IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 30, 2020, accepted July 18, 2020, date of publication August 5, 2020, date of current version August 19, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3014527

Hyperspectral Image Classification via Slice Sparse Coding Tensor Based Classifier With Compressive Dimensionality Reduction

LIXIA YANG[1,3], (Member, IEEE), RUI ZHANG[1,3], SHUYUAN YANG[2], (Senior Member, IEEE), AND LICHENG JIAO[2], (Fellow, IEEE)

[1]School of Mathematics and Statistics, Ningxia University, Yinchuan 750021, China
[2]School of Artificial Intelligence, Xidian University, Xi'an 710071, China
[3]Ningxia Key Laboratory of Scientific and Engineering Computing and Data Analysis, Ningxia University, Yinchuan 750021, China

Corresponding author: Rui Zhang (zhangrui1980@nxu.edu.cn)

This work was supported in part by the Science Research Project of Ningxia Higher Education under Grant NGY2018001, in part by the National Natural Science Foundation of China under Grant 61906102, in part by the Ningxia Key Research and Development Program (Special Talents Project) under Grant 2018BEB04025, in part by the Project funded by the Ningxia Natural Science Foundation under Grant 2019AAC03036 and Grant 2020AAC03063, in part by the Project funded by the Ningxia Key Research and Development Program under Grant 2019BEG03056, and in part by the Project funded by the China Postdoctoral Science Foundation under Grant 2018M633469.

ABSTRACT Tensor representation is the most natural and effective way to preserve the structural information of hyperspectral image (HSI), and thus is very beneficial to HSI processing. This paper represents the spectral features of each testing pixel and its spatial neighbors as a Spatial Neighbor Tensor (SNT), whose spectral vectors can be simultaneously sparse coded by the spectral vectors of a few common training pixels. The obtained sparse coding coefficients could be regarded as a Slice Sparse Coding Tensor (SSCT), which can be adaptively learnt and utilized to predict the labels of HSI pixels. Furthermore, to improve the efficiency, the Compressive Dimensionality Reduction (CDR) is introduced into tensor slice sparse coding for optimizing SSCT and thus an SSCT based Classifier with CDR (SSCTC-CDR) is proposed for hyperspectral image classification (HIC). The performance of SSCTC-CDR is evaluated on three real HSI data, and the results show that it can obtain high-accurate classification result with relatively low computation lost.

INDEX TERMS Hyperspectral image classification, compressive dimensionality reduction, slice sparse coding tensor.

I. INTRODUCTION

Benefited from the high spectral resolution of hyperspectral image (HSI), it can successfully discriminate many similar surface materials and phenomenons in various application areas [1], [2]. As a kind of effective way to extract useful information from HSI, Hyperspectral Image Classification (HIC) has become an active research area in remote sensing and many machine learning techniques have been applied to it in the past few decades [3], [5]–[8], [10], [14].

In the machine learning field, there are many classifiers with good classification performance, such as Support Vector Machine (SVM), Bayesian Network, Neural Network, etc. However, it can be found that directly using the spectral features as the input features of these existing classifiers is not appropriate to HIC. Because they suffer from the Hughes effect [4]: as the dimension of the data increases, the number of samples needs to be labeled rapidly increases to maintain the high classification accuracy. However, the labeled samples in HSI are always very limited due to the high labeling cost. Therefore, it is necessary to explore other prior information of hyperspectral data as a supplement to improve the classification performance of HSI.

It is well-known that as a type of image, the spatially neighboring pixels of the HSI are likely to belong to the same category. This prior can be used as spatial information. By combining it with spectral information, various Spatial-Spectral HIC (SS-HIC) approaches have been proposed for HIC to improve classification accuracy of HIC. The early SS-HIC methods, such as morphological transformation

The associate editor coordinating the review of this manuscript and approving it for publication was Miaohui Wang.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

based spatial-spectral classifiers [5] and composite kernel classifiers [6], usually express spatial information in the form of first-order feature vectors. These methods generate spatial features by computing the statistics of the spectral values, which are extracted from the spatial neighborhood of a certain pixel. By combing the generated spatial features together with the spectral features, a simple and effective solution is provided for HIC. But these SS-HIC approaches do not consider the specific spectral information of each spatial neighbor pixel, and thus the spatial features cannot fully capture the spatial local neighborhood variability of spectral features. To overcome this point, the spectral features of each pixel and its spatial neighbors are rearranged in the form of a second-order matrix, and the models such as Joint Sparse Coding based Classifier (JSCC) [7], regional-kernel-based SVM [8] are proposed for HIC. These methods fully considered the spectral features of all pixels in the spatial neighborhood. Thus they can obtain much better classification performance. However, the matriculated of hyperspectral data ignores the spatial position relationship of neighboring pixels. Therefore, a more reasonable data representation is needed to effectively maintain the inherent spatial-spectral structure of hyperspectral data. The essential form of HSI is a third-order tensors, so tensor representation is the most natural and effective way to preserve the spectral characteristics and spatial position relationship of HSI.

Furthermore, sparsity priors are often used in HIC. Sparsity is a main characteristic of human perception of the world. And there is an inevitable connection between the sparsity and the statistical characteristics of the natural environment [9]. HSI data also has its own sparsity. Although the spectral features of HSI data are high dimensional, the correlations among them are strong. For the pixels in the same class, the spectral features approximately lie in the same low-dimensional subspace. Thus the spectral features of HSI approximately lie in a few low-dimensional manifolds [7]. The hyperspectral pixels are sparse in the spectral feature space. This sparsity prior information has been applied to HIC in the tensor framework and a Hybrid Probabilistic Sparse Coding Classifier with Spatial Neighbor Tensor (HPSCC-SNT) was proposed in [11]. The tensor representation makes HPSCC-SNT well preserve the spectral characteristics and the spatial structural information of HSI. Furthermore, in HPSCC-SNT the misclassification of mixed-pixels is alleviated by the maximizing likelihood estimation. Thus the classification performance of HPSCC-SNT for HIC is very good, but it still has a few drawbacks:

(1) High computation and restoration cost: In HPSCC-SNT, predicting the labels of each testing pixel should compute its sparse representation by optimizing the hybrid of a tensor sparse coding and a maximizing likelihood estimation. Considering the high dimensionality of hyperspectral pixels, it is a complex optimization problem. Thus, both the computation cost and restoration cost of labeling (tens of) thousands of pixels in a HSI are very high.

