Stacked Tensor Subspace Learning for Hyperspectral Image Classification

Yantao Wei
School of Educational Information Technology
Central China Normal University
Wuhan 430079, China.
Email: yantaowei@mail.ccnu.edu.cn

Yicong Zhou
Department of Computer and Information Science
University of Macau
Macao, China.
Email: yicongzhou@umac.mo

***Abstract*—In this paper, we present a hierarchical feature learning method called Stacked Tensor Subspace Learning (STSL). It can jointly learn spectral and spatial features of hyperspectral images (HSIs) by iteratively abstracting neighboring regions. STSL is able to learn discriminative spectral-spatial features of the input HSI at different scales. In STSL, the joint spectral and spatial features are extracted using Marginal Fisher Analysis (MFA) and Tensor Principal Component Analysis (TPCA). Then Kernel-based Extreme Learning Machine (KELM), a shallow neural network, is embedded in the proposed method to classify image pixels. The important contributions to the success of STSL are exploiting local spatial structure of HSI by using tensor method and designing hierarchical architecture. Extensive experimental results on two challenging HSI data sets taken from the Airborne Visible-Infrared Imaging Spectrometer (AVIRIS) and Reflective Optics System Imaging Spectrometer (ROSIS) airborne sensors show that the proposed method can produce good classification accuracy with smaller training sets.**

I. Introduction

Hyperspectral image (HSI) classification, where each pixel is assigned to a predefined class, has been applied widely in many fields, including precision farming, water resources monitoring and geological survey [1]–[4]. Over the last few years, a lot of techniques have been developed to promote the HSI classification performance [2], [3], [5]–[8].

Subspace learning is an effective way to improve HSI classification accuracy [2], [9]–[14]. Many subspace learning methods have been adopted for HSI classification. For instance, Locally Linear Embedding (LLE) and Local Preserving Projection (LPP) have been utilized to reduce the dimensionality of HSIs [15], [16]. However, only spectral information are used in most of the subspace learning-based methods. Nowadays, spatial contexture information has been proven to increase classification accuracy [2], [3], [5]–[7], [17]. Many techniques have been adopted to make use of spatial information of HSIs [3], [8], [18], [19], such as morphological profiles [18], Markov fields [3] or Gabor filters [8]. Zhou et al. proposed a spatial and spectral regularized subspace learning method called Spatial and Spectral Regularized Local Discriminant Embedding (SSRLDE) for HSI classification [2]. Benediktsson et al. used Extended Morphological Profiles (EMPs) to model structural information [20]. Quesada-Barriuso et al. created an EMP from the wavelet features to obtain spectral-spatial features [21]. Kang et al. [22] proposed a spectral-spatial classification framework based on Edge-Preserving Filtering (EPF), where the filtering operation achieves a local optimization of the probabilities. Soltani-Farani et al. presented Spatial Aware Dictionary Learning [23] (SADL) method, which is a structured dictionary-based model for hyperspectral data incorporating both spectral and contextual characteristics of spectral samples. Maximizer of the Posterior Marginal by Loopy Belief Propagation (MPM-LBP) was proposed by Li et al. [24], it exploits the marginal probability distribution from both the spectral and spatial information. Li et al. proposed a framework for integrating multiple types of features extracted from both linear and nonlinear transformations [25].

However, most of these methods do not extract spectral-spatial features in a hierarchical fashion. Recent studies show that the hierarchical deep model can extract more abstract and invariant features of data [26]–[30]. Deep learning, inspired by the mechanism of human vision, recently attracted more and more attentions due to its good performance in many fields such as speech recognition, computer vision, and natural language processing [31]–[34]. In [26], a deep learning-based HSI classification method was proposed, where spectral and spatial information are extracted separately and then processed via stacked Autoencoders (AE). Similarly, Li et al. proposed to use Deep Belief Networks (DBN) for HSI classification [35]. But these methods directly use the conventional deep learning models and cannot take advantage of the characters of HSI. In [36], a hierarchical method called Spectral-Spatial Networks (SSN) was proposed by Zhou et al. However, they consider the local neighborhood as a long vector, which may cause the problem of curse of dimensionality under the high-dimensional and the losing of the structure information. Recently, tensor representation has attracted great interest and has been widely applied to problems with tensorial data. Consequently, a new hierarchical tensor-based method called Stacked Tensor Subspace Learning (STSL) is proposed for HSI classification in this paper. STSL can exploit spectral and spatial information efficiently. Apart from feature extraction, designing effective spectral classifiers is also a possible way to promote the classification accuracy. Support Vector Machine (SVM)-based approaches have been used successfully for HSI classification [37]. However, choosing suitable kernel functions, kernel specific parameters, regularization parameters and strategies for

978-1-5090-0620-5/16/$31.00 ©2016 IEEE

multiclass classification is the major concern in the design of SVM [38]. Demir et al. [39] proposed to use Relevance Vector Machine (RVM) for HSI classification. It is faster than SVM. Recently, Kernel-based Extreme Learning Machine (KELM) has been proposed by Huang et al. [40]. Compared with SVM, KELM is faster and has good generalization ability [40], [41]. It has been applied to HSI classification [36], [42], and the results confirm that it is comparable in accuracy with SVM and has lower computational complexity. Consequently, the KELM is used in this paper.

