International Journal of Wavelets, Multiresolution and Information Processing

Hyperspectral Image Classification Using Wavelet Transform-Based Smooth Ordering

Lina Yang[†], Hailong Su[*], Cheng Zhong[‡], and Zuqiang Meng[§]
School of Computer, Electronics and Information, Guangxi University, Nanning, 530004, China
[†]*linayang@gxu.edu.cn*
[*]*hailongsu@outlook.com*
[‡]*chzhong@gxu.edu.cn*
[§]*zqmeng@126.com*

Huiwu Luo
AI Research, Sichuan Changhong Electric Co., Ltd., Chengdu, 610000, China
luohuiwu@gmail.com

Xichun Li
Guangxi Normal University for Nationalities, Chongzuo, 532200, China
xichunli@yahoo.com

Yuan Yan Tang
Faculty of Science and Technology, University of Macau, Macau, 999078, China
yytang@umac.mo

Yang Lu
Department of Computer Science, Hong Kong Baptist University, Hongkong, 999077, China
lylylytc@gmail.com

To efficiently improve the accuracy of hyperspectral image (HSI) classification, the spatial information is usually fused with spectral information so that the classification performance can be enhanced. In this paper, we propose a new classification method called wavelet transform-based smooth ordering (WTSO). WTSO consists of three main components: wavelet transform for feature extraction, spectral-spatial based similarity measurement, smooth ordering based 1-D embedding, and construction of final classifier using interpolation scheme. Specifically, wavelet transform is firstly imposed to decompose the HSI signal into approximate coefficients (ACs) and details coefficients (DCs). Then, to measure the similar level of pairwise samples, a novel metric is defined on the ACs, where the spatial information is serving as the prior knowledge. Next, according to the measurement results, smooth ordering is applied so that the samples are aligned in a 1-D space (called 1-D embedding). Finally, since the reordering samples are smooth, the labels of test samples can be recovered using the simple 1-D interpolation method. In the last step, in order to reduce the bias and improve accuracy, the final classifier is constructed using multiple 1-D embeddings. The use of wavelet transform in WTSO

[*]Corresponding author

can also reduce the high dimensionality of HSI data. By converting the hight dimensional samples into a 1-D ordering sequence, WTSO can reduce the computational cost, and simultaneously perform classification for the test samples. Note that in WTSO, the smooth ordering based 1-D embedding and interpolation are executed in an iterative manner. And they will be terminated after a finite steps. The proposed method is experimentally demonstrated on two real HSI data sets: IndianPines and University of Pavia, achieving promising results.

Keywords: Wavelet transform; smooth ordering; HSI classification; smooth interpolation; feature extraction.

AMS Subject Classification: 22E46, 53C35, 57S20

1. Introduction

Over the past few decades, hyperspectral image (HSI), which is captured through remote sensing sensor, has become a hot topic in HSI research community.[25] A typical hyperspectral image has hundreds of narrow contiguous bands. Each pixel is a continuous spectrum that distinguishes different materials. Thus, it can be used for recognizing different ground objects with great accuracy and high precision.[3,7,11] The object of HSI classification is to assign a meaningful label to each pixel so that the pixel is interpretable. Although many methods have been developed and proposed to deal with this problem, the high similarity of different HSI classes bring many challenges to HSI classification. These challenges include high dimensionality and small sample size, which is also called *the curse of dimensionality* ("*Hughes phenomenon*"). Note that this problem is a normal phenomenon in HSI classification.[12]

In order to achieve high classification accuracy, many methods have been proposed.[22] Among these methods, support vector machine (SVM) performs state-of-the-art results. The strategy of SVM is to map the nonlinear separable data into a much higher dimensional space using a kernel trick, thus that the data can be linearly separated in the kernel-induced space.[5,6,14] Recently, sparse representation (SR), which was mainly used to process signal data, has been derived and applied to HSI classification. According to those technical reports published in recent years, SR-based classification methods can obtain excellent classification results.[8,13,16,23,24] Specifically, the idea of using SR for HSI classification was first introduced by Chen *et al.*,[36] where their proposed method is named *sparse representation classification* (SRC) method. According to our understanding, SRC method assumes that HSI pixels can be sparsely represented by a linear combination of a few training samples over a dictionary. The solution of this problem can be typically solved by a greedy algorithm, such as orthogonal associated pursuit (OMP) algorithm. The final test label is determined by a minimal residual principle.

Note that for HSI classification, the adjacent pixels trend to belong to the same class. It can serve as a prior knowledge, helping to decide the class label of the test sample. Thus, the spatial information is very helpful to determine the hardly classified samples. Motivated by this observation, many spectral-spatial approaches are

proposed.[15] For example, Tang *et al.*[38] proposed two kinds of sparse representation algorithms based on manifold to exploit the local structure of a test samples, where the spatial information is encoded as a regularization term in the objective function. Luo *et al.*[18] proposed a spectral-spatial one-dimensional manifold embedding (SS1DME), which utilizes the spectral-spatial information-based metric to learn the similarity of HSI pixels.

One of the main stream is to reduce the dimension of HSI data without losing the discriminant information.[1,10,30] Specifically, wavelet transform, which considers the problem in the frequency domain, is a powerful mathematical tool to extract the useful information for classification. Given an input signal, it provides time-frequency localization to distinguish different classes.[21] Wavelet transform provides an efficient way to reduce the dimensionality of HSI data. Considering this advantage, many wavelet transform based methods have been introduced for HSI classification. For example, Wang *et al.*[34] proposed stationary wavelet transform (SWT) to extract the spatial features. Each spectrum band image is firstly converted to SWT domain, then principal component analysis (PCA) is applied to reduce the dimensionality, finally k-nearest neighbors algorithm is employed for classification. Tang *et al.*[37] proposed a 3-D scattering wavelet transform, which filters the HSI cube data with a cascade of wavelet decompositions.[2] Qian *et al.*[26] proposed a method based on structured sparse logistic regression and 3-D discrete wavelet transform (3D-DWT) texture features. However, the aforementioned methods work on high-dimensional space. The complex theory makes the decision bound hard to determine.