(2) Performance is affected by parameters: In the optimization problem of HPSCC-SNT, there are two regularization parameters to be tuned. The experimental results show that HPSCC-SNT is robust to these two parameters in a particular range and drops rapidly out of the range. Thus to a certain extent, the HIC performance is affected by the values of these parameters.

To overcome these problems, a novel classifier under the tensor framework is developed for HIC in this paper. Following the tensor organization scheme in paper [10], the spectral features of each testing pixel and its spatial neighbors are represented as a Spatial Neighbor Tensor (SNT). This representation scheme can well preserve the spectral characteristics and spatial constraints of each testing pixel and its spatial neighbors, thus can highly improve the performance of proceeding classification. Taking the spatial homogeneity of hyperspectral image into consideration, it is assumed that the spectral vectors in each generated SNT are simultaneously sparse coded by the spectral vectors of a few common training pixels. The sparse coding coefficients associated with SNT could be regarded as a slice sparse tensor, which is denoted as Slice Sparse Coding Tensor (SSCT).

The SSCT can be adaptively learnt by a tensor slice sparse coding algorithm. Furthermore, the SSCT has a hidden competition among the subspaces in which each category of pixels lie. Thus it has the discriminant property and can be utilized to predict the label of HSI pixels. However, the computation and restoration cost of computing SSCTs are very high for classifying (tens of) thousands of high dimensional hyperspectral pixels. To improve the efficiency, the Compressive Dimensionality Reduction (CDR) is introduced into tensor slice sparse coding for optimizing SSCT and thus an SSCT based Classifier with CDR (SSCTC-CDR) is proposed for HIC.

Our proposed SSCTC-CDR has the following contributions: 1) Unlike the regularization methods whose performances are largely depend on the regularization parameters, there are very few parameters to be determined in SSCTC-CDR; 2) By introducing compressive dimensionality reduction into tensor sparse coding, SSCTC-CDR can remarkably reduce the size of the tensor to be sparsely encoded. The testing computation complexity of SSCTC-CDR is much reduced. The performance of SSCTC-CDR is evaluated on two real HSI data, and the results show that it can obtain high-accurate classification with relatively low computation lost.

The rest of this paper is framed as follows: The proposed SSCTC-CDR is depicted in section II. In section III, the performance of our proposed method is investigated by taking some experiments. Section IV summarizes the proposed methods.

II. SSCTC-CDR FOR HIC

Denoting $\mathbf{x} = [x_1, x_2, \ldots, x_B] \in \mathbb{R}^B$ as a hyperspectral pixel, the aim of HIC is labeling each pixel in the HIS

$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L, \mathbf{x}_{L+1}, \ldots, \mathbf{x}_{L+U}]$ as one of the known classes set $\{1, 2, \ldots, C\}$. In the supervised learning paradigm, supposing the first L pixels are labeled with their corresponding labels $y_l \in \{1, 2, \ldots, C\}$, and the remaining U pixels are unlabeled. Inferring the labels of hyperspectral pixels can be considered as a classification task. The labeled pixels and the other pixels are regarded as training samples and testing samples respectively. To obtain a relatively effective and efficient supervised HIC method, a novel classifier under the tensor framework is proposed in this section. Before describing the process of our proposed classifier, two relevant definitions about tensor [11] are simply introduced. For an M-order tensor $\underline{\mathbf{A}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_M}$, its m- mode product of with a matrix $\mathbf{D} \in \mathbb{R}^{J \times I_m}$ is also a tensor, which can be represented as

$$\underline{\mathbf{X}} = (\underline{\mathbf{A}} \times_m \mathbf{D}) \in \mathbb{R}^{I_1 \times \cdots \times I_{m-1} \times J \times I_{m+1} \times \cdots \times I_M}.$$

Element-wise, it has the following form,

$$x_{i_1 \cdots i_{m-1} j i_{m+1} \cdots i_M} = \sum_{i_m=1}^{I_m} a_{i_1 i_2 \cdots i_M} d_{j i_m}.$$

And its Frobenius norm is defined as

$$\left\|\underline{\mathbf{A}}\right\|_F = \sqrt{\sum_{i_1=1}^{I_1} \cdots \sum_{i_M=1}^{I_M} a_{i_1 \cdots i_M}^2}.$$

Our proposed classifier under the tensor framework can be realized by the following steps:

A. GENERATING SPATIAL NEIGHBOR TENSOR

Following most of the SS-HIC methods, for each unlabeled hyperspectral pixel $\mathbf{x}_u \in \mathbb{R}^B$ $(u = 1, 2, \ldots, U)$, its spatial neighbors are regulated as the pixels in a $T \times T$ spatial window centered by $\mathbf{x}_u$, which are denoted as $NB(\mathbf{x}_u)$. Then the SNT $\underline{\mathbf{X}}_u \in \mathbb{R}^{T \times T \times B}$ is generated by permuting the the spectral features of the T^2 spatial neighboring pixels according to their spatial locations, see **FIGURE 1.**

B. TENSOR SLICE SPARSE CODING

The tensor $\underline{\mathbf{X}}_u$ can be tucker decomposed as follows

$$\underline{\mathbf{X}}_u = \underline{\mathbf{A}}_u \times_1 \Psi_1 \times_2 \Psi_2 \times_3 \Psi_3 \tag{1}$$

here $\underline{\mathbf{A}}_u \in \mathbb{R}^{T \times T \times L}$ is the core tensor, the factor matrices $\Psi_1 = \Psi_2 = \mathbf{I}_T$, $\Psi_3 = \mathbf{D} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L] \in \mathbb{R}^{B \times L}$. In this situation, the right side of equation (1) is equivalent to $\underline{\mathbf{A}}_u \times_3 \mathbf{D}$. Thus equation (1) can be simplified as

$$\underline{\mathbf{X}}_u = \underline{\mathbf{A}}_u \times_3 \mathbf{D} \tag{2}$$

It is well-known that the spectral vector of each HSI pixel can be sparse coded by the atoms in dictionary matrix $\mathbf{D}$. Furthermore, it is assumed that the sparse coding vectors of spatial neighboring HSI pixels share a common sparse pattern under $\mathbf{D}$. Thus the tensor $\underline{\mathbf{A}}_u$ is assumed to be slice sparse in relation to the 3-mode dictionary $\mathbf{D}$. **FIGURE 2.** shows the tensor slice sparse coding visually.

