistically significant results as compared with the previous scheme. The McNemar's test is based upon the standardized normal test statistic⁵⁵, i.e.:

$$Z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}}$$

in which f_{12} indicates the samples correctly classified by the first scheme, and wrongly by the second one. The difference in accuracy between each pair of schemes is said to be statistically significant at 95% confidence level if $|Z| > 1.96$.

Section 3.1 describes the datasets used in the experiments and the configuration parameters of the WT-EMP and WTSS-EMP schemes. In Section 3.2, the analysis is focused on the accuracy produced by the schemes, and the execution time required for the classification (including the processing time involved by the scheme itself) when the ELM is incorporated as the classifier. This section also studies the improvements in the accuracy results obtained by the new proposal (WTSS-EMP scheme) by performing also an additional 1-D spectral denoising. Section 3.3 presents the robustness of the classification results in the presence of noise. Finally, in Section 3.4, the new WTSS-EMP scheme is compared to other published spectral-spatial classification schemes.

3.1 Datasets and experimental setup

The hyperspectral datasets used in the experiments are two urban areas taken by the Reflective Optics System Imaging Spectrometer (ROSIS-03) hyperspectral sensor, named Pavia Univ. and Pavia City, and two hyperspectral images of crop areas taken by the Airborne Visible-infrared Imaging Spectrometer (AVIRIS) sensor, named Indian Pines and Salinas.

The ROSIS-03 sensor provides a nominative spectral coverage ranging from 0.43 to 0.86 μm with a very high (1.3 m) spatial resolution per pixel. The spatial dimension of the Pavia Univ. hyperspectral image is 610×340 pixels with 103 spectral bands. The nine classes of interest available for this dataset are shown in Fig. 4(a). The second dataset from this sensor is the dense urban area of Pavia City, with spatial dimensions of 1096×715 pixels, and 102 spectral bands. The reference map in Fig. 4(b) contains nine classes of interest.

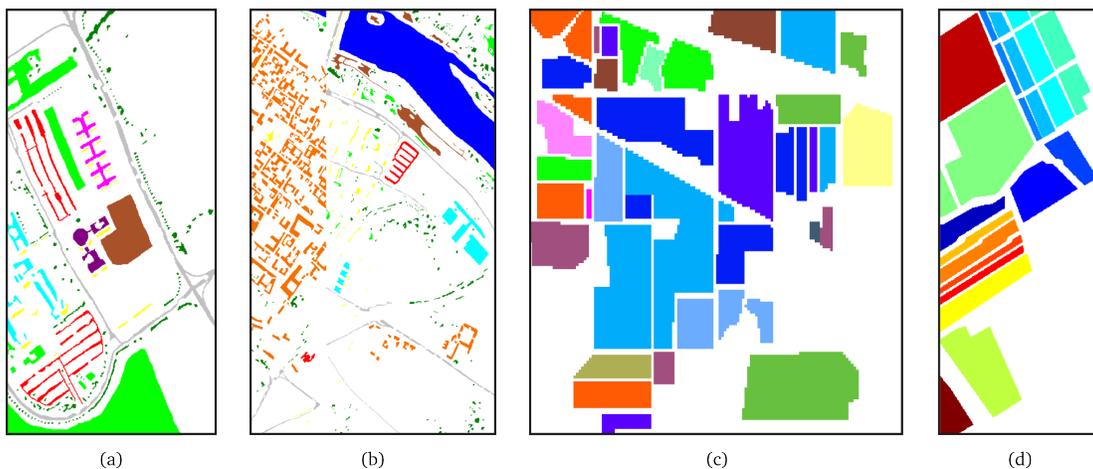


Figure 4. Reference maps for the hyperspectral datasets used in the experiments. Pavia Univ. (a), Pavia City (b), Indian Pines (c) and Salinas (d).

The AVIRIS sensor operates in the visible to mid infrared wavelength range, from 0.4 to 2.4 μm , collecting 224 spectral bands. The Indian Pines dataset consists of 145×145 pixels and 220 spectral bands (the four bands covering the region of water absorption were removed). It was acquired over a mixed agricultural/forested region with a moderate spatial resolution of 20 m/pixel. The Salinas dataset has 512×217 pixels with a higher spatial resolution (3.7 m) and none of the hyperspectral bands was removed. The reference maps for these last two images display sixteen classes as shown in Figs. 4(c) and 4(d).

The number of training samples used in the experiments are presented in Table 1 for these four datasets. These are chosen as in Refs. 5 and 20 to compare our scheme on equal terms. The accuracies are calculated excluding the samples used for training. Each result is obtained by executing the classification 50 times with different sets of randomly chosen training samples each time. The results were calculated as the average of the 50 executions.

Table 2 shows the configuration parameters used in the WT-EMP and WTSS-EMP schemes. The feature reduction by wavelets uses the low-pass and high-pass filters from the Cohen-Daubechies-Feauveau 9/7 (CDF97) wavelet.⁴⁸ The number of levels of 1-D CDF97 wavelet decomposition applied in the spectral domain (m), is calculated as $\lceil \log_2(N) - 2 \rceil$, with N the number of spectral bands. The EMP is built from 4 morphological profiles and each MP is created with 4 opening and 4 closing by reconstruction processes using a disk of radius of increasing size $r = \{1, 3, 5, 7\}$, as described in Section 2.