In order to deal with this issue, Wang[33] proposed a novel method called *smooth ordering* to classify the handwritten digits, achieving comparable results. Motivated by his pioneered work, in this paper, we try to integrate wavelet transform with smooth ordering to further explore the potential power of smooth ordering. We point out that our work is similar to the work taken by Ram *et al.*[9] According to their publication, their work is highly effective when used in image denoising, inpainting, and deblurring. However, in their work, smooth ordering is applied to overlapped patches, and the input signal is two dimensional, which is not suitable for HSI classification. In order to be able to fit the task of HSI classification, with a slight modification, we define a new distance measure before applying the scheme of smooth ordering. Then, according to the observation that the ACs contains the main information of the input signal, smooth ordering is applied to ACs only.

The main steps of our method can be explained as follows: Firstly, the wavelet transform is adopted to decompose the HSI image into ACs and DCs. Secondly, using a distance measurement on the ACs in our previous work,[35] the similarity of pairwise samples is calculated accordingly. Next, based on the similarity measurement, smooth ordering algorithm is applied to the coefficients to obtain a 1-D embedding of the input ACs. This calculation results in an ordered sequence of the input data in 1-D space. Finally, a group of classifiers are constructed using multiple 1-D embeddings, and the final class is assigned by the maximum voting rule.

The highlight of this work is to explore the capability of smooth ordering in frequency domain using wavelet transform. In details, we refine the following aspects to highlight our work:

(1) A new metric is defined to make our algorithm work properly in the wavelet domain.
(2) Smooth ordering is introduced to explicitly find out the decision bound, making the results more intuitive.

More importantly, the results of WTSO is superior when compared with other state-of-the-art methods.

The remainder of this paper is organized as follows. The proposed WTSO is described in Section 2. The effectiveness of the proposed WTSO method is experimentally evaluated and discussed in Section 3 using two real HSI data sets. Finally, we conclude our work in Section 4.

2. Methodology

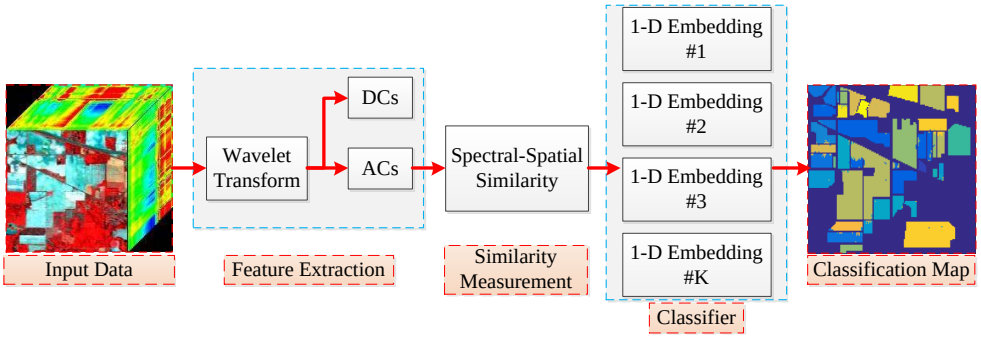


Fig. 1. Flowchart of the proposed WTSO classification scheme.

In this section, we introduce our proposed WTSO method. The flowchart of our method is shown in Fig. 1. The details are described in the following subsections.

2.1. *Feature Extraction Using Wavelet Decomposition*

In this section, we briefly review the concept of wavelet transform. Wavelet transform can be found in any articles that are related to the concept of wavelet, such as the one in Qian *et al.*[26] Wavelets are powerful mathematical tool to analyze 1-D/2-D signal data in time-frequency domain. An input signal can be decomposed from mother wavelets that are scaled and translated. This procedure can be formalized

as:

$$(W_{\psi f})(a,b) = < f(t), \psi_{a,b}(t) > = \int f(t)\psi_{a,b}(t)dt \tag{2.1}$$

where $\psi_{a,b}(t) = (1/\sqrt{|a|})\psi_{(t-b/a)}$, and $\int \psi(t)dt = 0$. The parameter a is a scale factor that dependently scales the frequency. Large $|a|$ means low frequency, whereas small $|a|$ means high frequency. On the other hand, parameter b is the time of input signal.

Benefiting from the two parameters, wavelet transform can observe signal structures by by discreting the parameters b and a (*i.e.*, time and frequency domain). In that case, the *discrete wavelet transform* is obtained:

$$(W_{\psi f})(m,n) = < f(t), \psi_{m,n}(t) > = \int f(t)\psi_{m,n}(t)dt \tag{2.2}$$

where $\psi_{m,n}(t) = a_0^{-m/2}\psi(t - nb_0a_0^m/a_0^m)$. Using discrete wavelet transform, the signal is decomposed into approximate coefficients (ACs) and detail coefficients (DCs), respectively.

If the wavelet functions and scaling functions are symbolized by $\psi(t)$ and $\varphi(t)$, the original signal $f(t)$ can be recovered by:

$$f(t) = \sum_k c_{j_0}(k)\varphi_{j_0,k}(t) + \sum_{j=j_0}^{\infty}\sum_k d_j(k)\psi_{j,k}(t) \tag{2.3}$$

where the ACs $c_{j_0,k}$ and DCs $d_{j,k}$ are given by:

$$c_{j_0,k} = < f(t), \varphi_{j_0,k}(t) > = \int f(t)\varphi_{j_0,k}(t)dt \tag{2.4}$$

$$d_{j,k} = < f(t), \psi_{j,k}(t) > = \int f(t)\psi_{j,k}(t)dt \tag{2.5}$$

In the proposed WTSO approach, we use a 2-D discrete wavelet transform to extract the descriptive features that are stored in the decomposed coefficients. Particularly, the symlets wavelet, which is a modified version of Daubechies wavelets by increasing the symmetry, is adopted. Because the information of an image is mainly represented in the ACs, we only use the approximate coefficients. The demonstration is visualized in Fig. 2.

