Shape-Adaptive Tensor Factorization Model for Dimensionality Reduction of Hyperspectral Images

ZHAOHUI XUE[ID], SIRUI YANG, AND MENGXUE ZHANG

School of Earth Sciences and Engineering, Hohai University, Nanjing 211100, China

Corresponding author: Zhaohui Xue (zhaohui.xue@hhu.edu.cn)

ABSTRACT Tensor-based dimensionality reduction (DR) of hyperspectral images is a promising research topic. However, patch-based tensorization usually adopts a squared neighborhood with fixed window size, which may be inaccurate in modeling the local spatial information in a hyperspectral image scene. In this work, we propose a novel *shape-adaptive tensor factorization (SATF)* model for dimensionality reduction and classification of hyperspectral images. Firstly, shape-adaptive patch features are extracted to build fourth-order tensors. Secondly, multilinear singular value decomposition (MLSVD) is adopted for tensor factorization and latent features are extracted via mode-i tensor-matrix product. Finally, classification is conducted by using a sparse multinomial logistic regression (SMLR) model. Experimental results, conducted with two popular hyperspectral data sets collected over the Indian Pines and the University of Pavia, respectively, indicate that the proposed method outperforms the other traditional and tensor-based DR methods.

INDEX TERMS Dimensionality reduction (DR), hyperspectral image (HSI), multilinear singular value decomposition (MLSVD), shape-adaptive (SA), tensor.

I. INTRODUCTION

Hyperspectral remote sensing sensors are capable of providing land cover images with unified spectral-spatial information, which has been motivating rapid developments for hyperspectral image (HSI) processing techniques [1], [2]. Among many processing tasks, HSI classification has attracted plenty of attention in the last decades [3]–[9]. The curse of dimensionality has posed great challenges for HSI classification since there is a high correlation between adjacent bands and the dimension of spectral features may be too high for classification purpose [10]. Dimensionality reduction (DR) methods have been commonly used to address this issue, and various DR approaches have been proposed in the literature [11]–[13].

DR aims at finding the intrinsic (or representative) low-dimensional features (or bands) spanning in the spectral domain. Traditional popular DR methods are matrix-based, including principal component analysis (PCA) [14], local linear embedding (LLE) [15], isometric feature mapping

(ISOMAP) [16], locality preserving projections (LPP) [17], linear graph embedding (LGE) [18], linear discriminant analysis (LDA) and its variants [19], nearest feature line embedding (NFLE) [20], sparse graph embedding (SGE) and its variants [21], [22], ensemble discriminative local metric learning (EDLML) [23], spatial-spectral hypergraph discriminant analysis (SSHGDA) [24], simultaneous spectral-spatial feature selection and extraction algorithm [25], ensemble manifold regularized sparse low-rank approximation (EMR-SLRA) algorithm [26], etc. However, vector- or matrix-based representation destroys the inherent spatial and spectral structure of HSI which can offer a physical interpretation of how spatial information and spectral bands contribute to the classification outcome [27].

Recently, tensor-based representation models are promising alternatives for dimensionality reduction of HSI [28]–[43]. Those methods can be roughly divided according to their mathematical formulations as follows.

1) Lower rank tensor approximation, a high-order extension of PCA, utilizes the global spectral and spatial correlation, respectively, to project HSI into a

---

The associate editor coordinating the review of this article and approving it for publication was Lefei Zhang.

low-rank space. For example, Renard et al. [28], [29] proposed a lower rank tensor approximation (LRTA) model which treats HSI as a third-order tensor with one spectral mode and two spatial modes, and performs a DR of the spectral mode and a projection onto a lower-dimensional subspace of the two spatial modes. Focusing on improving DR performance via spatial denosing, Bourennane et al. [30] proposed an adaptive multidimensional Wiener filtering for jointly performing spatial filtering for denoising and conducting LRTA for spectral DR. Considering the nonlocal spatial correlation of HSI, An et al. [35] proposed a group-based tensor model which segments the original 3-D HSI into blocks, then groups the blocks into clusters, and finally projects the clusters into low-rank space using LRTA to obtain a low-dimensionality feature space. Later, Velasco-Forero and Angulo [32] adopts additive morphological decomposition to build a fourth-order tensor and applied tensor PCA (TPCA) for DR. Similarly, Ren et al. [37] extended the traditional PCA to TPCA for HSI spectral-spatial feature extraction.

2) Patch alignment, represents the spatial information between local tensor samples and achieves global optimum. For example, Zhang et al. [31] proposed a tensor discriminative locality alignment (TDLA) for HSI spectral-spatial feature extraction and DR. Gao et al. [34] built a class-aware tensor neighborhood graph and adopted patch alignment for DR. Liu et al. [38] used the Gabor filter banks to extract spectral-spatial features and further introducing ℓ_1 and ℓ_2 norms into TDLA for DR.

3) Local tensor discriminant analysis, a tensor extension of local discriminant analysis. For example, Zhong et al. [33] applied spectral-spatial feature extraction methods to build third-order tensors and adopted a local tensor discriminant analysis for DR.

4) Tensor sparse and low-rank graph-based discriminant analysis, where the information from three perspectives (tensor representation, sparse and low-rank representation, and graph theory) is exploited to present the data structure for HSI. By regarding the hyperspectral data cube as a third-order tensor, Pan et al. [36] proposed tensor sparse and low-rank graph-based discriminant analysis framework which extracts small local patches centered at the training samples to maintain the structural information, resulting in a more discriminative graph. Further, considering the limited labeled samples in real application, An et al. [40] proposed a patch tensor-based sparse and low-rank graph method where the sparsity and low-rankness properties are jointly considered to capture the local and global intrinsic structures. For the objective of jointly exploiting intrinsic structure information and enhancing the discriminant ability, An et al. [41] also proposed a novel tensor-based low-rank graph with multimanifold regularization method, where a low-rank constraint is employed to preserve the global data structure while multimanifold information is utilized to enhance the discriminant ability, and tensor representation is used to preserve the spatial neighborhood information. To exploit the spatial correlations between pixels in different patches, Deng et al. [42] established a tensor low-rank discriminant embedding model which employs the tensor representation to preserve the intrinsic geometrical structure, uses low-rank reconstruction to uncover the potential relationship among the data points, and combines label information to enhance the discriminability of features. To better handle the gross errors (i.e., outliers), Sun et al. [43] proposed a lateral-slice sparse tensor robust PCA method which assumes that a third-order hyperspectral tensor has a low-rank structure, and gross errors or outliers are sparsely scattered in a 2-D space (i.e., lateral-slice) of the tensor, and DR is performed by formulating a low-rank and sparse tensor decomposition problem. Zhang et al. [44] proposed a unified low-rank matrix factorization to jointly perform the dimensionality reduction and data clustering for unsupervised HSI classification.

5) Tensor locality preserving projection, a tensor extension of LPP. For example, Deng et al. [39] used region covariance descriptor to build third-order tensors and adopted a tensor locality preserving projection (TLPP) for DR.

In summary, LRTA model paved the way for tensor-based DR in this community. Integrating advanced spectral-spatial feature extraction with LRTA model is the future research trend. Patch alignment, local tensor discriminant analysis, and tensor locality preserving projection obtain little attention recently. However, tensor sparse and low-rank graph-based DR attracted a lot of attention with the recent great progress in sparse representation, low-rank representation, and graph embedding. Although elegant dimensionality reduction and classification performances have been observed, most of those methods adopted squared-neighborhood with fixed size to generate the patch features when building high-order tensors for HSI. Traditional squared-neighborhood with fixed size can not accurately model the local spatial relationships among various objects characterized with different size and shape in real HSI scene.

