Hyperspectral Image Dimension Reduction Using Weight Modified Tensor-Patch-Based Methods

Boyu Feng, *Member, IEEE*, and Jinfei Wang, *Member, IEEE*

***Abstract*—Dimension reduction (DR) addresses the problem known as the curse of dimensionality in myriad hyperspectral imagery applications. Although the spatial pattern may assist in the distinction between different land covers that have close spectral signatures, it is often neglected by the current DR methods. In order to overcome this defect, two solutions: patch-based and tensor-patch-based, are studied in this article for a group of graph-based DR methods. To date, only a few attempts have been made in the patch- and tensor-patch-based variations for the graph-based DR methods. This article proposed two weight modified tensor-patch-based methods, namely weight modified tensor locality preserving projections and weight modified tensor neighborhood preserving embedding. Specifically, as graph-based DR methods heavily rely on the construction of adjacency graphs, this paper proposes a new use of the weighted region covariance matrix in the calculation of adjacency graphs. We found that the two proposed tensor-patch methods outperform the up-to-date methods.**

***Index Terms*—Dimension reduction (DR), hyperspectral imagery, patch, spectral-spatial, and tensor.**

I. Introduction

SPECTRAL variability is the main advantage of hyperspectral imaging over other existing remote sensors. In order to benefit from the abundant spectral information from the hyperspectral image, algorithms were designed to analyze the spectral bands [1], [2] and to explore the possibility of using the spectral variability with other data [3], [4]. On the other hand, the huge number of spectral bands also arose the problem known as the *curse of dimensionality* in myriad hyperspectral imagery applications [5], when dimension reduction (DR) will be conducted to obtain the useful underlying information. The traditional DR methods often consider a hyperspectral image as a 2-D matrix with rows of observations and columns of variables. However, the actual hyperspectral image is a 3-D cube with two spatial dimensions and one spectral dimension. By flattening the original image along the spectral dimension, the spatial information is lost. The motivation of this research is to improve the situation of treating the hyperspectral image as 2-D matrix in DR and include the spatial information in the original data through 3-D processes. Initial attempts have been made in supervised DR to preserve spatial information through image texture during DR process. A hybrid supervised DR method was performed in a raw spectral-spatial feature space, where spatial features are represented by the grey level co-occurrence matrix of each spectral channel [6]. Zhou *et al*. [7], proposed a spatial and spectral regularized local discriminant embedding method to include the spatial information through spatial filtering and spatial discriminant analysis. However, these initial attempts to preserve spatial information is by extracting the spatial information into additional layer of information and then still work with a flattened 2-D hyperspectral image. Starting in computer vision, the concepts of patch and tensor-patch were proposed to preserve information along each dimension during data processing. Briefly speaking, the patch-based framework considers each pixel along with its neighbors as an integrated patch containing both spectral and spatial information; and the tensor-patch-based framework considers the 3-D hyperspectral data as it is and performs data processing along each dimension. From the late 2000s, this idea sheds light on hyperspectral applications. The image patch distance (IPD) [8] is an initial patch-based proposal for spectral-spatial similarity measurement in hyperspectral study. It considers both the spectral and spatial information through patches from the hyperspectral image. The IPD patch-based framework has been adopted in supervised DR [9], [10]. Unfortunately, the patch-based framework can only preserve the local spatial information within the decided patch, while the global spatial information still lose in the flattening of the hyperspectral data. On the other hand, the idea of tensor-based DR appeared around the year of 2010. The tensor-based framework have been adopted into a few supervised DR methods [11]–[13], which requires ground truth data. Several unsupervised tensor-based DR methods for computer vision problems have been proposed, including multilinear principle component analysis (PCA) [14], concurrent subspaces analysis [15], and tensor canonical correlation analysis [16]. These methods usually work with a group of tensors. However, in the case of hyperspectral image DR, we only have one tensor. In [17], [18], the single hyperspectral image is first treated as a 3-D tensor, where DR is applied along the spectral dimension, and low-rank approximation is applied along the spatial dimensions as noise reduction. However, the DR and low-rank approximation are performed separately on the data and introduces complexity. Similarly, An *et al* [19], adopts a multiscale idea in the tensor-based low rank decomposition to automatically find the best rank for each dimensions. Since then, the concept of low rank tensor representation/approximation/decomposition

Manuscript received February 26, 2020; revised May 3, 2020; accepted June 1, 2020. Date of publication June 5, 2020; date of current version June 29, 2020. This work was supported by the NSERC Discovery Grant awarded to Dr. Jinfei Wang. *(Corresponding author: Boyu Feng.)*

The authors are with the Department of Geography, University of Western Ontario, London, ON N6A 5C2, Canada (e-mail: bfeng24@uwo.ca; jfwang@uwo.ca).

Digital Object Identifier 10.1109/JSTARS.2020.3000284

becomes popular, which aims to find the intrinsic data structure for better feature extraction and does not necessarily relate to DR [20]–[22]. Later, inspired by the patch alignment framework [23], a new tensor-patch-representation was proposed for hyperspectral image [24]–[26]. Under this tensor-patch-based framework, the input hyperspectral image needs to be first divided into local patches. In this way, the hyperspectral image becomes a 4-D data, composed of a group of 3-D tensors. In [24] and [25], the tensor-patch-representation has been first applied to supervised DR solutions (discriminant locality alignment and linear discriminant analysis). Unfortunately, the ground truth data asked by the supervised DR solution may not accessible for some study sites.

Although the concept of patch and tensor-patch have great potential in a group of unsupervised graph-based DR methods [27], related studies are limited [28], [29]. This group of methods learns the data local structure from adjacency graphs/weight matrices that only consider a certain number of the nearest pixels in the spectral space rather than all pixels in the image [30]. Four representative graph-based DR methods are locally linear embedding (LLE) [31], neighborhood preserving embedding (NPE) [32], Laplacian eigenmaps (LE) [33], and locality preserving projections (LPP) [34]. By introducing the patch-based weight calculation, the spatially coherent LLE was able to preserve spatial information [35]. Hong *et al*. have successfully introduced the spatial information through patch-based idea in LLE and LE [36]. Not until 2018, Deng *et al*. [28], adopted the patch alignment framework in the TLPP especially for hyperspectral DR. However, the use of the regional covariance matrix (RCM) in [28] treats each pair of data points in the region of interest equally, which may not reflect the locational-variated real-world situation. This article uses a weighted RCM (WRCM) to account for this shortage. Furthermore, the majority of existing papers tend to focus on only one method from the graph-based methods for possible improvements, while a comprehensive experiment on all the possibilities should be made. This article proposed two weight modified tensor-patch-based methods, namely weight modified tensor locality preserving projections (WMTLPPs) and weight modified tensor neighborhood preserving embedding (WMTNPE). Comprehensive experiments have been provided for the patch- and tensor-patch-variations in the graph-based DR methods. It has been found that the two proposed tensor-patch methods outperform the up-to-date methods.