2.2. *Spectral-Spatial Based Similarity Measurement*

Let $\mathcal{X} = \{x_i\}_{i=1}^N \in \mathbb{R}^n$ be a set of HSI pixels which are choose from M different classes. The data set $\mathcal{X}$ can be expressed as a matrix $X = [x_1, x_2, x_3, \ldots, x_N] \in \mathbb{R}^{n\times N}$, where each column stands for a pixel in $\mathcal{X}$. The notation $\mathcal{Y} \in \{1, 2, \ldots, M\}$ is the set of class labels.

In our previous work,[35] we proposed a spectral-spatial distance measure for smooth ordering. The quality of distance measurement is very critical for improving

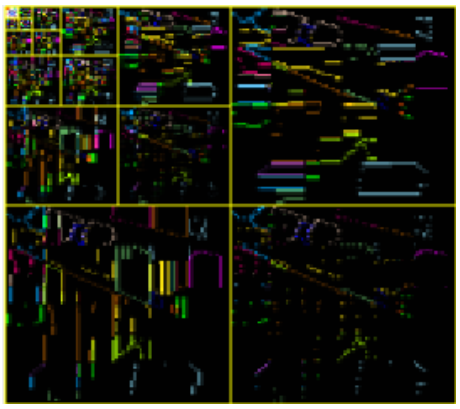


Fig. 2. A graphical demonstration of Wavelet decomposition using Indian Pines image (The decompose level is 5).

classification accuracy.[17] Here, a new distance measurement is developed to measure the similarity of different samples.

The spectral distance $D^w(a_i, a_j)$ between samples x_i and x_j is defined by:

$$D^w(a_i, a_j) = 1 - exp(-\frac{\|x_i - x_j\|^2}{\rho_i \rho_j}) \tag{2.6}$$

where $a_i \in \mathcal{S} = \{1, \cdots, N\}$ is the index of the a_i-th sample in $\mathcal{X}$, and ρ_i is a local scaling parameter defined by:

$$\rho_i = \left\| x_i - x_i^{K_{nn}} \right\|. \tag{2.7}$$

Here, $x_i^{K_{nn}}$ is the K_{nn}-th neighbor around x_i, and $\|*\|$ is a l_2 norm.

In order to improve classification accuracy, the spatial distance between two samples x_i and x_j is encoded according the the following equation:

$$D^s(a_i, a_j) = \begin{cases} -\mu, & \text{if } j \in \omega_i. \\ 0, & otherwise \end{cases} \tag{2.8}$$

where $\mu > 0$ is a weighing parameter that is used to enhance the similarity using a spatial prior. The symbol of ω_i is defined by:

$$\omega_i = \{j \in \mathcal{S} \mid dist(x_m, x_n) \leqslant \sqrt{r}\} \tag{2.9}$$

where $dist(x_m, x_n)$ is the distance between x_m and x_n. $\mathcal{S} = \{1, 2, 3, \ldots, N\}$.

Using the above notations, the spectral-spatial metric is given by:

$$D^{ws}(a_i, a_j) = D^w(a_i, a_j) + D^s(a_i, a_j). \tag{2.10}$$

2.3. *Smooth Ordering Based 1-D Embedding*

The goal of smooth ordering is to design an operator that is applied to approximate the coefficients and produces an ordering under the given distance metric. In the new ordering, pixels with similar spectral signature should be stay as close as possible.

After calculating a similarity measurement of pairwise instances, smooth ordering operated on the sequences $\{a_i\}_i^N$ can be considered as a permutation P on $A = [a_1, \cdots, a_N]$, where $a_i \in \mathcal{S}$. The effect is that all sequences will be reordered into a new sequence $A^P = [a_1^P, a_2^P, a_3^P, \ldots, a_N^P]$ $(a_i^P \in \mathcal{S})$. Clearly, there are many kinds of such sequences. But only the one which has the shortest traveling path is meaningful and closely related to our problem. Substantially, our problem is a in fact a TSP problem. Therefore, our problem is the well known TSP problem, which can be solved by the following optimization:

$$P = arg\min_P a_{TV}^P, \quad a_{TV}^P = \sum_{i=1}^{N-1} D^{ws}(a_{i+1}^P, a_i^P). \tag{2.11}$$

Let $a_{\pi(1)}$ be the first element after applying an ordering on A. The smooth ordering of A headed by $a_{\pi(1)}$ is a mapping presented by:

$$\begin{cases} h(a_{\pi(i)}) = t_i, \quad i = 1, 2, 3, \ldots, N. \\ \Delta t_i = t_{i+1} - t_i = \dfrac{D^{ws}(a_{\pi(i+1)}, a_{\pi_{(i)}})}{\sum_{j=0}^{N} D^{ws}(a_{\pi(j+1)}, a_{\pi(j)})}. \\ t_1 = 0 \\ t_N = 1 \end{cases} \tag{2.12}$$

Note that in the above equation, t_i has been normalized: $t_i \in [0, 1]$. After applying the above transform, the samples have been transformed to a new ordering in 1-D space, where the coefficients are represented by:

$$t_h = [t_1, t_2, t_3, \ldots, t_N]. \tag{2.13}$$

Considering that the ordering coefficients are close related to the original vectors, they can be viewed as an *1-D embedding* of the original signals.