In this work, for the first time, we propose a novel shape-adaptive tensor factorization (SATF) model for dimensionality reduction and classification of hyperspectral images. In the method, shape-adaptive patch features are extracted to build a fourth-order tensor which may be more suitable for modeling the complex spectral-spatial data structure of HSI compared to third-order or matrix. Tensor decomposition is performed via a multilinear singular value decomposition (MLSVD) model. The low-dimensional latent features are extracted inspired from the LRTA model by mode-i tensor-matrix product. A sparse multinomial logistic regression (SMLR) model is used for the subsequent classification. Experimental results demonstrate the good performance
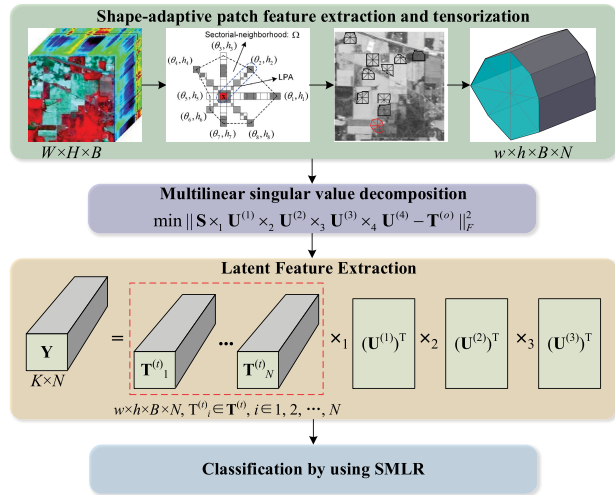
FIGURE 1. Flowchart of the proposed shape-adaptive tensor factorization model for dimensionality reduction and classification of hyperspectral images.

of the proposed method compared to other traditional and tensor-based dimensionality reduction methods. Note that our previous study has validated the good performance of using shape-adaptive-based neighborhood modeling method for sparse graph embedding based DR of HSI [22]. To the best of our knowledge, none of previous studies exploited shape-adaptive window for modeling neighborhood information in tensor-based DR methods for HSI.

II. PROPOSED METHOD

Let $X = [x_1, \ldots, x_N] \in \mathbb{R}^{W \times H \times B}$ be a hyperspectral image data cube with a B-dimensional signal for each pixel $\mathbf{x}_i = [x_1, \ldots, x_B]^T$, $i \in 1, \ldots, N$ ($N = W \times H$). Let $T \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_m}$ be a m-order tensor. Let $Y = [y_1, \ldots, y_N] \in \mathbb{R}^{K \times N}$ ($K \ll B$) be the latent features extracted from $\mathbf{X}$ with reduced dimension. The proposed method consists of three major steps: 1) shape-adaptive patch feature extraction and fourth-order tensorization; 2) multilinear singular value decomposition for tensor factorization and latent feature extraction via mode-i tensor-matrix product; 3) classification by using SMLR. A graphical illustration of the proposed method is shown in Fig. 1.

A. SHAPE-ADAPTIVE TENSORIZATION

In shape-adaptive method, an anisotropic local polynomial approximation (LPA)-intersection of confidence intervals (ICI), termed Anisotropic LPA-ICI, is used to provide multidirectional sectorial-neighborhood with adaptive size for each sector surrounding the central pixel [45]. A graphical illustration of the Anisotropic LPA-ICI method used to build the shape-adaptive neighborhoods (Ω) can be found in the top of Fig. 1.

A fast implementation of the Anisotropic LPA-ICI method was adopted with a set of candidate scales H (e.g., {1, 2, 3, 4, 5, 6, 7}) for eight directions θ_k (45° interval and $k = 1, 2, \ldots, 8$). First, a set of directional varying-scale estimates

are obtained by applying directional-LPA convolution kernels ($\{g_{h,\theta_k}\}_{h \in H}$) to a single-band image (i.e., the PC_1 of HSI), which is shown as follows

$$\widehat{X}^{PC_1}_{h,\theta_k} = X^{PC_1} \otimes g_{h,\theta_k}. \tag{1}$$

Then, these estimates are compared according to the ICI rule. As a result, an adaptive scale $h^+(X_p, \theta_k) \in H$ is defined for each image patch X_p (an initial squared window with length $2H_{\max} - 1$), and the adaptive neighborhood $U^+_{X_p}$ is finally obtained as

$$U^+_{X_p} = \bigcup_{k=1}^{K} \text{supp} g_{h^+(X_p, \theta_k), \theta_k}. \tag{2}$$

The shape-adaptive patch features can then be extracted based on the neighborhood $U^+_{X_p}$ built for each pixel. We model the extracted shape-adaptive patch features as fourth-order tensors, i.e., $T \in \mathbb{R}^{w \times h \times B \times N}$, where $w \times h$ is the initial window size of SA. Usually, the initial window size is $(2H_{\max} - 1) \times (2H_{\max} - 1)$, and those missing pixels for each tensor after SA are padded with zeros to uniform the shapes between different tensors.

B. TENSOR FACTORIZATION AND LATENT FEATURE EXTRACTION

In the factorization step, only the training samples (a total of $N^{(o)}$) were used to build shape-adaptive tensor $T^{(o)} \in \mathbb{R}^{w \times h \times B \times N^{(o)}}$, and it can be factorized by using multilinear singular value decomposition (MLSVD) [46] which factorizes the tensor as the multilinear tensor-matrix product of a core tensor $\mathbf{S}$ with four factors $U^{(i)}$

$$T^{(o)} \approx S \times_1 U^{(1)} \times_2 U^{(2)} \times_3 U^{(3)} \times_4 U^{(4)}, \tag{3}$$

where $S \in \mathbb{R}^{R_1 \times R_2 \times R_3 \times R_4}$, $U^{(i)} \in \mathbb{R}^{I_i \times R_i}$, $I_i \in \{w, h, B, N\}$, and $i \in \{1, 2, 3, 4\}$. The symbol "$\times_i$" denotes the i-mode product of tensor $\mathbf{T}$ with a matrix $\mathbf{U}$ along the mode-i. Equation (3) is transformed into a nonlinear non-convex ℓ_2-optimization problem which is iteratively solved by truncated higher-order singular value decomposition.

In the feature projection step, the latent features can be obtained by mode-i tensor-matrix product of test tensor $T^{(t)} \in \mathbb{R}^{w \times h \times B \times N^{(t)}}$ with the four factors $U^{(i)}$ calculated from (3)

$$Y = T^{(t)} \times_1 (U^{(1)})^T \times_2 (U^{(2)})^T \times_3 (U^{(3)})^T, \tag{4}$$

where $Y \in \mathbb{R}^{R_1 \times R_2 \times R_3 \times N^{(t)}}$.

The extracted latent features $\mathbf{Y}$ are rearranged back into matrix representations with dimension of $K \times N^{(t)}$, where $K = R_1 \times R_2 \times R_3$. For the objective of dimensionality reduction, the rank-one term $[R_1 \ R_2 \ R_3]$ is set to $[1 \ 1 \ K]$. In this setting, the first two ranks representing the spatial neighborhood domain are shrunk to one dimension for aggregating the neighborhood information, whereas the third rank representing the spectral domain is set to K for maintaining the spectral information. Here, K determines the reduced dimension.

C. CLASSIFICATION

In the last stage, the latent features are then embedded into a sparse multinomial logistic regression (SMLR) [47] model for training and prediction. We adopt the multinomial logistic regression via a variable splitting and augmented Lagrangian (LORSAL) [48] algorithm to optimize the model since it has yielded efficient and powerful performances for HSI classification in recent years [22], [49]–[53]. Algorithm 1 summarizes the proposed shape-adaptive tensor factorization (SATF) model for dimensionality reduction and classification of hyperspectral images.

Algorithm 1 Shape-adaptive tensor factorization model for hyperspectral image classification.

1: **Input:** Training set $X^{(o)}$, test set $X^{(t)}$, K, window size $(w \times h)$
2: **Output:** **Y** and classification map (CMap)
3: Shape-adaptive patch features extraction using (1)-(2).
4: Tensorization: $T^{(o)}$ and $T^{(t)}$.
5: Multilinear singular value decomposition using (3).
6: Latent feature extraction using (4).
7: Classification using SMLR optimized by LORSAL based on the features **Y**: CMap = LORSAL(**Y**)

III. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we evaluate the proposed method by using two popular hyperspectral data sets [1] collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) and the Reflective Optics Spectrographic Imaging System (ROSIS) instruments.

A. HYPERSPECTRAL DATA SETS

1) The first hyperspectral image was acquired by the AVIRIS sensor over the Indian Pines region in North-western Indiana in 1992. The image size in pixel is 145×145, with moderate spatial resolution of 20 m. The number of data channels in the acquired image is 220 (with spectral range from 0.4 to 2.5 μm). A total of 200 radiance channels are used in the experiments by removing several noisy and water absorbed bands. A three-band false color composite image and the ground-truth map are shown in Fig. 2. A total of 10366 samples containing 16 classes are available.

