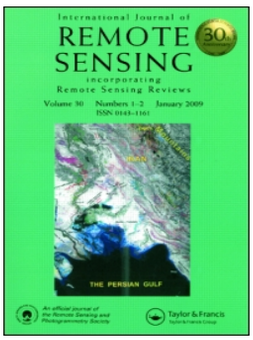

International Journal of Remote Sensing

ISSN: 0143-1161 (Print) 1366-5901 (Online) Journal homepage: https://www.tandfonline.com/loi/tres20

Hyperspectral image classification using graph-based wavelet transform

Nadia Zikiou, Mourad Lahdir & David Helbert

To cite this article: Nadia Zikiou, Mourad Lahdir & David Helbert (2020) Hyperspectral image classification using graph-based wavelet transform, International Journal of Remote Sensing, 41:7, 2624-2643, DOI: 10.1080/01431161.2019.1694194

To link to this article: https://doi.org/10.1080/01431161.2019.1694194

Published online: 06 Dec 2019.

INTERNATIONAL JOURNAL OF REMOTE SENSING
2019, VOL. 41, NO. 7, 2624–2643
https://doi.org/10.1080/01431161.2019.1694194

Taylor & Francis
Taylor & Francis Group

Hyperspectral image classification using graph-based wavelet transform

Nadia Zikiou[a], Mourad Lahdir[a] and David Helbert[b]

[a]Laboratory of Analysis and Modeling of Random Phenomena (LAMPA), UMMTO, Tizi-Ouzou, Algeria; [b]XLIM Research Institute (UMR CNRS 7252), University of Poitiers, Poitiers, France

ABSTRACT

Graph-based methods are developed to efficiently extract data information. In particular, these methods are adopted for high-dimensional data classification by exploiting information residing on weighted graphs. In this paper, we propose a new hyperspectral texture classifier based on graph-based wavelet transform. This recent graph transform allows extracting textural features from a constructed weighted graph using sparse representative pixels of hyperspectral image. Different measurements of spectral similarity between representative pixels are tested to decorrelate close pixels and improve the classification precision. To achieve the hyperspectral texture classification, Support Vector Machine is applied on spectral graph wavelet coefficients. Experimental results obtained by applying the proposed approach on Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) and Reflective Optics System Imaging Spectrometer (ROSIS) datasets provide good accuracy which could exceed 98.7%. Compared to other famous classification methods as conventional deep learning-based methods, the proposed method achieves better classification performance. Results have shown the effectiveness of the method in terms of robustness and accuracy.

ARTICLE HISTORY
Received 27 February 2019
Accepted 20 September 2019

1. Introduction

Hyperspectral images (HSI) provide very detailed information on spectral and spatial distributions of distinct materials. The rich spectrum information measured by hyperspectral sensors allows to distinguish different classes present in the image, which is very suitable for analysing and recognizing objects. However, high dimension of hyperspectral images is still problematic for classification because of the large number of contiguous spectral sub-bands and the small available labelled training. On the other hand, the high variations of hyperspectral texture to be classified require no stationary classification conditions. Many previous proposed methods have been used to overcome these problems. In Qian, Ye, and Zhou (2013), a hyperspectral classification method based on 3D wavelet transform, spectral-spatial texture descriptor and structured sparse logistic regression is proposed for feature extraction and pixel classification. The knowledge

CONTACT Nadia Zikiou zikiounadia@gmail.com Laboratory of Analysis and Modeling of Random Phenomena (LAMPA),UMMTO, BP 17 RP, Tizi-Ouzou 15000, Algeria

about the structure of features, allowed by this method and wavelet-based methods presented in He et al. (2015); Tang, Lu, and Yuan (2015), make the features selection more efficient. A sparse representation of hyperspectral features introducing a kernel function in Chen, Nasrabadi, and Tran (2013); Persello and Bruzzone (2016) has also shown good performances in hyperspectral image classification. Support Vector Machine (SVM) (Guo et al. 2008; Bruzzone and Persello 2009; Bovolo, Bruzzone and Carlin 2010) is another efficient approach in terms of robustness to the high dimensionality of the feature space and improving accuracy. A semi-supervised classification algorithm is proposed in Yang et al. (2014) in order to avoid the risk of over-fitting of the training samples by using an extension of SVM in the Laplacian Graph (LapSVM). To improve the classification accuracy of hyperspectral images, Tarabalka et al. (2010) introduced a method based on the fusion of supervised (SVM) and unsupervised (fuzzy *c*-means clustering) learning. In other works (Liu et al. 2012; Zhou et al. 2013) graph-based approaches have been suggested to solve classification problems. Deep learning methods also have particularly captured great interest in hyperspectral classification (Chen et al. 2014; Ma, Geng, and Wang 2015; Zou et al. 2015; Zhang, Zhang, and Du 2016) and have demonstrated excellent performances in the matter. The deep convolutional neural network (CNN) has been efficiently employed in hyperspectral texture features extraction (Hu et al. 2015; Mou, Ghamisi, and Zhu 2017; Li, Xie, and Li 2017b). Several CNN-based methods have been employed to improve the HIS classification performances such as in Li et al. (2017b); Ran et al. (2017); Zhang, Li, and Du (2018). The Spectral Graph Wavelets Transform (SGWT) is a prospective graph-based transform which has been developed (Hammond, Vandergheynst, and Gribonval 2011) and successfully used in many applications-based classification (Masoumi and Hamza 2017; Pham, Mercier and Michel 2014a), image denoising (Malek, Helbert, and Carre 2015) and Nonrigid image registration (Pham et al. 2018). This transform explores localized signals on the graph Fourier spectrum of an undirected graph. One of the reasons why the SGWT excels in feature extraction is the high flexibility of describing a large number of data. In this paper, we propose a very promising hyperspectral texture classifier based on SGWT and SVM learning. One of the main contributions of this work is to investigate adapted-SGWT for the hyperspectral data classification and evaluate the results of this method in terms of classification accuracy and robustness. The hyperspectral adapted SGWT is based on a structure which represents spectral and spatial information contained in a hyperspectral image. Another contribution is to observe the effect of different similarity measures used in the SGWT process, on the representative pixels extraction and hence the effectiveness of features selection. The proposed method is evaluated on both the airborne visible/infrared imaging spectrometer (AVIRIS) data sensor and the reflective optics system imaging spectrometer (ROSIS) data sensor.

2. Methodology

2.1. Proposed scheme

The diagram given in Figure 1 describes the different stages of our hyperspectral classification procedure. The first step is to apply the SGWT on the hyperspectral image to generate the spectral graph wavelet coefficients. In this stage, we vary five distance measures for the distance graph computation, i.e. Euclidian distance (ED), Kolmogorov

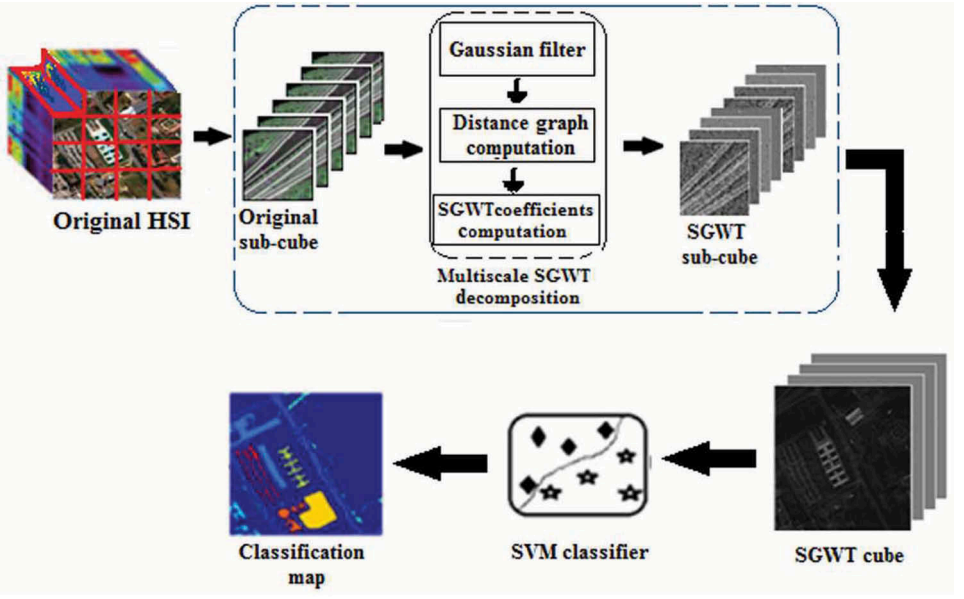

Figure 1. Schematic of the proposed approach.

distance (KD), Spectral angle (SA), Spectral information divergence (SID) and Normalized cross-correlation (NCC) to show the effect of each one on the classification accuracy. The next step includes the SVM classification of SGWT extracted coefficients.