2.4. *Final Classification Using Multiple Embeddings*

Now, we give the description of how to build the final classifiers after smooth ordering on the 1-D vectors. Smooth ordering is first proposed in Ref. 9, where the input is a image patch. After ordering, the sample 1-D data analysis tool (such as interpolation) is applied to process the 1-D vectors. And the experiments demonstrate that smooth ordering can obtain state-of-the-art results in many applications, such as image denoising,[28,29] deblurring, and impainting.[9] In Ref. 33 and Ref. 32, smooth ordering was integrated with semi-supervision learning.

According to the previous description, the smooth ordering operator P can be obtained by Eq. (2.11). Because we cannot directly solve it, a greedy algorithm is employed.[27] The main idea is describing as follows: First, the neighbor of a_j is defined by $N_a = \{a_i \in \chi \mid i \in \omega_j\}$, where ω_j is defined in Eq. (2.9). Second, a path-selection probability vector $\boldsymbol{p}^s = [p_1^s, p_2^s, \ldots, p_n^s](0 < p_i^s < 1)$ is defined. Its role is to select the consequent in the ordering sequence. The algorithm starts by picking a random point a_{j_0}. Assume $a_{\pi_(m)}$ is the current point, in order to find the optimization neighbor, we compute a selection probability q_k according to:

$$q_k = \frac{1}{1 + exp\left(\frac{D^{ws}(a_{\pi(k)}, a_{j1}) - D^{ws}(a_{\pi(k)}, a_{j2})}{n\epsilon}\right)} \tag{2.14}$$

where $\epsilon > 0$ is the path balance parameter. If $q_k < p^s_{\pi(k)}$, the second nearest neighbor a_{j_2} is chosen. In contrast, the first neighbor is chosen. For better description, this procedure is named as *building 1-D coefficients for smooth ordering*.

In this paper, we apply smooth ordering to the approximate coefficients, and produce the 1-D coefficients. Once the 1-D coefficients are obtained, and the sequence $\{t_1, t_2, t_3, \ldots, t_N\}$ is one-dimensional (1-D), therefore, the 1-D interpolation algorithm can be used to build the label function. In this paper, cubic spline interpolation is applied.

Only one interpolation is not sufficient to build a strong classier, to improve the performance, we use multiple interpolation functions to approximate the target function. That is, we repeat the procedure by selecting different starting points, each interpolation contribute equally to the final results. And the final label is decided according to the maximum voting.

Summarily, our proposed method WTSO for HSI classification is given in Algorithm 1.

Algorithm 1 Procedure of proposed WTSO method.

Input: Training data $\mathcal{X}_{train} = \{X_i\}_{i=1}^{n_0} \in \mathbb{R}^n$, Testing samples $\mathcal{X}_{test} = \{X_i\}_{i=n_0+1}^{N} \in \mathbb{R}^n$ and class labels $\mathcal{Y}_{label}$.

Output: Labels $\{y_i\}_{i=n_0+1}^{N}$ of all testing samples $\mathcal{X}_{test}$.

1: (**Wavelet transform**) Decompose HSI signal and obtain the approximate coefficients ACs.
2: (**Distance measurement**) Calculate pairwise similarity according to Eq. 2.10.
3: **for all** k = 1, 2,. . . , K **do**
4: (**1-D Embedding**) Compute 1-D embedding of ACs.
5: (**Classification**) Build classifier using cubic spline interpolation.
6: **end for**
7: (**Final classification**) Decide the class of $X : X \in \mathcal{X}_{test}$.

Table 1. The numbers of training and test samples for Indian Pines.

ID	Class name	Train	Test
1	Alfalfa	21	25
2	Corn-notill	135	1293
3	Corn-mintill	76	754
4	Corn	45	192
5	Grass-pasture	50	433
6	Grass-trees	57	673
7	Grass-pasture-mowed	18	10
8	Hay-windrowed	60	418
9	Oats	12	8
10	Soybean-notill	96	876
11	Soybean-mintill	210	2245
12	Soybean-clean	66	527
13	Wheat	30	175
14	Woods	105	1160
15	Buildings-Grass-Trees-Drives	38	348
16	Stone-Steel-Towers	21	72

3. Experiments

3.1. *Experimental Design*

In this section, we demonstrate the performance of our proposed method using two real word HSI data sets: Indian Pines and University of Pavia (PaviaU). The following methods are implemented and evaluated: SVM,[19] 3-D wavelet,[26] 3-D Gabor,[31] SVM-3DG,[4] Res. Conv-Deconv Net.[20] To be fair, the best performance of all methods are reported in this document. To be specific, SVM with 5-fold cross-validation is implemented. For the 3-D Gabor method, the size of filter is $7 \times 7 \times 7$. For the 3-D wavelet method, the size of filter is 2, where Haar wavelet is applied. In all experiments, 10% samples are used for training, and the remaining 90% ones are used for test.

3.1.1. *Data Sets*

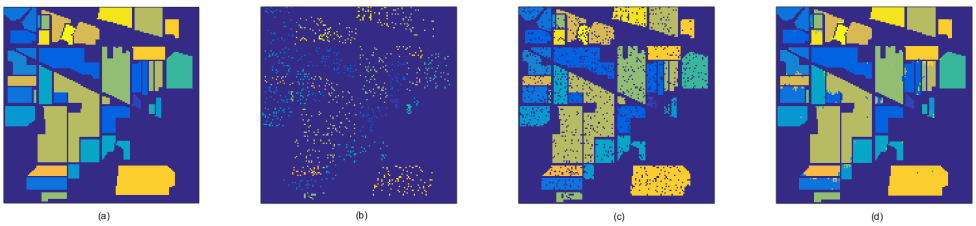


Fig. 3. Classification maps of Indian Pines scene. (a) Ground truth (b) Training set (c) Test set (d) WTSO (OA=97.75%)