2) The second hyperspectral image was acquired by the ROSIS sensor over the urban area of the University of Pavia, Italy. The image size in pixel is 610×340, with very high spatial resolution of 1.3 m. The number of data channels in the acquired image is 103 (with spectral range from 0.43 to 0.86 μm). A three-band false color composite image and the ground-truth map are

[1]Hyperspectral Remote Sensing Scenes: http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes

(a)   (b)

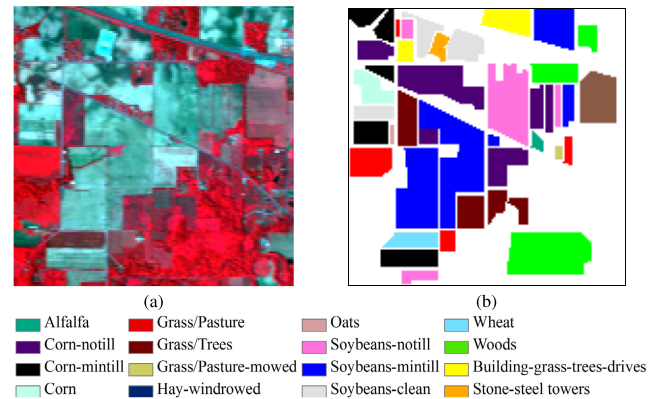|		
■ Alfalfa	■ Grass/Pasture	■ Oats
■ Corn-notill	■ Grass/Trees	■ Soybeans-notill
■ Corn-mintill	■ Grass/Pasture-mowed	■ Soybeans-mintill
■ Corn	■ Hay-windrowed	■ Soybeans-clean
■ Wheat	■ Woods	■ Building-grass-trees-drives
■ Stone-steel towers		

FIGURE 2. AVIRIS indian pines data set. (a) False color composite image (R: 50, G: 27, B: 17). (b) Ground-truth map containing 16 mutually exclusive land-cover classes.
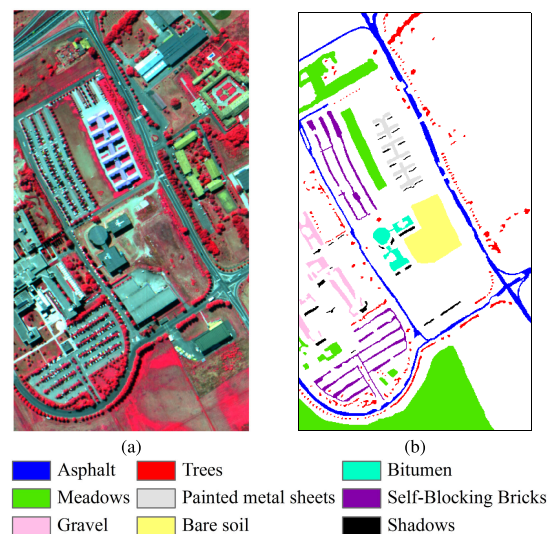

(a)   (b)

■ Asphalt	■ Trees	■ Bitumen
■ Meadows	■ Painted metal sheets	■ Self-Blocking Bricks
■ Gravel	■ Bare soil	■ Shadows

FIGURE 3. ROSIS university of pavia data set. (a) False color composite image (R: 102, G: 56, B: 31). (b) Ground-truth map containing 9 mutually exclusive land-cover classes.

shown in Fig. 3. A total of 42776 samples containing 9 classes are available.

B. EXPERIMENTAL SETTINGS

The corresponding parameter settings and notations adopted in our experiments are:

- For SATF, the candidate scales H range from $\{1, 2\}$, $\{1, \ldots, 3\}, \ldots, \{1, \ldots, 7\}$, leading to different window size of 3×3, 5×5, ..., 13×13, and the rank-one term is set to $[1 \ 1 \ K]$ where K denotes the reduced number of dimension.
- For other DR methods, we include principal component analysis (PCA) [14], locality preserving projections (LPP) [17], linear graph embedding (LGE) [18], multilinear principal component analysis (MPCA) [54], tensor locality preserving projections (TLPP) [55], and tensor linear graph embedding (TLGE) [56]. Different
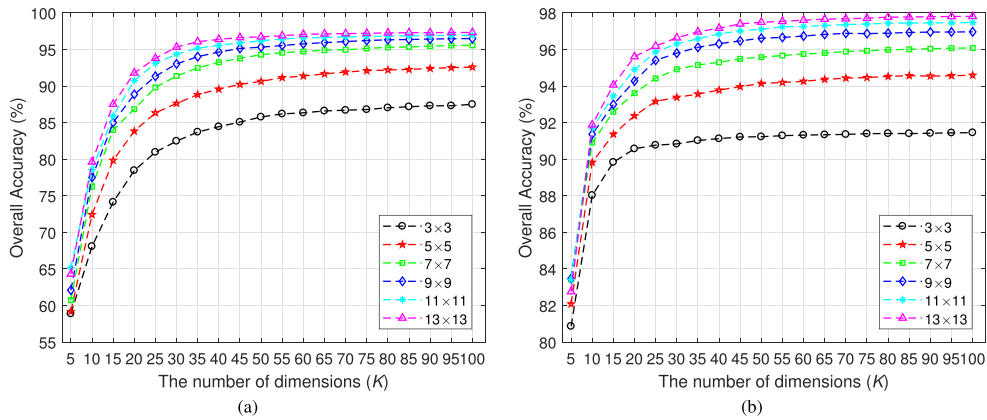
FIGURE 4. The impact of initial window size on classification accuracy obtained by the proposed method (SATF) for (a) AVIRIS Indian Pines and (b) ROSIS University of Pavia data sets. Different curves illustrate the overall accuracies as a function of the number of dimensions under different window size.
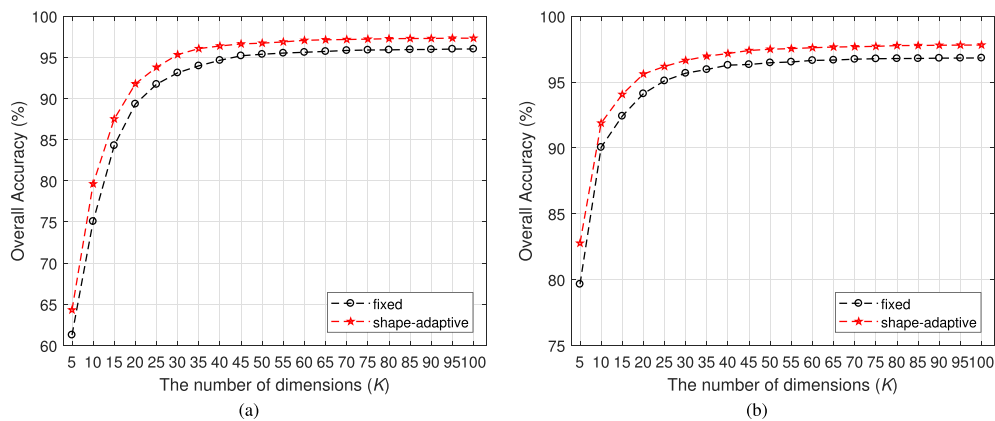

FIGURE 5. The comparison of shape-adaptive and fixed window on classification accuracy obtained by the proposed method for (a) AVIRIS Indian Pines and (b) ROSIS University of Pavia data sets. Different curves illustrate the overall accuracies as a function of the number of dimensions under the same window size of 13 × 13.

DR methods have produced the best performance by carefully optimizing the associate parameters.

- For LORSAL, the regularization parameter is set to 0.001 and the number of iterations is set to 100. The classification results are quantitatively evaluated by measuring the overall accuracy (OA), the average accuracy (AA), the individual class accuracy, and the Kappa statistic (κ). In addition, we also included k-nearest neighbor (k-NN) [57] and support vector machine (SVM) [58] for comparison.
- Finally, it should be noted that all the implementations were carried out using Matlab R2017b in a desktop PC equipped with an Intel Xeon E3 CPU (at 3.4GHz) and 32 GB of RAM.