II. Methods

A. Locally Preserving Projections and Neighborhood Preserving Embedding

The LPP and NPE methods are two of the early graph-based method attempts. They separately simplified the nonlinear LE and LLE methods and have great flexibility. They share the same solution procedure containing three steps listed later and shown in Fig. 1, except that they solve different eigenproblems due to their different objective functions.

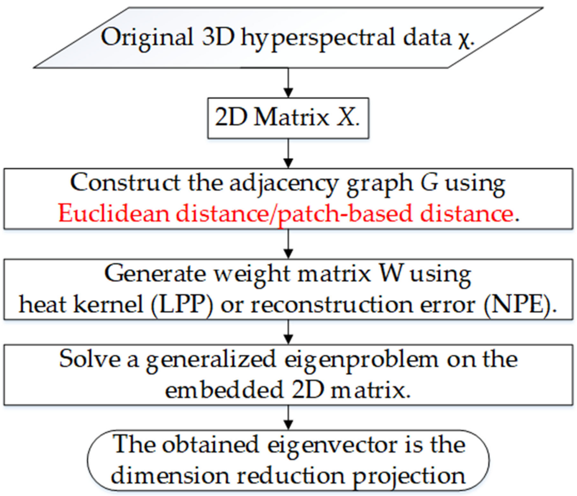

Fig. 1. Flowchart of the LPP/NPE DR methods.

1) *Constructing the adjacency graph G:* The adjacency graph *G* locates a group of adjacent pixels around the target pixel. The most common way to decide the adjacency of one pixel is the *k*-nearest neighbors *(k*NN) method [37], [38]. With a manually chosen *k* value, *k* nearest neighbors will be selected based on the Euclidean distance in traditional method or patch-based distance (e.g. IPD, RCM, and WRCM) in proposed methods. This patch-based modification shown as red font in Fig. 1 is the key change in the proposed methods and is discussed in depth in Section II-B.
2) *Generating the weight matrix W:* After obtaining the adjacency graph, the relation between each pixel and its adjacent pixels is decided using a weight matrix. The weight matrix is a sparse matrix, where each element depicts the relation between two pixels and equals to zero if the two pixels are not each other's adjacent pixel. This relation described by the weight matrix is the key criteria that the graph-based DR uses to preserve the original data structure during the DR process. Such relation is the key preserved through the graph-based DR transformation. Such relation is the key preserved through the graph-based DR transformation.
3) *Solving the eigenfunction:* The objectives of both the LPP and NPE can be described as two different minimization problems, which then were justified to be equivalent to solving two different eigenfunctions [32], [34]. The derived eigenvector is the projector that leads to the resulting DR projection.

In the step (1), the two methods of LPP and NPE are exactly the same. In the step (2), LPP and NPE have a different equation to calculate the weight matrix. For LPP, the value of weight matrix for any two adjacent pixels $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ is determined by heat kernel [34]

$$Wij = e - \frac{\| xi - xj\|^2}{t}. \tag{1}$$

where $\|\|$ is the Euclidean norm; and $\boldsymbol{t}$ is the power of weight set to be 1 and has little effect on the final results in the hyperspectral application of this manuscript. For NPE, the value of weight matrix for any two adjacent pixels $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ is determined by minimizing reconstruction error [32]

$$\min \sum_i \| x_i - \sum_j W_{ij} x_j \|^2 \tag{2}$$

with constraint $\sum_j \boldsymbol{W_{ij}} = 1,\ j = 1, 2, \ldots, n$, and n is total number of pixels in the hyperspectral image.

Regarding the different weight matrices, the objective functions in step (3) are different between the two methods LPP and NPE. For LPP, the objective is to ensure that if two data points are close to each other in the original space, they should stay close in the projected feature space. If two pixels are not each other's adjacent pixel and their weight is zero, their relation will not be considered in the DR transformation. It can be realized through the minimization problem

$$\min \sum_{ij} \| y_i - y_j \|^2 W_{ij}, \tag{3}$$

where $\boldsymbol{y_i}$ and $\boldsymbol{y_j}$ are data points in the projected feature space; and $\boldsymbol{W_{ij}}$ is the weight matrix obtained in the step (2). For adjacent pixels, their closeness will be scaled by their weight, and larger weights exert more penalty on their difference. He and Niyogi justified that the above minimization problem is equivalent to solve the eigenfunction [34]

$$XLX^T a = \lambda XDX^T a, \tag{4}$$

where the Laplacian matrix $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$; $\boldsymbol{D_{ii}} = \sum_j \boldsymbol{W_{ji}}$ is a diagonal matrix with column sums of $\mathbf{W}$; and λ and a are the eigenvalue and eigenvector we want to solve.

For NPE, the objective function is to minimize the reconstruction error among each neighborhood

$$\min \sum_{ij} \| y_i - \sum_j W_{ij} y_j \|^2 \tag{5}$$

where $\boldsymbol{y_i}$ and $\boldsymbol{y_j}$ are pixels in the projected feature space. Thus, if pixel $\boldsymbol{y_j}$ is not $\boldsymbol{y_i}$'s adjacent pixel and their weight is zero, $\boldsymbol{y_j}$ will be not included in the reconstruction of $\boldsymbol{y_i}$. For $\boldsymbol{y_j}$ that is adjacent to $\boldsymbol{y_i}$, it will be scaled by their weight, and larger weight exert more influence in the reconstruction. This minimization problem can be also simplified as an eigenfunction:

$$XMX^T a = \lambda XX^T a \tag{6}$$

where $\boldsymbol{M} = (\boldsymbol{I} - \boldsymbol{W})^T (\boldsymbol{I} - \boldsymbol{W})$ and $\boldsymbol{I} = \mathrm{diag}(1, \ldots, 1)$; and λ and a are the eigenvalue and eigenvector we want to solve.

B. Tensor Locality Preserving Projections and Tensor Neighborhood Preserving Embedding

In the LPP and NPE process, the 3-D hyperspectral image needs to be raster-scanned along the spectral dimension, when the spatial information are lost. Although the introduction of patch-based distance calculation in the adjacency graph construction considers the local spatial information, the eigenproblem in the LPP and NPE methods is based on the 2-D data and fail to include the global spatial information. There is a need to upgrade them into tensor versions [39], [40]. The TNPE and TLPP methods keep the original dimensions of the data and solve three eigenproblems on each dimension sequentially. Three eigenvectors are obtained and are used to project the data along each of the three dimensions.

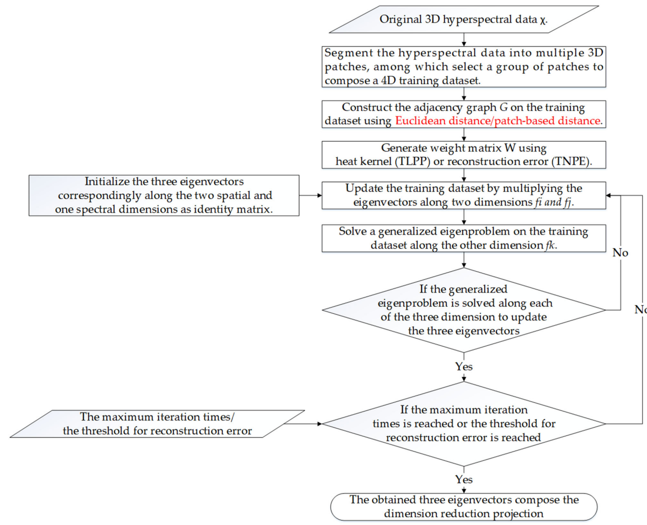

Fig. 2. Flowchart of the TLPP/TNPE DR methods.