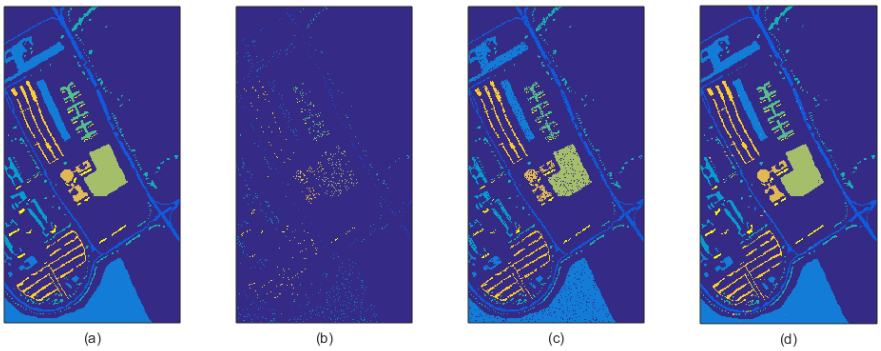


Fig. 4. Classification maps of University of Pavia scene. (a) Ground truth (b) Training set (c) Test set (d) WTSO (OA=97.76%)

- The first data set is the Indian Pines which is captured by AVIRIS sensor over the northwestern Indian Pines test site. The size of this data set is 145 × 145, and it consists of 224 spectral with in the wavelength from 0.4×10^{-6} to 0.6×10^{-6}m. In this scene, the water absorption bands ([104-108],[150-163],220) are removed. There are approximate 10249 labeled pixels with 16 classes. The ground truth of Indian Pines is shown in Fig. 3(a), where the numbers used for training and test are shown in Table 1.
- The second data set used in our experiments is University of Pavia, which is gathered by the Reflective Optics System Imaging Spectrometer (ROSIS-03) in Pavia, northern Italy. It contains 115 bands, and the size is 610 × 340, along with 9 classes. The spectral coverage ranges from 0.43μm to 0.86 μm. The ground truth is shown in Fig. 4(a), where the numbers of each class used for training and test are shown in Table 2.

Table 2. The numbers of training and test samples for University of Pavia.

ID	Class name	Train	Test
1	Asphalt	552	6079
2	Meadows	1160	17489
3	Gravel	303	1796
4	Trees	327	2737
5	Painted metal sheets	260	1085
6	Bare Soil	439	4590
7	Bitumen	262	1068
8	Self-Blocking Bricks	378	3304
9	Shadows	231	716

3.1.2. *Performance Indicators*

The results are measured by overall accuracy (OA), average accuracy (AA), and Kappa coefficient (Kappa). In details, OA is the ratio of those correctly classified samples to the total samples. AA is the average of all classes. Finally, Kappa coefficient describes the agreement of correct percent classification by performing a random classification test, in which the percentage accuracy is ignored.

3.2. *Experimental Results*

(1) Indian Pines data set. Table 3 shows the classification results of Indian Pines scene of all baseline methods. One can observe that the proposed WTSO obtains the best results in all metrics. The OA obtained by 3-D Gabor and WTSO exceeds 90%, yet the proposed WTSO produces an OA value that exceeds 3-D Gabor by 4.33%, which is the best. For Alfalfa, Corn-notill, Corn, Soybean-mintill and Woods classes, the proposed WTSO obtains the best results. Moreover, Res.Conv-Deconv Net produces the best results in Corn-mintill, Grass-pasture, Grass-trees, Soybean-notill, Wheat, Buildings-Grass-Trees-Drives and Stone-Steel-Towers classes. In all, comparing with the results of other baseline methods, the results produced by our WTSO are significantly improved.
(2) University of Pavia (PaviaU) data set. In this experiment, we randomly chose 3912 labeled samples, where the ratio is about 8.2% of all samples. Table 4 shows the classification accuracy of the proposed method when compared with other methods. It can be observed that our proposed method produces the best results in OA, whereas 3-D Gabor obtains the best results in AA and Kappa agreement. Particularly, WTSO obtains the best performance in Gravel, Bitumen and Self-Blocking Bricks classes. However, the 3-D wavelet method produces the best result in Painted metal sheets class (100%). Res.Conv-Deconv Net produces the worst results, due to the complexity of neural network.

3.3. *Impact of Free Parameters*

Now, we analyze the impact of three free parameters on the classification performance in the proposed WTSO method. Three different experiments are designed on two HSI scenes to evaluate the impact of the free parameters. The final values of these parameters are selected based on the highest score criterion.

(1) Firstly, we analyze the influence of K_{nn} in Eq. 2.7.
(2) Secondly, we discuss the impact of μ in Eq. 2.8. The spatial parameter is very important for the proposed method.
(3) In the third experiment, we analyze the effect of ϵ in Eq. 2.14.
(4) In the fourth experiment, the decomposition level of wavelet transform is studied.

Firstly, we discuss the impact of K_{nn} in Eq. 2.7. We step its value from 1 to 9

Table 3. Classification accuracy (%) for Indian Pines using different methods. Best results are marked bold.