C. INFLUENCE OF WINDOW SIZE PARAMETER

In the first experiment, we evaluate the impacts of initial window size on classification accuracy obtained by the proposed method. Note that the initial window size is set the same to fixed window size to ensure a fair comparison. As shown in Fig. 4, the classification accuracy increases as

the window size also increases in different cases. This is due to the fact that a larger window size can model more detailed information to represent the central pixel. The OAs become stable when the window size is larger than 11 × 11. Considering the balance between classification accuracy and computational complexity, we experimentally set the window size to 13 × 13. Another observation is that the OAs increase as the dimensions also increase, and the peak values occur when $K \geq 35$ for Indian Pines and $K \geq 45$ for University of Pavia data sets.

D. THE COMPARISON OF SHAPE-ADAPTIVE WITH FIXED WINDOW

In the second experiment, we evaluate the performance improvement of SATF by using shape-adaptive method to generate the patch features. To this end, we compare shape-adaptive window with fixed window. As exhibited in Fig. 5, shape-adaptive-based method significantly improves the classification accuracy obtained by using fixed window, with the increases of OAs are ~2% for Indian Pines and ~1% for University of Pavia data sets. This observation can be
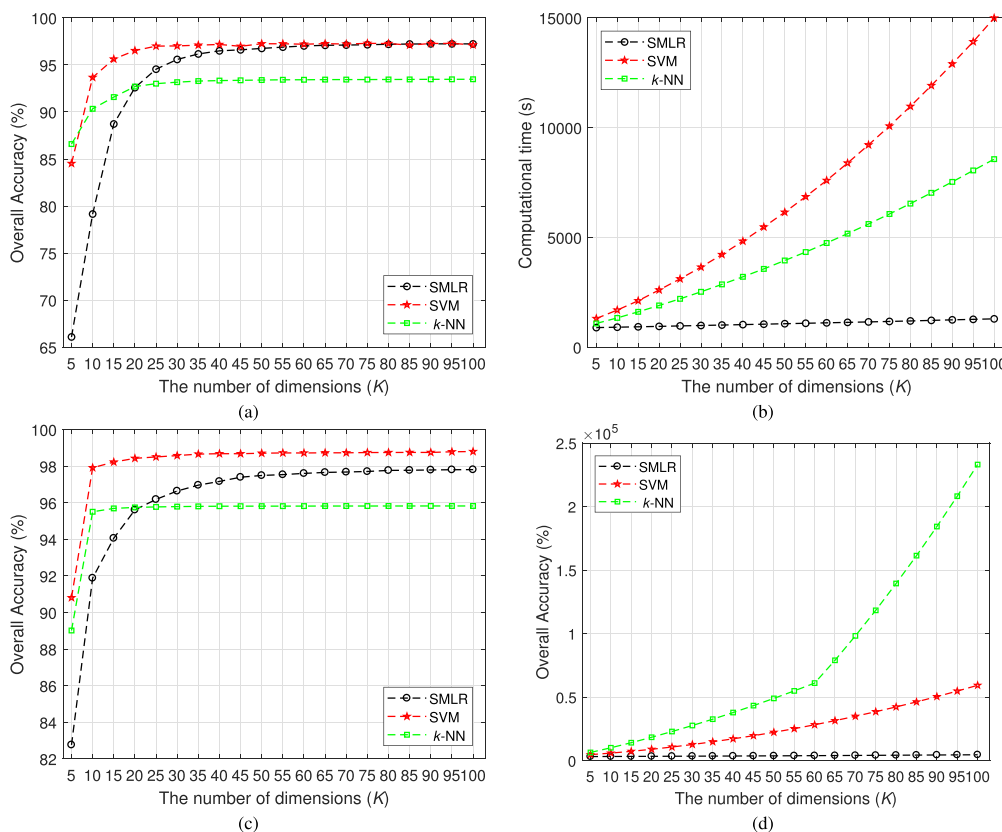
FIGURE 6. The comparison of different classifiers based on the latent features extracted by SATF for AVIRIS Indian Pines [(a) OA and (b) Computational time] and ROSIS University of Pavia [(c) OA and (d) Computational time] data sets. Different curves illustrate the OAs or computational times as a function of the number of dimensions under the same window size of 13 × 13.
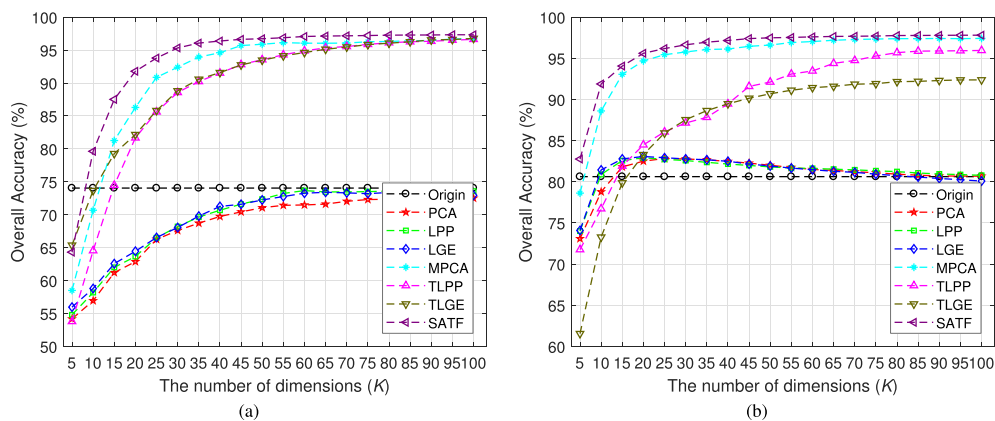

FIGURE 7. The comparison of different DR methods for (a) AVIRIS Indian Pines and (b) ROSIS University of Pavia data sets. Different curves illustrate the overall accuracies as a function of the number of dimensions under the same window size of 13 × 13.

explained that shape-adaptive method can accurately model the spatial information by adaptively adjust the window shape and size, yielding much purer patch features.

E. THE COMPARISON OF DIFFERENT CLASSIFIERS

In this experiment, we evaluate the classification performance of different classifiers based on the extracted latent features. As shown in Fig. 6, SVM produces the best classification

accuracies in different cases for the two data sets, and SMLR is better than k-NN. However, the computational times of SVM and k-NN are significantly high than SMLR. It's worth noting that the computational complexities of SVM and k-NN are exponentially increased with the increase of dimensions, whereas the complexity of SMLR is very low and insensitive to dimension. This observation demonstrates the powerful performance of SMLR.

TABLE 1. Overall (OA), average (AA) and individual class accuracies (%), kappa statistic (κ), standard deviation, and computational time obtained by different methods for the AVIRIS Indian Pines data sets (10% labeled samples used for training, $K = 35$).