2.2. Characteristics of weighted graphs

We use the Spectral Graph Wavelet Transform to efficiently extract information from high-dimensional data localized on the vertices of weighted graphs. A weighted graph G consists of two sets: vertices set V representing the nodes, edges set E representing connection between nodes and a weight function $w : \mathrm{E} \rightarrow \mathrm{R}^{+}$ which assigns a positive weight to each edge of connecting vertices (i, j) (Hammond, Vandergheynst, and Gribonval 2011; Leonardi and Van De Ville 2011; Pham, Mercier and Michel 2014b; Kerola, Inoue, and Shinoda 2014; Shuman et al. 2015). The graph Laplacian matrix of a signal processing is positive, symmetric and semi-definite matrix, defined as:

$$\mathbf{L} = \mathbf{D} - \mathbf{A}, \tag{1}$$

where **D** is a matrix of size $N \times N$. It is defined as a diagonal matrix with diagonal elements equal to the degrees, and zeros elsewhere. The degrees represent the sum of the weights of all the incident edges of each vertex. **A** is the symmetric adjacency matrix of G of size $N \times N$. **D** and **A** are defined as:

$$\mathbf{A}_{i,j} = \begin{cases} w_{i,j} & \text{if } (i,j) \in \mathrm{E} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

$$\mathbf{D}_{ii} = \sum_{j} \mathbf{A}_{i,j} \tag{3}$$

For every real function $f : V \rightarrow \mathbb{R}$ given on the vertices of graph G is considered as a vector in $\mathbb{R}^N$. The value of the function f on each vertex of the graph defines each coordinate. It is notified as $f \in \mathbb{R}^N$, where $f(i)$ is the value on the i^{th} vertex. Equation (1) can be satisfied for any $f \in \mathbb{R}^N$, it is given by:

$$(\mathbf{L}f)(i) = \sum_{i \sim j} w_{i,j} \times (f(i) - f(j)) \tag{4}$$

where the sum over $i \sim j$ indicates the summation of j vertices connected to the vertex i and $w_{i,j}$ denotes the edge weight connecting i and j The Laplacian matrix $\mathbf{L}$ has a complete set of eigenvectors which are denoted by $\boldsymbol{X_l}$ for $l = 0, \ldots, N-1$ with eigenvalues λ_l.

$$\mathbf{L}_{X_l} = \lambda_l \boldsymbol{X_l} \tag{5}$$

The eigenvalues are non-negative.

2.3. Spectral graph wavelet transform

A translation of smooth graph spectral filters is accomplished by multiplying each filter by a graph Laplacian eigenvector in the Fourier domain. The graph Fourier transform $\hat{f}$ for any function $f \in \mathrm{R}^N$, defined on the vertices of weighted graph G, is given by the formula:

$$\hat{f}(l) = \langle \boldsymbol{X}_l, f \rangle = \sum_{n=1}^{N} \boldsymbol{X}_l^*(n) f(n) \tag{6}$$

To generate the spectral graph wavelet transform, a kernel function $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ in spectral domain is introduced, such as: $g(0) = 0$ and $|\lim_{x \rightarrow \infty} g(x) = 0$. The spectral representation of a wavelet operator $T_g = g(\mathbf{L})$ is obtained by modulating each Fourier mode, as

$$\hat{T}_g f(l) = g(\lambda l) f(l) \tag{7}$$

At scale J, this operator is obtained by $T_g^J = g(J\mathbf{L})$. The scaling function $\phi_{J,n}(m)$ localized in vertex n at scale J to capture the residual low-pass h components is defined as:

$$\phi_{J,n}(m) = \sum_{l=0}^{N-1} h(J\lambda_l) \boldsymbol{X}_l^*(n) \boldsymbol{X}_l(m) \tag{8}$$

The SGWT of a function on vertices $f \in \mathrm{R}^N$ at scale J and vertex n is thus defined as follows:

$$W_f(J, n) = \sum_{k=0}^{N-1} g(J\lambda_k) \hat{f}_{(k)} \boldsymbol{X}_k(n) \tag{9}$$

Hence, a linear function W_f is generated with N representative pixels presented in $(J+1)$ sets of scale coefficients. This corresponds to a low-pass filter h and J band-pass kernels. To highlight that, the direct decomposition needs high time consumption. To address this issue, authors in Hammond, Vandergheynst, and Gribonval (2011) proposed to use Chebyshev polynomial approximation. The generated coefficients performed by the

multi-scale analysis allow a good knowledge of local features and graph similarity between vertices at different scales.

Authors in Hammond, Vandergheynst, and Gribonval (2011) also pointed out the difference between the classical orthogonal wavelet transforms and the SGWT. This is principally summarized in the selection of a mask of vertices at each scale of SGWT subsampling, unlike the classical wavelet transforms which decomposes the transform to coarser spatial levels. Note that in the current work, the whole original data is divided in sub-cubes, to avoid the problem of memory capacity caused by the high dimension data computation.

2.4. Similarity measures of connected components

To build HSI data weighted graphs, we measure similarity between each spectrum pixel and its *T*-neighbourhood spectrum pixels. Hence, representative pixels are selected based on similarity measure between these pixels. In this work, we use edge weights formula based on similarity computation between vertices given by Pham, Mercier and Michel (2014b), defined as follows:

$$w(i,j) = \begin{cases} e^{-\gamma[\text{dist}(\delta(i),\delta(j))]^2} & \text{if } j \in N_k(i) \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

where $\text{dist}(\delta(i),\delta(j))$ is the distance between compared spectrum pixels $(\delta(i)),(\delta(j))$ of vertices i and j and γ is set to 1 (Pham, Mercier and Michel 2014a). Figure 2 shows a schematic of selection of representative pixels.

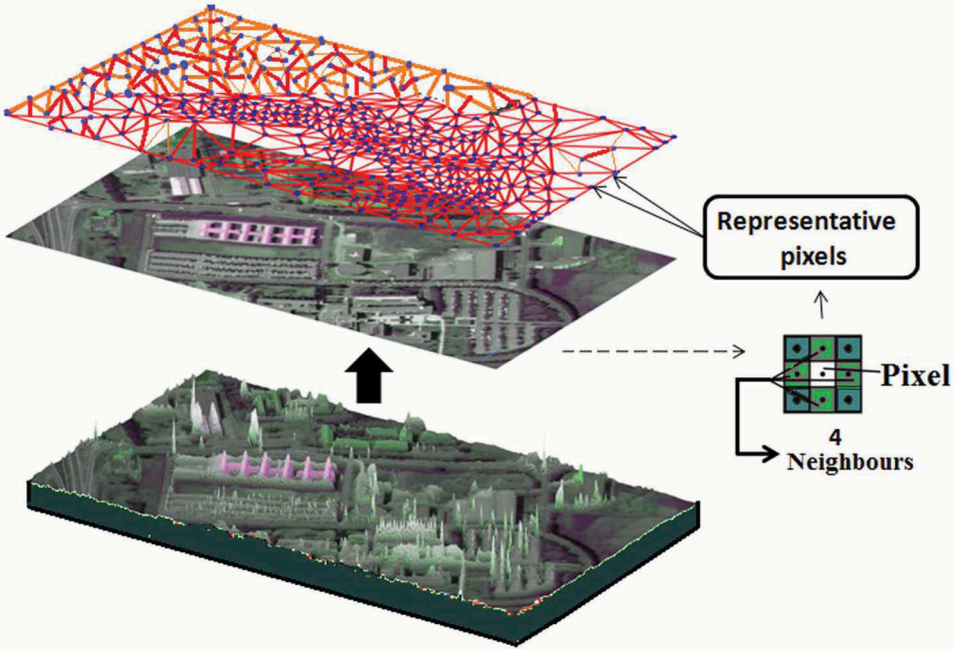

Figure 2. Selection of representative pixels.

