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. To effectively alleviate the general HSI classification problem, in this paper, we propose a new classification method called *wavelet transform-based smooth ordering* (WTSO). Specifically, WTSO consists of three main parts: wavelet decomposition, smooth ordering, and one-dimensional (1-D) interpolation. In this approach, the wavelet transform is firstly imposed to decompose the HSI signal into approximate coefficients (ACs) and details coefficients (DCs). Then, smooth ordering is applied to the ACs so that the coefficients is aligned in a 1-D space. In this procedure, the spatial information is encoded. Finally, the simple 1-D signal interpolation tool is adopted to build the final classifier. The use of wavelet transform in WTSO is to reduce the high dimensionality of HSI on one hand and capture the intrinsic property of each input signal on the other hand. Besides, by converting the hight dimensional samples into a 1-D ordering sequence, WTSO can reduce the computational cost and simultaneously perform prediction for the samples without labels. Note that in TWSO, the smooth ordering and

[*]Corresponding author

1-D interpolation are used in an iterative manner, and will be terminated after a finite number. 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 func-

tion. 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 in our previous work,[35] smooth ordering is performed on the ACs, where the dissimilarity of each sample is calculated accordingly. Next, based on the measurement, smooth ordering algorithm is applied to the coefficients, resulting in an ordered sequence of these data in 1-D space. Finally, a group of classifiers are constructed using 1-D interpolation, 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 preliminary 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

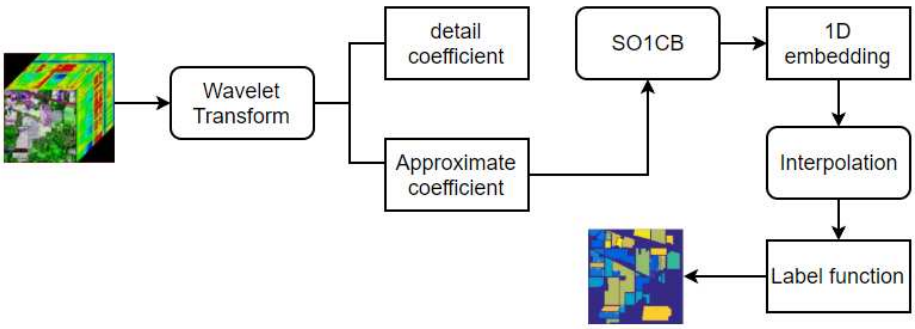


Fig. 1. Flowchart of the proposed WTSO classification scheme.

In this section, we introduce our proposed WTSO method. The flowchart of our method is shown in Fig. 1, where SO1CB is the abbreviation of the developed smooth ordering-based scheme. The details are described in the following subsections.

2.1. *Descriptive Transformation by 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. *Smooth Ordering of Input Signals*

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.

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

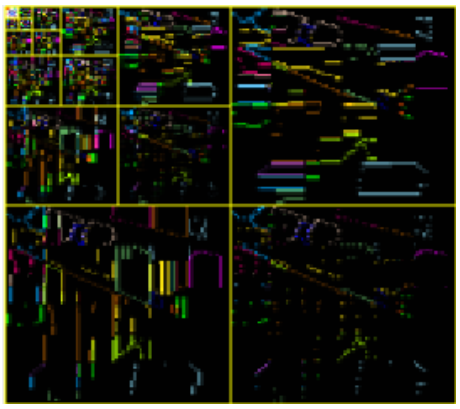


Fig. 2. A graphical demonstration of Wavelet decomposition using Indian Pines image (The decompose level is 5).

the new ordering, pixels with similar spectral signature should be stay as close as possible.

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 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:

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

In fact, 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 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.

2.3. *Building Classifiers by Iterative Scheme*

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 Our proposed method WTSO.