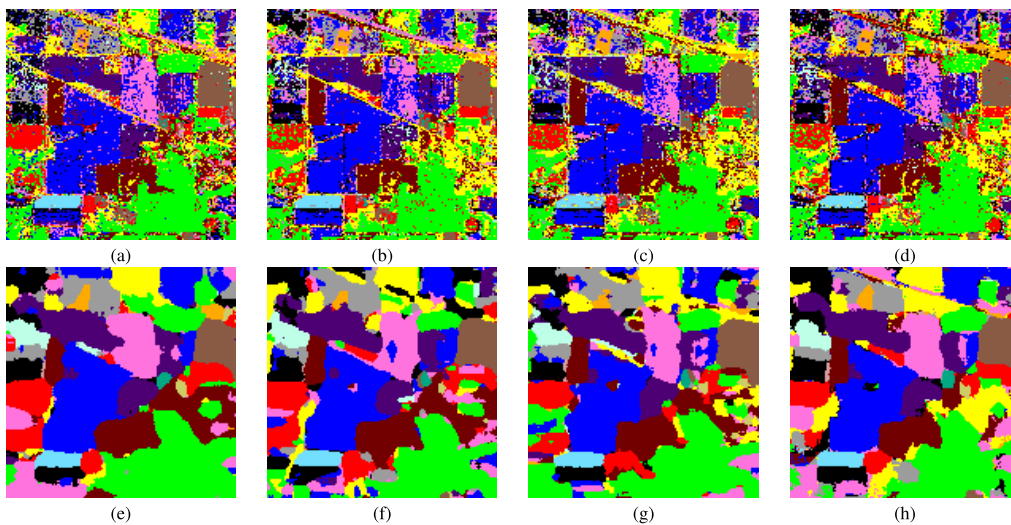| Class	#Samples		Methods							
	Train	Test	Origin	PCA	LPP	LGE	MPCA	TLPP	TLGE	SATF
Alfalfa	5	41	15.61±9.84	27.07±16.80	22.44±10.34	28.54±14.41	83.66±7.63	79.02±10.42	80.24±12.03	**96.10±1.26**
Corn-notill	143	1285	74.88±2.05	65.34±2.44	65.27±3.44	66.40±3.37	90.19±1.90	85.34±3.51	85.90±3.32	**93.35±2.35**
Corn-mintill	83	747	53.13±4.28	44.79±3.55	50.12±2.09	51.46±3.22	89.29±2.11	84.87±4.95	89.29±2.63	**94.19±1.39**
Corn	24	213	28.64±3.84	26.29±3.31	28.36±6.40	34.46±7.84	88.64±6.36	85.73±6.12	86.20±7.59	**94.46±2.24**
Grass-pasture	48	435	83.98±3.25	70.23±3.02	75.15±3.87	74.87±5.18	92.90±2.42	93.01±3.43	93.17±5.28	**93.49±3.04**
Grass-trees	73	657	94.66±3.09	95.19±2.51	94.38±1.64	93.29±1.01	97.20±1.07	99.30±0.57	97.32±1.66	**99.47±0.39**
Grass-pasture-mowed	3	25	24.40±15.37	32.80±16.95	37.20±12.66	30.80±16.66	**95.60±9.13**	72.80±11.12	81.60±18.20	75.20±15.64
Hay-windrowed	48	430	96.09±1.74	98.07±1.55	98.09±1.81	97.53±1.97	98.00±1.50	98.98±2.04	99.30±1.07	**99.95±0.15**
Oats	2	18	5.56±5.86	12.22±5.74	8.89±8.76	11.67±8.86	**59.44±16.37**	15.00±13.37	48.33±15.28	51.67±17.58
Soybeans-notill	97	875	60.67±3.51	47.21±5.36	47.71±3.74	47.57±3.69	89.86±2.16	78.42±5.00	83.07±2.10	**92.10±1.71**
Soybeans-mintill	246	2209	75.95±1.64	75.79±1.62	76.71±1.04	76.12±1.32	96.35±0.87	93.37±2.47	91.44±1.67	**97.35±0.53**
Soybean-clean	59	534	57.23±4.99	39.04±3.58	37.51±4.52	37.02±6.90	89.98±2.54	83.84±5.94	83.30±3.92	**94.61±1.95**
Wheat	21	184	94.46±3.62	93.32±2.98	89.62±5.13	94.51±3.86	98.15±2.21	97.17±1.11	96.03±7.64	**98.64±1.21**
Woods	127	1138	92.89±1.93	91.49±1.95	92.93±1.25	92.10±1.31	99.31±0.47	98.51±0.95	98.20±0.98	**99.74±0.17**
Bldg-grass-tree-drives	39	347	56.77±6.46	44.99±3.72	44.52±3.62	45.71±7.07	97.03±1.90	96.69±3.03	94.55±3.66	**98.33±0.91**
Stone-steel-towers	9	84	62.26±11.64	80.71±3.01	83.93±3.90	82.38±2.62	86.67±5.61	47.02±24.25	59.40±20.36	**92.02±2.10**
Overall accuracy	-	-	74.05±0.66	68.72±1.10	69.64±0.62	69.79±0.86	93.94±0.41	90.22±2.07	90.57±1.32	**96.06±0.42**
Average accuracy	-	-	61.07±1.49	59.04±1.93	59.55±1.26	60.28±1.44	90.77±1.31	81.82±3.12	85.46±3.50	**91.92±1.65**
κ statistic	-	-	0.702±0.01	0.639±0.01	0.650±0.01	0.652±0.01	0.931±0.00	0.888±0.02	0.892±0.02	**0.955±0.00**
Time (Seconds)	-	-	**0.27±0.04**	1.43±0.38	1.33±0.24	1.41±0.28	22.94±5.17	95.32±31.09	84.52±30.98	8.70±0.35

FIGURE 8. Classification maps obtained by different methods for the AVIRIS Indian Pines data sets (10% labeled samples used for training, $K = 35$). (a) Origin (OA = 74.05%), (b) PCA (OA = 68.72%), (c) LPP (OA = 69.64%), (d) LGE (OA = 69.79%), (e) MPCA (OA = 93.94%), (f) TLPP (OA = 90.22%), (g) TLGE (OA = 90.57%), (h) SATF (OA = 96.06%).

F. THE COMPARISON OF SATF WITH OTHER DR METHODS

In this experiment, we compare the proposed method with other traditional and tensor-based DR methods. Fig. 7 plots the evolution of classification accuracies obtained by different DR methods as a function of the number of dimensions under the same window size of 13×13. The proposed method significantly outperforms the other counterparts in different cases, and the obtained OAs are higher than 95% when $K \geq 30$ for the two data sets, which demonstrates that the proposed method can produce more latent and compact features for classification. In addition, MPCA provides competitive performance followed by TLPP and TLGE. Another observation is that tensor-based DR methods are better than the traditional methods, i.e., PCA, LPP, LGE. Generally, the above results validate the superiority of SATF compared to other related DR methods.

G. CLASSIFICATION RESULTS

1) AVIRIS INDIAN PINES DATA SETS

The OAs, AAs, individual class accuracies, κ statistic, standard deviations, and computational time obtained by different methods are reported in Table 1 for the Indian Pines data sets. The proposed method significantly outperforms the other tensor-based methods. SATF obtains an OA of 96.06%, which is 2.12%-5.84% higher than others. As for AA and κ, SATF respectively yields 91.92% and 0.955, which are 1.15%-10.10% and 0.024-0.067 higher compared to others. As for the individual class accuracy, SATF also obtains the highest accuracies for most of the classes (14 of 16 classes). As for computational time, our method only cost 8.7s, which is shorter than the other three tensor-based methods. In addition, tensor-based methods including MPCA, TLPP, TLGE, and SATF obtained higher accuracies than matrix-

TABLE 2. Overall (OA), average (AA) and individual class accuracies (%), kappa statistic (κ), standard deviation, and computational time obtained by different methods for the ROSIS University of Pavia data sets (5% labeled samples used for training, $K = 45$).