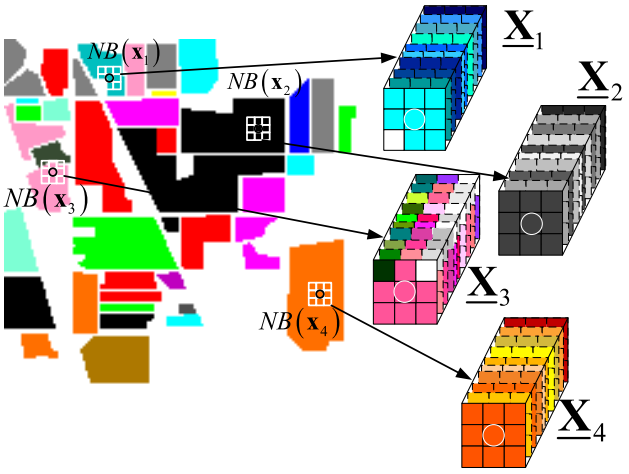

FIGURE 1. Spatial neighbor tensor.

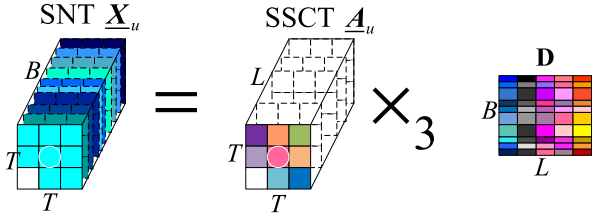

FIGURE 2. Tensor slice sparse coding.

The SSCT $\underline{\mathbf{A}}_u$ can be calculated by solving the following optimization problem:

$$\begin{cases} \min \left\|\underline{\mathbf{X}}_u - \underline{\mathbf{A}}_u \times_3 \mathbf{D}\right\|_F \\ \text{s.t. } \left\|\underline{\mathbf{A}}_u\right\|_0 \leq S \end{cases} \tag{3}$$

where $\left\|\underline{\mathbf{A}}_u\right\|_0$ represents the number of nonzero frontal slices of tensor $\underline{\mathbf{A}}_u$, and S is the pre-set sparsity level.

Furthermore, it can be inferred from paper [12] that equation (1) is equivalent to a Kronecker representation

$$\mathbf{x}_u = (\Psi_3 \otimes \Psi_2 \otimes \Psi_1)\mathbf{a}_u \tag{4}$$

where $\mathbf{x}_u = \text{vec}(\underline{\mathbf{X}}_u)$, $\mathbf{a}_u = \text{vec}(\underline{\mathbf{A}}_u)$ and $\otimes$ represents the Kronecker product. Thus, SSCT $\underline{\mathbf{A}}_u$ can be obtained by the following recovery algorithm.

C. COMPRESSIVE DIMENSIONALITY REDUCTION

Due to the high dimension of HSI data, the recovery of SSCT $\underline{\mathbf{A}}_u$ is costly when the number of pixels is great. A feasible way of reducing computation cost is adopting the compressive dimensionality reduction. To realize this point, a compressive measurement matrix $\Phi \in \mathbb{R}^{n \times B}$ $(n < B)$ is multiplied on the both sides of the formula (1), and the representation is rewritten as

$$\underline{\mathbf{X}}_u \times_3 \Phi = \underline{\mathbf{A}}_u \times_3 \Phi\mathbf{D} \tag{5}$$

While maintaining the proceeding HIC accuracy, to remarkably reduce the optimization cost of (5), the size of $\underline{\mathbf{X}}_u \times_3 \Phi \in \mathbb{R}^{T \times T \times n}$ should be as small as possible.

Thus we should choose the measurement matrix Φ which has low incoherence with dictionary matrix $\mathbf{D}$. In general, the random Gaussian matrix is an available choice of matrix Φ, because it is highly incoherent with any fixed basis. Due to the randomness of the random Gaussian matrix, the method will be unstable for the HIC problem. A feasible way of overcoming this problem is adopting the coupled measurement matrix that comes from the singular value decomposition of the dictionary matrix $\mathbf{D} = \mathbf{U}\Lambda\mathbf{V}$. Compared with the random Gaussian matrix, the coupled measurement matrix $\Phi = \mathbf{U}_M^T$ has higher incoherence with the dictionary. Thus this letter employs the matrix $\Phi = \mathbf{U}_M^T$ for compressive dimensionality reduction, and the SSCT $\underline{\mathbf{A}}_u^r$ corresponding to the compressive dimensionality reduced SNT $\underline{\mathbf{X}}_u^r = \underline{\mathbf{X}}_u \times_3 \Phi$, can be recovered by Algorithm in TABLE 1.

TABLE 1. The procedure of recovery SSCT.