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: **(Decompose scheme)** Decompose HSI and obtained the approximate coefficients.
2: **(Distance measure)** Define new distance measure for smooth ordering using Eq. 2.10.
3: **for all** k = 1, 2,. . . , M **do**
4: Compute 1-D embedding of ACs.
5: Built classifier using cubic spline interpolation.
6: **end for**
7: Decide class of $X : X \in \mathcal{X}_{test}$

Table 1. The numbers of training and test samples of 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 image: Indian Pines and University of Pavia (PaviaU). For fairness, other methods are carried out to compare with us: SVM,[19] 3-D wavelet,[26] 3-D Gabor,[31] SVM-3DG,[4] Res. Conv-Deconv Net.[20] All the experiments are performed on windows system desktop PC with Inter(R) Core(TM) i5-7400 CPU at 3.00GHz, 64bit system, and 8.00 GB RAM using MATLAB language. The optimal parameter of SVM is achieved via fivefold cross-validation. For the 3-D Gabor, the size of the filter is $7 \times 7 \times 7$. The size of wavelet filter is 2 and Haar wavelet is applied for 3-D wavelet. In this paper, 10% samples is used for the training set, and remaining 90% is used as the testing set for classification.

3.1.1. *Data set*

- 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 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 for Indian Pines is shown in Fig. 7(a). Numbers of the training and testing samples of Indian Pines is 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 with 9

classes. The spectral coverage range from 0.43-0.86 μm. The Ground truth for Indian Pines is shown in Fig. 8(a). Numbers of the training and testing samples of Indian Pines is shown in Table 2.

Table 2. The numbers of training and test samples of 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 metrics*

In our experiments, the results are measured by the overall accuracy (OA), average accuracy (AA) and Kappa coefficient. In general, an accuracy is used to measure the probability of each classes classified correctly. For detail, OA measures the ratio between the correct and total number of predictions across all test sets. AA measures the average accuracy of each class. It describes the consistency of each category by determining the degree to which a particular class matches between two data sets.

3.2. *Experimental results*

Table 3 shows the classification results of Indian Pines scene using our proposed method and the methods being compared. The best results are marked bold. It is obvious that the proposed method produced the best classification accuracy of all OA, AA, and Kappa. Table 4 shown the classification accuracy of the proposed method compared with other approaches for University of Pavia, we can see that the proposed method yielded the best result in metric criterion OA. 3-D Gabor achieved the best accuracy in metric criterion AA and Kappa. Fig. 7 (d) is the classification map using SOWT for Indian Pines scene. Fig. 8 (d) shown the classification map for PaviaU using proposed method SOWT.

3.3. *The impact of parameter*

Now, we analyze the impact of the parameter on the HSI classification. Three different experiments are designed on two HSI scenes to evaluate the impact of the

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	SOWT
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	SOWT
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

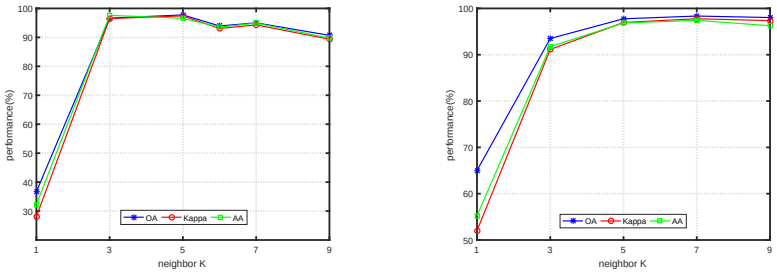
parameter for the proposed method SOWT. The proper values of the parameters are selected based on the highest accuracy criterion.

(i) Firstly, we will analyse the influence of neighbor K_{nei} in Eq. 2.7.
(ii) Secondly, we discuss the impact of parameter μ in Eq. 2.8. The spatial parameter is very important for the proposed method.
(iii) In the third experiment, we analysze the effect of parameter ϵ in Eq. 2.14.

First, let parameter K_{nei} in Eq. 2.7 increased from 1 to 9. Fig. 3(a) and Fig. 3(b)