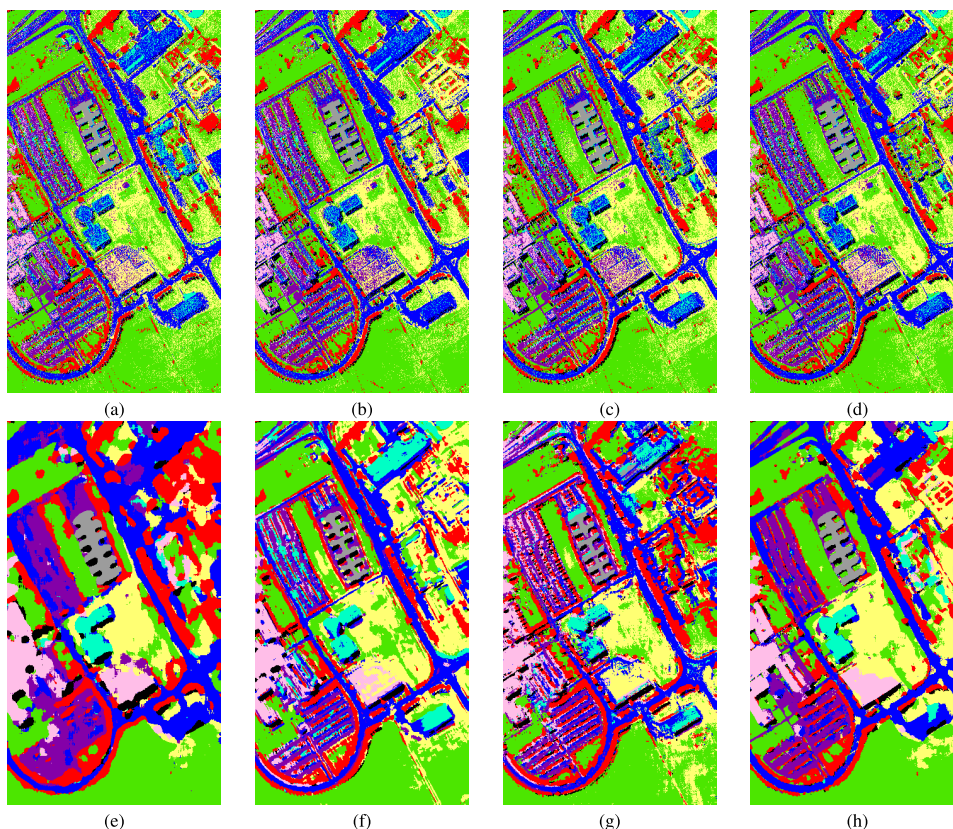| Class	#Samples		Methods							
	Train	Test	Origin	PCA	LPP	LGE	MPCA	TLPP	TLGE	SATF
Asphalt	332	6299	79.22±1.67	82.63±0.82	83.08±0.70	82.64±1.14	95.15±0.85	94.48±0.64	87.24±2.56	**95.96±0.71**
Meadows	932	17717	92.92±0.55	93.28±0.76	93.47±0.72	92.93±1.03	99.62±0.12	96.77±1.20	97.11±0.65	**99.85±0.08**
Gravel	105	1994	62.65±2.41	65.57±1.33	63.80±1.97	64.73±2.20	92.53±2.27	87.00±2.15	75.95±2.00	**94.37±1.80**
Trees	153	2911	87.49±2.29	89.23±0.86	87.99±0.58	88.90±1.34	91.07±1.10	95.67±0.89	95.78±1.09	**96.26±0.96**
Painted metal sheets	67	1278	99.68±0.12	99.77±0.09	99.69±0.10	99.92±0.09	**99.98±0.03**	98.79±1.32	96.57±1.75	99.83±0.14
Bare soil	251	4778	58.59±2.12	58.44±2.75	58.85±1.74	58.87±2.21	**98.47±0.34**	68.83±4.46	81.19±2.55	97.28±0.79
Bitumen	67	1263	37.67±3.48	39.87±3.93	36.71±3.58	40.02±5.32	99.09±0.98	90.85±3.01	82.34±4.12	**99.55±0.37**
Self-Blocking Bricks	184	3498	68.38±2.02	72.00±2.25	69.81±1.70	72.07±2.73	**98.47±1.42**	87.46±1.42	75.59±3.05	97.28±0.79
Shadows	47	900	63.94±6.42	78.48±4.08	77.63±4.60	80.28±5.31	85.01±2.00	92.34±1.89	**92.64±1.40**	85.08±5.08
Overall accuracy	-	-	80.63±0.45	82.27±0.42	81.99±0.35	82.16±0.46	96.46±0.28	91.58±0.89	90.15±0.94	**97.39±0.24**
Average accuracy	-	-	72.28±1.01	75.48±0.56	74.56±0.63	75.60±1.18	94.26±0.43	90.29±0.63	87.16±1.18	**95.56±0.67**
κ statistic	-	-	0.740±0.01	0.762±0.01	0.758±0.00	0.761±0.01	0.953±0.00	0.887±0.01	0.869±0.01	**0.965±0.00**
Time (Seconds)	-	-	**0.81±0.07**	6.01±0.45	6.34±0.75	6.17±0.33	197.48±32.26	342.86±68.86	442.85±141.13	36.54±1.02

FIGURE 9. Classification maps obtained by different methods for the ROSIS University of Pavia data sets (5% labeled samples used for training, $K = 45$). (a) Origin (OA = 80.63%), (b) PCA (OA = 82.27%), (c) LPP (OA = 81.99%), (d) LGE (OA = 82.16%), (e) MPCA (OA = 96.46%), (f) TLPP (OA = 91.58%), (g) TLGE (OA = 90.15%), (h) SATF (OA = 97.39%).

based methods, which validates the effectiveness of tensor factorization for hyperspectral image feature extraction and classification. The classification maps can be visually inspected from Fig. 8, where the tensor-based methods produce more smooth and accurate results. Whereas, the traditional matrix-based methods obtain noisy results. This is due to the fact that we use patch features to represent tensors. This tensorization technique considers each pixel as a third-order tensor where the spectral and spatial information are jointly exploited by using tensor factorization for feature extraction. Therefore, the within-class regions are smooth while the between-class regions have clear boundaries.

2) ROSIS UNIVERSITY OF PAVIA DATA SETS

Table 2 reports the details of classification accuracies obtained by different methods for the University of Pavia data sets. Again, the proposed method significantly outperforms the other tensor-based methods in this scene with the highest OA of 97.39%, which is 0.93%-7.24% higher than other tensor-based methods. In addition, SATF obtains 95.56% and 0.965 respectively for AA and κ, which are 1.30%-8.40% and 0.012-0.096 higher than others. As for the individual class accuracy, SATF also obtains the highest accuracies for six classes among the total of nine classes. As for computational time, our method only costs 36s, which
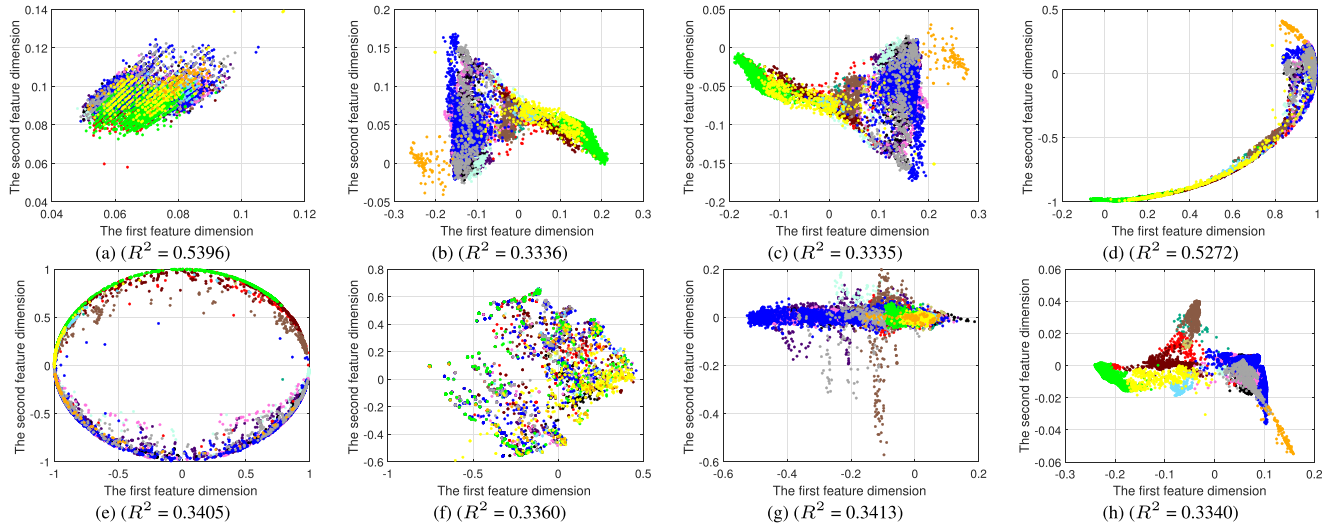
FIGURE 10. Scattering plots considering the first two bands of (a) origin and the first two features derived from (b) PCA, (c) LPP, (d) LGE, (e) MPCA, (f) TLPP, (g) TLGE, and (h) SATF using all the ground-truth data belonging to different classes in the AVIRIS Indian Pines scene. Different colors represent different classes and the correlation coefficient (R^2) in each case is reported inside parentheses.
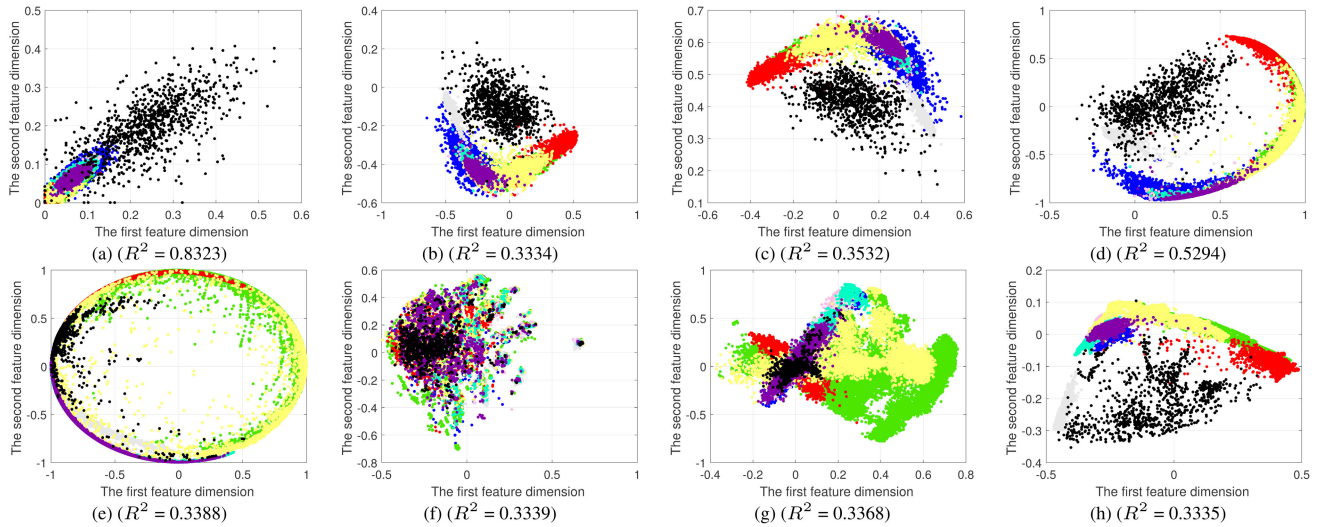