Input: SNT $\underline{\mathbf{X}}_u$, dictionary $\mathbf{D} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_L] \in \mathbb{R}^{B \times L}$, sparsity level S (or tolerance ε);

Output: Index set $\{I\}$, SSCT $\underline{\mathbf{A}}_u$ whose nonzero entries in mode-3 indexed by $\{I\}$ are tensor $\hat{\underline{\mathbf{A}}}_u$;

1: Index set $I = [\varnothing], \underline{\mathbf{R}} = \underline{\mathbf{X}}_u, l = 1$;

2: While $|I| \le S$ and $\|\underline{\mathbf{R}}\|_F \ge \varepsilon$ do

3: $[i_l] = \arg\max_{i=1,\cdots,I} \left| \underline{\mathbf{R}} \times_3 \mathbf{D}^T(:,i) \right|$;

4: $I = [I, i_l]$;

5: $\hat{\mathbf{a}}_u = \arg\min_{\mathbf{v}} \left\| \left(\mathbf{D}(:,I) \otimes \mathbf{I}_T \otimes \mathbf{I}_T \right) \mathbf{v} - \mathbf{x}_u \right\|_2^2$;

6: $\underline{\mathbf{R}} = \underline{\mathbf{X}}_u - \hat{\underline{\mathbf{A}}}_u \times_3 \mathbf{D}(:,I)$;

7: $l = l + 1$;

8: End while;

D. CLASSIFYING

Once obtaining the SSCT corresponding to the compressive dimensionality reduced SNT $\underline{\mathbf{X}}_u^r$, the unknown pixel $\mathbf{x}_u$ can be classified by comparing the residual between tensor $\underline{\mathbf{X}}_u^r$ and the approximate values from the c-th class training pixels

$$r_c(\underline{\mathbf{X}}_u^r) = \left| \underline{\mathbf{X}}_u^r - \underline{\mathbf{A}}_{u,c}^r \times_3 \Phi \mathbf{D}_c \right| \tag{6}$$

where $\mathbf{D}_c \in \mathbb{R}^{B \times L_c}$ (with $L = \sum_{c=1}^{C} L_c$) is a class sub-dictionary consisting of the training pixels in the c-th class, and $\underline{\mathbf{A}}_{u,c}^r$ consists of the L_c frontal slices of the SSCT corresponding to $\underline{\mathbf{X}}_u^r$. Then the class label of $\mathbf{x}_u$ can be predicted by

$$\text{Class}(\mathbf{x}_u) = \arg\min_{c=1,\ldots,C} r_c(\underline{\mathbf{X}}_u^r) \tag{7}$$

And thus an SSCT based Classifier with CDR (SSCTC-CDR) is obtained for HIC.

III. EXPERIMENTAL RESULTS AND DISCUSSION

A. EXPERIMENTAL DATASETS

The performance of our proposed SSCTC-CDR is evaluated on three real hyperspectral images. The first experimental dataset is the Indian Pines image, which includes 16 classes of 10366 interest pixels. This image is an imbalanced dataset, the number of pixels each class ranges from 20 to 2468 pixels. The second experimental dataset is the Kennedy Space Center (KSC) Image. It contains 13 classes representing the various land cover types. There are 5211 interested pixels sparsely distributed in the whole image. These two datasets are downloaded from [13]. The last experimental dataset is the Houston University Image provided by Prof. Saurabh Prasad. There are 15 classes of 15029 pixels in this image. The detail class information of the three experimental dataset is shown in TABLE 2.

B. EXPERIMENTAL SETUP

To evaluate the performance of our proposed SSCTC-CDR, it is compared with some related methods, including: 1) sparse learning approaches Sparse Coding based Classifier (SCC) [14], SVM [3], Composite-Kernel SVM (CK-SVM) [7], JSCC [8] Spatial-Aware Dictionary Learning (SADL) [15], HPSCC-SNT [11]; 2) deep learning approaches Spectral-Spatial Networks (SSN) [16], Rolling guidance filter and Vertex Component Analysis Network (R-VCANet) [17]. The parameters settings for SCC, JSCC, SVM and CK-SVM following paper [8]. To eliminate the impact of random sampling on the performance of various methods, we run each experiment 20 times under the same conditions and report the average results. The simulations of different algorithms are carried out in MATLAB7.10.0 (R2010a) environment running in Core 2 Quad, 3.20-GHZ CPU with 8.00-GB RAM.

C. INVESTIGATION OF THE ALGORITHMIC PERFORMANCES

This subsection evaluates the performance of our proposed SSCTC-CDR on the two experimental datasets. For the Indian Pines image, the ground-truth dataset is randomly split into two parts, 1% for training and the other for testing. The performance of SSCTC-CDR on it is not only visually compared but also quantitatively investigated. The ground-truth dataset of both KSC and Houston University image is randomly split into training set, which containing 3% pixels of each class, and testing set which contains the other pixels. For KSC and Houston University image, we only quantitatively investigate the performance of SSCTC-CDR. Classification maps are adopted to show visual results. And the quantitative classification performance of each model is assessed by accuracy of each class, Overall Accuracy (OA), Average Accuracy (AA) and Kappa coefficient [18].

1) INVESTIGATION OF THE PERFORMANCES ON INDIAN PINES

In this experiment, the sparsity level S and the spatial neighborhood window width T in our proposed SSCTC-CDR is set as 20 and 9 respectively. $Ratio = {}^{n}/_{B} \times 100\%$ is utilized to measure the compressive dimensionality reduction ratio of SSCTC-CDR, which is set as 50% in

TABLE 2. Details of the experimental datasets.

Class	Indian Pines		KSC		Houston University	
	Name	#Samples	Name	#Samples	Name	#Samples
1	Alfalfa	54	Scrub	761	Grass_healthy	1251
2	Corn-no till	1434	Willow swamp	243	Grass_stressed	1254
3	Corn-min till	834	Cabbage palm hammock	256	Grass_synegic	697
4	Corn	234	Cabbage palm/oak hammock	252	Tree	1244
5	Grass/Pasture	497	Slash pine	161	Soil	1242
6	Grass/Trees	747	Oak/broadleaf hammock	229	Water	325
7	Grass/Pasture-mowed	26	Hardwood swamp	105	Residential	1268
8	Hay-windrowed	489	Graminoid marsh	431	Commercial	1244
9	Oats	20	Spartina marsh	520	Road	1252
10	Soy beans-no till	968	Cattail marsh	404	Highway	1227
11	Soy beans-min till	2468	Salt marsh	419	Railway	1235
12	Soy beans-clean till	614	Mud flats	503	Parking lot 1	1233
13	Wheat	212	Water	927	Parking lot 2	469
14	Woods	1294	-	-	Tennis court	428
15	Bldg-Grass-Tree-Drives	380	-	-	Running track	660
16	Stone-steel Towers	95	-	-	-	-
Total	-	10366	-	5211	-	15029