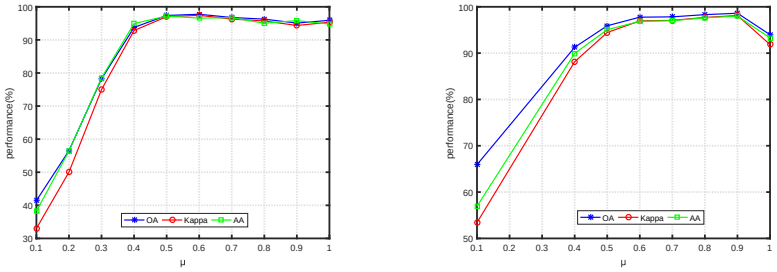
shown the impact of K_{nei} for Indian Pines and PaviaU respectively. It is obviously in the picture, for Indian Pines, when K_{nei} range from 1 to 3, the performance is sharply increased from 30% to 96%. When K_{nei} rises from 5 to 7, the accuracy obtained the maximum 97.75%, and more smooth. When K_{nei} is greater than 7, the classification accuracy is declined. For PaviaU, When K_{nei} ranges from 1 to 7, the performance is increased from 65% to 98%, but when $K_{nei} = 9$, the accuracy is lower than $K_{nei} = 7$. For Indian Pines, when $K_{nei} = 5$, the highest accuracy is achieved. For PaviaU in Fig. 3(b), when $1 \leqslant K_{nei} \leqslant 7$, the result is increased and the best accuracy is obtained when $K_{nei} = 7$, but almost the same as when $K_{nei} = 5$. When $K_{nei} > 7$, the performance become smooth. Hence, in this paper, based on the above analysis, 5 is chosen for K_{nei} in SOWT.

second, parameter μ in Eq. 2.8, which to balance spectral information and spatial prior, is very important to improve classification accuracy. Fig. 4(a) and Fig. 4(b) shown the OA, AA and Kappa coefficient using proposed method SOWT when μ rises from 0.1 to 1 for Indian Pines and PaviaU respectively. For Indian Pines, when μ is in the range of [0.1,0.6], the OA, AA, and Kappa coefficient are increased and



(a) (b)

Fig. 3. The impact of parameter K_{nei}. (a) Indian Pines (b) PaviaU

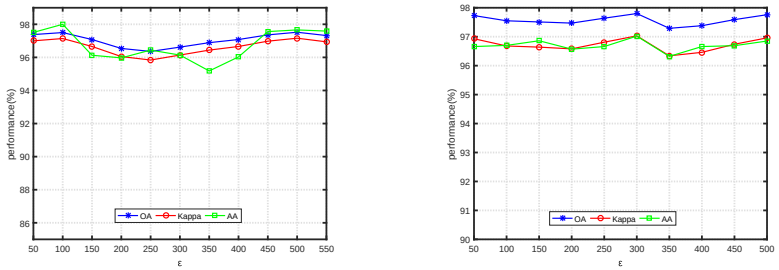


(a) (b)

Fig. 4. The impact of parameter μ. (a) Indian Pines (b) PaviaU

accepted. However, when μ rises from 0.7 to 1, all of OA, AA, Kappa are declined compared with a range of [0.1,0.6]. Particularly, when $\mu = 0.6$, the performance reached the maximum for all OA, AA, and Kappa. For PaviaU, when μ rises from 0.1 to 0.9, the accuracy is increased, and reached maximum value when $\mu = 0.9$, about OA = 98.61%, Kappa = 98.12%, AA = 97.98%. However, when $\mu = 0.9$, the accuracy is almost same as when $\mu = 0.6$. When $\mu > 0.9$, the classification accuracy is decreased. Therefore, for computationally, we choose μ be equals 0.6 in this paper.

Third, we have shown the influence of parameter ϵ in Equ. 2.14 in the range from 50 to 550 for Indian Pines, 50 to 500 for PaviaU with the step 50. Fig. 5(a) and Fig. 5(b) shown the parameter impact for Indian Pines and PaviaU respectively. From Fig. 5(a), we can see that when ϵ is increased from 50 to 200. OA, AA,



(a) (b)

Fig. 5. The impact of parameter ϵ. (a) Indian Pines (b) PaviaU

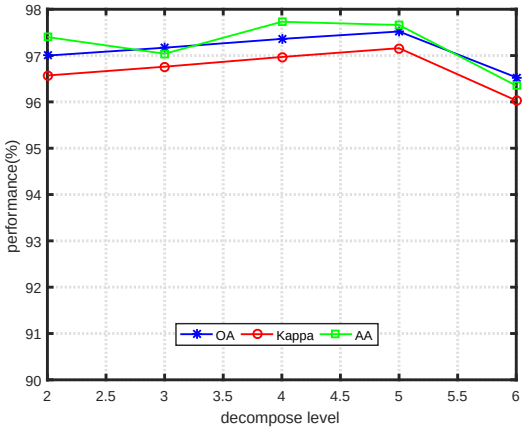


Fig. 6. The impact of decompose level for Indian Pines.