In TLPP and TNPE, preprocess is needed for the given 3-D hyperspectral image $\mathcal{X} \in \mathbb{R}^{a \times b \times d}$, where a and b are the row and column numbers representing the spatial location and d is the spectral band number. The image is first spatially segmented into n 3-D patches with the same window size $\mathcal{X}_{\boldsymbol{q}} \in \mathbb{R}^{w \times w \times d}$, where $n = a \times b$ is the number of pixels in the image and w is the size of the window. This process creates a new 4-D dataset from the original 3-D data when one pixel vector along the spectral dimension becomes a cube composed by a group of pixel vectors in a spatial neighborhood. Then, a certain number, $c \ll n$, of training patches are randomly selected as input to the TLPP/TNPE algorithms with the following three steps and shown in Fig. 2.

1) *Constructing the adjacency graph G:* The TLPP and TNPE use the kNN methods. With a manually chosen k value, k nearest neighbors will be selected based on the Frobenius distance in traditional method or patch-based distance (e.g. IPD, RCM, and WRCM) in proposed methods. This patch-based modification shown as red font in Fig. 2 is the key change in the proposed methods and is discussed in depth in Section II-B.
2) *Generating the weight matrix W:* The TLPP and TNPE use the similar methods correspondingly as the above LPP and NPE methods. Specifically, in the heat kernel and reconstruction error methods, the Euclidean norm in LPP/NPE used for the vectors will be replaced by the Frobenius norm in TLPP and TNPE for the matrices, as the 3-D node is matrixed as a 2-D matrix along the spectral dimension.
3) *Solving the eigenfunction:* The TLPP and TNPE has the same objectives accordingly with LPP and NPE, but their objective functions are in 3-D version. Specifically, three eigenvectors were solved sequentially through the tensor

calculations discussed below along each dimension for the original 3-D hyperspectral data.

In the step (3), a basic tensor terminology and a tensor operation are used and explained later.

Terminology 1: Given a *m*-mode tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_m}$, the *f*-mode unfolding of tensor $\mathcal{A}$ is denoted as $\mathcal{A}^{(f)}$. It flattens the tensor $\mathcal{A}$ into a matrix $\mathcal{A}^{(\boldsymbol{f})} \in \mathbb{R}^{I_f \times I_1 \ldots I_{f-1} I_{f+1} \ldots I_m}$. The columns of $\mathcal{A}^{(\boldsymbol{f})}$ is obtained by fixing all but one mode.

Operation 1: A *f*-mode product between a *m*-mode tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_m}$ and a matrix $\mathrm{U} \in \mathbb{R}^{J_f \times I_f}$ gives a tensor $\mathcal{A} \times_f U \in \mathcal{R}^{I_1 \times \ldots I_{f-1} \times J_f \times I_{f+1} \ldots \times I_m}$. The element presentation of the *f*-mode product is

$$(\mathcal{A} \times_f U)_{i_1 \ldots I_{f-1} j_f i_{f+1} \ldots i_m} = \sum_{i_f=1}^{I_f} \mathcal{A}_{i_1 \ldots i_{f-1} i_f i_{f+1} \ldots i_m} U_{j_f i_f}. \tag{7}$$

Using the Terminology 1, the *f*-mode unfolding of the *f*-mode product results can be expressed as

$$(\mathcal{A} \times_f U)^{(f)} = U \mathcal{A}^{(f)}. \tag{8}$$

In the step (3), the objective function of TLPP is the minimization problem [39]:

$$\min \sum_{ij} \| \mathcal{Y}_i - \mathcal{Y}_j \|_F^2 \, W_{ij} \tag{9}$$

where $\mathcal{Y}_{\boldsymbol{i}}$ and $\mathcal{Y}_{\boldsymbol{j}} \in \mathbb{R}^{w \times w \times L}$ are hyperspectral training patches in the projected feature space transformed from the original hyperspectral training patches $\mathcal{X}_{\boldsymbol{i}}$ and $\mathcal{X}_{\boldsymbol{j}} \in \mathbb{R}^{w \times w \times K}$ in the image space, where w is the size of the window, L is reduced dimension, and the K is the spectral band number. This minimization problem can be solved by the three eigenfunction

$$\left(\sum_{ij} W_{ij} \left(\mathcal{Y}_i^{(f)} - \mathcal{Y}_j^{(f)} \right) \left(\mathcal{Y}_i^{(f)} - \mathcal{Y}_j^{(f)} \right)^T \right) \mathrm{U}_f$$
$$= \lambda \left(\sum_i \mathcal{Y}_i^{(f)} \mathcal{Y}_i^{(f)T} D_{ii} \right) \mathrm{U}_f, \mathrm{f} = 1,\ 2,\ 3, \mathrm{f} = 1,\ 2,\ 3 \tag{10}$$

where $\mathbf{D}_{\boldsymbol{ii}} = \sum_j \boldsymbol{W_{ji}}; \mathcal{Y}_{\boldsymbol{i}}^{(\boldsymbol{f})}$ and $\mathcal{Y}_{\boldsymbol{i}}^{(\boldsymbol{f})}$ are the *f*-mode unfolding of the hyperspectral training patches $\mathcal{Y}_{\boldsymbol{i}}$ and $\mathcal{Y}_{\boldsymbol{j}}$ respectively, and $\mathbf{U}_{\boldsymbol{f}}, \mathrm{f} = 1,\ 2,\ 3$ are the three target eigenvectors.

The objective function of TNPE is to keep the reconstruction error among each neighborhood small

$$\min \sum_i \left\| \mathcal{Y}_i - \sum_j W_{ij} \mathcal{Y}_j \right\|_F^2 \tag{11}$$

The earlier minimization problem can be solved by the three eigenfunction

$$\left(\sum_i \left(\mathcal{Y}_i^{(f)} - \sum_j W_{ij} \mathcal{Y}_j^{(f)} \right) \left(\mathcal{Y}_i^{(f)} - \sum_j W_{ij} \mathcal{Y}_j^{(f)} \right)^T \right) \mathrm{U}_f$$
$$= \lambda \left(\sum_i \mathcal{Y}_i^{(f)} \mathcal{Y}_i^{(f)T} \right) \mathrm{U}_f, \mathrm{f} = 1,\ 2,\ 3. \tag{12}$$

For both TLPP and TNPE, the eigenfunction is solved along each of the three dimensions to obtain the three eigenvectors $\mathbf{U}_{\boldsymbol{f}}, \mathrm{f} = 1,\ 2,\ 3$ as projectors.