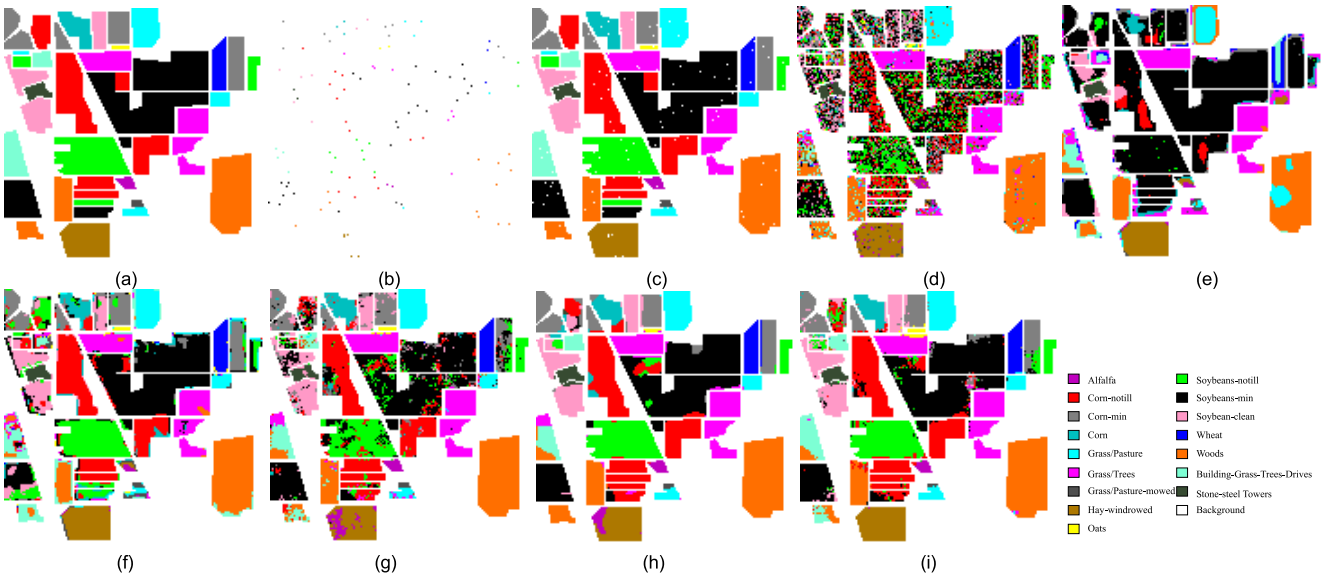

FIGURE 3. (a) Ground-truth. (b) Training pixels. (c) Testing pixels and Classification map of (d) SCC (e) SVM (f) CK-SVM (g) JSCC (h) HPSCC-SNT (i) SSCTC-CDR on the Indian Pines image.

this experiment. By training the models of SCC, SVM, CK-SVM, JSCC, HPSCC-SNT, SSCTC-CDR on the pixels shown in **FIGURE 3.** (b) respectively, the learned models can predict the labels of testing pixels (see **FIGURE 3.**(c)). And the visual classification maps of these five methods are presented in **FIGURE 3.** (d)-(i), respectively. A visualized impression can be obtained from **FIGURE 3.** (d)-(i) that our proposed SSCTC-CDR has much better classification result than the compared methods except HPSCC-SNT. The classification map of our proposed SSCTC-CDR has very well spatial homogeneity. It can be found that SSCTC-CDR avoids most of the salt-and-pepper misclassifications in SCC and JSCC. Furthermore, most of the speckle-like misclassifications in SSCTC-CDR, but for HPSCC-SNT they are almost speckle-like misclassifications. This is because for the mixed pixels near boundary, their SNTs contain

TABLE 3. Classification accuracy (\%) and consumed time(second) of different methods on the Indian Pines image.

Class	SCC	SVM	CK-SVM	JSCC	SADL	SSN	HPSCC-SNT	SSCTC-CDR
1	44.34	27.74	42.45	**94.53**	69.81	81.32	95.47	94.15
2	45.13	44.54	76.53	81.13	68.73	80.51	**89.13**	83.59
3	26.51	36.97	67.22	67.92	65.58	79.75	**82.74**	78.63
4	18.92	21.21	41.21	53.98	62.81	68.57	**71.39**	64.29
5	65.63	59.67	68.03	72.11	**82.34**	79.25	73.82	76.16
6	90.09	84.37	85.72	98.54	93.84	95.18	**100**	97.2
7	40	76.8	79.6	78.8	**100**	95.2	92.8	79.2
8	86.63	85.93	90.87	97.5	97.48	93.08	98.57	**99.9**
9	40.53	32.63	86.32	57.37	81.05	89.47	65.26	**90**
10	30.55	49.32	72.37	70.56	74.11	77.85	79.41	**82.72**
11	56.66	62.03	75.06	84.15	81.34	85.62	88.14	**88.4**
12	27.41	31.91	53.46	69.59	57.71	69.95	72.59	**75.82**
13	94.98	88.52	85.12	98.09	98.66	96.36	**100**	96.65
14	90.87	84.45	90.82	96.28	93.71	96.28	**97.89**	96.49
15	20.21	21.49	43.7	66.17	68.7	**83.16**	64.1	71.3
16	83.94	48.62	93.3	90.43	79.57	96.06	**100**	93.4
OA	55.71 ±1.14	57.86 ±2.86	74.57 ±2.44	78.85 ±1.41	78.95 ±**1.09**	84.70 ±2.28	**87.08** ±1.61	86.72 ±1.28
AA	53.90 ±2.14	53.51 ±1.58	71.99 ±3.05	81.32 ±**1.49**	79.72 ±2.78	85.48 ±2.10	**85.70** ±2.28	85.49 ±1.75
Kappa×100	49.23 ±1.40	52.00 ±3.09	71.12 ±2.78	75.81 ±1.66	75.98 ±**1.27**	82.57 ±2.55	**85.26** ±1.83	84.85 ±1.44
Time	17.68	-	15.64	2004.2	-	-	5549.32	3109.31

TABLE 4. Classification accuracy (\%) and consumed time(second) of different methods on the KSC image.