STSL consists of spectral and spatial feature learning units. These units are serially connected to form a hierarchical architecture. In the proposed method, Marginal Fisher Analysis (MFA) and Tensor Principal Component Analysis (TPCA) are used to learn discriminative features from the 1st-order tensors and 2nd-order tensors, respectively. Finally, the learned features are input into KELM classifier. There are several advantages to be highlighted as follows:

- STSL provides a new way to exploit discriminative spectral-spatial information in a hierarchical fashion. The tensor-based method can help to make use of the spatial information.
- The proposed method has low sample complexity. That is, it can achieve high classification accuracy with a small number of training samples.

The remainder of this paper is structured as follows. The related works are briefly reviewed in Section II. The STSL method for HSI classification is proposed in Section III. The experimental results and discussions are presented in Section IV. Finally, Section V summarizes the paper and future works are suggested.

II. Related Work

The HSI classification problem has been approached with a wide variety of methods in recent years. Recently, some methods relying on hierarchical structure have been proposed to make use of the spectral and spatial information of HSIs. Bruzzone et al. firstly exploited hierarchical analysis for context-based classification of high-resolution remote sensing images [43]. Recently, deep learning methods have been used in the HSI classification. Li et al. [35] proposed a DBN approach for classification of HSIs. The model is a stack of restricted Boltzmann machines, which are trained by greedy layer-wise unsupervised learning. However, by reducing the data to the first three PCA components, the spectral characteristics of the images have not been used in a principal manner by the DBN model. In the field of remote sensing imagery classification, stacked AE was firstly introduced into HSI classification by Chen et al. [26] However, the AE can only handle spectral features, but over looks spatial patterns lying in images. To overcome this problem, Contextual Deep Learning (CDL) was introduced into HSI classification by considering contextual information in spatial domain [44]. Hu et al. used Convolutional Neural Networks (CNN) to classify HSIs directly in spectral domain [45]. Furthermore, spectral-spatial classification method based on CNN is proposed by Slavkovikj et al. [46]. However, these methods usually need more training samples. In [36], Spectral-Spatial Feature Learning (SSFL) was designed to obtain robust features of HSIs. It combines the spectral feature learning and spatial feature learning in a hierarchical fashion. Stacking a set of SSFL units, a deep hierarchical model called SSN is further proposed for HSI classification. SSN can exploit both discriminative spectral and spatial information simultaneously. Specifically, SSN learns useful high-level features by alternating between spectral and spatial feature learning operations. Then KELM is embedded in SSN to classify image pixels. Compared with state-of-the-art methods, SSN with a deep hierarchical architecture obtains higher classification accuracy, especially when the number of the training samples is small. However, it only takes vectors as input. Unlike [36] however, the proposed method uses tensor as input and thus preserves more structure information.

III. Stacked Tensor Subspace Learning

In tensor analysis, vector is the 1st-order tensor and matrix is the 2nd-order tensor. In this paper, we intend to learn spectral and spatial information from 1st-order tensor and 2nd-order tensor, respectively. The flowchart of the proposed method is given in Figure 1. We can find that the STSL consists of spectral feature learning unit, spatial feature learning unit and classification unit. The alternating between spectral feature learning unit and spatial feature learning unit forms a hierarchical architecture. Let $\boldsymbol{I}_{tr} = \{\boldsymbol{I}_1, \boldsymbol{I}_2, \ldots, \boldsymbol{I}_N\}$ be the training set, where the training samples belong to $\mathfrak{C}$ classes. The KELM is briefly reviewed as follows [36], [40], [47].

A. Kernel-based Extreme Learning Machine

Let $\{\mathbf{x}_i, \mathbf{y}_i\}(i = 1, \ldots, N)$ be the N training samples, where $\mathbf{x}_i \in R^d$ and $\mathbf{y}_i = (y_{i,1}, \ldots, y_{i,\mathfrak{C}}) \in R^{\mathfrak{C}}$ indicates the class label. Assuming that the training samples belong to $\mathfrak{C}$ classes and

$$\mathbf{y}_{i,j} = \begin{cases} 1, & \mathbf{x}_i \text{ belongs to the } j\text{th class;} \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

The output function of ELM with L hidden neurons is given by

$$f_L(\mathbf{x}_i) = \sum_{j=1}^{L} \beta_j h_j(\mathbf{x}_i) = \mathbf{h}(\mathbf{x}_i)\beta = \mathbf{y}_i, \tag{2}$$

where $\beta = \left(\beta_1 \beta_2 \ldots \beta_L\right)^T$ is vector of output weights, and $\mathbf{h}(\mathbf{x}_i) = (h_1(\mathbf{x}_i), \ldots, h_L(\mathbf{x}_i))$ is the activation function corresponding to the input $\mathbf{x}_i$. The training pixels are mapped onto the L-dimensional feature space by $\mathbf{h}(\mathbf{x}_i)$. So N equations coming from (2) can be rewritten in a compact form $\mathbf{H}\beta = \mathbf{Y}$, where $\mathbf{H}$ is the output matrix of the hidden layer

$$\mathbf{H} = \begin{pmatrix} \mathbf{h}(\mathbf{x}_1) \\ \vdots \\ \mathbf{h}(\mathbf{x}_N) \end{pmatrix} = \begin{pmatrix} \mathbf{h_1}(\mathbf{x}_1) \ldots \mathbf{h_L}(\mathbf{x}_1) \\ \vdots \\ \mathbf{h_1}(\mathbf{x}_N) \ldots \mathbf{h_L}(\mathbf{x}_N) \end{pmatrix} \tag{3}$$

and

$$\mathbf{Y} = \left(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N\right)^T. \tag{4}$$