Kappa are declined for Indian Pines, However, when its ranges from 250 to 500, classification accuracy are increased except AA range from 250 to 350. But when ϵ is range of [350,500], all of AA, OA, Kappa are increased, and when $\epsilon = 500$ obtained maximum, is 97.52% for OA, 97.16% for Kappa and 97.66% for AA. When $\epsilon > 500$, the accuracy become smooth even decreased. In Fig. 5(b), we can see that all of AA, OA, Kappa are smooth when ϵ ranges from 50 to 300, however, when ϵ from 300 to 350, all of OA, AA, Kappa are declined. When ϵ rises from 350 to 500, all of AA, OA, Kappa are increased, and highest accuracy is obtained when $\epsilon = 500$. Therefore, for computationally, we choose ϵ in this paper is 500.

Finally, we analyze the impact of decompose level for Indian Pines range from 2 to 6. Fig. 6 shown the results. It is obvious that when decompose level increased from 2 to 5, OA and Kappa are increased and AA declined when decomposing level from 2 to 3. The best result is obtained when decompose level is 5 except AA. However, when decompose level is greater than 5, all of AA, OA and Kappa are sharply declined. Hence, we choose to decompose level equals 5 in this paper.

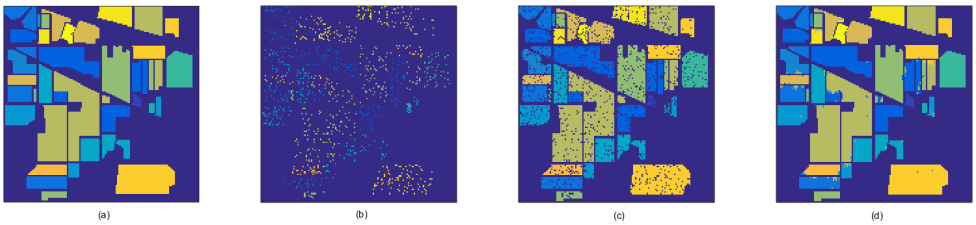


Fig. 7. Classification maps for Indian Pines scene using training sample and testing samples from Table 2. (a) Ground truth (b) Training set (c) Test set (d) SOWT(OA=97.75%)

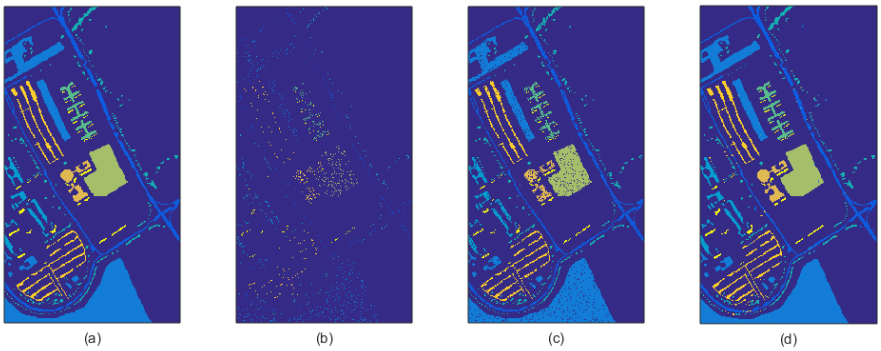


Fig. 8. Classification maps for University of Pavia scene using training sample and testing samples from Table 2. (a) Ground truth (b) Training set (c) Test set (d) SOWT(OA=97.76%)

4. Conclusion

A novel and effective method SOWT has been proposed for HSI classification in this paper. In the SOWT method, the wavelet transform is first applied to the input signal so that the signal has been decomposed into ACs and DCs. Then, a new distance measure is defined for smooth ordering. Based the smooth ordering, in SOWT, the ACs are smoothly embedding onto 1-D space, so that simple 1-D tools, such as interpolation can be used for feature extract. Finally, the classifier is established using interpolation. The SOWT method can process ACs in low-dimensional space, which becomes more effectively. To demonstrate the effectiveness of SOWT, two real HSI data are chosen for experiments. Compared with other recently proposed state-of-the-art wavelet transform based methods, our approach can obtain high 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.