C. Patch-Based Adjacency Maps

In the traditional graph-based methods, the single pixels are used to calculate the adjacency map, which only considers the spectral information. Three different methods considering both spectral and spatial information in the data were used in this article to produce adjacency maps: IPD, RCM, and WRCM. The three methods are based on patch representation of the data. Before introducing the three spatial-spectral methods, we first consider the three different patches according to three different pixel locations: at the corner, on the edge and in the image. For a given pixel, a group of $w \times w$ surrounding pixels is decided as the patch, where w is the window size. If the pixel is located in the image, the moving window can naturally cover its spatial neighbors. If the pixel lies on the edges or at the corner of the image, we used a reflection transformation to fill the non-existing spatial neighbors.

The IPD is especially proposed for adjacency-graph based methods like LPP and NPE. For any two pixels in the hyperspectral image, the IPD calculate their similarity based on the small neighborhood (spatial window) of the two pixels. Given two pixels x_i and x_j from a hyperspectral image, a spatial window of size w is used to find the neighborhoods for the two pixels x_i and x_j. The neighborhoods are correspondingly denoted as $\Omega(x_i) = \{a_1,\ a_2, \ldots, a_{w^2}\}$ and $\Omega(x_j) = \{b_1,\ b_2, \ldots, b_{w^2}\}$. The IPD is defined as

$$d_{IPD}\left(x_i, x_j\right) = \sum_{l=1}^{w^2} d_u\left(a_l, b_l\right)$$
$$d_u\left(a_l, b_l\right) = \max\left(\min_{b \in \Omega(x_j)} d\left(a_l, b\right), \min_{b \in \Omega(x_i)} d\left(b_l, a\right) \right) \tag{13}$$

where $d(a, b)$ is a spectral similarity function comparing a to b. Using this equation, the similarity between two pixels is measured with their surrounding neighbors. Thus, the IPD incorporates both the spatial and spectral information.

The RCM is a region descriptor proposed by Tuzel *et al.* [41]. In the hyperspectral image $\mathcal{X} \in \mathbb{R}^{\mathbf{a} \times \mathbf{b} \times \mathbf{d}}$, a region of interest $\mathcal{R} \in \mathbb{R}^{w \times w \times d}$ (w is the window size) is represented by a $d \times d$ covariance matrix

$$C_R = \frac{1}{n} \sum_{i=1}^{n} \left(r_i - \mu\right) \left(r_i - \mu\right)^T \tag{14}$$

where $n = w \times w$ is the amount of pixels in the patch; r_i is a single d-dimensional pixel in $\mathcal{R}$; and $\mu = \frac{1}{n} \sum_{i=1}^{n} r_i$ is the mean value. It has a few properties: if two feature bands (d_1 and d_2) tend to increase together, then $\boldsymbol{C_R}[d_1, d_2] > 0$; if feature band d_1 tends to decrease when feature band d_2 increases, then $\boldsymbol{C_R}[d_1, d_2] < 0$; and if two feature bands are independent, then $\boldsymbol{C_R}[d_1, d_2] = 0$. The measure of distances between weighted covariance matrices are adopted from the original RCM method

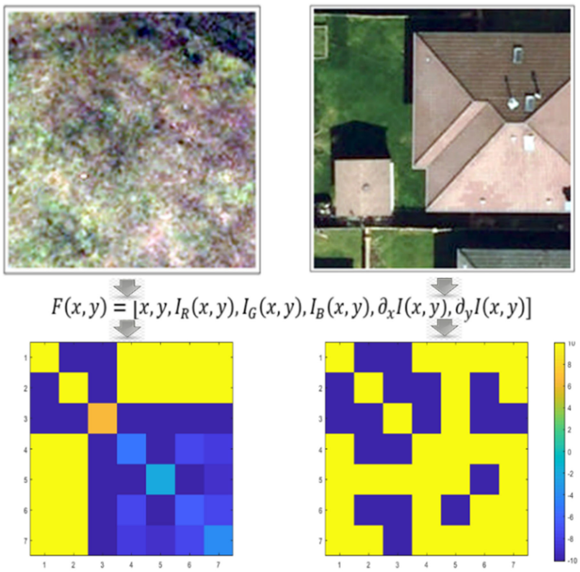

Fig. 3. Example of RCM/WRCM.

[41] and is based on an eigenproblem

$$\boldsymbol{D}\left(C_R^p, C_R^g\right) = \sqrt{\sum\nolimits_{i=1}^{d} ln^2 \lambda_i \left(\boldsymbol{C_R^p}, \boldsymbol{C_R^g}\right)} \tag{15}$$

where $\lambda_i(\boldsymbol{C_R^p}, \boldsymbol{C_R^g})$ is the generalized eigenvalues of $\boldsymbol{C_R^p}$ and $\boldsymbol{C_R^g}$, computed from

$$\lambda_i C_R^p a_i - C_R^g a_i = 0, \; i = 1, \ldots, d \tag{16}$$

where x_i are the generalized eigenvectors.

The WRCM is developed from the RCM. It rewrites (14) from RCM as [42]

$$C_R = \frac{1}{2}\frac{1}{n \times n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(r_i - r_j\right)\left(r_i - r_j\right)^T \tag{17}$$

by replacing μ as

$$\mu = \frac{1}{n}\sum_{i=1}^{n} r_i. \tag{18}$$

In (14), each pair of pixels in the current patch contributes equally to the final results. The WRCM method introduces a weight term in the calculation

$$C_R = \frac{1}{2}\frac{1}{n \times n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(r_i - r_j\right)\left(r_i - r_j\right)^T W_{ij} \tag{19}$$

where $\boldsymbol{W_{ij}}$ is calculated as (1). The measure of distances between weighted covariance matrices are adopted from the original RCM method [41] shown in (15) and (16). Based on the RCM, the WRCM reflects the weighted feature correlation, which assigns more weights to points pairs that are spatially closer. Fig. 3 shows an easy example in computer vision, with the commonly used image features of pixel locations *(x, y)*, color (RGB) intensity, and the norm of the first order derivatives of the intensities with respect to *x* and *y*. After performing RCM/WRCM, the 3-D image patch turns into a matrix of size $d \times d$, where d is the band number.

TABLE I
NAMING OF THE PROPOSED GRAPH-BASED DR METHOD

Graph-based method	Adjacency graph	Name	Graph-based method	Adjacency graph	Name
LPP	Euclidea	LPP	TLPP	Euclidea	TLPP
LPP	IPD	IPD-LPP	TLPP	IPD	IPD-TLPP
LPP	RCM	MLPP*	TLPP	RCM	MTLPP
LPP	WRCM	WMLPP	TLPP	WRCM	WMTLPP
NPE	Euclidea	NPE	TNPE	Euclidea	TNPE
NPE	IPD	IPD-NP	TNPE	IPD	IPD-TNP
NPE	RCM	MNPE*	TNPE	RCM	MTNPE*
NPE	WRCM	WMNPE	TNPE	WRCM	WMTNP

D. Weight Modified Graph-Based DR Methods

According to the abovementioned workflow of the graph-based DR methods and the patch-based adjacency graph calculation, a group of patch- and tensor-patch-based methods can be derived. Deng *et al*. [28] has modified the TLPP method by replacing the Euclidean distance with RCM method in the adjacency graph calculation and called it modified TLPP (MTLPP). This manuscript further improve this replacement with the weighted RCM and called this modified new method the weight WMTLPP and weight modified tensor neighborhood preserving embedding (WMTNPE). Based on the naming strategy, the tested graph-based DR methods are given in Table I. Names with one asterisk sign suggest methods proposed in this manuscript.