The calculation of spectral similarity between the connecting pixels allows the identification of various classes in the image. To show the effect of these distance measures on the classification results, we have chosen five similarity measures which are frequently used for classification purposes. Thus, we introduce the Euclidian distance (ED) (Keshava 2004; Robila 2005), Kolmogorov distance (KD) (Franchi and Angulo 2015), Spectral angle (SA) (Robila 2005; Achalakul and Taylor 2001; Franchi and Angulo 2015), Spectral information divergence (SID) (Qin et al. 2009; Franchi and Angulo 2015) and Normalized cross-correlation (NCC) (Nakhmani and Tannenbaum 2013).

2.4.1. Euclidean distance

The Euclidean distance measures the similarity between two observations in the same moment t. In the context of our work, the measure is used to estimate the distance between two spectra, i.e. the spectrum to classify and the T-neighbourhood pixels spectrum. The Euclidean distance (ED) of two n -dimensional vectors d_1 and d_2 is defined as:

$$d_{\mathrm{ED}}\ (d_1, d_2) = \sqrt{\sum_{i=1}^{n} (d_1 - d_2)^2} \tag{11}$$

2.4.2. Kolmogorov distance

The Kolmogorov-Smirnov distance is a common distance measure which calculates the maximal difference between the cumulative distributions of two spectra. The Kolmogorov-Smirnov distance of two n-dimensional vectors d_1 and d_2 is defined as:

$$d_{\mathrm{KD}}\ (d_1, d_2) = \max(|\mathrm{P}_{d_1,1}, \mathrm{P}_{d_2,1}|, \ldots, |\mathrm{P}_{d_1,2}, \mathrm{P}_{d_2,2}|) \tag{12}$$

where P_{d_1} and P_{d_2} are the cumulative distributions of d_1 and d_2.

2.4.3. Spectral angle

The spectral angle is another distance measure which is used in hyperspectral data classification for assessing the similarity between two spectra. The spectral angle (SA) gives the angle formed between the spectrum to classify and the neighbourhood spectrum. The SA of two n-dimensional vectors d_1 and d_2 is defined as:

$$\mathrm{SA}(d_1, d_2) = \arccos\ \left(\frac{\langle d_1 . d_2 \rangle}{\| d_1 \|_2\ \| d_2 \|_2}\right) \tag{13}$$

where $< . >$ represents the dot product of the vectors and $\| . \|$ represents the Euclidean norm.

2.4.4. Spectral information divergence

The spectral information divergence (SID) measures spectral similarity between two n-dimensional pixel vectors d_1 and d_2. The larger the values of SID, the larger are the differences between the two spectra. It is defined as:

$$\mathrm{SID}(d_1, d_2) = D(d_1 \parallel d_2) + D(d_2 \parallel d_1) \tag{14}$$

D is the information divergence, $D(d_1 \parallel d_2)$ and $D(d_2 \parallel d_1)$ represent the relative entropy of d_1 respect to d_2 and of d_2 respect to d_1 respectively.

2.4.5. Normalized cross-correlation

Normalized cross-correlation (NCC) measures the degree of similarity between two compared images. The NCC is not invariant to linear changes in the amplitude of illumination and contrast variations. The NCC values obtained lead to values in interval $[0, 1]$, where 1 indicates best similarity. The NCC of two n-dimensional vectors d_1 and d_2 is defined as:

$$\mathrm{NCC} = \frac{\sum_{i=1}^{n} d_1 d_2}{\sqrt{\sum_{i=1}^{n} d_1^2 \sum_{i=1}^{n} d_2^2}} \tag{15}$$

Algorithm 1 resumes steps of implementation of SGWT for hyperspectral data images.

Algorithm 1 SGWT

Input: Hyperspectral data cube $(N' \times M' \times S')$
Output: SGWT coefficients $(N' \times M' \times S' \times (l'+1))$
Divide original data cube in $(N_1 \times M_1 \times S')$ sub-cubes
For each sub-cube
  Compute weighted graph adjacency matrix **A**
  Compute graph Laplacian **L**
  Design function filters for spectral graph transform
  For Each element of computed Chebyshev polynomials
    Compute all SGWT coefficients
  End For
End For

In Figure 3, an example of SGWT multi-scale decomposition applied on a region of the 5 [th] band of Pavia University image with $J = 3$ and four closest neighbours T-neighbourhood $= 4$.

The obtained SGWT coefficients will be classified using SVM classifier, which will be discussed in Subsection 2.5.

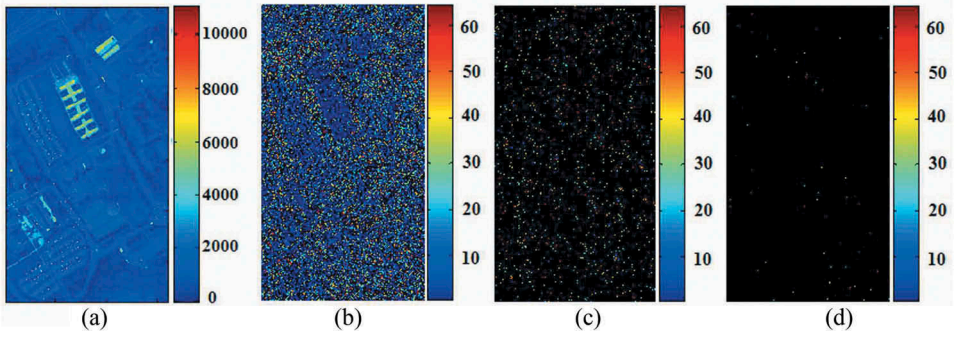

Figure 3. Multi-scale decomposition over the 5 [th] band of Pavia University image. (a) Scaling function. (b–d) SGWT coefficients, scales 1–3.

2.5. Classification

The SVM is an efficient supervised classifier which is based on small samples selected from the dataset. It was proposed by Vapnik (1995) as a kind of statistical learning method. The SVM classification allows the optimal separation hyperplane by the training of appropriate samples and the test of the input variables dataset. SVM is basically designed for binary classification but still has been extended to treat multiclass classification. On the other hand, if the training data are not linearly separable, a kernel function is employed to classify the samples. In our case, the SGWT extracted features are not linearly separable and so a kernel-based SVM is used for the classification. Furthermore, SVM decomposes extracted features of dataset on sub-regions to realize good classification by SVM kernel function. The classification is done by minimizing the following function:

$$\frac{1}{2}(a^{\mathsf{T}}a + b^2) - C\sum_{I=1}^{M}\xi_I \tag{16}$$

subject to constraints:

$$y_I(a^{\mathsf{T}}\phi(\mathrm{S}_I) + b) \geq 1 - \xi_I, \quad \xi_I \geq 0, I = 1, .., M \tag{17}$$

where C is a positive regularization parameter, a is the coefficient vector, b is the bias and ξ_I are called slack variables introduced to measure the non-separability of data. The class label $y \in \pm 1$ and S_I is the features space set for index I from 1 to M training cases. The radial basis function (RBF) ϕ is the SVM kernel function used for our approach. It is defined as follows:

$$\phi = \exp(-\gamma|X_I - X_J|^2) \tag{18}$$

2.6. Proposed algorithm

The SGWT-SVM classification performances are computed for the SVM classifier applied on SGWT coefficients. SVM model is constructed by selecting representative pixels of ground truth considered classes, which provides testing and training data groups. All steps of our proposed SGWT-SVM classifier are given in Algorithm 2.

Algorithm 2 Proposed SGWT-SVM classifier

Input: Hyperspectral data cube $(N' \times M' \times S')$
Output: Classified image $(N' \times M')$
Compute SGWT data cube $(N' \times M' \times S')(l' + 1)$
Load ground truth data $\mathbf{G}_{\mathrm{t}}(N' \times M')$
Scale data image, where scaled data $\in [-1, 1]$
Create training and testing data groups
Construct SVM model for SGWT coefficients
Compute classification performance parameters

3. Datasets and experiments

3.1. The proposed classifier's parameters

For our experiments, we use 200 training samples per class and we eliminate the classes which contain small training samples (Li et al. 2017b; Ran et al. 2017). For the proposed SGWT-SVM approach, all programs are implemented using Matlab and GNU Octave languages. SGWT toolbox (http://wiki.epfl.ch/sgwt) and LibSVM (http:https://www.csie.ntu.edu.tw/cjlin/libsvm/) are used for numerical computation. The SGWT was implemented with a number of 3 scales and a Gaussian filter. For the construction of graph, we set 4-neighbourhood pixels and Chebyshev polynomial degree of 100. The SVM model is used with a radial Basis Function kernel, a regularization factor $C = 1$ and 10 cross-validation. Note that these parameters are empirically set to arrive classification accuracy and computational time.