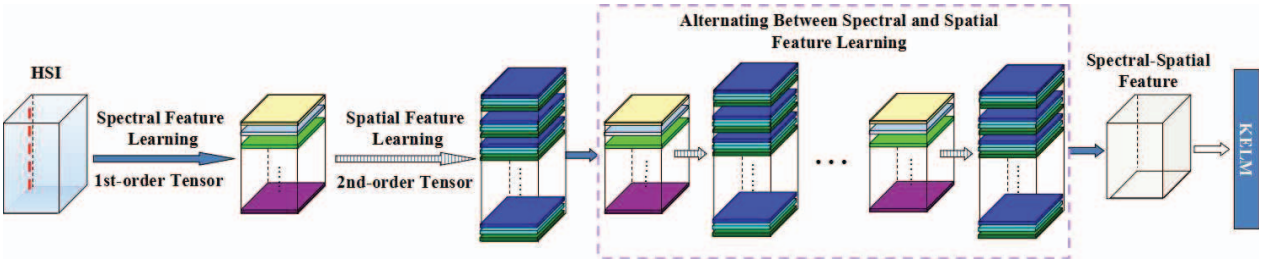


Fig. 1. Flowchart of the proposed method.

is the expected output matrix. Then we have

$$f(\mathbf{x}_i) = \mathbf{h}\beta^T = \mathbf{h}(\mathbf{x}_i)\mathbf{H}^T(\frac{\mathbf{I}}{\rho} + \mathbf{H}\mathbf{H}^T)^{-1}\mathbf{Y}. \quad (5)$$

As point out by Huang et al. in [40], if the feature mapping $\mathbf{h}(\mathbf{x}_i)$ is unknown, a kernel matrix for ELM can be defined as follows:

$$\mathbf{K}(\mathbf{X},\mathbf{X}) = \mathbf{H}\mathbf{H}^T : k(\mathbf{x}_i,\mathbf{x}_j) = \mathbf{h}(\mathbf{x}_i) * \mathbf{h}(\mathbf{x}_j)^T. \quad (6)$$

In this way, the output of the KELM classifier can be written as

$$\begin{aligned} f(\mathbf{x}_i) &= \mathbf{h}(\mathbf{x}_i)\mathbf{H}^T\left(\frac{\mathbf{I}}{\rho} + \mathbf{H}\mathbf{H}^T\right)^{-1}\mathbf{Y} \\ &= \begin{pmatrix}\mathbf{K}(\mathbf{x}_i,\mathbf{x}_1)\\ \vdots \\ \mathbf{K}(\mathbf{x}_i,\mathbf{x}_N)\end{pmatrix}^T\left(\frac{\mathbf{I}}{\rho} + \mathbf{K}(\mathbf{X},\mathbf{X})\right)^{-1}\mathbf{Y}. \quad (7)\end{aligned}$$

Finally, the label of the input data is determined by the index of the output node with the highest output value [47].

B. The Proposed Method

The detailed description of each processing unit is given as follows.

(1) **Spectral Feature Learning Unit**
In HSI classification, we expect that the learned feature can minimize the intra-class compactness and maximize the inter-class separability. Consequently, MFA [48] is used to learn discriminative spectral features in this paper, where the input are the 1st-order tensors. In this way, pixels within the same class can have the similar features despite they may be far away. These features can be obtained by projecting the pixels on the learned projection directions, where the projection directions can be given by

$$\widetilde{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}} \frac{\boldsymbol{w}^T\boldsymbol{I}_{tr}\boldsymbol{L}\boldsymbol{I}_{tr}^T\boldsymbol{w}}{\boldsymbol{w}^T\boldsymbol{I}_{tr}\boldsymbol{L}_p\boldsymbol{I}_{tr}^T\boldsymbol{w}} \quad (8)$$

where $\boldsymbol{L}$ and $\boldsymbol{L}_p$ are the Laplacian matrix of intra-class graph and inter-class graph, respectively [48]. Once the projection directions are obtained, the HSI can be projected to them pixel by pixel. In this way, the new representation of each pixel can be obtained.

(2) **Spatial Feature Learning Unit**
HSIs are usually smooth in the sense that each pixel share similar spectral characteristics or have the same class membership with its neighboring pixels. In order to exploit the contextual correlation within HSI, the spatial feature learning methods based on TPCA is given in this paper.
For the output of the spectral feature learning unit, we crop $l \times l$ image patches centered at each training pixel from each feature map (represented by the color rectangle in the Figure 1). If there are N training pixels, then we can collect N image patches from each map. The cropped image patches form a 3rd-order tensor. Then the new feature map can be obtained by projecting the image patches on the projection matrices. These matrices are obtained by maximizing the total scatter tensor of the projected low-dimensional feature as follows:

$$\begin{aligned} & f\{U^{(n)}, n = 1,2,\ldots,\hat{N}\} \\ =\ & \arg\max_{U^{(n)}} \Psi_y \\ =\ & \arg\max_{U^{(n)}} \sum_{m=1}^{M} \|\mathfrak{Y}_m - \bar{\mathfrak{Y}}\|^2, \quad (9)\end{aligned}$$

where $\mathfrak{Y}_m = \mathfrak{X}_m \times_1 U^{(1)T} \times_2 U^{(2)T} \cdots \times_{\hat{N}} U^{(\hat{N})T}$, $\bar{\mathfrak{Y}}$ is the mean of samples, and $\mathfrak{X}_m$ is the $\hat{N}$-th order tensor. For each patch, the new representation can be vectorized. In this way, a new feature cube can be obtained. Similarly, higher layer features can be obtained by expanded this architecture to n. Consequently, the proposed method can extract hierarchical features. The tensor method can preserve the spatial information.