In the WTSS-EMP scheme, the 1-D CDF97 filters are used for the 1-D wavelet denoising and a 1×3 window is used to shrink the wavelet coefficients using the universal threshold λ . For the 2-D wavelet denoising, the set of filters for perfect reconstruction from Ref. 56 are used and a 3×3 window is considered along with the universal threshold that is automatically calculated for each hyperspectral band, unlike the WT-EMP scheme²³

Table 1. Number of training and test samples for the hyperspectral datasets used in the experiments.

| Pavia Univ. | | | | Indian Pines | | | | | | | |
|-------------|-----------|-------|-------|--------------|---------------|-------|------|-----|-----------------|-------|------|
| # | Classes | Train | Test | # | Classes | Train | Test | # | Classes | Train | Test |
| 1. | Asphalt | 548 | 6083 | 1. | Alfalfa | 15 | 39 | 10. | Soybean-notill | 50 | 918 |
| 2. | Meadows | 540 | 18109 | 2. | Corn-notill | 50 | 1384 | 11. | Soybean-mintill | 50 | 2418 |
| 3. | Gravel | 392 | 1707 | 3. | Corn-mintill | 50 | 784 | 12. | Soybean-clean | 50 | 564 |
| 4. | Trees | 524 | 2540 | 4. | Corn | 50 | 184 | 13. | Wheat | 50 | 162 |
| 5. | Metal | 265 | 1080 | 5. | Grass-pasture | 50 | 447 | 14. | Woods | 50 | 1244 |
| 6. | Bare Soil | 532 | 4497 | 6. | Grass-trees | 50 | 697 | 15. | Bld-Grass-Trees | 50 | 330 |
| 7. | Bitumen | 375 | 955 | 7. | Grass-mowed | 15 | 11 | 16. | Stone-Steel | 50 | 45 |
| 8. | Bricks | 514 | 3168 | 8. | Hay-windrowed | 50 | 439 | | | | |
| 9. | Shadows | 231 | 716 | 9. | Oats | 15 | 5 | | | | |

| Pavia City | | | | Salinas | | | | | | | |
|------------|-----------|-------|-------|---------|---------------------|-------|-------|-----|--------------------|-------|------|
| # | Classes | Train | Test | # | Classes | Train | Test | # | Classes | Train | Test |
| 1. | Asphalt | 548 | 6083 | 1. | Brocoli gr. weeds 1 | 40 | 1969 | 10. | Corn gr. weeds | 65 | 3213 |
| 2. | Meadows | 540 | 18109 | 2. | Brocoli gr. weeds 1 | 74 | 3652 | 11. | Lettuce rom 4wk | 21 | 1047 |
| 3. | Gravel | 392 | 1707 | 3. | Fallow | 39 | 1937 | 12. | Lettuce rom 5wk | 38 | 1889 |
| 4. | Trees | 524 | 2540 | 4. | Fallow rough plow | 27 | 1367 | 13. | Lettuce rom 6wk | 18 | 898 |
| 5. | Metal | 265 | 1080 | 5. | Fallow smooth | 53 | 2625 | 14. | Lettuce rom 7wk | 21 | 1049 |
| 6. | Bare Soil | 532 | 4497 | 6. | Stubble | 79 | 3880 | 15. | Vinyard untrain | 145 | 7123 |
| 7. | Bitumen | 375 | 955 | 7. | Celery | 71 | 3508 | 16. | Vinyard ver trelli | 36 | 1771 |
| 8. | Bricks | 514 | 3168 | 8. | Grapes untrain | 225 | 11046 | | | | |
| 9. | Shadows | 231 | 716 | 9. | Soil vinyard dev. | 124 | 6079 | | | | |

Table 2. Configuration parameters used in the WT-EMP and the WTSS-EMP schemes: number of wavelet decomposition levels applied for feature reduction (m), threshold applied for 2-D denoising (λ), best parameters for SVM (C, γ), and number of neurons in the hidden layer for ELM (L).

| Dataset | m | λ^* | C | γ | L |
|--------------|-----|-------------|------|----------|-----|
| Pavia Univ. | 5 | 4 | 128 | 0.125 | 500 |
| Pavia City | 5 | 4 | 128 | 0.125 | 500 |
| Indian Pines | 6 | 0.01 | 1024 | 0.5 | 350 |
| Pavia City | 6 | 0.01 | 256 | 0.125 | 350 |

* Only for the WT-EMP scheme.

where it was set manually with the values shown in Table 2. If the window is too large, a lot of noise will remain, so small windows are recommended⁴⁷ for neighbouring shrinkage.

The classification is carried out using the LIBSVM library and the ELM algorithm. In the case of SVM the Gaussian radial basis function (RBF) is used as the activation function. For ELM the sigmoidal activation function $g(\mathbf{x}) = \frac{1}{1 + e^{-x}}$ is used. Table 2 also includes the two parameters (C, γ) corresponding to the penalty term and the width of the radius of the Gaussian function for the SVM classifier, and the number neurons in the hidden layer (L) for the ELM classifier. The best parameters (C, γ) for training the SVM are determined in the range $C = [1, 4, 16, 64, 128]$, $\gamma = [0.5, 0.25, 0.125, 0.0625]$ by 5-fold cross-validation for each dataset. The number of neurons in the hidden layer for the ELM classifier was selected in the range $L = [300, 800]$ with an increasing step of 50 neurons.