Class	SCC	SVM	CK-SVM	JSCC	SADL	R-VCANet	HPSCC-SNT	SSCTC-CDR
1	90.26	92.75	98.51	99.15	99.73	99.36	**100**	99.84
2	81.84	81.88	80.9	89.1	82.48	94.08	**97.26**	96.45
3	76.76	89.23	89.11	91.34	95.95	96.35	**99.23**	98.38
4	41.69	62.76	55.68	65.56	**90.95**	87.84	90.45	84.94
5	42.97	55.94	47.35	71.68	63.23	90.51	98.13	**99.48**
6	39.82	55.18	75	66.59	86.82	93.42	**96.36**	94.32
7	70.79	78.42	83.37	89.8	73.27	99.57	**100**	99.11
8	74.01	85.38	95.29	93.46	98.56	**98.96**	96.59	96.75
9	89.44	94.75	99.96	98.63	96.62	97.97	**100**	**100**
10	91.28	90.69	95.91	99.05	96.16	99.97	**100**	**100**
11	95.31	93.23	97.46	98.25	94.57	99.62	**100**	**100**
12	70.04	81.13	94.46	92.81	85.01	**99.44**	94.95	96.47
13	98.86	98.57	**100**	**100**	98.22	**100**	**100**	**100**
OA	81.45 ±0.76	86.95 ±0.78	91.72 ±0.82	93.16 ±0.83	93.31 ±0.78	97.90 ±**0.66**	**98.39** ±0.88	98.11 ±0.82
AA	74.08 ±1.35	81.53 ±1.42	85.61 ±1.39	88.88 ±1.67	89.35 ±**1.01**	96.70 ±1.10	**97.92** ±1.41	97.36 ± 1.22
Kappa×100	79.34 ±0.85	85.45 ±0.87	90.78 ±0.91	92.38 ±0.93	92.55 ±0.97	97.66 ±**0.74**	**98.20** ±0.98	97.90 ±0.91
Time	23.69	-	23.79	271.59	-	-	618.13	497.07

pixels belonging to different categories, and thus they provide negative effects for classifying the central pixel in SSCTC-CDR. But this kind of negative effect is suppressed in HPSCC-SNT.

Besides the visualized comparison, the performance of SSCTC-CDR is quantitatively investigated by comparing with SADL, SSN and the above mentioned five algorithms. The experimental results of SVM, SADL, and SSN use the results in [16] immediately. Other parameters setting are the same as in the last experiment. The classification results and consumed time of these algorithms are shown in TABLE 3. The value in bold of each row indicates the best result among

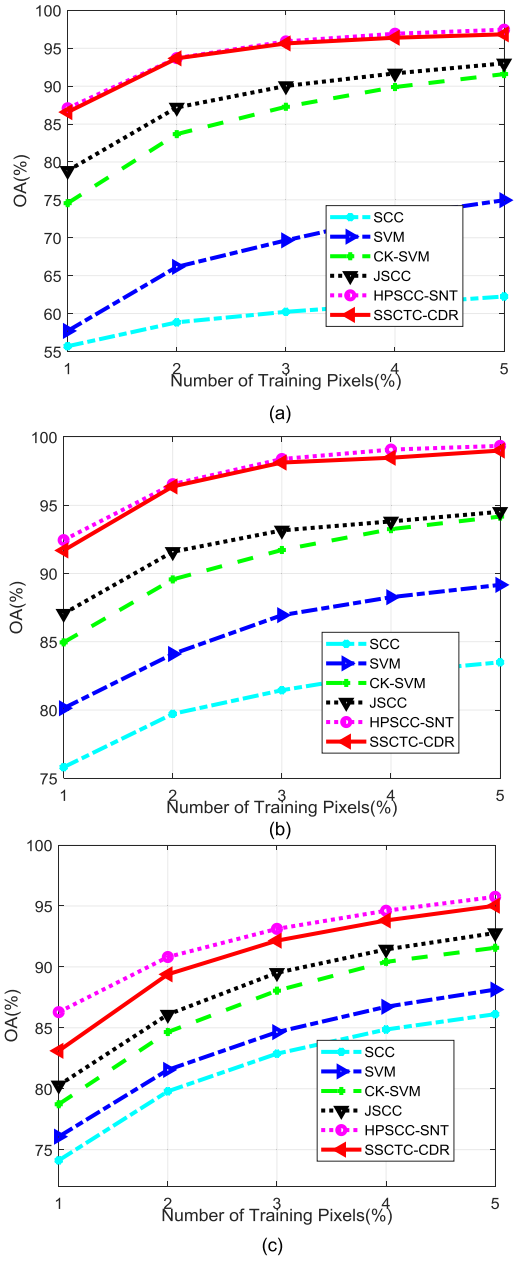

FIGURE 4. Performance of SSCTC-CDR with different number of training pixels on (a) Indian Pines (b) KSC (c).

the seven methods. TABLE 3 shows that the OA, AA and Kappa coefficient for SSCTC-CDR reaches 86.72%, 85.49% and 84.68% respectively. Furthermore, SSCTC-CDR obtains the highest accuracy for 5 classes and the second highest accuracy for 4 classes. Our proposed SSCTC-CDR obtains much higher classification accuracy than the compared methods except HPPSCC-SNT. Furthermore, the standard deviations of classification accuracy obtained by our proposed SSCTC-CDR are lower than most of the compared methods. This shows that benefiting from the tensor coding mechanism, our method can maintain relatively high classification robustness while improving the classification accuracy. Moreover, although the classification performance