(3) **Classification Unit**
After several alternations of the spectral and spatial feature learning process, the output of the last spatial feature learning unit is input to the KELM for classification. This is due to KELM's fast learning speed and strong generalization [40], [42].

The pseudo codes of the training and testing procedures of STSL are given in Algorithm 1 and Algorithm 2, respectively. In Algorithm 1, $\mathbb{I} \in R^{m' \times n' \times d}$ is the input image, L is the layer number, K_1^l and K_2^l are the number of project directions in lth spectral and spatial learning stages, respectively. In Algorithm 2, y_t is the label of the test sample.

Algorithm 1 STSL: Training procedure

Require: $\mathbb{I} \in R^{m' \times n' \times d}$, $\boldsymbol{I}_{tr} = [\boldsymbol{I}_1 \dots \boldsymbol{I}_N]$.
Ensure: Project directions, KELM classifier.
for $l = 1 : L$ **do**
  Obtain spectral projection directions by MFA.
  for $i = 1 : m$ **do**
    for $j = 1 : n$ **do**
      Using spectral projection directions to process the ith row and jth column of the input image.
    end for
  end for
  for $i = 1 : N$ **do**
    for $j = 1 : K_1^l$ **do**
      Crop a neighboring region for the ith training pixel on the jth map, and the cropped image patches form a 3rd-order tensor.
    end for
  end for
  Obtain K_2^l projection matrices by TPCA.
  for $j = 1 : K_1^l$
    Crop a neighboring region for the each pixel on the jth map and process them using obtained projection matrices. **do**
  end for
end for
Training KELM classifier.

Algorithm 2 STSL: Test procedure

Require: $\mathbb{I} \in R^{m' \times n' \times d}$.
Ensure: y_t.
Image preprocessing.
for $l = 1 : L$ **do**
  for $i = 1 : m$ **do**
    for $j = 1 : n$ **do**
      Process the pixel in the ith row and jth column by learned K_1^l project directions.
    end for
  end for
  for $j = 1 : K_1^l$ **do**
    Crop a neighboring region for each pixel on the jth map and process them using learned K_2^l project matrices.
  end for
end for
Feed the output to the KELM.

IV. Experimental Results

In our experiments, two hyperspectral data sets were used to demonstrate the effectiveness of the proposed method. The performance is evaluated in terms of overall accuracy (OA), average accuracy (AA), and the Kappa coefficient of agreement (κ). Let $\mathfrak{C}$ be the number of the classes, N_i be the number of the samples in the ith class, and n_i be the number of correctly classified samples in the ith class, then we have

$$OA = \frac{\sum_{i=1}^{\mathfrak{C}} n_i}{\sum_{i=1}^{\mathfrak{C}} N_i} \tag{10}$$

and

$$AA = \frac{1}{\mathfrak{C}} \sum_{i=1}^{\mathfrak{C}} \frac{n_i}{N_i}. \tag{11}$$

Let

$$\mathbf{M} = \begin{pmatrix} m_{11} \dots m_{1\mathfrak{C}} \\ \vdots \\ m_{\mathfrak{C}1} \dots m_{\mathfrak{C}\mathfrak{C}} \end{pmatrix} \tag{12}$$

be the confusion matrix, then

$$\kappa = \frac{p_0 - p_c}{1 - p_e}, \tag{13}$$

where

$$p_o = \frac{1}{\sum_{i=1}^{\mathfrak{C}} N_i} \sum_{i=1}^{\mathfrak{C}} m_{ii} \tag{14}$$

and

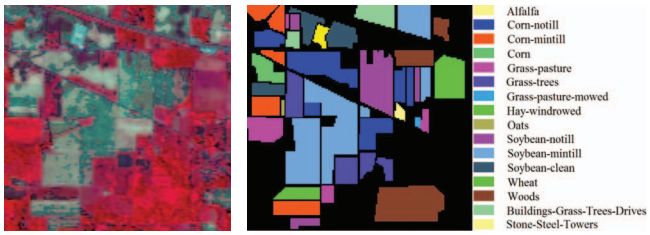
$$p_e = (\frac{1}{\sum_{i=1}^{\mathfrak{C}} N_i})^2 \sum_{i=1}^{\mathfrak{C}} m_{i.} m_{.i}. \tag{15}$$

The STSL is compared to support vector machine (SVM) [49], KELM [50], PCA plus Gabor filter (PCA+Gabor), EPF [22], and MPM-LBP [24]. We use MATLAB 2014a on a computer with Intel Corei7 3.6GHz CPU and 12GB memory.

A. Hyperspectral Data Sets

Two hyperspectral data sets collected by two different instruments are used in our experiments.

- Indian Pines data set was acquired by the AVIRIS sensor over the Indian Pines test site in 1992 [51]. The image scene contains 145×145 pixels and 220 spectral bands. The ground truth available is designated into 16 classes. The class descriptions and sample distributions for this image are given in Table I. In the experiments, the number of bands has been reduced to 200 due to atmospheric affection. This scene is challenging because of the significant presence of mixed pixels and the unbalanced number of available labeled pixels in each class [52], [53]. A three-band false color image and the ground-truth image are shown in Fig. 2.
- The second data set was gathered by the ROSIS sensor over Pavia, Italy. This image has $610\times$ 340 pixels (covering the wavelength range from 0.4 to $0.9\mu m$) and 115 bands. In our experiments, 12 bands are removed due to the noise. So there are 103 bands are retained. There are



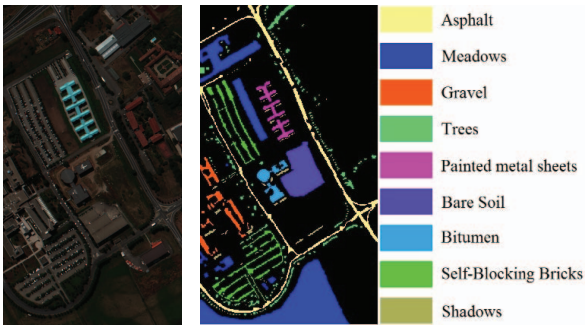
(a) (b)

Fig. 2. (a) False color composition of the AVIRIS Indian Pines scene. (b) Reference map containing 16 mutually exclusive land-cover classes.