3.2. Datasets

To evaluate the effectiveness of the proposed method, four hyperspectral datasets available online (http:www.ehu.eus/ccwintco/index.php ? title = Hyperspectral Remote Sensing Scenes) are employed, i.e. Indian Pines, KSC, University and centre of Pavia data sets, as illustrated in Figure 4 (a), 5 (a), 6 (a) and 7 (a). For this evaluation, we adopt the classification metrics: overall accuracy (OA), average accuracy (AA), and kappa coefficient (κ) and we compare the performances of our method to other known classification methods in term of accuracy.

3.2.1. Indian pines dataset

This dataset was acquired in Northwestern Indiana in 1992 by the National Aeronautics and Space Administration (NASA) with Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor. It generates 220 spectral bands from 0.38 to 2.5 μm, which in 20 noisy bands are removed for experiments. The spatial image size is 145 $\times$ 145 with a resolution of 20 m per pixel. This hyperspectral dataset includes 16 ground-truths and 10,366 labelled pixels, we randomly use 9 classes, as shown in Table 1 (Qian, Ye, and Zhou 2013; Ham et al. 2005). Figure 4(a,b) shows the Indiana original image and the ground truth map, respectively.

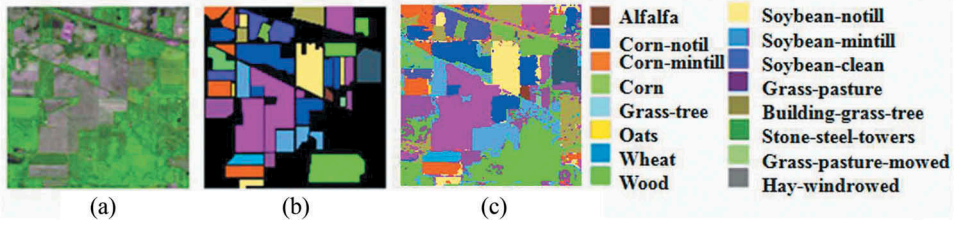

Figure 4. Results on Indian Pines data. (a) Pseudo colour original image. (b) Ground truth map. (c) Classification results of SGWT-SVM.

Table 1. Ground truth classes of training Indian Pines scene and their respective sample number.

	Class	Samples	Training	Test
1	Corn-notill	1434	200	1228
2	Corn-mintill	834	200	630
3	Grass-pasture	497	200	283
4	Grass-trees	747	200	530
5	Hay-windrowed	489	200	278
6	Soybean-notill	968	200	772
7	Soybean-mintill	2468	200	2255
8	Soybean-clean	614	200	393
9	Woods	1294	200	1065

3.2.2. KSC dataset

Kennedy Space Centre database was acquired in Florida in 1996 by NASA with AVIRIS sensor. After removing noisiest spectral bands, 176 bands are used for experiments with spatial size image of 512 × 614. It contains 13 land-cover classes with 5211 labelled pixels (Qian, Ye, and Zhou 2013; Tarabalka, Benediktsson, and Chanussot 2009). As in Indian Pines data, we use only the classes with the largest number of samples; they are given in Table 2. Figure 5(a,b) shows the KSC original image and the ground truth map, respectively.

3.2.3. Pavia university dataset

This third dataset used in our experiment, Pavia U, was acquired over Pavia University by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor in 2001. It generates 115 spectral bands from 0.43 to 0.8 μm, which in 12 noisy bands are removed for experiments. The size of each band is 641 × 340 with a resolution of 1.3 m per pixel. This base contains 9 classes shown in Table 3 (Qian, Ye, and Zhou 2013; Tarabalka,

Table 2. Ground truth classes of KSC scene and their respective sample number.

	Class	Samples	Training	Test
1	Scrub	761	200	561
2	Graminoid marsh	431	200	231
3	Spartina marsh	520	200	320
4	Cattail marsh	404	200	203
5	Salt marsh	419	200	218
6	Mud flats	503	200	302
7	Water	927	200	726

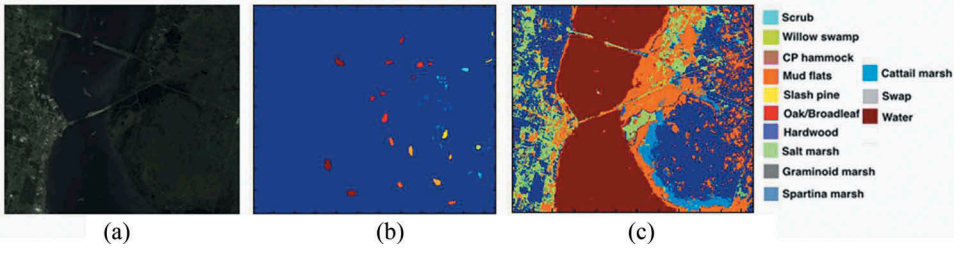

Figure 5. Results on KSC data. (a) Pseudo colour original image. (b) Ground truth map. (c) Classification results of SGWT-SVM.

Table 3. Ground truth classes of training Pavia University scene and their respective sample number.

	Class	Samples	Training	Test
1	Asphalt	6631	200	6431
2	Meadows	18,649	200	18,449
3	Gravel	2099	200	1899
4	Trees	3064	200	2864
5	Painted metal sheets	1345	200	1145
6	Bare soil	5029	200	4829
7	Bitumen	1330	200	1130
8	Self-blocking bricks	3682	200	3482
9	Shadows	947	200	747

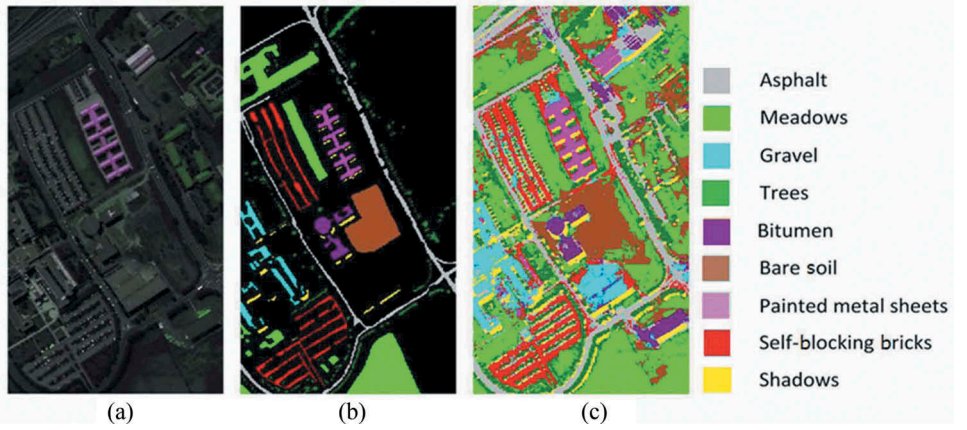

Figure 6. Results on Pavia University data. (a) Pseudo colour original image. (b) Ground truth map. (c) Classification results of SGWT-SVM.

Benediktsson, and Chanussot 2009). Figure 6(a,b) shows the Pavia U original image and the ground truth map, respectively.

3.2.4. Pavia centre dataset

Similar to Pavia U data, Pavia centre data is acquired by ROSIS sensor covering the centre city of Pavia in Italy. It also contains 115 spectral bands from 0.43 to 0.86 μm, with 13 bands removed. Only a size of 1096 $\times$ 715 of Pavia Centre scene is generally used for test experiments. The rest of information is discarded. In Table 4, we show the 9 classes of this

Table 4. Ground truth classes of training Pavia Centre scene and their respective sample number.