The computer platform used was an Intel Core i7 860 CPU at 2.80 GHz with 8 GB of RAM using a 64-bit Linux operating system and the experiments are computed in MATLAB*.

3.2 Accuracy and execution time assessment

In order to study the effect of introducing the ELM classifier instead of the SVM in the WT-EMP scheme, Table 3 summarizes the classification results with the two classifiers in terms of OA, κ , McNemar's test (Z) and execution times obtained for the four hyperspectral images used in this work. This execution time includes the training time required for the supervised classification. It can be observed that the OA improves in all the cases by using the ELM. Although the differences in κ between the classifiers is only 0.5 percentage points in the closest case, see the Pavia City column in Table 3, the differences are statistically significant according to

*The underlying research materials for this article can be accessed at <https://wiki.citius.usc.es/hiperespectral:wtss-emp>.

Table 3. Classification results obtained by the WT-EMP scheme in terms of OA indicating the standard deviation (std), Kappa coefficient of agreement (κ), McNemar's test (Z) and execution times in seconds. The Z value compares the results produced by the scheme using the SVM and the ELM classifiers.

| | Pavia Univ. | | Pavia City | | Indian Pines | | Salinas | |
|----------------|-------------|--------|------------|--------|--------------|-------|---------|--------|
| | SVM | ELM | SVM | ELM | SVM | ELM | SVM | ELM |
| OA (%) | 98.65 | 99.79 | 99.54 | 99.85 | 88.93 | 90.45 | 91.67 | 97.14 |
| std (\pm) | 0.168 | 0.054 | 0.034 | 0.026 | 1.138 | 0.953 | 0.453 | 0.208 |
| κ (%) | 98.20 | 99.70 | 99.30 | 99.80 | 87.40 | 89.10 | 90.70 | 96.80 |
| Z | -18.95 | | -17.42 | | -4.97 | | -45.79 | |
| Time (seconds) | 59.009 | 20.366 | 184.100 | 64.293 | 7.785 | 5.258 | 34.550 | 15.162 |

Table 4. Classification results obtained by the schemes (using the ELM classifier in all cases) in terms of OA indicating the standard deviation (std), Kappa coefficient of agreement (κ) and McNemar's test (Z). The Z value compares the results produced by the first and the new scheme.

| | Pavia Univ. | | Pavia City | | Indian Pines | | Salinas | |
|---------------|-------------|----------|------------|----------|--------------|----------|---------|----------|
| | WT-EMP | WTSS-EMP | WT-EMP | WTSS-EMP | WT-EMP | WTSS-EMP | WT-EMP | WTSS-EMP |
| OA (%) | 99.79 | 99.78 | 99.85 | 99.86 | 90.45 | 93.62 | 97.14 | 98.44 |
| std (\pm) | 0.054 | 0.056 | 0.026 | 0.022 | 0.953 | 0.979 | 0.208 | 0.200 |
| κ (%) | 99.70 | 99.70 | 99.80 | 99.80 | 89.10 | 92.70 | 96.80 | 98.30 |
| Z | 0.44 | | -0.84 | | -10.46 | | -16.93 | |

McNemar's test. The negative sign of Z shown in Table 3 ($Z < 0$) indicates that using the ELM classifier is more accurate than using the SVM classifier.

It can be observed in Table 3 that the highest difference in OA between the WT-EMP scheme using SVM and ELM is obtained for the Salinas image, with 5.47 percentage points in favour of the ELM. The lowest values of OA and κ are obtained for the Indian Pines image, taken from the AVIRIS sensor. In general, worse results are usually reported for this dataset, mainly owing to the low spatial resolution (20 m/pixel unlike the 3.7m of Salinas²).

Regarding the execution time, it can be observed in Table 3 that the smallest times are required for classifying the hyperspectral images with the ELM classifier. The classification time is reduced between 1.4 \times and 2.9 \times for the Indian Pines and Pavia Univ. scenes, respectively. The ELM is considered computationally more efficient. So, the results obtained by our schemes are presented from now on only for this classifier.

When an additional 1-D denoising is added to the WT-EMP scheme, we obtain the WTSS-EMP one. As described in Section 2, this scheme incorporates neighbouring wavelet coefficients in the threshold formula, indicated by Eq. (4). See the parameters used for the scheme in Table 2 for more details.