TABLE I
TOTAL NUMBER OF SAMPLES IN EACH GROUND-TRUTH CLASS IN THE INDIAN PINES DATA SET.

No	Class Name	Subtotal
1	Alfalfa	54
2	Corn-notill	1434
3	Corn-mintill	834
4	Corn	234
5	Grass-pasture	497
6	Grass-trees	747
7	Grass-pasture-mowed	26
8	Hay-windrowed	489
9	Oats	20
10	Soybean-notill	968
11	Soybean-mintill	2468
12	Soybean-clean	614
13	Wheat	212
14	Woods	1294
15	Buildings-Grass-Trees-Drives	380
16	Stone-Steel-Towers	95
Total		**10366**

9 ground-truth classes, in total 43923 labeled samples. The class descriptions and sample distributions for this image are given in Table II. Fig. 3 shows a three-band false color image and the ground-truth image.



(a) (b)

Fig. 3. (a) False color composition of the ROSIS University of Pavia scene. (b) Reference map containing 9 mutually exclusive land-cover classes.

B. Experiments with the AVIRIS Indian Pines Data Set

For the first data set, the training samples were selected proportionally from different classes, where the number of the training samples is 1% of the available data. And the samples of the rest were used for testing. In this experiment,

TABLE II
TOTAL NUMBER OF SAMPLES IN EACH GROUND-TRUTH CLASS IN THE PAVIA UNIVERSITY DATA SET.

No	Class Name	Total
1	Asphalt	6852
2	Meadows	18686
3	Gravel	2207
4	Trees	3436
5	Painted metal sheets	1378
6	Bare Soil	5104
7	Bitumen	1356
8	Self-Blocking Bricks	3878
9	Shadows	1026
Total		**43923**

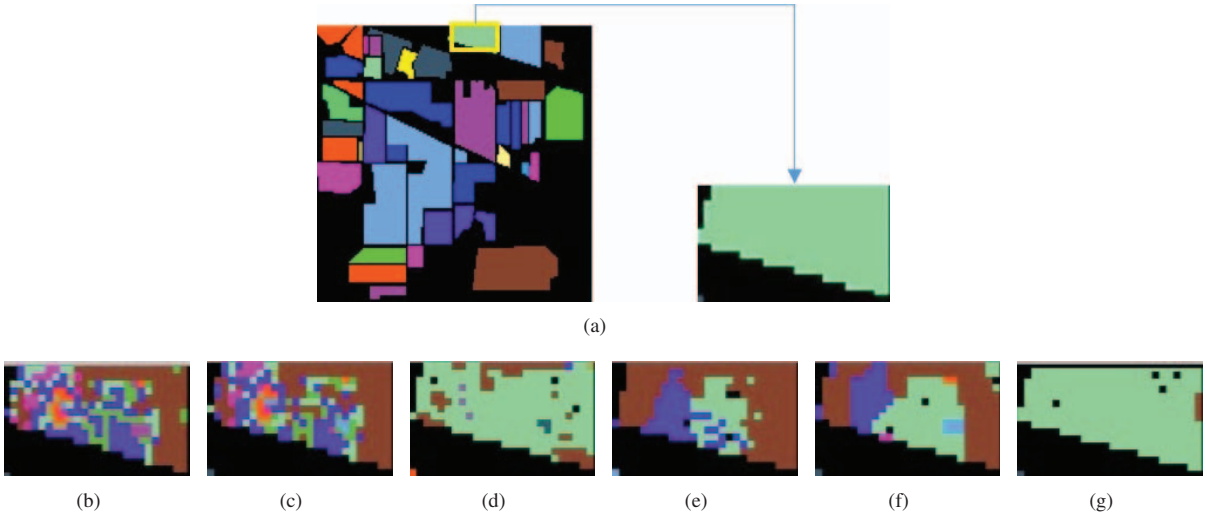
the proposed method used 8 layers to learn features. The experimental results are given in Table III, where experimental results are averaged over 10 runs. The experimental results show that the proposed method performs the best. Fig. 4 illustrates the classification maps obtained by the different methods. It corresponds to one of the random runs. Sub-figures (in the second row) from left to right are results from SVM, KELM, PCA+Gabor, EPF, MPM-LBP and STSL, respectively. It is obvious from Fig. 4 that STSL gives plausible results on smooth homogeneous regions by making use of the spatial information hierarchically. As can be seen, the maps obtained by SVM and KELM have very noisy appearance. One of the reasons for this maybe that spectral-based method cannot make use of the spatial correlation of the HSI. In addition, the affect of the number of training samples to the accuracy of different methods is given in Figure 5. It is consistent with the results in Table III. We can find that STSL performs better than other methods with a small number of training samples. This demonstrates that our method has low sample complexity. Finally, the comparison of time consumed by different methods is given in Table IV. It shows that the proposed method is time consuming.