E. Evaluation of the DR Method

In order to evaluate the DR results, classification was performed on the series of dimension-reduced images. The support vector machine (SVM) classifier was used in this article. As comparison, we also applied the SVM classifier on the original image, the PCA dimension-reduced result and the dimension-reduced results from two up-to-date spectral-spatial DR methods: robust local manifold representation for DR (RLMR) [36], and spatial and spectral regularized local discriminant embedding for DR (SSRLDE) [7]. We used two testing hyperspectral images with different spatial resolution. The two images are separately located in urban and rural environments, focusing on targets of different scales. The differences between the two images allow us to compare the impact of different images on the choice of window size in the calculation of the patch-based adjacency maps and weight matrix. In the results, the classification overall accuracy (OA), average accuracy (AA), and Kappa coefficient are provided. The OA is the number of pixels correctly classified divided by the number of total pixels. The AA is the sum of producer class accuracies divided by the

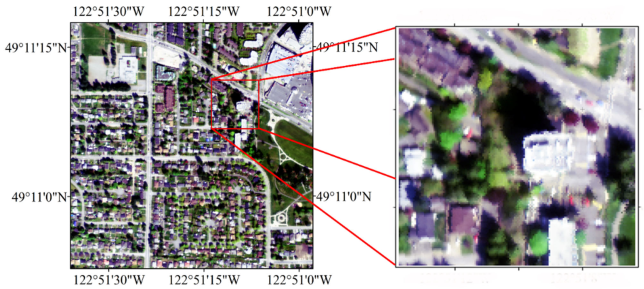

Fig. 4. Surrey, BC, CASI hyperspectral image in RGB (selected study area is in red rectangle).

Fig. 5. Indian Pines, AVIRIS hyperspectral image in RGB.

Fig. 6. University of Pavia, ROSIS hyperspectral image in RGB.

number of classes, where the producer's accuracy reflects the omission error when a class A pixel fails to be classified as class A. The Kappa coefficient measures inter-rater agreement for classified items. Statistically, the Kappa coefficient considers the possibility of the agreement occurring by chance. For all the three indexes, larger values suggest better classification results.

III. Experiments and Results

A. Studied Hyperspectral Images and Setup

This article used two real hyperspectral images. The first hyperspectral image depicts the urban area in City of Surrey, BC, Canada (see Fig. 4). It was obtained by the airborne CASI-1500 sensor during April 2013. The image contains 72 spectral bands from visible to near-infrared portion (0.36 to 1.05 μm) with a 9.6 nm band interval. The spatial resolution of the images is 1 m. The original image size is 1671×1646, which is too large considering the associated computation complexity for our current computing unit. Thus, we selected a small part from the CASI hyperspectral data of size 150×150. The size of 150×150 was selected as it was similar to the second widely-used study hyperspectral image (Indian Pines) that used a size of 145×145. The different patterns of these two images may have impact on the efficiency of the proposed methods. Thus in order to make the experiments on the two images comparable in terms of computation complexity, their size was set to be similar. In regards to the location of the cropped area, it is a typical urban scene in residential area, containing large amount of impervious surface area as well as green spaces.

The second image is the widely-used test data: Indian Pines (see Fig. 5). The Indian Pines dataset is collected by the AVIRIS sensor in 1992 over an agricultural area in Northern Indiana, IN, USA. The spectral bands of the Indian Pines data span from 400-2500 nm. After deleting the 20 bands affected by water absorption, 200 bands were used in this study. The Indian Pines image has a size of 145×145 and contains 10249 samples in 16 different classes.

The third image is also a widely-used test data: University of Pavia (see Fig. 6). This scene is acquired by the ROSIS sensor during a flight campaign over Pavia, northern Italy. It has 103 spectral bands, and a spatial resolution of 1.3 m. The size of the data is 610×340. The image ground truths differentiate nine classes.

To fulfill the objectives of this research, we designed our experiment as follows. A total of 16 DR methods were tested: two of them were traditional methods (LPP and NPE), six of them were patch-based methods (IPD-LPP, MLPP, WMLPP,

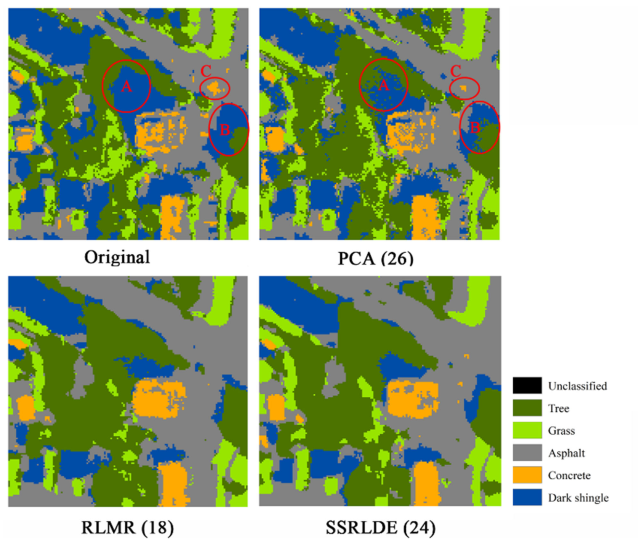

Fig. 7. Original image and PCA (preserved dimensions) image classification for Surrey, CASI image.

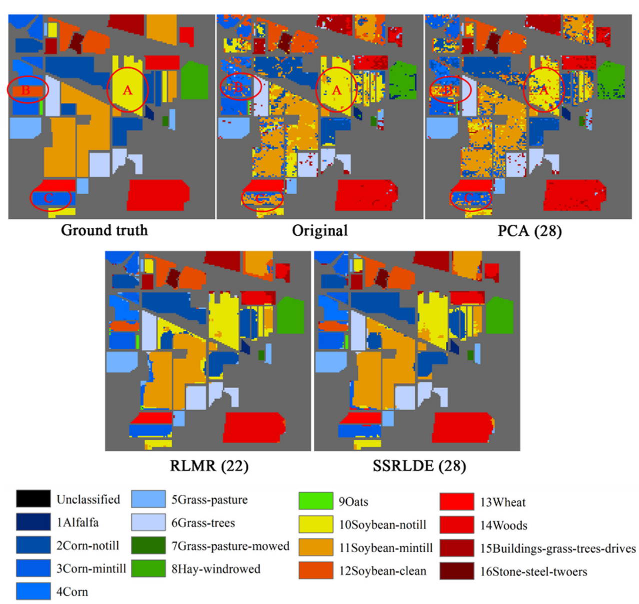

Fig. 8. Ground truths, original image classification, and PCA (preserved dimensions) image classification for Indian Pines, AVIRIS image.