	Class	Samples	Training	Test
1	Water	65,971	200	65,771
2	Trees	7598	200	7398
3	Asphalt	3090	200	2890
4	Self-blocking bricks	2685	200	2485
5	Bitumen	6584	200	6384
6	Tiles	9248	200	9048
7	Shadows	7287	200	7087
8	Meadows	42,826	200	42,626
9	Bare soil	2863	200	2663

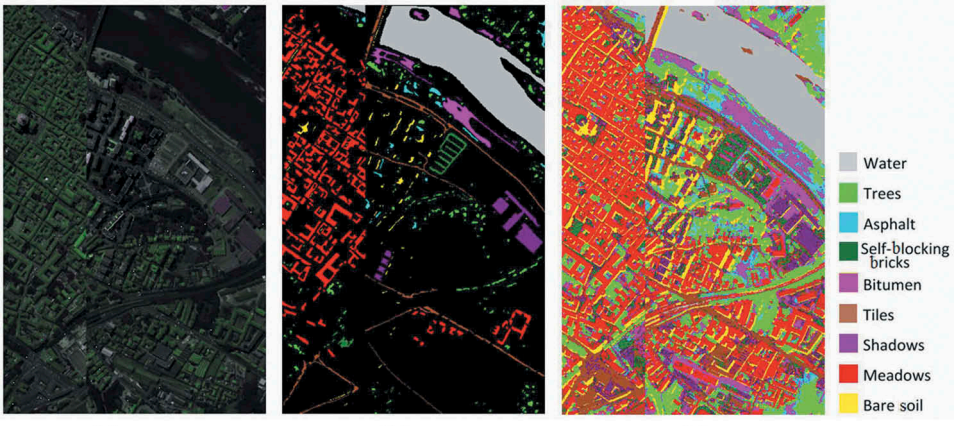

Figure 7. Results on Pavia Centre data. (a) Pseudo colour original image. (b) Ground truth map. (c) Classification results of SGWT-SVM.

dataset (Mei et al. 2017). In Figure 7(a,b), the Pavia Centre original image and the ground truth map, respectively, are shown.

3.3. Discussion on distance measures

In order to verify the impact of the different distance measures presented in Section 2.4 on the classification performances, we compare the overall accuracy (OA), average accuracy (AA) and κ given by each distance on the four datasets. Results presented in Table 5, show high classification performances with the tested distance measurements for all datasets. However, one can observe that SA, SID, and NCC measures outcome over ED and KD measures. Indeed, the three evaluation parameters for ED and KD are less than the

Table 5. Experimental results for Euclidian (ED), Kolmogorov (KD) Spectral angle (SA), Spectral information divergence (SID), and Normalized cross correlation (NCC) distance measures.

Dataset	Samples number	Results	ED	KD	SA	SID	NCC
Indian Pines	200 of training samples	OA	85.08	85.19	94.50	**95.21**	95.09
		AA	78.07	79.94	96.21	**96.61**	95.96
		κ	82.32	82.48	93.42	**94.27**	94.12
	20% of training samples	OA	87.82	87.99	96.65	**98.77**	96.58
		AA	88.39	87.00	96.35	**98.90**	96.52
		κ	0.85	0.85	0.94	**0.98**	0.94
KSC	200 of training samples	OA	99.20	99.56	99.65	99.65	**99.73**
		AA	99.60	99.49	99.75	**99.98**	99.80
		κ	0.99	0.99	0.99	**1.00**	0.99
Pavia U	200 of training samples	OA	90.82	91.74	96.14	96.68	**97.72**
		AA	92.82	96.16	96.62	97.41	**98.01**
		κ	0.88	0.87	0.95	0.95	**0.97**
	20% of training samples	OA	97.23	98.07	**99.65**	99.48	99.41
		AA	97.53	98.13	**99.47**	99.42	99.24
		κ	0.97	0.98	**0.99**	0.99	0.99
Pavia C	200 of training samples	OA	98.13	98.37	98.14	**99.72**	98.47
		AA	95.55	96.17	96.30	**99.60**	97.35
		κ	0.97	0.97	0.97	**0.99**	0.98

other distances. Especially for the Indian Pines data, these give about 10% less accuracy than the others. For SA, SID, and NCC, the results are very appreciable and represent competitive performances for each according to the data distributions. The SID outperforms the other computed distances for all classification measures (OA, AA and κ) over two data distributions (Indian Pines and Pavia C). Therefore, we further our experiments, for the next Section 3.4, using this distance. To highlight that in Indian Pines and Pavia University datasets, the selection of 200 training samples did not achieve the best classification accuracies. The selection of about 20% training samples for each class of these datasets considerably improves the classification performances. The test classification results for all datasets with all distance measurements are shown in Table 5. Note that the best obtained results are shown in this table using bold formatting.

3.4. Discussion on neighbouring components

To evaluate the influence of the number of T-neighbourhood taken pixels on the classification results, we select different T-neighbourhood pixels tested over the fourth HSI datasets using SID distance measure. Thus, we have tested the influence of four selected T-neighbourhood pixels (T = 1, 2, 4, 8). We also measure the computational time taken by each selection. Note that experiments were carried out using GPU computation provided by The Matlab Image Processing Toolbox to achieve the SGWT computation process of each divided original data cube. All experiments are executed with Intel Xeon-E5620 with 2.40 GHz and 32-GB RAM. From the results shown in Table 6, in all tested datasets, the classification performances increase with the number of taken neighbourhood pixels. Indeed, one can observe an improvement of at least 0.14 % of OA with increasing T for all datasets. Except for Pavia University dataset, for which one can highlight that the larger number of selected T-neighbourhood pixels does not yield to the better performance, i.e. the 8-neighbourhood pixels give less classification results than those given by 4-neighbourhood selected pixels. Similarly, the computational complexity of the proposed classifier increases when the T-neighbourhood pixels increase. As shown in Table 6, the

Table 6. OA (%), AA (%), κ, and Computational Time (s) of SGWT-SVM with different T-neighbourhood pixels.

		T-neighbourhood			
Dataset	Measures	$T = 1$	$T = 2$	$T = 4$	$T = 8$
Indian Pines	OA	81.59	97.56	98.77	99.30
	AA	86.69	97.77	98.90	99.24
	κ	0.78	0.97	0.98	0.99
	Time	187.68	171.76	237.16	352.84
KSC	OA	97.49	98.35	99.85	99.87
	AA	94.37	95.62	99.25	99.90
	κ	0.94	0.96	0.99	0.99
	Time	252.66	265.68	339.09	441.74
Pavia U	OA	97.91	98.29	99.65	99.10
	AA	97.57	97.75	99.47	98.62
	κ	0.97	0.97	0.99	0.98
	Time	417.36	445.70	469.10	556.55
Pavia C	OA	93.51	98.8	99.72	99.86
	AA	71.21	76.65	99.60	99.75
	κ	0.61	0.79	0.96	0.99
	Time	584.95	625.85	655.98	793.30

8-neighbourhood require a larger amount of computational time. One can conclude that the computational time is dominated by the number of selected T-neighbourhood pixels. Therefore, in our experiments, 4-neighbourhood pixels are used as input of our SGWT-SVM proposed classifier to minimize consumption-time with preserving classification performance.

3.5. Classification performance

To evaluate the effectiveness of the proposed SGWT-SVM, we use for each distribution the outperforming distance measure, i.e. SID for Indian Pines and Pavia centre datasets, NCC for KSC dataset and SA for Pavia University dataset. We compare results with those given by several other methods such as SVM, ELM, SPPF Framework (Ran et al. 2017), 3D-WT-SVM-rbf, 3D-WT-Mixed-lasso (Qian, Ye, and Zhou 2013), CNN (Li et al. 2017b), CNN-PPF (Li et al. 2017b; Mei et al. 2017), C-CNN (Mei et al. 2017) and R-PCA CNN (Mei et al. 2017; Makantasis et al. 2015) applied on the Indian Pines and Pavia U datasets. For the KSC HSI data, we compare our results with those presented in Qian, Ye, and Zhou (2013); Zhong et al. (2018). For Pavia Centre, results are compared with Hu et al. (2015); Li et al. (2017b); Mei et al. (2017); Makantasis et al. (2015). The tuning parameters of the outperforming methods compared to the proposed SGWT-SVM classifier are listed in Table 7.

Comparison results are shown in Tables 8–11 for the Indian Pines, KSC and Pavia U and Centre datasets, respectively. One can observe that the proposed SGWT-SVM classifier overcomes the other presented classifiers in terms of classification accuracies. For example, in Indian Pines, SGWT-SVM achieves an OA of 98.90%, which is almost 1% and 2% over 3DWT-SVM and C-CNN classifiers (97. 99 % and 96.76%) respectively. As in Indian Pines, in the Pavia University and KSC datasets, the proposed classifier performed better than all other classifiers and achieves higher classification accuracies. Similar to these datasets, Pavia Centre data achieves higher accuracies than the four other compared methods (SVM-RBF, CNN, R-PCA-CNN, and CNN-PPF) and presents competitive performances with C-CNN method.