TABLE III
OA (IN PERCENTAGE), AA (IN PERCENTAGE), AND KAPPA COEFFICIENT FOR THE AVIRIS INDIAN PINES DATA SET.

Method	OA	AA	κ
SVM	57.86	53.51	0.520
KELM	58.52	50.67	0.524
PCA+Gabor	68.78	68.06	0.641
EPF	71.75	62.35	0.671
MPM-LBP	77.16	65.97	0.736
STSL	**80.71**	**79.12**	**0.779**

TABLE IV
COMPARISON OF TIME CONSUMED BY DIFFERENT METHODS.

Method	SVM	KELM	PCA+Gabor	EPF	MPM-LBP	STSL
Time(S)	0.558	0.423	0.483	4.985	78.576	265.409



(a) (b) (c) (d) (e) (f) (g)

Fig. 4. Magnified parts of different classification maps for the Indian Pines data set. (a)Ground truth; (b) SVM; (c) KELM; (d) PCA+Gabor;(e) EPF; (f) MPM-LBP; (g) STSL.

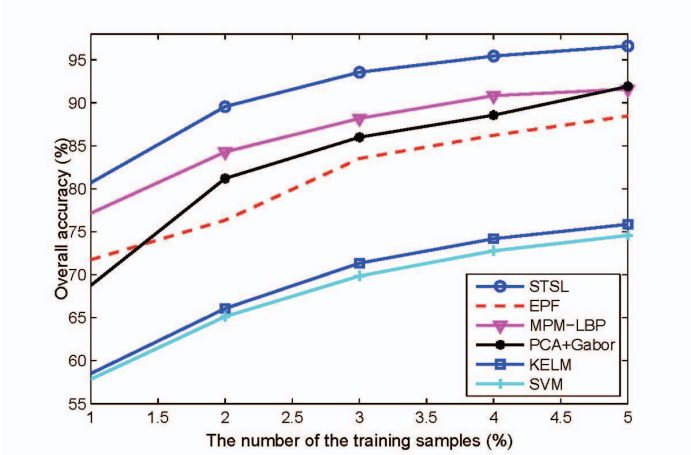


Fig. 5. OAs of different methods with different numbers of training samples for the Aviris Indian Pines Data Set.

C. Experiments with the ROSIS University of Pavia Data Set

For the second data set, labeled samples are randomly divided into training set and test set with a ratio of 1:99. For this data set, the number of feature learning layers is 4. The same conclusion can be made for this data set. The experimental results are given in Table V, where experimental results are averaged over 10 runs too. Experimental results given in Table V show that methods making use of the spectral-spatial information perform better than spectral-based methods. This phenomenon demonstrates the advantage of exploiting spatial information in HSI classification. It is easy to find that STSL performs the best on this data set. Fig. 6 illustrates the classification maps of the methods listed in Table V. From the visual inspection of the images, we find that the proposed method outperforms other methods as the resulting classification maps are smoother. It also confirms that the spectral-spatial classification method is able to prevent salt-and-pepper noise in the classification maps.

TABLE V
OA (IN PERCENTAGE), AA (IN PERCENT), AND KAPPA COEFFICIENT FOR THE PAVIA UNIVERSITY DATA SET.

Method	OA	AA	κ
SVM	89.19	85.64	0.857
KELM	89.08	85.40	0.855
PCA+Gabor	94.18	91.75	0.923
EPF	95.06	**96.16**	0.935
MPM-LBP	95.42	92.31	0.940
STSL	**96.49**	92.45	**0.954**

V. CONCLUSION

In this paper, a hierarchical method called STSL is proposed for HSI classification. This method integrates tensor subspace learning into hierarchical architecture to learn discriminative spectral-spectral features. Experimental results on two widely-used HSI data sets have demonstrated the effectiveness of STSL. Comparing with previously developed methods, the accuracy is improved. The experimental results also show that STSL has low sample complexity. We also note that the higher order tensor can be used to learn spectral-spatial features, which is our future work.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China under Grants 61502195, 61472155 and 61075116, Macau Science and Technology Development Fund under Grant FDCT/106/2013/A3, Research Committee at University of Macau under Grants MYRG2014-00003-FST, MRG017/ZYC/2014/FST, MYRG113(Y1-L3)-FST12-ZYC and MRG001/ZYC/2013/FST, and the Funda-