The classification results obtained by the WTSS-EMP scheme are presented in Table 4. Although the results for Pavia Univ. and Pavia City are similar for both schemes, the OA and κ obtained by the WTSS-EMP significantly improve for the images of Indian Pines and Salinas, as indicated by McNemar's test (see the last row in Table 4). The results for Indian Pines are 3.17 percentage points better reaching an OA of 93.62%. Figure 5 shows the reference map for this scene, and the classification maps for the pixel-wise ELM classifier, the WT-EMP and the WTSS-EMP schemes. It can be observed that the classification map is more homogeneous for the last scheme. In particular, two dashed boxes have been introduced in Fig. 5(a) enclosing two areas with

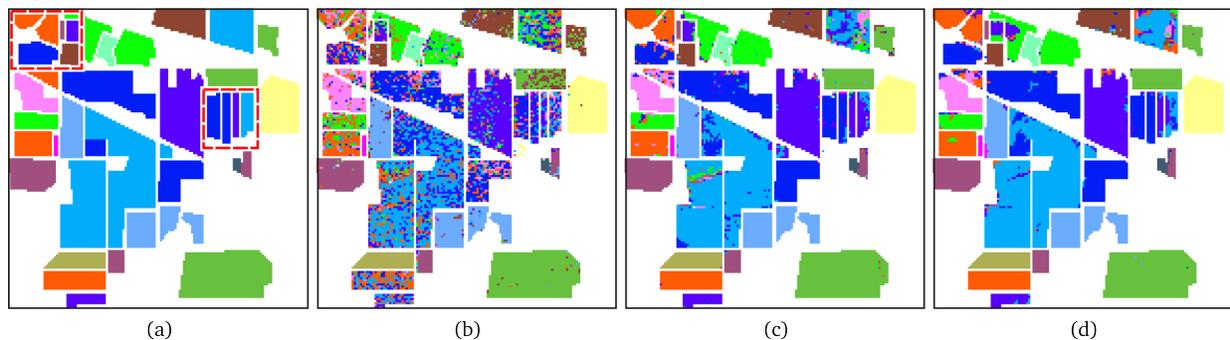


Figure 5. Reference map for the Indian Pines image with two dashed boxes highlighting areas with very close and small regions (a), and classification maps for ELM (b), WT-EMP (c) and WTSS-EMP (d).

Table 5. OA values (%) in presence of additive white gaussian noise with PSNR of 16, 20 and 25 dB for the hyperspectral images of Pavia Univ. and Indian Pines.

| | Pavia Univ. | | | Indian Pines | | |
|-------|-------------|--------|----------|--------------|--------|----------|
| PSNR | ELM | WT-EMP | WTSS-EMP | ELM | WT-EMP | WTSS-EMP |
| 16 dB | 47.96 | 99.62 | 99.40 | 14.27 | 15.38 | 85.70 |
| 20 dB | 56.22 | 99.77 | 99.62 | 17.31 | 20.51 | 89.42 |
| 25 dB | 62.76 | 99.82 | 99.75 | 22.62 | 31.56 | 92.04 |

very close and small regions. For these regions the proposed scheme, WTSS-EMP, improves the classification results illustrating that the new 1-D denoising stage is especially efficient for close regions in the reference map.

3.3 Experimental results in the presence of noise

The following section presents an experiment performed in presence of additive noise that was carried out with the same configuration as described in the previous section. The hyperspectral images of Pavia Univ. and Indian Pines are corrupted by AWGN with a peak signal-to-noise ratio (PSNR) of 16, 20 and 25 dB.

Table 5 compares the results obtained by the WT-EMP and the WTSS-EMP spatial-spectral schemes to the SVM and the ELM pixel-wise classifiers. WT-EMP and WTSS-EMP for the case of the Pavia Univ. image improve the OA from 56.22% achieved by the ELM to 99.77% achieved by the WT-EMP and 99.62% for WTSS-EMP, for a PSNR of 20dB. These results are close to the accuracies obtained over Pavia Univ. without noise as shown in Table 4, indicating the effectiveness of the filters used for denoising and the quality of the developed schemes.

The results over Indian Pines show the low accuracy obtained by the WT-EMP in the presence of additive noise. Only 20.51% of OA is reached in the case of a PSNR of 20 dB. The WTSS-EMP scheme outperforms those results by using a 1-D spectral and 2-D spatial wavelet-based denoising, achieving a higher overall accuracy, 89.42%, as indicated in Table 5. The same behavior is observed for the case of a PSNR of 16 and 25 dB.

3.4 Comparison with other schemes

The classification results are compared in this section to those obtained by other classification schemes available in the literature. The comparisons have been made with those schemes that perform a supervised classification on two well-known datasets used in remote sensing classification: Pavia Univ. and Indian Pines. In particular, the following schemes have been used for the comparison: Hseg+MV,⁵ EMP-KPCA,⁵ DB-DB,²⁰ NW-NW,²⁰ WIMF1³¹ and FORPDN.⁴² A brief description of these schemes is given below. Hseg+MV⁵ is a scheme based on hierarchical image segmentation and majority vote. The segmented regions created by the HSEG algorithm are combined with the classes obtained by the pixel-wise SVM classification via majority vote. EMP-KPCA⁵ is a spectral-spatial scheme that incorporates the spatial information with morphological profiles built from the first principal components extracted by kernel PCA. DB-DB²⁰ is based on Extended Multi-Attributes Profiles (EMAPs), in particular, the attributes of area and standard deviation. The profiles are created from features extracted by DBFE and concatenated via stacked vectors to create the EMAP, which is reduced a second time by the same technique. NW-NW²⁰ is based on area and standard deviation profiles. The profiles are created from features extracted by NWFE and concatenated via stacked vectors, which are reduced once again by the same feature extraction technique. WIMF1³¹ is a scheme that uses 2-D Empirical Mode Decomposition (EMD) and 1-D wavelets for spatial and spectral processing, respectively. Data fusion is implicitly performed by applying the wavelet transform along each intrinsic mode function (IMF) found by the 2D-EMD. FORPDN⁴² is a scheme based on first order roughness penalty applied in the wavelet domain for hyperspectral image denoising. This scheme, which only applies denoising, outperforms other denoising methods, such as BiShri2D,³⁸ BiShri3D,³⁹ PCABiShr⁴⁰ and BM4D³⁵ in terms of SNR and classification accuracy.