Table 7. Parameters of the compared methods.

Method	Parameters			
DWT-based method	Wavelet filter	Model parameters selection	Sparsity parameter ($\lambda = 0.5^x$)	Training data
3DDWT SVM-rbf (Qian et al 2013)	Haar	With cross validation	$x = 7$ (IN) $x = 5$ (KSC) $x = 4$ (PU)	25 (%) (IN) 25 (%) (KSC) 100 (%)(PU)
CNN-based methods	Learning rate	Dimensionality of features	Window size	Training data
PPF-CNN (Li et al 2017b)	0.001 (IN) 0.01 (PU)	10	5×5	200
SPPF-Framework (Ran et al. 2017)	–	10	3×3	200
C-CNN (Mei et al. 2017)	0.01	100	5×5	200
SSRN (Zhong et al. 2018)	0.0003 (IN) 0.0003 (PU) 0.0001(KSC)	24 (IN) 24 (PU) 16 (KSC)	7×7	20% (IN) 10% (PU) 70% (KSC)

Table 8. Classification performances of Indian Pines using different techniques.

	SVM [1]	ELM [1]	CNN [2]	R-PCA- CNN [3, 4]	CNN- PPF [2, 3]	SPPF Frame-work[1]	3DDWT Mixed-Lasso[5]	C-CNN [3]	3DDWT SVM-rbf [5]	SGWT -SVM
1	78.26	79.40	78.58	82.39	92.99	94.22	–	96.28	–	96.26
2	81.27	85.08	85.23	85.41	96.66	97.94	–	92.26	–	100.0
3	98.59	96.47	95.75	95.24	98.58	100.0	–	99.3	–	98.61
4	98.68	99.06	99.81	99.25	100.0	99.43	–	99.25	–	100.0
5	100.0	100.0	99.64	100.0	100.0	100	–	100.0	–	100.0
6	76.94	86.66	89.63	82.76	96.24	95.85	–	92.84	–	97.24
7	65.10	69.84	81.55	96.2	87.80	92.20	–	98.21	–	99.18
8	84.99	89.31	95.42	82.14	98.98	98.47	–	92.45	–	98.86
9	98.78	98.40	98.59	99.81	99.81	99.81	–	98.98	–	100.0
AA	86.96	89.36	90.60	91.47	96.78	97.55	95.90	96.62	97.35	**98.77**
AO	80.72	83.80	86.44	91.09	94.34	95.92	96.78	96.76	97.99	**98.90**

[1]: Ran et al. (2017), [2]: Li et al. (2017b), [3]: Mei et al. (2017), [4]: Makantasis et al. (2015), [5]: Qian, Ye, and Zhou (2013)

Table 9. Classification performances of KSC using different techniques.

	SVM [6]	SAE [6]	3DDWTSVM-rbf [5]	CNNL [6]	CNN [6]	3DDWTMixedLasso [5]	SPC [6]	SPA [6]	SSRN [6]	SGWT -SVM
1	86.16	92.04	–	95.20	98.48	–	99.19	99.18	99.88	100.0
2	42.55	85.59	–	87.53	92.16	–	92.60	95.39	99.00	100.0
3	67.69	72.12	–	73.35	81.84	–	85.49	93.45	98.26	100.0
4	65.12	94.10	–	97.28	98.21	–	98.09	98.67	99.54	100.0
5	67.82	94.57	–	98.05	99.04	–	99.53	99.43	99.70	100.0
6	93.40	98.91	–	99.40	99.85	–	99.96	99.96	99.96	99.02
7	100	98.39	–	98.72	98.89	–	99.86	99.63	99.80	99.59
AA	65.64	89.76	93.23	92.57	95.09	96.74	96.56	97.81	99.33	**99.80**
AO	80.29	92.99	94.05	95.45	97.08	97.65	97.90	98.63	99.61	**99.73**

[5]: Qian, Ye, and Zhou (2013), [6]: Zhong et al. (2018)

Table 10. Classification performances of Pavia University using different techniques.

	SVM [1]	ELM [1]	CNN [2]	R-PCA-CNN [3, 4]	CNN- PPF [2, 3]	SPPF Frame-work [1]	3DDWT Mixed-Lasso [5]	C-CNN [3]	3DDWT-SVM-rbf[5]	SGWT -SVM
1	82.69	82.71	88.38	92.43	93.89	97.42	–	97.4	–	98.29
2	87.65	91.23	91.27	94.84	91.71	95.76	–	99.4	–	99.62
3	79.36	79.20	85.88	90.89	83.46	94.05	–	94.84	–	98.95
4	94.24	93.02	97.24	93.99	97.07	97.52	–	99.16	–	99.83
5	99.83	99.30	99.91	100.0	100.0	100.0	–	100.0	–	100.0
6	89.36	91.51	96.41	92.86	95.92	99.13	–	98.7	–	99.91
7	89.38	92.92	93.62	93.89	97.35	96.19	–	100.0	–	99.76
8	81.68	86.79	87.45	91.18	87.71	93.62	–	94.57	–	98.69
9	99.87	99.87	99.57	99.33	99.73	99.60	–	99.87	–	100.0
AA	89.34	90.73	93.36	94.38	94.09	97.03	97.56	98.22	98.63	**99.47**
AO	87.25	89.55	92.27	93.87	92.73	96.48	98.15	98.41	98.81	**99.65**

[1]: Ran et al. (2017), [2]: Li et al. (2017b), [3]: Mei et al. (2017), [4]: Makantasis et al. (2015), [5]: Qian, Ye, and Zhou (2013)

Figures 4(c), 5(c), 7(c) and 9(c) show the classification results of the proposed SGWT-SVM over Indian Pines, KSC, Pavia University, and Pavia Centre datasets. Results consistently support those given in Tables 8–11, where the ground cover maps of entire image scenes are greatly generated. The method achieves well-separated regions including those with unlabelled pixels.

To test the robustness of the proposed approach, we vary the number of training samples per class from 50 to 200 with an interval of 50. Results illustrated in Figure 8, show that the classification accuracies increase with the number of training data. Note that even

Table 11. Classification performances of Pavia Centre using different techniques.

	SVM [3]	CNN [7, 3]	R-PCA-CNN [3, 4]	CNN-PPF [2, 3]	C-CNN [3]	SGWT-SVM
1	99.81	99.94	99.91	98.94	100.0	100.0
2	92.23	93.34	94.58	98.04	99.18	99.30
3	95.22	93.01	95.36	97.44	99.45	99.20
4	94.85	90.34	96.02	99.11	99.64	99.7
5	96.01	93.00	94.94	98.75	99.81	100.0
6	96.66	95.09	97.13	98.82	99.27	99.07
7	90.63	92.30	93.64	93.69	99.28	99.22
8	98.59	99.33	99.03	99.72	99.92	100.0
9	99.96	99.55	99.47	100.0	100.0	100.0
AA	95.99	95.10	96.68	98.28	99.62	**99.60**
AO	98.09	98.15	98.53	98.85	99.83	**99.72**

[2]: Li et al. (2017a), [3]: Mei et al. (2017), [4]: Makantasis et al. (2015), [7]: Hu et al. (2015)

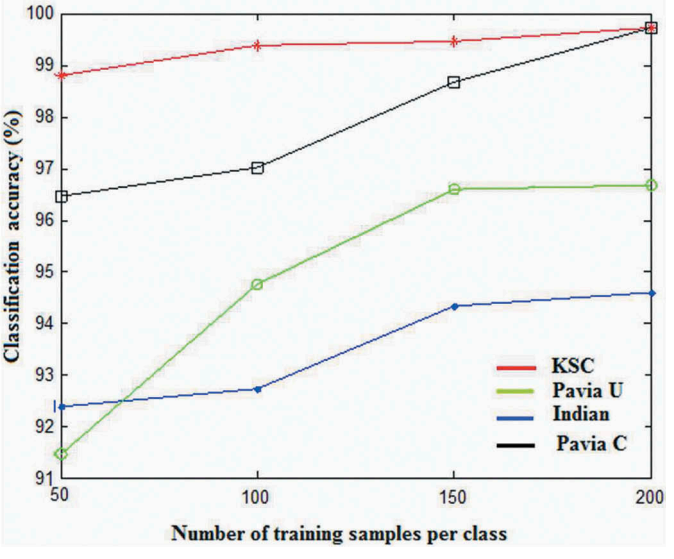

Figure 8. OA of SGWT-SVM varying the number of training samples per class.

with a small number of training samples (50, 100), the approach yields to high accuracy with all HSI used data. This demonstrates clearly the robustness of the SGWT-SVM classifier.