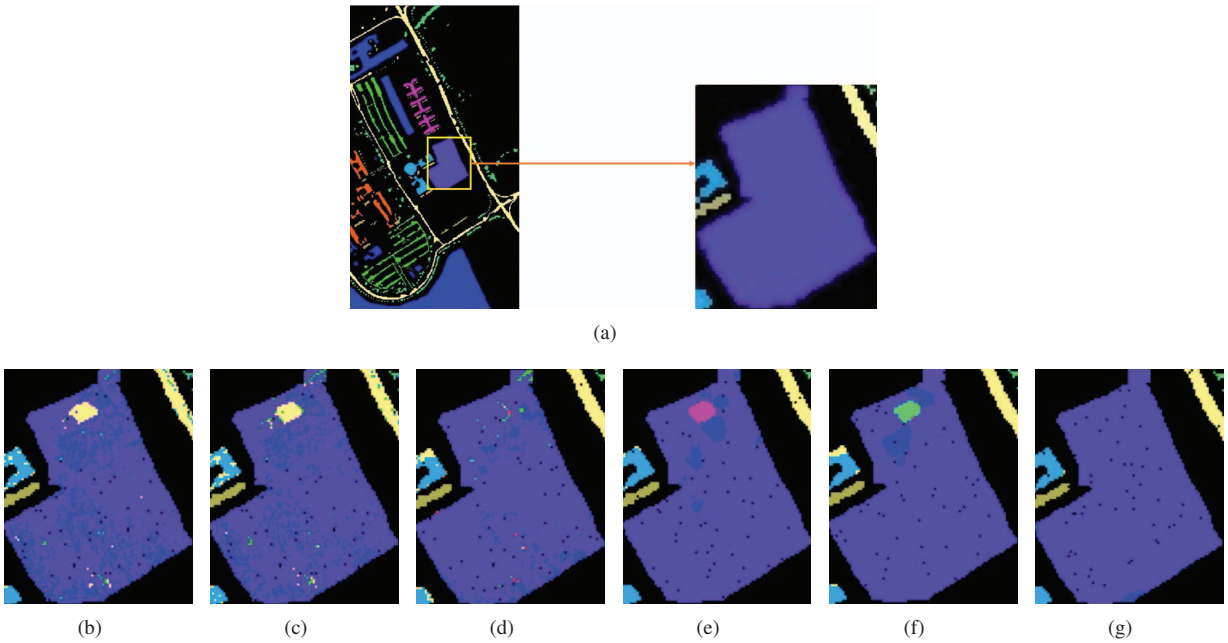


Fig. 6. Magnified parts of different classification maps for the University of Pavia data set using different methods. (a)Ground truth; (b) SVM; (c) KELM; (d) PCA+Gabor;(e) EPF; (f) MPM-LBP; (g) STSL.

mental Research Funds for the Central Universities under Grants CCNU14A05023 and CCNU16A05022.

REFERENCES

[1] F. Zhang, B. Du, and L. Zhang, "Saliency-guided unsupervised feature learning for scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 4, pp. 2175–2184, 2015.

[2] Y. Zhou, J. Peng, and C. L. P. Chen, "Dimension reduction using spatial and spectral regularized local discriminant embedding for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 2, pp. 1082–1095, 2015.

[3] P. Ghamisi, J. A. Benediktsson, and M. O. Ulfarsson, "Spectral-spatial classification of hyperspectral images based on hidden markov random fields," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 5, pp. 2565–2574, 2014.

[4] Y. Wei, G. Xiao, H. Deng, H. Chen, M. Tong, G. Zhao, and Q. Liu, "Hyperspectral image classification using fpca-based kernel extreme learning machine," *Optik-International Journal for Light and Electron Optics*, vol. 126, no. 23, pp. 3942–3948, 2015.

[5] S. Bernabe, P. R. Marpu, A. Plaza, M. D. Mura, and J. A. Benediktsson, "Spectral-spatial classification of multispectral images using kernel feature space representation," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 1, pp. 288–292, 2014.

[6] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral-spatial hyperspectral image classification via multiscale adaptive sparse representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 12, pp. 7738–7749, 2014.

[7] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proceedings of the IEEE*, vol. 101, no. 3, pp. 652–675, 2013.

[8] O. Rajadell, P. Garcia-Sevilla, and F. Pla, "Spectral-spatial pixel characterization using gabor filters for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 4, pp. 860–864, 2013.

[9] Q. Shi, L. Zhang, and B. Du, "Semisupervised discriminative locally enhanced alignment for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 9, pp. 4800–4815, 2013.

[10] Y. L. Chang, J. N. Liu, C. C. Han, and Y. N. Chen, "Hyperspectral image classification using nearest feature line embedding approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 1, pp. 278–287, 2013.

[11] H. Y. Huang and B. C. Kuo, "Double nearest proportion feature extraction for hyperspectral-image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 11, pp. 4034–4046, 2010.

[12] S. T. Tu, J. Y. Chen, W. Yang, and H. Sun, "Laplacian eigenmaps-based polarimetric dimensionality reduction for sar image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 1, pp. 170–179, 2011.

[13] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-preserving dimensionality reduction and classification for hyperspectral image analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 4, pp. 1185–1198, 2012.

[14] R. B. Luo, W. Z. Liao, and Y. G. Pi, "Discriminative supervised neighborhood preserving embedding feature extraction for hyperspectral-image classification," *Telkomnika*, vol. 10, no. 5, pp. 1051–1056, 2012.

[15] T. Han and D. G. Goodenough, "Nonlinear feature extraction of hyperspectral data based on locally linear embedding (lle)," *Proceedings of Geoscience and Remote Sensing Symposium*, vol. 2, pp. 1237–1240, 2005.

[16] Z. Wang and B. He, "Locality perserving projections algorithm for hyperspectral image dimensionality reduction," in *Proceeding of the 19th International Conference on Geoinformatics*, China, 2010, pp. 1–4.

[17] R. Ji, Y. Gao, R. Hong, Q. Liu, D. Tao, and X. Li, "Spectral-spatial constraint hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 3, pp. 1811–1824, 2014.

[18] M. Fauvel, J. Benediktsson, J. Chanussot, and J. Sveinsson, "Spectral and spatial classification of hyperspectral data using svms and morphological profiles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 10, no. 4, pp. 1688–1697, 2008.

[19] C. Chen, W. Li, H. Su, and K. Liu, "Spectral-spatial classification of hyperspectral image based on kernel extreme learning machine," *Remote Sensing*, vol. 6, no. 6, pp. 5795–5814, 2014.

[20] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 43, no. 3, pp. 480–491, 2005.