IPD-NPE, MNPE, and WMNPE), and eight of them were tensor-patch-based methods (TLPP, IPD-TLPP, MTLPP, WMTLPP, TNPE, IPD-TNPE, MTNPE, and WMTNPE). In order to obtain an appropriate local window size and reduced dimensionality, several cross-validation experiments were performed respectively. For the 14 patch- and tensor-patch-based methods, the window sizes of 3×3, 5×5, 7×7, 9×9, 11×11, and 13×13 were tested. A set of preserved dimensionality (2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, and 30) was chosen. The dimension-reduced images were then classified using SVM classifier. The optimal preserved dimensionality is the one that derives the best overall accuracy value and will be used to represent the performance of the tested DR methods. For the

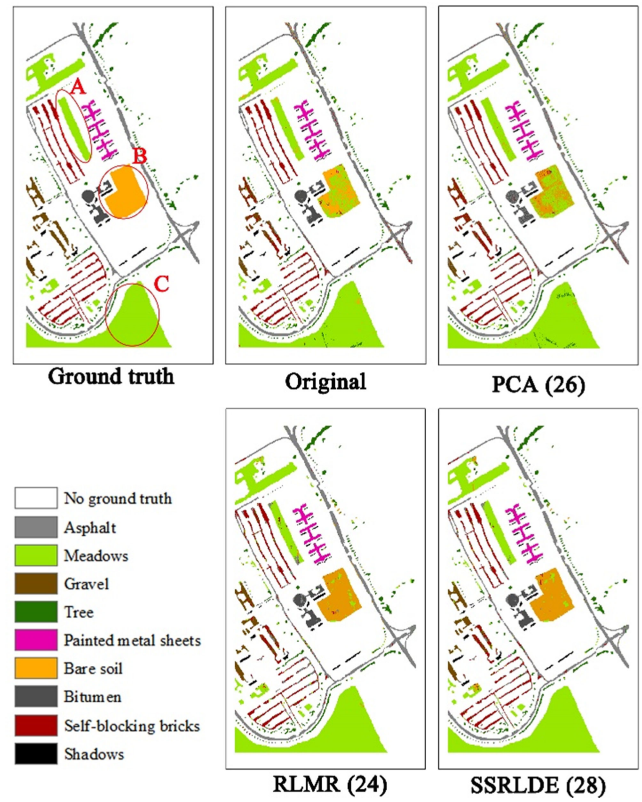

Fig. 9. Ground truths, original image classification, and PCA (preserved dimensions) image classification for University of Pavia, ROSIS image.

classification purpose, we manually selected training samples for the first studied image (Surrey, BC, CASI), covering eight land cover/land-use types: grass, bush, tree, concrete, asphalt, dark shingle, dark roof panel, and roof paint. For the Surrey, BC, CASI image with 1 m spatial resolution, we used the orthophoto data taken one month earlier as reference data. The orthophoto data has a 10 cm spatial resolution and contains three bands of RGB. In the case of the second studied image, reference data are available on GIC's website.[1] In the rural scene, 16 land cover/land-use types were identified: alfalfa, corn-no-till, corn-min-till, corn, grass-pasture, grass-tree, grass-pasture-mowed, hay-windrowed, oats, soybean-no-till, soybean-min-till, soybean-clean, wheat, woods, building-grass-tree-drive, and stone-steel-tower. A group of 1071 training samples, a group of 2610 training samples, and a group of 2000 training samples were separately generated for the Surrey, BC, CASI, Indian Pines, AVIRIS, and University of Pavia, ROSIS images. For comparison, the SVM classification results of the original image and the PCA, RLMR, and SSRLDE dimension-reduced results were also generated. In the case of PCA, RLMR, and SSRLDE DR, classification was performed with 15 sets of preserved bands (2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, and 30). To evaluate the classification results, one set of 9,913, one set of 7,288, and one set of 40,000 random testing points were

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

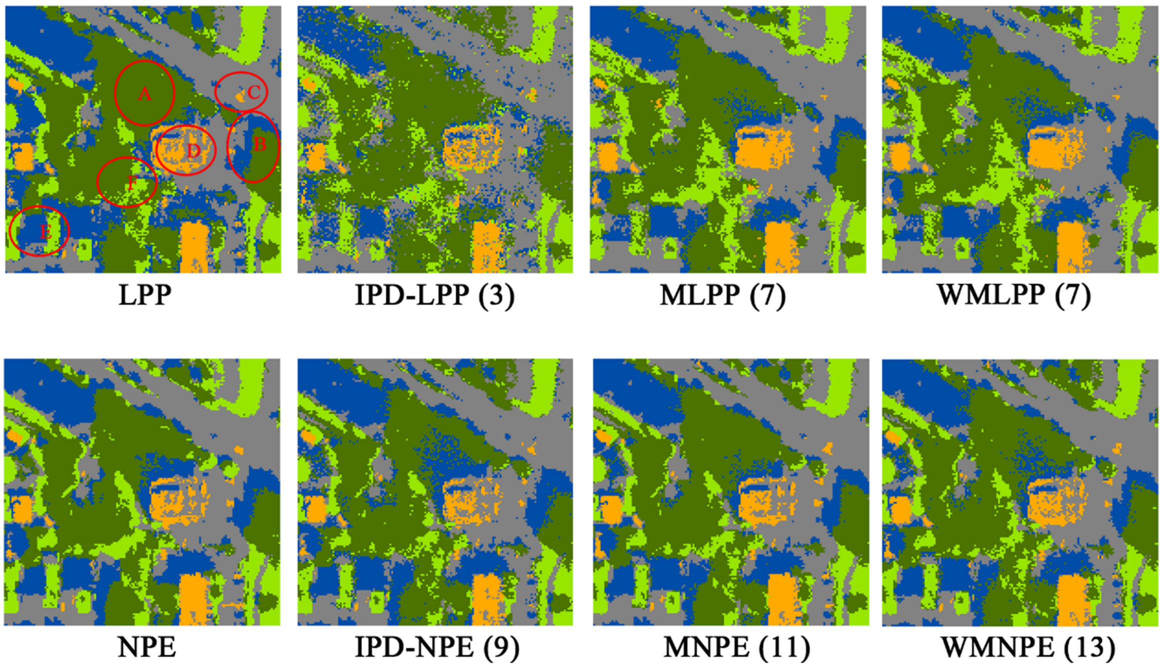

Fig. 10. Classification maps for patch-based DR results with highest overall accuracy for Surrey, BC, CASI image (numbers in parentheses are the corresponding window sizes for the highest overall accuracy).

TABLE II
OVERALL ACCURACY OF EXISTING UNSUPERVISED/SUPERVISED DR METHODS

Overall accuracy	Original image	PCA	RLMR	SSRLDE
Surrey, CASI	79.1%	79.0%	87.1%	88.3%
Indian Pines, AVIRIS	80.1%	83.4%	95.8%	96.9%
University of Pavia, ROSIS	82.3%	85.5%	90.7%	91.8%

generated separately for the Surrey, BC, CASI, Indian Pines, AVIRIS, and University of Pavia, ROSIS images.