FIGURE 11. Scattering plots considering the first two bands of (a) origin and the first two features derived from (b) PCA, (c) LPP, (d) LGE, (e) MPCA, (f) TLPP, (g) TLGE, and (h) SATF using all the ground-truth data belonging to different classes in the ROSIS University of Pavia scene. Different colors represent different classes and the correlation coefficient (R^2) in each case is reported inside parentheses.

is very fast considering the quite long time of the other three tensor-based methods, e.g., TLGE costs more than 400s in this scene. In this scene, the tensor-based methods are also superior to matrix-based methods. Fig. 9 visually exhibits the classification maps. According to the results, the tensor-based methods produce more smooth and accurate results, whereas the traditional matrix-based methods obtain noisy results. The above observations validated the effectiveness of the proposed method for hyperspectral feature extraction and classification.

H. ANALYSIS OF FEATURE SEPARABILITY
Finally, we analyzed the separability of the latent features extracted by different DR methods. To this end, we compare

the scattering plots considering the first two bands or features in Fig. 10 and Fig. 11, respectively for the Indian Pines and University of Pavia data sets. We use all the labeled samples from the ground-truth data to plot the figures, and different colors represent different classes. As shown in these two figures, different pixels are highly mixed and hardly to be separated in the spectral domain. On the other hand, in the low-dimensional domain, the separability is greatly enhanced. However, it is hard to rank the separability level for different methods by visual inspection. We then conduct a quantitative evaluation of the feature separability by calculating the correlation coefficient (R^2) for different methods. As we can see from the results, the correlations are greatly reduced by DR methods, and the proposed method obtains the lowest correlation, i.e., $R^2 = 0.3340$ for Indian Pines and

$R^2 = 0.3335$ for University of Pavia data sets, illustrating a good separability of the derived features.

IV. CONCLUSIONS

In this paper, we focus on the limitations of current tensor-based DR methods by presenting a novel shape-adaptive tensor factorization model for HSI classification. In particular, the method builds fourth-order tensor based on spectral-spatial features modeled by shape-adaptive neighborhood, yields more latent features via MLSVD, and conducts classification by using SMLR. On the above analysis of the experimental results based on the two real data sets, we can conclude that shape-adaptive neighborhood modeling significantly improves the performance of using traditional fixed window in tensor-based DR, and the proposed method outperforms traditional matrix-based and some tensor-based DR methods in terms of feature separability and classification accuracy.

Although our experimental results are encouraging, further work on additional scenes and comparison methods should be conducted in future. In our work, for the first time, we have introduced shape-adaptive tensor factorization model in the literature. Our next work will focus on adapting our method with multi-scale neighborhood information modeling strategy since detailed data structures and more complementary information may be captured and beneficial to the classification performance.

ACKNOWLEDGMENT

REFERENCES

[1] W.-K. Ma, J. M. Bioucas-Dias, J. Chanussot, and P. Gader, "Signal and image processing in hyperspectral remote sensing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 22–23, Jan. 2014.

[2] P. Ghamisi, N. Yokoya, J. Li, W. Liao, S. Liu, J. Plaza, B. Rasti, and A. Plaza, "Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 37–78, Dec. 2017.

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

[4] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 45–54, Jan. 2014.

[5] P. Ghamisi, J. Plaza, Y. Chen, J. Li, and A. J. Plaza, "Advanced spectral classifiers for hyperspectral images: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 1, pp. 8–32, Mar. 2017.

[6] L. He, 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.*, vol. 56, no. 3, pp. 1579–1597, Mar. 2017.

[7] P. Ghamisi, E. Maggiori, S. Li, R. Souza, Y. Tarablaka, G. Moser, A. De Giorgi, L. Fang, Y. Chen, M. Chi, S. B. Serpico, and J. A. Benediktsson, "New frontiers in spectral-spatial hyperspectral image classification: The latest advances based on mathematical morphology, Markov random fields, segmentation, sparse representation, and deep learning," *IEEE Geosci. Remote Sens. Mag.*, vol. 6, no. 3, pp. 10–43, Sep. 2018.

[8] J. Peng, W. Sun, and Q. Du, "Self-paced joint sparse representation for the classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1183–1194, Feb. 2019.

[9] J. Peng, L. Li, and Y. Y. Tang, "Maximum likelihood estimation-based joint sparse representation for the classification of hyperspectral remote sensing images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 6, pp. 1790–1802, Jun. 2019.

[10] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. M. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.

[11] X. Jia, B.-K. Kuo, and M. M. Crawford, "Feature mining for hyperspectral image classification," *Proc. IEEE*, vol. 101, no. 3, pp. 676–697, Mar. 2013.

[12] D. Lunga, S. Prasad, M. M. Crawford, and O. Ersoy, "Manifold-learning-based feature extraction for classification of hyperspectral data: A review of advances in manifold learning," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 55–66, Jan. 2014.

[13] W. Sun and Q. Du, "Hyperspectral band selection: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 118–139, Jun. 2019.

[14] I. T. Jolliffe, "Principal component analysis and factor analysis," in *Principal Component Analysis*. New York, NY, USA: Springer, 1986, ch. 7, pp. 115–128.

[15] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[16] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.

[17] X. He and P. Niyogi, "Locality preserving projections," in *Proc. 17th Annu. Conf. Neural Inf. Process. Syst. (NIPS)*, in Advances in Neural Information Processing Systems, 2003, pp. 153–160.

[18] D. Cai, X. He, and J. Han, "Spectral regression for efficient regularized subspace learning," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2007, pp. 1–8.

[19] T. V. Bandos, L. Bruzzone, and G. Camps-Valls, "Classification of hyperspectral images with regularized linear discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 862–873, Mar. 2009.

[20] Y.-L. Chang, J.-N. Liu, C.-C. Han, and Y.-N. Chen, "Hyperspectral image classification using nearest feature line embedding approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 278–287, Jan. 2014.

[21] N. H. Ly, Q. Du, and J. E. Fowler, "Sparse graph-based discriminant analysis for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 3872–3884, Jul. 2014.

[22] Z. Xue, P. Du, J. Li, and H. Su, "Simultaneous sparse graph embedding for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 6114–6133, Jun. 2015.

[23] Y. Dong, B. Du, L. Zhang, and L. Zhang, "Dimensionality reduction and classification of hyperspectral images using ensemble discriminative local metric learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2509–2524, May 2017.

[24] F. Luo, B. Du, L. Zhang, L. Zhang, and D. Tao, "Feature learning using spatial-spectral Hypergraph discriminant analysis for hyperspectral image," *IEEE Trans. Cybern.*, vol. 49, no. 7, pp. 2406–2419, Jul. 2019.

[25] L. Zhang, Q. Zhang, B. Du, X. Huang, Y. Y. Tang, and D. Tao, "Simultaneous spectral-spatial feature selection and extraction for hyperspectral images," *IEEE Trans. Cybern.*, vol. 48, no. 1, pp. 16–28, Jan. 2018.