[21] P. Quesada-Barriuso, F. Arguello, and D. B. Heras, "Spectral–spatial classification of hyperspectral images using wavelets and extended

morphological profiles," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 7, no. 4, pp. 1177–1185, 2014.

[22] X. Kang, S. Li, and J. A. Benediktsson, "Spectral–spatial hyperspectral image classification with edge-preserving filtering," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 5, pp. 2666–2677, 2014.

[23] A. Soltani-Farani, H. R. Rabiee, and S. A. Hosseini, "Spatial-aware dictionary learning for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, DOI:10.1109/TGRS.2014.2325067, 2014.

[24] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral–spatial classification of hyperspectral data using loopy belief propagation and active learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 2, pp. 844–856, 2013.

[25] J. Li, X. Huang, P. Gamba, J. M. Bioucas-Dias, L. Zhang, J. Atli Benediktsson, and A. Plaza, "Multiple feature learning for hyperspectral image classification," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 53, no. 3, pp. 1592–1606, 2015.

[26] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2094–2107, 2014.

[27] G. E. Hinton, S. Osindero, and Y. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[28] Y. Tang, T. Xia, Y. Wei, H. Li, and L. Li, "Hierarchical kernel-based rotation and scale invariant similarity," *Pattern Recognition*, vol. 47, no. 4, pp. 1674–1688, 2014.

[29] H. Li, Y. Wei, L. Li, and C. P. Chen, "Hierarchical feature extraction with local neural response for image recognition," *IEEE Transactions on Cybernetics*, vol. 43, no. 2, pp. 412–424, 2013.

[30] R. Salakhutdinov, J. B. Tenenbaum, and A. Torralba, "Learning with hierarchical-deep models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1958–1971, 2013.

[31] G. Hinton and R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[32] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*. Curran Associates, Inc., 2012, pp. 1097–1105.

[33] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning hierarchical features for scene labeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1915–1929, 2013.

[34] B. Chen, G. Polatkan, G. Sapiro, D. Blei, D. Dunson, and L. Carin, "Deep learning with hierarchical convolutional factor analysis," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1887–1901, 2013.

[35] T. Li, J. Zhang, and Y. Zhang, "Classification of hyperspectral image based on deep belief networks," in *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2014, pp. 5132–5136.

[36] Y. Zhou and Y. Wei, "Learning hierarchical spectral-spatial features for hyperspectral image classification," 2015. In press.

[37] Y. Bazi and F. Melgani, "Toward an optimal svm classification system for hyperspectral remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 11, pp. 3374–3385, 2006.

[38] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: a review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 3, pp. 247–259, 2011.

[39] B. Demir and S. Ertürk, "Hyperspectral image classification using relevance vector machines," *Geoscience and Remote Sensing Letters*, vol. 4, no. 4, pp. 586–590, 2007.

[40] G. B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 42, no. 2, pp. 513–529, 2012.

[41] R. Zhang, Y. Lan, G.-B. Huang, Z.-B. Xu, and Y. C. Soh, "Dynamic extreme learning machine and its approximation capability," *IEEE Transactions on Cybernetics*, vol. 43, no. 6, pp. 2054–2065, 2013.

[42] M. Pal, A. E. Maxwell, and T. A. Warner, "Kernel-based extreme learning machine for remote-sensing image classification," *Remote Sensing Letters*, vol. 4, no. 9, pp. 853–862, 2013.

[43] L. Bruzzone and L. Carlin, "A multilevel context-based system for classification of very high spatial resolution images," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 44, no. 9, pp. 2587–2600, 2006.

[44] X. Ma, J. Geng, and H. Wang, "Hyperspectral image classification via contextual deep learning," *EURASIP Journal on Image and Video Processing*, vol. 2015, no. 1, pp. 1–12, 2015.

[45] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," *Journal of Sensors*, vol. 2015, pp. 1–12, 2015.

[46] V. Slavkovikj, S. Verstockt, W. De Neve, S. Van Hoecke, and R. Van de Walle, "Hyperspectral image classification with convolutional neural networks," in *Proceedings of the 23rd Annual ACM Conference on Multimedia Conference*. ACM, 2015, pp. 1159–1162.

[47] G. B. Huang, D. H. Wang, and Y. Lan, "Extreme learning machines: a survey," *International Journal of Machine Learning and Cybernetics*, vol. 2, pp. 107–122, 2011.

[48] S. Yan, D. Xu, B. Zhang, H. J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: a general framework for dimensionality reduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, pp. 40–51, 2007.

[49] M. Marconcini, G. Camps-Valls, and L. Bruzzone, "A composite semisupervised svm for classification of hyperspectral images," *Geoscience and Remote Sensing Letters, IEEE*, vol. 6, no. 2, pp. 234–238, 2009.

[50] M. Pal, A. E. Maxwell, and T. A. Warner, "Kernel-based extreme learning machine for remote-sensing image classification," *Remote Sensing Letters*, vol. 4, no. 9, pp. 853–862, 2013.

[51] H. Li, G. Xiao, T. Xia, Y. Tang, and L. Li, "Hyperspectral image classification using functional data analysis," *IEEE transactions on Cybernetics*, vol. 44, no. 9, pp. 1544–1555, 2014.

[52] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectralcspatial classification of hyperspectral data using loopy belief propagation and active learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 2, pp. 844–856, 2013.

[53] I. Dópido, J. Li, P. R. Marpu, A. Plaza, J. M. Bioucas-Dias, and J. A. Benediktsson, "Semi-supervised self-learning for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 7, pp. 4032–4044, 2013.