Unfortunately, there are no published results of all the aforementioned methods for the two datasets, so we will include for each dataset the available results. The first comparison is conducted on the urban area taken

Table 6. Classification results (accuracies in %) obtained by the pixel-wise ELM classifier, EMP-KPCA,⁵ DB-DB,²⁰ WIMF1,³¹ and WTSS-EMP schemes for the Pavia Univ. dataset.

| # | ELM | EMP-KPCA ⁵ | DB-DB ²⁰ | WIMF1 ³¹ | WTSS-EMP |
|----------|-------|-----------------------|---------------------|---------------------|--------------|
| 1. | 82.56 | 96.0 | 95.43 | – | 99.67 |
| 2. | 90.58 | 97.5 | 95.88 | – | 99.83 |
| 3. | 76.49 | 81.1 | 100 | – | 99.79 |
| 4. | 96.61 | 99.3 | 90.94 | – | 99.43 |
| 5. | 99.55 | 99.4 | 100 | – | 99.45 |
| 6. | 92.89 | 99.2 | 98.41 | – | 99.94 |
| 7. | 91.96 | 98.8 | 99.18 | – | 99.96 |
| 8. | 89.56 | 99.4 | 98.65 | – | 99.79 |
| 9. | 99.65 | 98.0 | 99.96 | – | 99.55 |
| OA | 89.73 | 96.3 | 97.89 | 99.04 | 99.78 |
| AA | 91.08 | 95.7 | 97.60 | — | 99.71 |
| κ | 86.20 | 95.0 | 97.20 | 97.0 | 99.70 |

by the ROSIS-03 sensor: Pavia Univ. Table 6 shows the classification results obtained by the pixel-wise ELM classifier, EMP-KPCA,⁵ DB-DB,²⁰ WIMF1,³¹ as well as the proposed WTSS-EMP scheme. The best results are indicated in bold. As expected, all the schemes improve the OA as compared to the pixel-wise ELM classifier. The scheme based on EMP created from KPCA reaches 96.3% but better results are obtained by the DB-DB scheme with 97.89% in overall accuracy. DB-DB is based on two attribute profiles which usually produce better

Table 7. Classification results (accuracies in %) obtained by the pixel-wise SVM classifier, Hseg+MV,⁵ NW-NW,²⁰ FORPDN⁴² and WTSS-EMP schemes for the Indian Pines dataset.

| | SVM | Hseg+MV ⁵ | FORPDN ⁴² | NW-NW ²⁰ | WTSS-EMP |
|----------|-------|----------------------|----------------------|---------------------|--------------|
| 1. | 85.23 | 92.3 | 97.44 | 94.87 | 98.26 |
| 2. | 74.75 | 90.5 | 83.31 | 90.20 | 89.49 |
| 3. | 75.56 | 83.0 | 92.73 | 98.85 | 92.89 |
| 4. | 90.75 | 97.5 | 94.02 | 97.28 | 99.18 |
| 5. | 92.89 | 94.4 | 96.42 | 95.52 | 95.75 |
| 6. | 94.13 | 97.6 | 95.12 | 99.56 | 98.41 |
| 7. | 90.18 | 100 | 100 | 100 | 96.91 |
| 8. | 97.60 | 99.5 | 98.18 | 99.54 | 99.92 |
| 9. | 98.40 | 100 | 100 | 100 | 100 |
| 10. | 79.41 | 92.1 | 89.87 | 86.27 | 93.88 |
| 11. | 66.05 | 84.1 | 77.79 | 94.58 | 89.20 |
| 12. | 85.16 | 95.4 | 87.06 | 93.61 | 94.22 |
| 13. | 99.28 | 98.2 | 98.77 | 99.38 | 99.53 |
| 14. | 88.38 | 98.6 | 94.94 | 92.76 | 98.01 |
| 15. | 71.41 | 82.1 | 98.18 | 99.09 | 97.76 |
| 16. | 97.51 | 100 | 97.78 | 100 | 99.33 |
| OA | 79.49 | 90.8 | 88.28 | 94.17 | 93.62 |
| AA | 86.67 | 94.0 | 93.85 | 96.35 | 96.42 |
| κ | 76.80 | 90.0 | 86.60 | 93.30 | 92.70 |

results than morphological profiles. In addition DB-DB is built from DBFE components, a feature reduction method that is effective with a sufficient set of available training samples, as it is the case for this dataset. The OA obtained by WIMF1 reaches 99.04%. The advantage of this method stems from the ability to represent further detailed information and more features of the image by using IMFs and wavelets. The WTSS-EMP scheme produces the best classification in terms of OA (99.78%), AA (99.71%) and κ (99.70%) with a high regularity among classes.