To further demonstrate the effectiveness and generalizability of the proposed method, we have illustrated in Figure 9, the overall accuracies of different classifiers with different numbers of training samples. A selection of 1 to 10 of training samples per class is effectuated for Indian Pines, Pavia Centre, and Pavia University datasets. The obtained results are compared to those given in Mei et al. (2017) (Figure 9(a,c)). For KSC dataset, and similar to Zhong et al. (2018), 1 % to 20 % of training samples are selected (Figure 9(d)). For all datasets, results show that the SVM-SGWT outperforms the other compared methods in terms of classification performance, even with small number of training samples. Indeed, an improvement of at least 0.3 % is observed with all datasets.

It is demonstrated that deep learning-based classification methods are vigorously effective for HSI image classification. However, results given by the proposed approach,

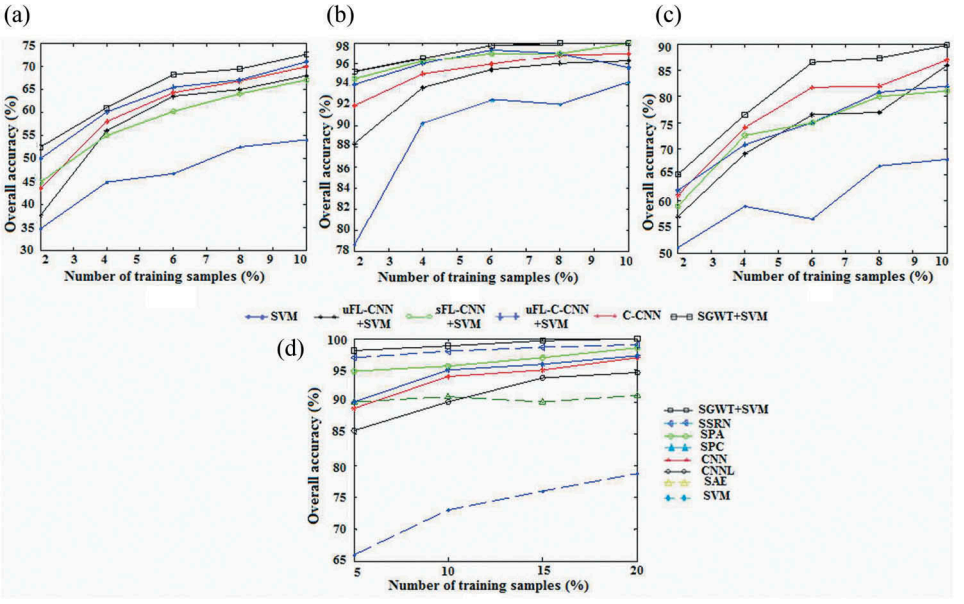

Figure 9. OA of different classifiers with different numbers of training samples. (a) Indian pines, (b) Pavia centre, (c) Pavia University, (d) KSC.

which consider the data structure with an appropriate selection of connected components overcome clearly several famous CNN-based HSI image classification methods. Note that only five similarity measurements are tested in this work. The use of other test similarity may improve the classification results. Considering the obtained classification performance and based on the fact that the proposed classifier can extract more discriminant features, one can conclude that the SGWT-SVM classifier can excel in HSI image classification fields.

4. Conclusion

In this paper, a Spectral Graph Wavelet Transform based 3D classifier is developed for hyperspectral classification. The proposed classifier is based on the classification of the SGWT obtained coefficients with SVM model. In our procedure, we also evaluated different measures of spectral similarity for better selection of representative pixels, and so identify classes in the image. Also, according to similarity measure's results, the choice of the distance measured between connected pixels can affect the classification performance. It is observed that SA, SID and NCC similarity measures outcome over ED and KD measures and lead to appreciable classification results. Results of the proposed approach with SID distance applied on four selected hyperspectral image datasets of AVIRIS and ROSIS sensors demonstrate the effectiveness and the robustness of this one. Thus, classification accuracies obtained with our method even with a modest number of training simples, show competitive performances with those of the best recent techniques presented in state-of-the-art of hyperspectral image classification. Future works on denoising, unmixing and compression problems can be proposed. Another interesting

perspective is to use a new classifier other than the SVM such as Random multi-graphs (RMG) classifiers (Zhang et al. 2017; Gao et al. 2018). This may improve classification performances.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Achalakul, T., and S. Taylor. 2001. "Real-time Multi-spectral Image Fusion." *Concurecncy and Computing: Practice and Experience* 13: 1063–1081. doi:10.1002/cpe.591.

Bovolo, F., L. Bruzzone, and L. Carlin. 2010. "A Novel Technique for Subpixel Image Classification Based on Support Vector Machine." *IEEE Transactions on Image Processing* 19 (11): 2983–2999. doi:10.1109/TIP.2010.2051632.

Bruzzone, L., and C. Persello. 2009. "Approaches Based on Support Vector Machine to Classification of Remote Sensing Data." *In Handbook of Pattern Recognition and Computer Vision C. Chen, Ed. Singapore: World Scientific* 3 (2): 329–352.

Chen, Y., N. M. Nasrabadi, and T. D. Tran. 2013. "Hyperspectral Image Classification via Kernel Sparse representation." *IEEE Transactions on Geoscience and Remote Sensing* 51 (1): 217–231. doi:10.1109/TGRS.2012.2201730.

Chen, Y., Z. Lin, X. Zhao, G. Wang, and Y. Gu. 2014. "Deep Learning-based Classification of Hyperspectral Data." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7 (6): 2094–2107. doi:10.1109/JSTARS.2014.2329330.

Franchi, G., and J. Angulo 2015. "Quantization of Hyperspectral Image Manifold Using Probabilistic Distances". In *International Conference on Networked Geometric Science of Information*, 406–414. Springer, Cham, Germany..

Gao, F., Q. Wang, J. Dong, and Q. Xu. 2018. "Spectral and Spatial Classification of Hyperspectral Images Based on Random Multi-Graphs." *Remote Sensing* 10 (8): 1271. doi:10.3390/rs10081271.

Guo, B., S. R. Gunn, R. I. Damper, and J. D. B. Nelson. 2008. "Customizing Kernel Functions for SVM-based Hyperspectral Image Classification." *IEEE Transactions on Image Processing* 17 (4): 622–629. doi:10.1109/TIP.2008.918955.

Ham, J., Y. Chen, M. M. Crawford, and J. Ghosh. 2005. "Investigation of the Random Forest Framework for Classification of Hyperspectral Data." *IEEE Transactions on Geoscience and Remote Sensing* 43 (3): 492–501. doi:10.1109/TGRS.2004.842481.

Hammond, D. K., P. Vandergheynst, and R. Gribonval. 2011. "Wavelets on Graphs via Spectral Graph Theory." *Applied and Computational Harmonic Analysis* 30 (2): 129–150. doi:10.1016/j.acha.2010.04.005.

He, L., Y. Li, X. Li, and W. Wu. 2015. "Spectral-Spatial Classification of Hyperspectral Images via Spatial Translation-invariant Wavelet-based Sparse representation." *IEEE Transactions on Geoscience and Remote Sensing* 53 (5): 2696–2712. doi:10.1109/TGRS.2014.2363682.

Hu, W., Y. Huang, W. Li, F. Zhang, and H. Li. Jan 2015. "Deep Convolutional Neural Networks for Hyperspectral Image Classification." *Journal of Sensors* 2015 Art. no. 258619: 1–12. doi:10.1155/2015/258619.

Kerola, T., N. Inoue, and K. Shinoda 2014. "Spectral Graph Skeletons for 3D Action Recognition". In *Asian Conference on Computer Vision*, 417–432. Springer, Cham, Germany..

Keshava, N. 2004. "Distance Metrics and Band Selection in Hyperspectral Processing with Applications to Material Identification and Spectral Libraries." *IEEE Transactions in Geosciences and Remote Sensing* 42: 1552–1565. doi:10.1109/TGRS.2004.830549.