[26] L. Zhang, Q. Zhang, L. Zhang, D. Tao, X. Huang, and B. Du, "Ensemble manifold regularized sparse low-rank approximation for multiview feature embedding," *Pattern Recognit.*, vol. 48, no. 10, pp. 3102–3112, Oct. 2015.

[27] K. Makantasis, A. D. Doulamis, N. D. Doulamis, and A. Nikitakis, "Tensor-based classification models for hyperspectral data analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 12, pp. 6884–6898, Dec. 2018.

[28] N. Renard, S. Bourennane, and J. Blanc-Talon, "Denoising and dimensionality reduction using multilinear tools for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 5, no. 2, pp. 138–142, Apr. 2008.

[29] N. Renard and S. Bourennane, "Dimensionality reduction based on tensor modeling for classification methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 4, pp. 1123–1131, Apr. 2009.

[30] S. Bourennane, C. Fossati, and A. Cailly, "Improvement of classification for hyperspectral images based on tensor modeling," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 801–805, May 2010.

[31] L. Zhang, L. Zhang, D. Tao, and X. Huang, "Tensor discriminative locality alignment for hyperspectral image spectral-spatial feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 242–256, Jan. 2013.

[32] S. Velasco-Forero and J. Angulo, "Classification of hyperspectral images by tensor modeling and additive morphological decomposition," *Pattern Recognit.*, vol. 46, no. 2, pp. 566–577, Feb. 2013.

[33] Z. Zhong, B. Fan, J. Duan, L. Wang, K. Ding, S. Xiang, and C. Pan, "Discriminant tensor spectral–spatial feature extraction for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 1028–1032, May 2015.

[34] Y. Gao, X. Wang, Y. Cheng, and Z. Wang, "Dimensionality reduction for hyperspectral data based on class-aware tensor neighborhood graph and patch alignment," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 8, pp. 1582–1593, Aug. 2015.

[35] J. An, X. Zhang, and L. C. Jiao, "Dimensionality reduction based on group-based tensor model for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 10, pp. 1497–1501, Oct. 2016.

[36] L. Pan, H.-C. Li, Y.-J. Deng, F. Zhang, X.-D. Chen, and Q. Du, "Hyperspectral dimensionality reduction by tensor sparse and low-rank graph-based discriminant analysis," *Remote Sens.*, vol. 9, no. 5, p. 452, May 2017.

[37] Y. Ren, L. Liao, S. J. Maybank, Y. Zhang, and X. Liu, "Hyperspectral image spectral-spatial feature extraction via tensor principal component analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 9, pp. 1431–1435, Sep. 2017.

[38] Z. Liu, B. Tang, X. He, Q. Qiu, and H. Wang, "Sparse tensor-based dimensionality reduction for hyperspectral spectral–spatial discriminant feature extraction," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1775–1779, Oct. 2017.

[39] Y.-J. Deng, H.-C. Li, L. Pan, L.-Y. Shao, Q. Du, and W. J. Emery, "Modified tensor locality preserving projection for dimensionality reduction of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 2, pp. 277–281, Feb. 2018.

[40] J. An, X. Zhang, H. Zhou, J. Feng, and L. Jiao, "Patch tensor-based sparse and low-rank graph for hyperspectral images dimensionality reduction," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 7, pp. 2513–2527, Jul. 2018.

[41] J. An, X. Zhang, H. Zhou, and L. Jiao, "Tensor-based low-rank graph with multimanifold regularization for dimensionality reduction of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4731–4746, Aug. 2018.

[42] Y.-J. Deng, H.-C. Li, K. Fu, Q. Du, and W. J. Emery, "Tensor low-rank discriminant embedding for hyperspectral image dimensionality reduction," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 12, pp. 7183–7194, Dec. 2018.

[43] W. Sun, G. Yang, J. Peng, and Q. Du, "Lateral-slice sparse tensor robust principal component analysis for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, to be published.

[44] L. Zhang, L. Zhang, B. Du, J. You, and D. Tao, "Hyperspectral image unsupervised classification by robust manifold matrix factorization," *Inf. Sci.*, vol. 485, pp. 154–169, Jun. 2019.

[45] A. Foi, V. Katkovnik, and K. Egiazarian, "Pointwise shape-adaptive DCT for high-quality denoising and deblocking of grayscale and color images," *IEEE Trans. Image Process.*, vol. 16, no. 5, pp. 1395–1411, May 2007.

[46] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1253–1278, 2000.

[47] B. Krishnapuram, L. Carin, M. A. T. Figueiredo, and A. J. Hartemink, "Sparse multinomial logistic regression: Fast algorithms and generalization bounds," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 6, pp. 957–968, Jun. 2005.

[48] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Hyperspectral image segmentation using a new Bayesian approach with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3947–3960, Oct. 2011.

[49] Z. Xue, J. Li, L. Cheng, and P. Du, "Spectral–spatial classification of hyperspectral data via morphological component analysis-based image separation," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 70–84, Jan. 2015.

[50] P. Du, Z. Xue, J. Li, and A. Plaza, "Learning discriminative sparse representations for hyperspectral image classification," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 1089–1104, Sep. 2015.

[51] Z. Xue, P. Du, J. Li, and H. Su, "Sparse graph regularization for hyperspectral remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 4, pp. 2351–2366, Apr. 2017.

[52] Z. Xue, P. Du, J. Li, and H. Su, "Sparse graph regularization for robust crop mapping using hyperspectral remotely sensed imagery with very few *in situ* data," *ISPRS J. Photogramm. Remote Sens.*, vol. 124, pp. 1–15, Feb. 2017.

[53] S. Zhou, Z. Xue, and P. Du, "Semisupervised stacked autoencoder with cotraining for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3813–3826, Jun. 2019.

[54] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "MPCA: Multilinear principal component analysis of tensor objects," *IEEE Trans. Neural Netw.*, vol. 19, no. 1, pp. 18–39, Jan. 2008.

[55] X. He, D. Cai, and P. Niyogi, "Tensor subspace analysis," in *Proc. Neural Inf. Process. Syst. (NIPS)*, in Advances in Neural Information Processing Systems, 2005, pp. 499–506.

[56] C. Deng, X. He, and J. Han, "Subspace learning based on tensor analysis," Dept. Comput. Sci., UIUC, Tech. Rep. UIUCDCS-R-2005-2572, 2005.

[57] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *Amer. Statist.*, vol. 46, no. 3, pp. 175–185, 1992.

[58] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

ZHAOHUI XUE received the B.S. degree in geomatics engineering from Shandong Agricultural University, Tai'an, China, in 2009, the M.E. degree in remote sensing from the China University of Mining and Technology, Beijing, China, in 2012, and the Ph.D. degree in cartography and geographic information system from Nanjing University, Nanjing, China, in 2015.

He is currently an Associate Professor with the School of Earth Sciences and Engineering, Hohai University, Nanjing. His research interests include hyperspectral image classification, time-series image analysis, pattern recognition, and machine learning.

Dr. Xue has been honored as an outstanding graduate for the B.S., M.E., and Ph.D. degrees, in 2009, 2012, and 2015, respectively. He was a recipient of the National Scholarship for Doctoral Graduate Students granted by the Ministry of Education of China, in 2014. He received the Best Reviewer Award for the IEEE Geoscience and Remote Sensing Society. He has been a Reviewer for the IEEE Transactions on Geoscience and Remote Sensing, the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, the IEEE Geoscience and Remote Sensing Letters, the *ISPRS Journal of Photogrammetry and Remote Sensing*, and the *International Journal of Remote Sensing*.

SIRUI YANG received the B.S. degree in geomatics engineering from Hohai University, Nanjing, China, in 2017, where she is currently pursuing the M.E. degree in remote sensing. Her research interest includes hyperspectral image fusion and classification. She was a recipient of the National Scholarship for Master Graduate Students granted by the Ministry of Education of China, in 2018.

MENGXUE ZHANG received the B.S. degree in geomatics engineering from Hohai University, Nanjing, China, in 2018, where he is currently pursuing the M.E. degree in geographical information engineering. His research interests include hyperspectral image classification, radar automatic target recognition, and machine learning.

● ● ●