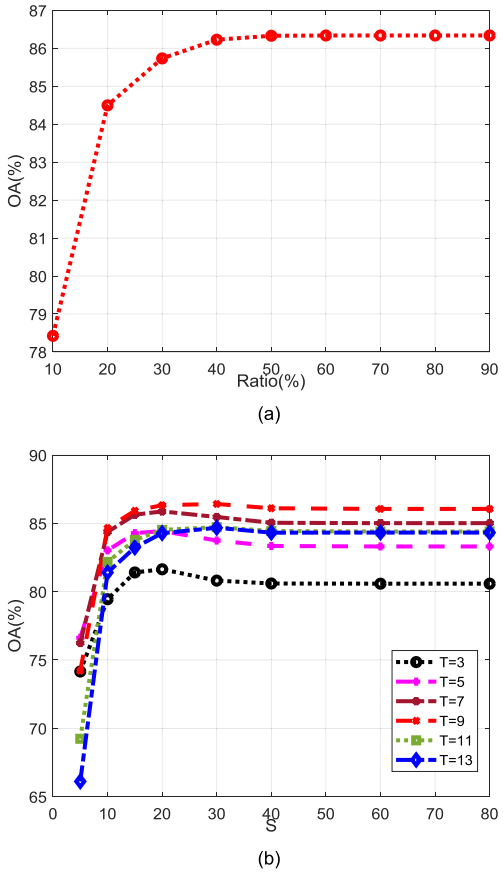

FIGURE 5. Performance of SSCTC-CDR with the variation of (a) *Ratio* (b) *T* and *S* on Indian Pines.

of SSCTC-CDR is slightly worse than HPSCC-SNT, its consumed time is much less than HPSCC-SNT.

2) INVESTIGATION OF THE PERFORMANCES ON KSC

This experiment quantitatively evaluates the performances of SSCTC-CDR and six compared methods. The sparsity level S, the spatial neighborhood window width T and the tolerance ε in our proposed SSCTC-CDR is set as 20, 7 and 0.001 respectively. The other experimental settings follow last experiment. The experimental result of R-VCANet comes from [17]. The results are reported in TABLE 4, from which it can be found that our proposed SSCTC-CDR obtains the highest OA, AA, Kappa coefficient and accuracy for most of the classes among these algorithms other than HPSCC-SNT. And SSCTC-CDR obtains slightly worse classification results than HPSCC-SNT by consuming less time.

3) INVESTIGATION OF THE PERFORMANCES ON HOUSTON UNIVERSITY

In this experiment, the performance of SSCTC-CDR is compared with five related methods. The sparsity level S and the spatial neighborhood window width T is set as 10 and 5 respectively. The other experimental settings follow

TABLE 5. Classification accuracy (\%) and consumed time(second) of different methods on the Houston University image.

Class	SCC	SVM	CK-SVM	JSCC	HPSCC-SNT	SSCTC-CDR
1	94.69	95.70	96.94	96.49	96.70	**97.03**
2	97.79	97.77	98.35	**99.09**	98.44	99.07
3	99.70	99.67	98.80	**100.00**	**100.00**	**100.00**
4	88.70	95.00	96.69	93.45	**98.16**	96.08
5	97.51	98.40	99.68	99.57	**100.00**	99.78
6	95.05	88.72	84.73	95.94	**98.08**	95.24
7	74.78	84.00	88.34	81.54	**97.31**	84.39
8	76.82	77.37	78.02	80.46	**88.69**	81.69
9	68.67	74.91	72.04	77.39	76.96	**81.95**
10	87.60	86.98	88.19	94.41	96.39	**97.58**
11	69.19	71.80	82.93	86.09	87.07	**91.90**
12	70.43	65.59	76.22	83.44	84.34	**89.72**
13	22.58	26.04	60.20	55.07	**82.31**	67.05
14	95.74	96.37	98.96	99.88	**99.96**	99.90
15	98.53	98.34	96.64	99.48	**100.00**	99.98
OA	82.88 ±0.52	84.65 ±1.34	88.05 ±1.26	89.52 ±0.66	**93.12 ±0.34**	92.15 ±0.62
AA	82.52 ±**0.20**	83.78 ±1.22	87.78 ±1.25	89.49 ±0.63	**93.63** ±0.42	92.09 ± 9.55
Kappa×100	81.48 ±0.56	83.39 ±1.45	87.08 ±1.36	88.67 ±0.71	**92.56 ±0.37**	91.52 ±0.67
Time	30.54	28.76	46.55	305.53	2387.80	1014.10

last experiment. TABLE 5 reports the classification results and the consumed time. The best result among these six algorithms are in bold each row. TABLE 5 shows that SSCTC-CDR obtains the highest accuracy for 6 classes and the second highest accuracy for 6 classes. The OA, AA, Kappa coefficient of SSCTC-CDR is slightly worse than HPSCC-SNT but much better than the other methods. And SSCTC-CDR consumes much less time than HPSCC-SNT.

D. INVESTIGATION OF THE NUMBER OF TRAINING PIXELS

This experiment investigates how the number of training pixels influences the classification accuracy. For both the two imageries, 1%, 2%, 3%, 4%, 5% pixels from each class are randomly chosen as training pixels, and the other pixels are used for testing respectively. Other settings are the same as in last subsection. **FIGURE 4.** (a),(b),(c) shows the average OAs of different methods on testing pixels of the three experimental imageries respectively. From **FIGURE 4.**, it can be found that using the same training pixels, the performance of our proposed SSCTC-CDR is slightly worse than HPSCC-SNT but much better than SCC, SVM, CK-SVM and JSCC. Furthermore, as the number of training pixels increasing, the performance of each method improves, but the improvement of SSCTC-CDR below compared methods except HPSCC-SNT. The reason can be that the tensor structure conveys much spatial information, which is beneficial to obtain high performance by using few training pixels.