B. Classification Results for Comparison

Figs. 7–9 show the classification results of the original image, and the best classification results from the PCA, RLMR, and SSRLDE correspondingly for the Surrey, BC, CASI, the Indian Pines, AVIRIS, and University of Pavia, ROSIS images. The overall accuracies of the four comparison classification results are given in Table II. The newer spectral-spatial-based DR methods (RLMR and SSRLDE) provide much higher accuracies for both the studied images. In the classification results of the Surrey, BC, and CASI image, the most noticeable error occurs in shadow areas in circles A and B, where trees are misclassified as dark shingle. The other noticeable error is that the painted asphalt areas are misclassified as concrete, which is the case in circle C. In the classification results of the Indian Pines, AVIRIS image, both the original- and PCA-derived classification maps have small-misclassified areas within most croplands. Yet, the PCA works better in several crop lands in circles A, B, and C. The results from the RLMR and SSRLDE provide much smoother classification maps. In the classification results of the University of Pavia, ROSIS image, circles A, B, and C contain a lot of errors when the classification is performed on the original and PCA images. These errors greatly eliminated by the spatial-spectral methods RLMR and SSRLDE.

C. Patch-Based DR Results

Four LPP-based and four NPE-based DR methods were used to derive a group of dimension-reduced images. The results were then classified by the SVM classifier The best classification results are shown in Fig. 10 for Surrey, BC, CASI image, in Fig. 11. for Indian Pines, AVIRIS image and in Fig. 12. for University of Pavia, ROSIS image. Compared to the classification results from the original and PCA-derived images in Fig. 7, the tree area in shadow (circle A) in Fig. 10. is better classified and the painted asphalt areas are less frequently misclassified as concrete (circle C). Yet, the tree areas in shadow (circle B) still suffer from misclassification. In the Surrey, BC, and CASI image classification maps (see Fig. 10.), the LPP-based methods appear to be more accurate in determining concrete roof (circle D) then the NPE-based methods. On the other hand, NPE-based methods appear to be more accurate in determining dark shingle roof (circle E). Further, LPP-based methods tend to misclassify tree areas as grass areas (circle F). Considering the three different patch-based adjacency map calculation methods, the IPD-based method provides the worst results in both LPP- and NPE-based DR. In the Indian Pines, AVIRIS image classification maps (see Fig. 11), the small portion of misclassified areas were eliminated, compared to the classification maps in Fig. 8 and the RCM- and WRCM-based methods seem to provide better results than the

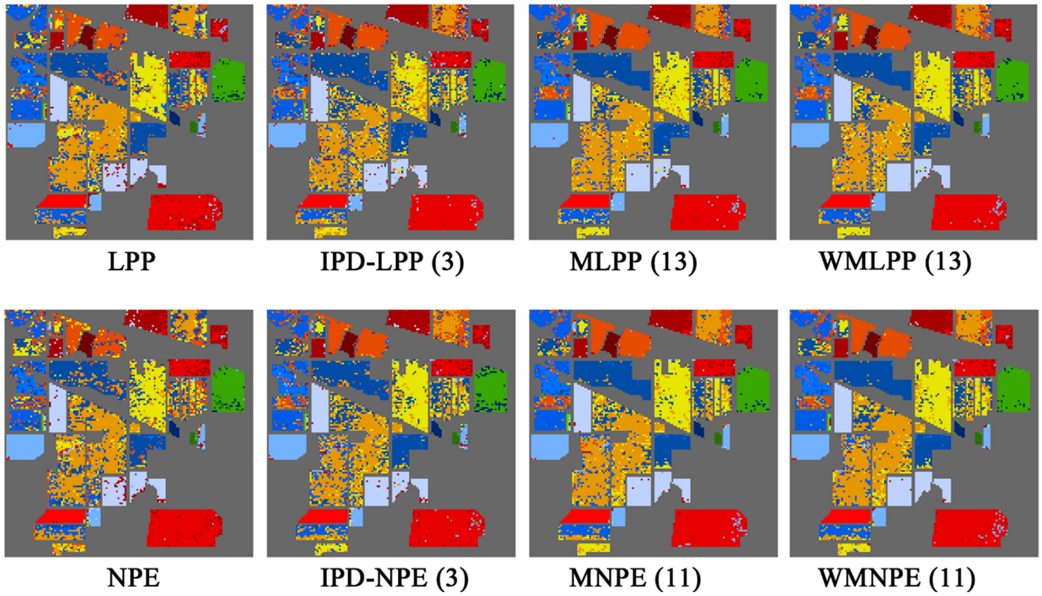

Fig. 11. Classification maps for patch-based DR results with highest overall accuracy for Indian Pines, AVIRIS image (numbers in parentheses are the corresponding window sizes for the highest overall accuracy).