The second comparison is performed on a dataset taken by the AVIRIS sensor: Indian Pines. For this image, the pixel-wise SVM classifier produces better results than the ELM, so it is included as the base for the comparison. The results published for the schemes Hseg+MV,⁵ NW-NW,²⁰ FORPDN,⁴² as well as the pixel-wise SVM classifier are shown in Table 7. The Hseg+MV scheme improves by 11.31 percentage points the OA as compared to the pixel-wise SVM. The results obtained by FORPDN, a scheme performing only denoising without any additional spatial processing, shows the relevance of a denoising stage. For this image, the results are greatly improved reaching an OA of 88.28%, as indicated in Table 7 for the aforementioned scheme. The combination of the 1-D spectral and the 2-D spatial denoising with the data fusion of the EMP (last column in Table 7) outperforms the other two methods. In particular, the OA is improved by 14.13 percentage points. The best OA (94.17%) is obtained by the NW-NW scheme, that is based on standard deviation and area attribute profiles and classification by SVM, but the proposed WTSS-EMP scheme has a higher AA (96.42%). As for the previous dataset the WTSS-EMP scheme presents a good regularity among classes.

4. CONCLUSIONS

This paper analyzes the effect of wavelet denoising of remote sensing hyperspectral images when a spatial-spectral classification scheme is applied. The study is performed over the classification scheme called WT-EMP that was previously presented by the authors. The scheme extracts the features of the image by wavelets and then an EMP is built. In addition, the hyperspectral image is denoised also by wavelets in the 2-D spatial domain, band by band. The EMP is combined with the denoised hyperspectral image in a new vector of features, and the classification is carried out by an SVM classifier. In this paper the SVM classifier is replaced by ELM achieving a relevant reduction in execution time, that reaches 2.8 times for the Pavia City image, while preserving the classification accuracy. After that, a new 1-D denoising stage is applied to each pixel vector before the 2-D spatial denoising. The resulting classification scheme including ELM and the new wavelet-based denoising stage is called WTSS-EMP.

The proposed scheme has been tested over four hyperspectral images from the ROSIS-03 and the AVIRIS sensors over urban and agriculture areas showing an increase in accuracy for the images that present close regions in the classification reference map. In these cases, increases of 3.17 and 1.30 percentage points in OA are achieved, in particular, for the Indian Pines and Salinas images respectively. The robustness of the WTSS-EMP scheme in the presence of additive white Gaussian noise has also been shown. Finally, the scheme has proven to be competitive compared to other schemes in the classification of the Pavia Univ. and Indian Pines images, with the best regularity among classes in both cases.

ACKNOWLEDGMENTS

This work was supported in part by the Ministry of Science and Innovation, Government of Spain, cofounded by the FEDER funds of European Union, under contract TIN 2013-41129-P, and by Xunta de Galicia, Program for Consolidation of Competitive Research Groups ref. 2014/008. The authors would like to thank Prof. D. Landgrebe, Prof. P. Gamba, Dr. L. Johnson and Dr. J. A. Gualtieri for releasing the Indian Pines, the areas of the Pavia University and City of Pavia, Italy, and the Salinas datasets to the community. The authors would also like to thank Prof. D. Manuel Graña Romay of the University of the Basque Country for providing a collection of hyperspectral scenes on his research website.

REFERENCES

- [1] Shaw, G. A. and Burke, H.-H. K., "Spectral imaging for remote sensing," *Lincoln Laboratory Journal* **14**(1), 3–28 (2003).