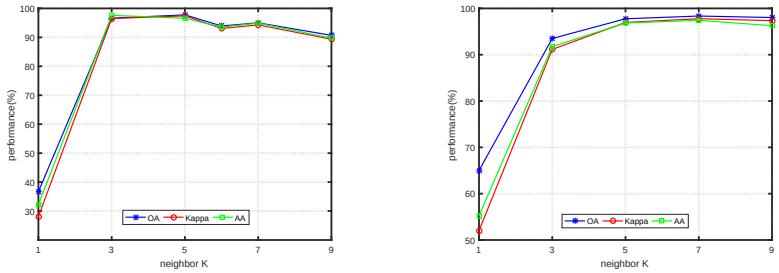
Class	SVM	3-D wavelet	3-D Gabor	SVM-3DG	Res.Conv-Deconv Net	WTSO
1	89.29	61.22	82.65	96.77	74.86	**100.00**
2	69.92	85.33	93.60	58.46	95.28	**97.85**
3	57.78	80.70	90.33	93.37	**100.00**	95.60
4	71.43	54.27	92.04	96.40	95.08	**98.50**
5	90.39	94.57	91.58	86.11	**96.56**	95.51
6	94.52	95.05	92.31	95.80	**99.09**	97.89
7	85.71	55.83	72.92	**100.00**	84.42	**100.00**
8	96.98	97.65	99.37	**100.00**	74.57	99.77
9	88.89	55.79	61.05	**100.00**	80.14	80.00
10	75.17	80.27	90.03	68.86	**100.00**	97.76
11	84.57	86.30	95.51	78.57	95.74	**98.66**
12	74.95	82.06	85.74	**96.89**	96.06	92.57
13	97.16	95.76	90.16	94.21	**100.00**	96.57
14	96.62	92.95	98.58	77.84	84.62	**99.31**
15	54.94	55.22	95.19	95.42	**100.00**	99.71
16	93.65	80.70	87.33	98.72	**100.00**	95.52
AA	83.83	78.35	88.65	89.94	92.28	**96.58**
OA	80.74	85.31	93.42	81.12	85.76	**97.75**
Kappa	78.03	83.25	92.50	78.64	83.85	**97.42**

Table 4. Classification accuracy (%) for University of Pavia using different methods. Best results are marked bold.

Class	SVM	3-D wavelet	3-D Gabor	SVM-3DG	Res.Conv-Deconv Net	WTSO
1	92.40	97.18	**98.48**	97.39	78.99	96.00
2	97.83	97.64	**99.78**	97.27	97.16	99.56
3	87.44	89.44	93.62	89.41	61.46	**99.67**
4	96.03	95.73	96.42	**97.25**	95.76	85.23
5	99.72	**100.00**	99.18	99.61	97.77	99.82
6	91.00	89.23	**99.60**	98.41	59.46	99.23
7	90.20	93.45	90.49	98.20	79.50	**99.72**
8	86.94	93.42	95.88	84.00	96.82	**98.00**
9	99.86	99.09	82.22	**99.89**	92.40	94.40
AA	93.59	95.02	**97.98**	95.71	84.37	96.85
OA	94.32	95.65	95.07	96.06	87.39	**97.76**
Kappa	92.37	94.23	**97.32**	94.78	83.08	96.96

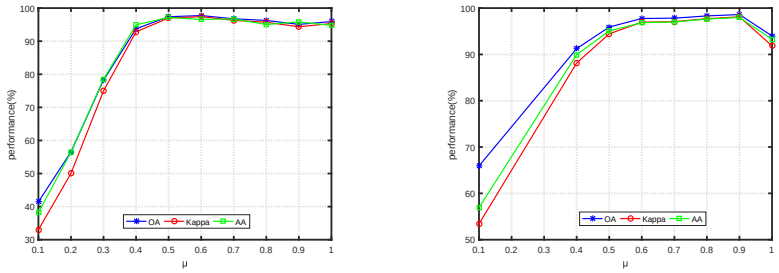
using the size of 2. Fig. 5 (a) and Fig. 5(b) display the impact of K_{nn} for Indian Pines and PaviaU, respectively. It can be observed that, for Indian Pines, when K_{nn} ranges from 1 to 3, the performance is sharply increased. When K_{nn} rises from 5 to 7, the obtained OA is maximum. When K_{nn} is greater than 7, the classification accuracy is declined. For PaviaU, When K_{nn} ranges from 1 to 7, the OA value is increased from 65% to 98%, but when $K_{nn} = 9$, the OA value is lower than the one when $K_{nn} = 7$. For Indian Pines, when $K_{nn} = 5$, the highest accuracy is

achieved. For PaviaU in Fig. 5(b), when $1 \leqslant K_{nn} \leqslant 7$, the result is increasing and the best accuracy is obtained when $K_{nn} = 7$. Hence, in this paper, based on the above analysis, we use a fixed value in WTSO: $K_{nn} = 5$.



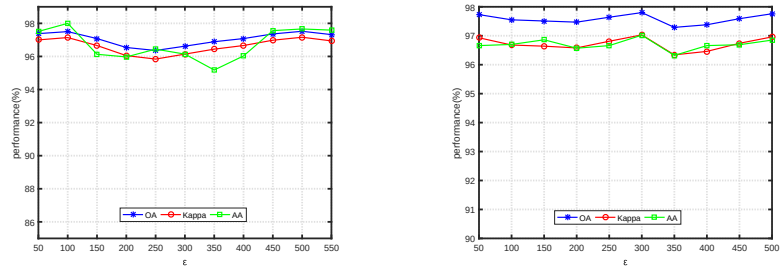
(a) (b)

Fig. 5. The impact of K_{nn}. (a) Indian Pines (b) PaviaU



(a) (b)

Fig. 6. The impact of parameter μ. (a) Indian Pines (b) PaviaU



(a) (b)

Fig. 7. The impact of parameter ϵ. (a) Indian Pines (b) PaviaU

Secondly, the parameter μ used in Eq. 2.8, whose role is to balance the spectral information and spatial prior, is very important. Fig. 6 (a) and Fig. 6 (b) show the OA, AA and Kappa coefficient when μ rises from 0.1 to 1, respectively. For Indian Pines, when μ is in the range of [0.1,0.6], the OA, AA, and Kappa coefficient are increased. However, when μ rises from 0.7 to 1, all of OA, AA, Kappa are declined. Particularly, when $\mu = 0.6$, the performance reaches the maximum for all metrics. For PaviaU, when μ rises from 0.1 to 0.9, the accuracy is increased, and it reaches the maximum value when $\mu = 0.9$(OA = 98.61%, Kappa = 98.12%, AA = 97.98%). However, when $\mu = 0.9$, the accuracy is almost the same as $\mu = 0.6$. When $\mu > 0.9$, the classification accuracy is decreased. Therefore, for computationally, we choose μ to be equal to 0.6 in this paper.