Leonardi, N., and D. Van De Ville 2011. "Wavelet Frames on Graphs Defined by fMRI Functional Connectivity". In *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on, Chicagi, IL, IEEE*, 2136–2139.

Li, W., G. Wu, F. Zhang, and Q. Du. 2017a. "Hyperspectral Image Classification Using Deep Pixel-pair features." *IEEE Transactions on Geoscience and Remote Sensing* 55 (2): 844–853. doi:10.1109/TGRS.2016.2616355.

Li, Y., W. Xie, and H. Li. 2017b. "Hyperspectral Image Reconstruction by Deep Convolutional Neural Network for Classification." *Pattern Recognition* 63: 371–383. doi:10.1016/j.patcog.2016.10.019.

Liu, L., P. Fieguth, D. Clausi, and G. Kuang. 2012. "Sorted Random Projections for Robust Rotation-invariant Texture Classification." *Pattern Recognition* 45 (6): 2405–2418. doi:10.1016/j.patcog.2011.10.027.

Ma, X., J. Geng, and H. Wang. 2015. "Hyperspectral Image Classification via Contextual Deep Learning." *EURASIP Journal on Image and Video Processing* 20 (1): 1–12.

Makantasis, K., K. Karantzalos, A. Doulamis, and N. Doulamis 2015. "Deep Supervised Learning for Hyperspectral Data Classification through Convolutional Neural Networks". In *Proc. IGARSS*, 4959–4962. Milan, Italy.

Malek, M., D. Helbert, and P. Carre. 2015. "Color Graph Based Wavelet Transform with Perceptual Information." *Journal of Electronic Imaging* 24 (5): 053004. doi:10.1117/1.JEI.24.5.053004.

Masoumi, M., and A. B. Hamza. 2017. "Shape Classification Using Spectral Graph Wavelets." *Applied Intelligence* 47 (4): 1256–1269. doi:10.1007/s10489-017-0955-7.

Mei, S., J. Ji, J. Hou, X. Li, and Q. Du. 2017. "Learning Sensor-specific Spatial-spectral Features of Hyperspectral Images via Convolutional Neural Networks." *IEEE Transactions on Geoscience and Remote Sensing* 55 (8): 4520–4533. doi:10.1109/TGRS.2017.2693346.

Mou, L., P. Ghamisi, and X. X. Zhu. 2017. "Deep Recurrent Neural Networks for Hyperspectral Image Classification." *IEEE Transactions on Geoscience and Remote Sensing* 55 (7): 3639–3655. doi:10.1109/TGRS.2016.2636241.

Nakhmani, A., and A. Tannenbaum. 2013. "A New Distance Measure Based on Generalized Image Normalized Cross-correlation for Robust Video Tracking and Image Recognition." *Pattern Recognition Letters* 34 (3): 315–321. doi:10.1016/j.patrec.2012.10.025.

Persello, C., and L. Bruzzone. 2016. "Kernel-based Domain-invariant Feature Selection in Hyperspectral Images for Transfer Learning." *IEEE Transactions on Geoscience and Remote Sensing* 54 (5): 2615–2626. doi:10.1109/TGRS.2015.2503885.

Pham, M. T., G. Mercier, and J. Michel. 2014a. "Textural Features from Wavelets on Graphs for Very High Resolution Panchromatic Pléiades Image Classification." *Revue Française De Photogrammétrie Et De Télédétection* 208: 131–136.

Pham, M. T., G. Mercier, and J. Michel 2014b. "Wavelets on Graphs for Very High Resolution Multispectral Image Texture Segmentation". In *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International*, Quebec, Canada, 2273–2276. IEEE. doi:10.1021/es4049626.

Pham, N., D. Helbert, P. Bourdon, and P. Carre 2018. "Spectral Graph Wavelet Based Nonrigid Image Registration". In *2018 25th IEEE International Conference on Image Processing (ICIP)*, Athens, Greece, 3348–3352. IEEE.

Qian, Y., M. Ye, and J. Zhou. 2013. "Hyperspectral Image Classification Based on Structured Sparse Logistic Regression and Three-dimensional Wavelet Texture Features." *IEEE Transactions on Geoscience and Remote Sensing* 51 (4): 2276–2291. doi:10.1109/TGRS.2012.2209657.

Qin, J., T. F. Burks, M. A. Ritenour, and W. G. Bonn. 2009. "Detection of Citrus Canker Using Hyperspectral Reflectance Imaging with Spectral Information Divergence." *Journal of Food Engineering* 93 (2): 183–191. doi:10.1016/j.jfoodeng.2009.01.014.

Ran, L., Y. Zhang, W. Wei, and Q. Zhang. 2017. "A Hyperspectral Image Classification Framework with Spatial Pixel Pair Features." *Sensors* 17 (10): 2421. doi:10.3390/s17102421.

Robila, S. A. 2005. "Using Spectral Distances for Speedup in Hyperspectral Image Processing." *International Journal of Remote Sensing* 26 (24): 5629–5650. doi:10.1080/01431160500168728.

Shuman, D. I., C. Wiesmeyr, N. Holighaus, and P. Vandergheynst. 2015. "Spectrum-adapted Tight Graph Wavelet and Vertex-frequency Frames." *IEEE Transactions on Signal Processing* 63 (16): 4223–4235. doi:10.1109/TSP.2015.2424203.

Tang, Y. Y., Y. Lu, and H. Yuan. 2015. "Hyperspectral Image Classification Based on Three-dimensional Scattering Wavelet Transform." *IEEE Transactions on Geoscience and Remote Sensing* 53 (5): 2467–2480. doi:10.1109/TGRS.2014.2360672.

Tarabalka, Y., J. A. Benediktsson, and J. Chanussot. 2009. "Spectral-spatial Classification of Hyperspectral Imagery Based on Partitional Clustering Techniques." *IEEE Transactions on Geoscience and Remote Sensing* 47 (8): 2973–2987. doi:10.1109/TGRS.2009.2016214.

Tarabalka, Y., J. A. Benediktsson, J. Chanussot, and J. C. Tilton. 2010. "Multiple Spectral-spatial Classification Approach for Hyperspectral Data." *IEEE Transactions on Geoscience and Remote Sensing* 48 (11): 4122–4132.

Vapnik, V. N. 1995. *The Nature of Statistical Learning Theory*. Berlin: Springer-Verlag.

Yang, L., S. Yang, P. Jin, and R. Zhang. 2014. "Semi-supervised Hyperspectral Image Classification Using Spatio-spectral Laplacian Support Vector Machine." *IEEE Geoscience and Remote Sensing Letters* 11 (3): 651–655. doi:10.1109/LGRS.2013.2273792.

Zhang, L., L. Zhang, and B. Du. 2016. "Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art." *IEEE Geoscience and Remote Sensing Magazine* 4 (2): 22–40. doi:10.1109/MGRS.2016.2540798.

Zhang, M., W. Li, and Q. Du. 2018. "Diverse Region-Based CNN for Hyperspectral Image Classification." *IEEE Transactions on Image Processing* 27 (6): 2623–2634. doi:10.1109/TIP.2018.2809606.

Zhang, Q., J. Sun, G. Zhong, and J. Dong. 2017. "Random Multi-graphs: A Semi-supervised Learning Framework for Classification of High Dimensional Data." *Image and Vision Computing* 60: 30–37. doi:10.1016/j.imavis.2016.08.006.

Zhong, Z., J. Li, Z. Luo, and M. Chapman. 2018. "Spectral-Spatial Residual Network for Hyperspectral Image Classification: A 3-D Deep Learning Framework." *IEEE Transactions on Geoscience and Remote Sensing* 56 (2): 847–858. doi:10.1109/TGRS.2017.2755542.

Zhou, S., J. Shi, J. Zhu, Y. Cai, and R. Wang. 2013. "Shearlet-based Texture Feature Extraction for Classification of Breast Tumor in Ultrasound Image." *Biomedical Signal Processing and Control* 8 (6): 688–696. doi:10.1016/j.bspc.2013.06.011.

Zou, Q., L. Ni, T. Zhang, and Q. Wang. 2015. "Deep Learning Based Feature Selection for Remote Sensing Scene classification." *IEEE Geoscience and Remote Sensing Letters* 12 (11): 2321–2325. doi:10.1109/LGRS.2015.2475299.