- [2] Gualtieri, J. A. and Crompton, R. F., “Support vector machines for hyperspectral remote sensing classification,” in [27th AIPR Workshop: Advances in Computer-Assisted Recognition], *Proc. SPIE* **3584**, 221–232 (Jan 1999).
- [3] Fauvel, M., Chanussot, J., and Benediktsson, J. A., “Evaluation of kernels for multiclass classification of hyperspectral remote sensing data,” in [2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings], **2**, II813–II816 (May 2006).
- [4] Giorgos, M., Jungo, I., and Caesar, O., “Support vector machines in remote sensing: A review,” *ISPRS Journal of Photogrammetry and Remote Sensing* **66**(3), 247–259 (2011).
- [5] Fauvel, M., Tarabalka, Y., Benediktsson, J. A., Chanussot, J., and Tilton, J. C., “Advances in spectral-spatial classification of hyperspectral images,” *Proceedings of the IEEE* **101**(3), 652–675 (2013).
- [6] Huang, G.-B., Zhu, Q.-Y., and Siew, C.-K., “Extreme learning machine: Theory and applications,” *Neurocomputing* **70**(1-3), 489–501 (2006).
- [7] Pal, M., Maxwell, A. E., and Warner, T. A., “Kernel-based extreme learning machine for remote-sensing image classification,” *Remote Sensing Letters* **4**(9), 853–862 (2013).
- [8] Moreno, R., Corona, F., Lendasse, A., Graña, M., and Galvão, L. S., “Extreme learning machines for soybean classification in remote sensing hyperspectral images,” *Neurocomputing* **128**, 207–216 (2014).
- [9] Heras, D. B., Argüello, F., and Quesada-Barriuso, P., “Exploring ELM-based spatial-spectral classification of hyperspectral images,” *International Journal of Remote Sensing* **35**(2), 401–423 (2014).
- [10] Huang, G.-B., “An insight into extreme learning machines: Random neurons, random features and kernels,” *Cognitive Computation* **6**(3), 376–390 (2014).
- [11] López-Fandiño, J., Quesada-Barriuso, P., Heras, D. B., and Argüello, F., “Efficient ELM-based techniques for the classification of hyperspectral remote sensing images on commodity GPUs,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **8**(6), 2884–2893 (2015).
- [12] Tarabalka, Y., Chanussot, J., and Benediktsson, J. A., “Segmentation and classification of hyperspectral images using watershed transformation,” *Pattern Recognition* **43**(7), 2367–2379 (2010).
- [13] Priego, B., Bellas, F., and Duro, R. J., “ECAS-II: A hybrid algorithm for the construction of multidimensional image segmenters,” in [2015 International Joint Conference on Neural Networks (IJCNN)], 1–8 (July 2015).
- [14] Tarabalka, Y., Benediktsson, J. A., and Chanussot, J., “Spectral-spatial classification of hyperspectral imagery based on partitioning clustering techniques,” *Geoscience and Remote Sensing, IEEE Transactions on* **47**(8), 2973–2987 (2009).
- [15] Tarabalka, Y., Chanussot, J., and Benediktsson, J. A., “Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* **40**(5), 1267–1279 (2010).
- [16] Kang, X., Li, S., and Benediktsson, J. A., “Spectral-spatial hyperspectral image classification with edge-preserving filtering,” *IEEE Transactions on Geoscience and Remote Sensing* **52**(5), 2666–2677 (2014).
- [17] Plaza, A., Benediktsson, J. A., Boardman, J. W., Brazile, J., Bruzzone, L., Camps-Valls, G., Chanussot, J., Fauvel, M., Gamba, P., Gualtieri, A., Marconcini, M., Tilton, J. C., and Trianni, G., “Recent advances in techniques for hyperspectral image processing,” *Remote Sensing of Environment* **113**, **Supplement 1**, S110–S122 (2009). Imaging Spectroscopy Special Issue.
- [18] Palmason, J., Benediktsson, J. A., Sveinsson, J. R., and Chanussot, J., “Classification of hyperspectral data from urban areas using morphological preprocessing and independent component analysis,” in [Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS '05.], **1**, 176–179 (July 2005).
- [19] Benediktsson, J. A., Pesaresi, M., and Amason, K., “Classification and feature extraction for remote sensing images from urban areas based on morphological transformations,” *IEEE Transactions on Geoscience and Remote Sensing* **41**, 1940–1949 (Sept 2003).
- [20] Marpu, P., Pedergnana, M., Mura, M. D., Peeters, S., Benediktsson, J. A., and Bruzzone, L., “Classification of hyperspectral data using extended attribute profiles based on supervised and unsupervised feature extraction techniques,” *International Journal of Image and Data Fusion* **3**(3), 269–298 (2012).