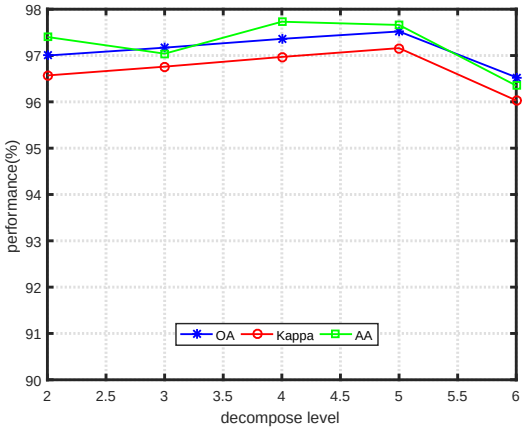


Fig. 8. The impact of decompose level for Indian Pines.

In the third experiment, we show the influence of ϵ in Eq. 2.14 by ranging it from 50 to 550 for Indian Pines, and 50 to 500 for PaviaU, with the step 50 for both images. Fig. 7(a) and Fig. 7(b) show the impact for Indian Pines and PaviaU, respectively. From Fig. 7(a), we can see that when ϵ is increased from 50 to 200, all of OA, AA, Kappa are declined for Indian Pines. However, when it ranges from 250 to 500, the classification accuracy is overall increased (AA is an exception in the rang of 250 to 350). When ϵ is in the range of [350,500], all metrics are increased, where the best result is obtained when $\epsilon = 500$. When $\epsilon > 500$, the accuracy becomes smooth and decreasing. In Fig. 7(b), we can observe that all of AA, OA, Kappa are smooth when ϵ ranges from 50 to 300. However, when ϵ ranges from 300 to 350, all of OA, AA, Kappa are declined. And when ϵ rises from 350 to 500, all of AA, OA, Kappa are increased again. The highest accuracy is obtained when $\epsilon = 500$. Therefore, we choose $\epsilon = 500$ in this paper.

Finally, in the fourth experiment, we analyze the impact of decomposition level using Indian Pines, where the level ranges from 2 to 6. Fig. 8 shows the results. It is obvious that when increasing the decomposition level from 2 to 5, OA and Kappa are improved, and AA is decreasing when in the range of [2, 3]. The best results are obtained when the decomposition level is 5 except AA. However, when the decomposition level is greater than 5, all of AA, OA and Kappa are sharply declined. Hence, we choose the decomposition level to be 5 in this paper.

4. Conclusion

A novel and effective method WTSO is proposed for HSI classification in this paper. In the WTSO method, the wavelet transform is first applied to the input signal so that the signal can be decomposed into ACs and DCs. Then, a new distance measurement is applied before smooth ordering to obtain the similarity of different samples. The obtained ACs are smoothly embedded onto a 1-D space, after that, the simple 1-D tools, such as interpolation, can be used for constructing the candidate classifiers. And the final classifier are constructed using multiple 1-D embeddings. The WTSO method processes ACs in low-dimensional space, which is more explicit. To demonstrate the effectiveness of WTSO, two real HSI data sets are experimentally evaluated. Compared with other recently proposed state-of-the-art wavelet-based methods, our approach can obtain higher accuracy in most cases.

5. Acknowledgement

This work was supported by the National Natural Science Foundation with No. 61862005, and Natural Science Foundation of Guangxi with No. 2017GXNSFBA198226. This work was also supported by National Key R&D Program of China under Grant 2017YFA0700800.

References

1. T. V. Bandos, L. Bruzzone and G. Camps-Valls, Classification of hyperspectral images with regularized linear discriminant analysis, *IEEE Trans. Geosci. Remote Sens.* **47**(3) (2009) 862–873.
2. J. Bruna and S. Mallat, Classification with scattering operators, *arXiv preprint arXiv:1011.3023* (2010).
3. G. Camps-Valls, D. Tuia, L. Bruzzone and J. Benediktsson, Advances in hyperspectral image classification: Earth monitoring with statistical learning methods, *Signal Processing Magazine, IEEE* **31** (10 2013) 45–54.
4. X. Cao, L. Xu, D. Meng, Q. Zhao and Z. Xu, Integration of 3-dimensional discrete wavelet transform and markov random field for hyperspectral image classification, *Neurocomputing* **226** (2017) 90–100.
5. P. Chen, L. Jiao, L. Fang, S. Gou, J. Zhao and Z. Zhao, Dimensionality reduction of hyperspectral imagery using sparse graph learning, *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **10**(3) (2017) 1165–1181.
6. C. K. Chui and J. Wang, Randomized anisotropic transform for nonlinear dimensionality reduction, *GEM - International Journal on Geomathematics* **1**(1) (2010) 23–50.