E. INFLUENCES OF THE PARAMETERS RATIO, T AND S

In this subsection, the Indian Pines image is taken as an example to investigate the influences of parameters *Ratio*, T and S on the classification performance of the proposed SSCTC-CDR. We range the compressive dimensionality reduction ratio *Ratio*, the spatial neighborhood window width T and the sparsity level S from 10% to 90%, 3 to 13 and 5 to 80 respectively and preset the other parameters as last subsection. The average OAs are presented in **FIGURE 5.** (a),(b),(c) respectively. The following results can be observed from **FIGURE 5.**:

1) As the *Ratio* increases, the OA of SSCTC-CDR increases steadily until *Ratio* reaches 50%, after which the OA remains steady. This shows that compressive dimensionality reduction in our proposed method can greatly reduce the computation cost and storage cost while maintaining classification accuracy.
2) For SSCTC-CDR, when S is less than 10, the performance is poor due to the over-low sparsity level. The OA increases with the increase of S, while it is essentially unchanged when S exceeds 20. This shows that too high sparsity level cannot improve the classification performance of SSCTC-CDR.
3) When T is set to 3 or 5, the neighborhood information contained in the window is insufficient, thus the classification accuracy is very low. As T increases, OA peaks at 9, but starts to decrease when T is greater than 9. It may be for too large T, the pixels belonging to other categories are selected into the

spatial neighbor window, which results in classification errors.

Therefore, *Ratio*, T and S on the Indian Pines image is set to 50%, 9 and 20 respectively in the former experiments.

IV. CONCLUSION

This paper proposes a new tensor organization based classifier, called SSCTC-CDR, for hyperspectral image classification. Our proposed tensor representation preserves as many original spatial constraints of spatial neighboring pixels as possible by utilizing their tensor structural information. What's more, the computation cost and storage cost in our proposed method can be greatly reduced by compressive dimensionality reduction. Furthermore, unlike HPSCC-SNT whose classification result is affected by the range of two regularization parameters, there are very few parameters in SSCTC-CDR. The experiment results over two real hyperspectral imageries show that our algorithm can achieve relatively high-accurate classification result with lower computation and restoration lost. However, our proposed SSCTC-CDR does not consider the effect of mixed-pixels, thus the classification result of SSCTC-CDR is slightly worse than the methods such as HPSCC-SNT which alleviates the misclassification of mixed pixels. This drawback will be overcome in future works.

REFERENCES

[1] Y. Kim and Y. Kim, "Hyperspectral image classification based on spectral mixture analysis for crop type determination," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Valencia, Spain, Jul. 2018, pp. 5304–5307.

[2] J. Xia, B. Huang, Y. W. Yang, H. X. Cao, W. Zhang, L. Xu, Q. Wan, Y. Ke, W. Zhang, and D. Ge, "Hyperspectral identification and classification of oilseed rape waterlogging stress levels using parallel computing," *IEEE Access*, vol. 6, pp. 57663–57675, 2018.

[3] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

[4] B. Demir and S. Ertürk, "Hyperspectral image classification using relevance vector machines," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 4, pp. 586–590, Oct. 2007.

[5] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*. Hoboken, NJ, USA: Wiley, 2005.

[6] K. Tan, E. Li, Q. Du, and P. Du, "Hyperspectral image classification using band selection and morphological profiles," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 1, pp. 40–48, Jan. 2014.

[7] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. Vila-Frances, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.

[8] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.

[9] J. Peng, Y. Zhou, and C. L. P. Chen, "Region-kernel-based support vector machines for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 4810–4824, Sep. 2015.

[10] H. B. Barlow, "Single units and sensation: A neuron doctrine for perceptual psychology?" *Perception*, vol. 1, no. 4, pp. 371–394, Dec. 1972.

[11] L. Yang, M. Wang, S. Yang, H. Zhao, L. Jiao, and X. Feng, "Hybrid probabilistic sparse coding with spatial neighbor tensor for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2491–2502, May 2018.

[12] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, Aug. 2009.

[13] [Online]. Available: http://www.ehu.eus/ccwintco/index.php?title=HyperspectralRemoteSensingScenes

[14] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[15] A. Soltani-Farani, H. R. Rabiee, and S. A. Hosseini, "Spatial-aware dictionary learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 527–541, Jan. 2015.

[16] Y. Zhou and Y. Wei, "Learning hierarchical spectral–spatial features for hyperspectral image classification," *IEEE Trans. Cybern.*, vol. 46, no. 7, pp. 1667–1678, Jul. 2016.

[17] B. Pan, Z. Shi, and X. Xu, "R-VCANet: A new deep-learning-based hyperspectral image classification method," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 1975–1986, May 2017.

[18] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis: An Introduction*, 4th ed. New York, NY, USA: Springer-Verlag, 2006.

LIXIA YANG (Member, IEEE) received the B.S. degree in mathematics education and the M.S. degree in applied mathematics from Ningxia University, Yinchuan, China, in 2004 and 2007, respectively, and the Ph.D. degree in intelligent information processing from Xidian University, Xi'an, China, in 2016. She is currently an Associate Professor with the School of Mathematics and Statistics, Ningxia University. Her main current research interests include machine learning and image processing.

RUI ZHANG received the B.S. degree in mathematics education and the M.S. degree in pure mathematics from Ningxia University, Yinchuan, China, in 2002 and 2006, respectively, and the Ph.D. degree in applied mathematics from Xidian University, Xi'an, China, in 2018. He is currently an Instructor with the School of Mathematics and Statistics, Ningxia University. His main current research interest includes image processing.

SHUYUAN YANG (Senior Member, IEEE) received the B.S. degree in electrical engineering and the M.S. and Ph.D. degrees in circuits and systems from Xidian University, Xi'an, China, in 2000, 2003, and 2005, respectively. She is currently a Professor with the School of Artificial Intelligence, Xidian University. Her main current research interests include intelligent signal processing, machine learning, and image processing.

LICHENG JIAO (Fellow, IEEE) received the B.S. degree from Shanghai Jiao Tong University, Shanghai, China, in 1982, and the M.S. and Ph.D. degrees from Xian Jiaotong University, Xi'an, China, in 1984 and 1990, respectively. He is currently a Professor with the School of Artificial Intelligence, Xidian University, Xi'an. His research interests include neural networks, data mining, nonlinear intelligence signal processing, and communication.