- [21] Benediktsson, J. A., Palmason, J. A., and Sveinsson, J. R., "Classification of hyperspectral data from urban areas based on extended morphological profiles," *Geoscience and Remote Sensing, IEEE Transactions on* **43**(3), 480–491 (2005).
- [22] Licciardi, G., Marpu, P. R., Benediktsson, J. A., and Chanussot, J., "Extended morphological profiles using auto-associative neural networks for hyperspectral data classification," in *[2011 3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)]*, 1–4 (June 2011).
- [23] Quesada-Barriuso, P., Argüello, F., and Heras, D. B., "Spectral-spatial classification of hyperspectral images using wavelets and extended morphological profiles," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **7**(4), 1177–1185 (2014).
- [24] Mura, M. D., Benediktsson, J. A., Waske, B., and Bruzzone, L., "Morphological attribute profiles for the analysis of very high resolution images," *Geoscience and Remote Sensing, IEEE Transactions on* **48**(10), 3747–3762 (2010).
- [25] Mura, M. D., Villa, A., Benediktsson, J. A., Chanussot, J., and Bruzzone, L., "Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis," *Geoscience and Remote Sensing Letters, IEEE* **8**(3), 542–546 (2011).
- [26] Tarabalka, Y., Fauvel, M., Chanussot, J., and Benediktsson, J. A., "SVM- and MRF-based method for accurate classification of hyperspectral images," *IEEE Geoscience and Remote Sensing Letters* **7**(4), 736–740 (2010).
- [27] Fauvel, M., Benediktsson, J. A., Chanussot, J., and Sveinsson, J. R., "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *Geoscience and Remote Sensing, IEEE Transactions on* **46**(11), 3804–3814 (2008).
- [28] Camps-Valls, G., Gomez-Chova, L., Muñoz Marí, J., Vila-Francés, J., and Calpe-Maravilla, J., "Composite kernels for hyperspectral image classification," *Geoscience and Remote Sensing Letters, IEEE* **3**(1), 93–97 (2006).
- [29] Fauvel, M., Chanussot, J., and Benediktsson, J. A., "A spatial-spectral kernel-based approach for the classification of remote-sensing images," *Pattern Recognition* **45**(1), 381–392 (2012).
- [30] Demir, B. and Ertürk, S., "Empirical mode decomposition of hyperspectral images for support vector machine classification," *IEEE Transactions on Geoscience and Remote Sensing* **48**(11), 4071–4084 (2010).
- [31] Gormus, E. T., Canagarajah, N., and Achim, A., "Dimensionality reduction of hyperspectral images using empirical mode decompositions and wavelets," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of* **5**(6), 1821–1830 (2012).
- [32] Kerekes, J. P. and Baum, J. E., "Hyperspectral imaging system modeling," *Lincoln Laboratory Journal* **14**(1), 117–130 (2003).
- [33] Collet, C., Chanussot, J., and Chehdi, K., [*Multivariate image processing*], London, UK : ISTE, Hoboken, NJ : J. Wiley (2010).
- [34] Demir, B. and Ertürk, S., "Improved hyperspectral image classification with noise reduction pre-process," in *[Signal Processing Conference, 2008 16th European]*, 1–4 (Aug 2008).
- [35] Maggioni, M., Katkovnik, V., Egiazarian, K., and Foi, A., "Nonlocal transform-domain filter for volumetric data denoising and reconstruction," *IEEE Transactions on Image Processing* **22**(1), 119–133 (2013).
- [36] Atkinson, I., Kamalabadi, F., and Jones, D. L., "Wavelet-based hyperspectral image estimation," in *[Geoscience and Remote Sensing Symposium, 2003. IGARSS '03. Proceedings. 2003 IEEE International]*, **2**, 743–745 (July 2003).
- [37] Rasti, B., Sveinsson, J. R., Ulfarsson, M. O., and Benediktsson, J. A., "Hyperspectral image denoising using 3d wavelets," in *[2012 IEEE International Geoscience and Remote Sensing Symposium]*, 1349–1352 (July 2012).
- [38] Sendur, L. and Selesnick, I. W., "Bivariate shrinkage functions for wavelet-based denoising exploiting inter-scale dependency," *IEEE Transactions on Signal Processing* **50**(11), 2744–2756 (2002).
- [39] Chen, G., Bui, T. D., and Krzyzak, A., "Denoising of three-dimensional data cube using bivariate wavelet shrinking," *International Journal of Pattern Recognition and Artificial Intelligence* **25**(03), 403–413 (2011).
- [40] Chen, G. and Qian, S. E., "Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage," *IEEE Transactions on Geoscience and Remote Sensing* **49**(3), 973–980 (2011).

- [41] Selesnick, I. W., Baraniuk, R. G., and Kingsbury, N. C., “The dual-tree complex wavelet transform,” *IEEE Signal Processing Magazine* **22**(6), 123–151 (2005).
- [42] Rasti, B., Sveinsson, J. R., Ulfarsson, M. O., and Benediktsson, J. A., “Hyperspectral image denoising using first order spectral roughness penalty in wavelet domain,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* , 1–10 (2013).
- [43] Donoho, D. L., “De-noising by soft-thresholding,” *Information Theory, IEEE Transactions on* **41**(3), 613–627 (1995).
- [44] Chang, S. G., Yu, B., and Vetterli, M., “Adaptive wavelet thresholding for image denoising and compression,” *IEEE Transactions on Image Processing* **9**(9), 1532–1546 (2000).
- [45] Donoho, D. L. and Johnstone, I. A., “Adapting to unknown smoothness via wavelet shrinkage,” *Journal of the American Statistical Association* **90**(432), 1200–1224 (1995).
- [46] Cai, T. T. and Silverman, B. W., “Incorporating information on neighbouring coefficients into wavelet estimation,” *Sankhyā: The Indian Journal of Statistics, Series B (1960-2002)* **63**(2), 127–148 (2001).
- [47] Chen, G. Y., Bui, T. D., and Krzyzak, A., “Image denoising using neighbouring wavelet coefficients,” *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing* **2**, 917–920 (2004).
- [48] Daubechies, I., [*Ten lectures on wavelets*], vol. 61, Society for Industrial and Applied Mathematics (1992).
- [49] Mallat, S., [*A wavelet tour of signal processing*], Academic press (1999).
- [50] Mahesh, P., “Extreme-learning-machine-based land cover classification,” *International Journal of Remote Sensing* **30**(14), 3835–3841 (2009).
- [51] Courrieu, P., “Fast computation of moore-penrose inverse matrices,” *Neural Information Processing* **8**(2), 25–29 (2005).
- [52] Serra, J., [*Image Analysis and Mathematical Morphology*], Academic Press (1982).
- [53] Soille, P., [*Morphological Image Analysis: Principles and Applications*], Springer-Verlag New York, Inc., 2nd ed. (2003).
- [54] Viera, A. J. and Garrett, J. M., “Understanding interobserver agreement: the kappa statistic,” *Fam Med* **37**(5), 360–363 (2005).
- [55] F., G. M., “Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy,” *Photogramm. Eng. Remote Sens.* **7**(5), 627–633 (2004).
- [56] Selesnick, I. W., “The Double Density DWT,” in [*Wavelets in Signal and Image Analysis: From Theory to Practice*], ch. 2, 36–66, Kluwer Academic Publishers (2001).