7. M. Cui and S. Prasad, Class-dependent sparse representation classifier for robust hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* **53**(5) (2015) 2683–2695.
8. L. Fang, W. Cheng, S. Li and J. A. Benediktsson, Hyperspectral image classification via multiple-feature-based adaptive sparse representation, *IEEE Trans. Instrum. Meas.* **66**(7) (2017) 1646–1657.
9. R. Idan, E. Michael and C. Israel, Image processing using smooth ordering of its patches, *IEEE Trans. Image Process.* **22**(7) (2013) 2764–2774.
10. B. C. Kuo and C. H. Li, Kernel nonparametric weighted feature extraction for classification., *IEEE Trans. Geosci. Remote Sens.* **42**(5) (2004) 1096–1105.
11. C. Li, Y. Ma, J. Huang, X. Mei and J. Ma, Hyperspectral image denoising using the robust low-rank tensor recovery, *Journal of the Optical Society of America A* **32** (09 2015) 1604–1612.
12. H. Li, G. Xiao, T. Xia, Y. Y. Tang and L. Li, Hyperspectral image classification using functional data analysis, *IEEE Trans. Cybern.* **44**(9) (2014) 1544–1555.
13. J. Li, H. Zhang, Y. Huang and L. Zhang, Hyperspectral image classification by nonlocal joint collaborative representation with a locally adaptive dictionary, *IEEE Trans. Geosci. Remote Sens.* **52**(6) (2014) 3707–3719.
14. W. Li, S. Prasad, J. E. Fowler and L. M. Bruce, Locality-preserving dimensionality reduction and classification for hyperspectral image analysis, *IEEE Trans. Geosci. Remote Sens.* **50**(4) (2012) 1185–1198.
15. H. Lin, J. Li, C. Liu and S. Li, Recent advances on spectral-spatial hyperspectral image classification: An overview and new guidelines, *IEEE Trans. Geosci. Remote Sens.* **PP**(99) (2018) 1–19.
16. J. Liu, Z. Wu, Z. Wei, X. Liang and S. Le, Spatial-spectral kernel sparse representation for hyperspectral image classification, *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **6**(6) (2013) 2462–2471.
17. H. Luo, Y. Y. Tang, Y. Wang, J. Wang, R. P. Biuk-Aghai, J. Pan, R. Liu and L. Yang, Hyperspectral image classification using metric learning in one-dimensional embedding framework, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **10** (May 2017) 1987–2001.
18. H. Luo, Y. T. Yuan, Y. Wang, J. Wang, L. Yang, C. Li and T. Hu, Hyperspectral image classification based on spectral–spatial one-dimensional manifold embedding, *IEEE Trans. Geosci. Remote Sens.* **54**(9) (2016) 5319–5340.
19. G. Mercier and M. Lennon, Support vector machines for hyperspectral image classification with spectral-based kernels, *Geoscience and Remote Sensing Symposium, 2003. IGARSS'03. Proceedings. 2003 IEEE International*, **1**, IEEE2003, pp. 288–290.
20. L. Mou, P. Ghamisi and X. X. Zhu, Unsupervised spectral–spatial feature learning via deep residual conv–deconv network for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* **56**(1) (2018) 391–406.
21. D. R. Nayak, R. Dash and B. Majhi, Brain mr image classification using two-dimensional discrete wavelet transform and adaboost with random forests, *Neurocomputing* **177** (2016) 188–197.
22. M. E. Newman, K. P. Mclaren and B. S. Wilson, Comparing the effects of classification techniques on landscape-level assessments: pixel-based versus object-based classification, *Int. J. Remote Sens.* **32**(14) (2011) 4055–4073.
23. L. Pan, H. C. Li, H. Meng, W. Li, Q. Du, W. J. Emery, L. Pan, H. C. Li, H. Meng and W. Li, Hyperspectral image classification via low-rank and sparse representation with spectral consistency constraint, *IEEE Geosci. Remote Sens. Lett.* **PP**(99) (2017) 1–5.

24. J. Peng and D. Qian, Robust joint sparse representation based on maximum correntropy criterion for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* **55**(12) (2017) 7152–7164.
25. A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba and A. Gualtieri, Recent advances in techniques for hyperspectral image processing, *Remote Sens. Environ.* **113**(1) (2009) S110–S122.
26. Y. Qian, M. Ye and J. Zhou, Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features, *IEEE Trans. Geosci. Remote Sens.* **51**(4) (2013) 2276–2291.
27. I. Ram, M. Elad and I. Cohen, Generalized tree-based wavelet transform, *IEEE Trans. Signal Process.* **59**(9) (2011) 4199–4209.
28. I. Ram, M. Elad and I. Cohen, Redundant wavelets on graphs and high dimensional data clouds, *IEEE Signal Process. Lett.* **19**(5) (2012) 291–294.
29. I. Ram, M. Elad and I. Cohen, Image denoising using nl-means via smooth patch ordering, *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, IEEE2013, pp. 1350–1354.
30. S. B. Serpico and G. Moser, Extraction of spectral channels from hyperspectral images for classification purposes, *IEEE Trans. Geosci. Remote Sens.* **45**(2) (2007) 484–495.
31. L. Shen and S. Jia, Three-dimensional gabor wavelets for pixel-based hyperspectral imagery classification, *IEEE Trans. Geosci. Remote Sens.* **49**(12) (2011) 5039–5046.
32. J. Wang, Semi-supervised learning using ensembles of multiple 1d-embedding-based label boosting, *Int. J. Wavelets Multiresolution Inf. Process.* **14**(02) (2016) p. 1640001.
33. J. Wang, Semi-supervised learning using multiple one-dimensional embedding based adaptive interpolation, *Int. J. Wavelets Multiresolution Inf. Process.* **14**(02) (2016) p. 1640002.
34. Y. Wang and S. Cui, Hyperspectral image feature classification using stationary wavelet transform, *2014 International Conference on Wavelet Analysis and Pattern Recognition*, IEEE2014, pp. 104–108.
35. L. Yang, H. Su, C. Zhong, L. Bai, P. Wei, X. Dang and H. Luo, Hyperspectral image classification based on different affinity metrics, *2018 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR)*, IEEEJuly 2018, pp. 203–208.
36. C. Yi, N. M. Nasrabadi and T. D. Tran, Hyperspectral image classification using dictionary-based sparse representation, *IEEE Trans. Geosci. Remote Sens.* **49**(10) (2011) 3973–3985.
37. Y. T. Yuan, L. Yang and H. Yuan, Hyperspectral image classification based on three-dimensional scattering wavelet transform, *IEEE Trans. Geosci. Remote Sens.* **53**(5) (2015) 2467–2480.
38. Y. T. Yuan, H. Yuan and L. Li, Manifold-based sparse representation for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.* **52**(12) (2014) 7606–7618.