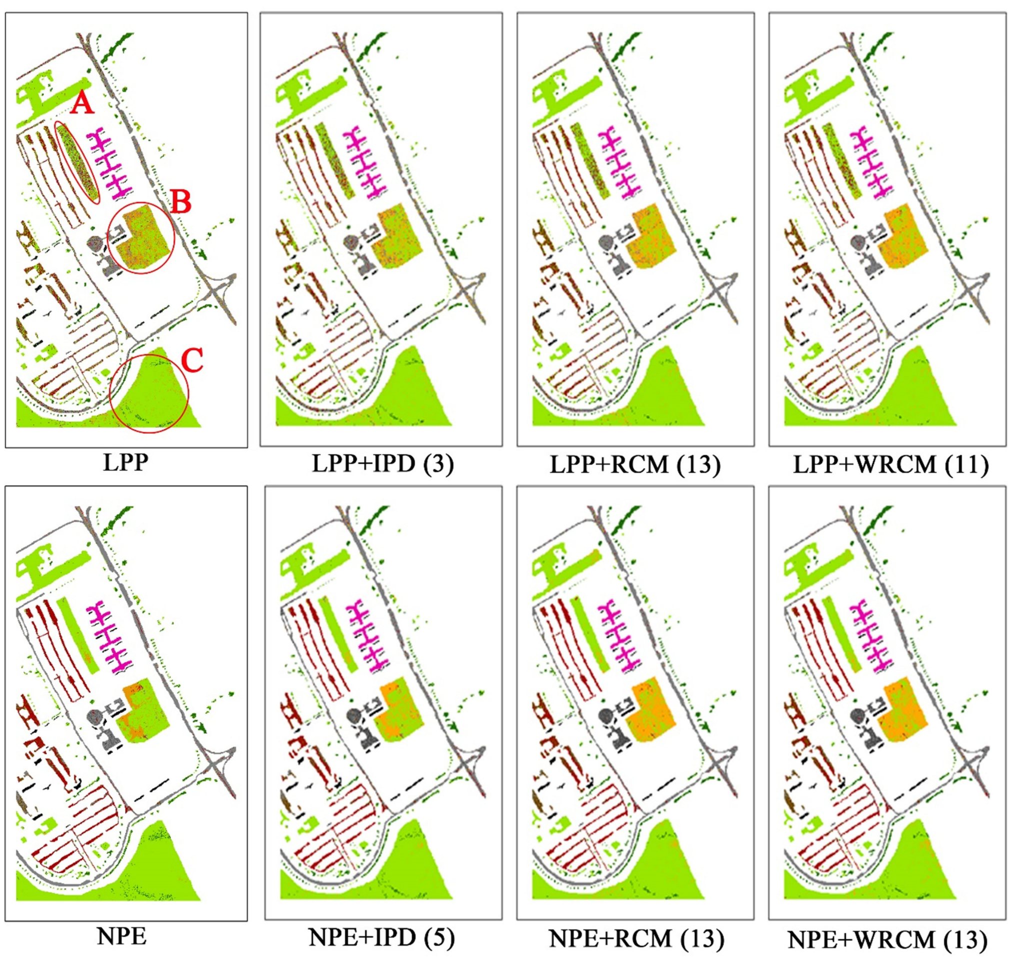

Fig. 12. Classification maps for patch-based DR results with highest overall accuracy for University of Pavia, ROSIS image (numbers in parentheses are the corresponding window sizes for the highest overall accuracy).

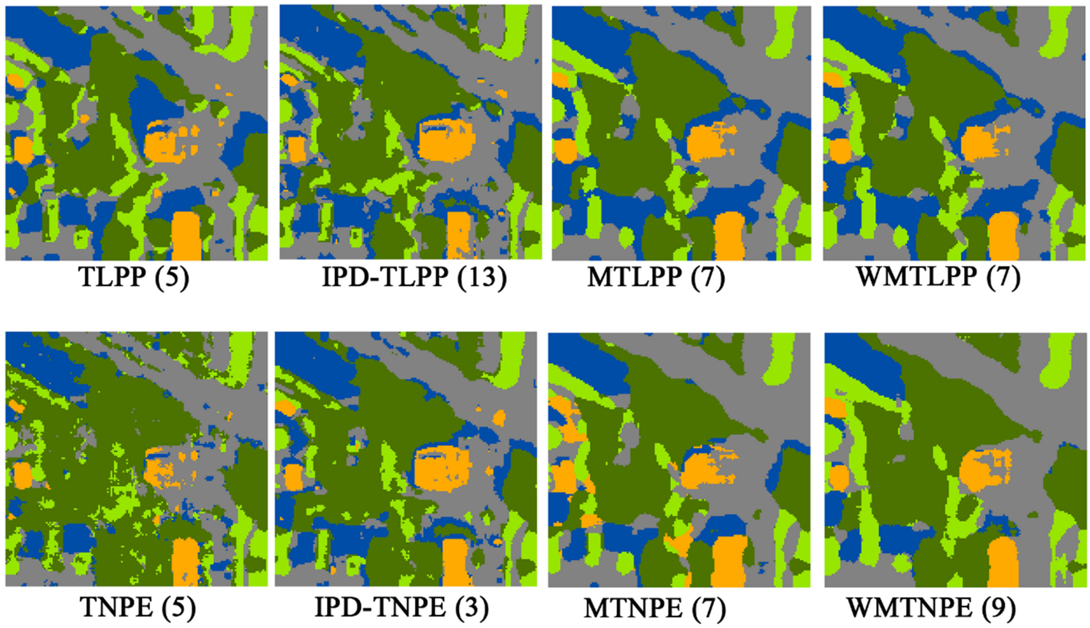

Fig. 13. Classification maps for tensor-patch-based DR results with highest overall accuracy for Surrey, BC, CASI image (numbers in parentheses are the corresponding window sizes for the highest overall accuracy).

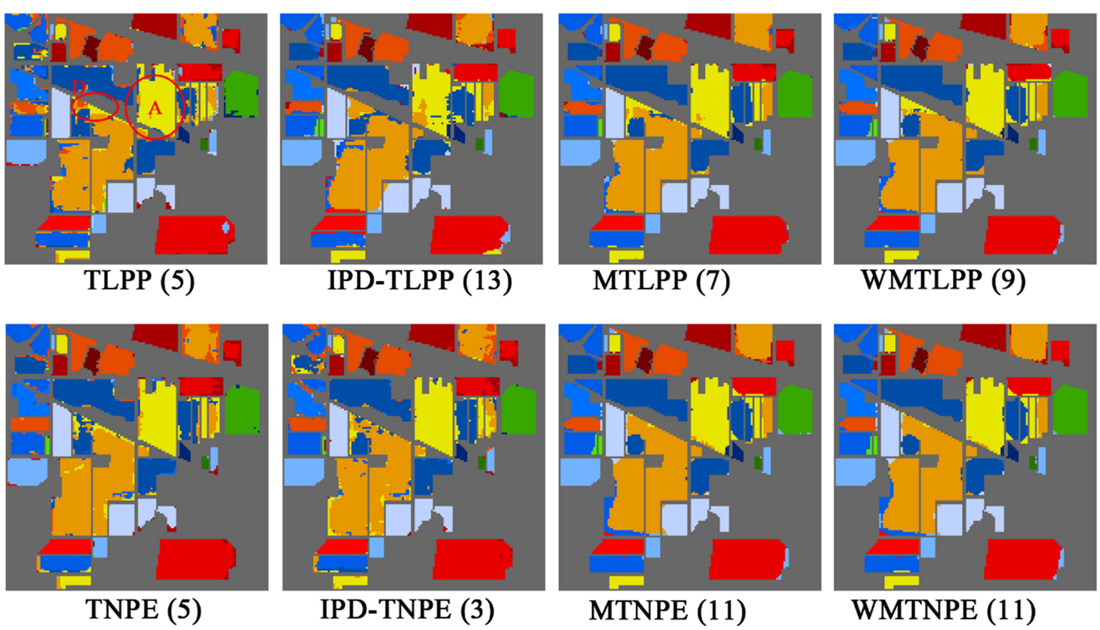

Fig. 14. Classification maps for tensor-patch-based DR results with highest overall accuracy for Indian Pines, AVIRIS image (numbers in parentheses are the corresponding window sizes for the highest overall accuracy).

IPD-based method. In the University of Pavia, ROSIS image classification maps (see Fig. 12), errors in Circle A and B have been mitigated on different levels compared to the classification maps in Fig. 9. However, errors in circle C stay.

C. Tensor-Patch-Based DR Results

Four TLPP-based and four TNPE-based DR methods were used to derive a group of dimension-reduced images. The results were then classified by the SVM classifier The best classification results are shown in Fig. 13. for Surrey, BC, and CASI image, in Fig. 14. for Indian Pines, AVIRIS image, and in Fig. 15. for University of Pavia, ROSIS image. Compared to the above classification results from patch-based methods, the classification results from tensor-patch-based methods suffer much less from small misclassified areas in all of the studied images. In Fig. 13, the problem of misclassified trees under shadows that appears at different levels in the results of PCA, LPP, and NPE methods

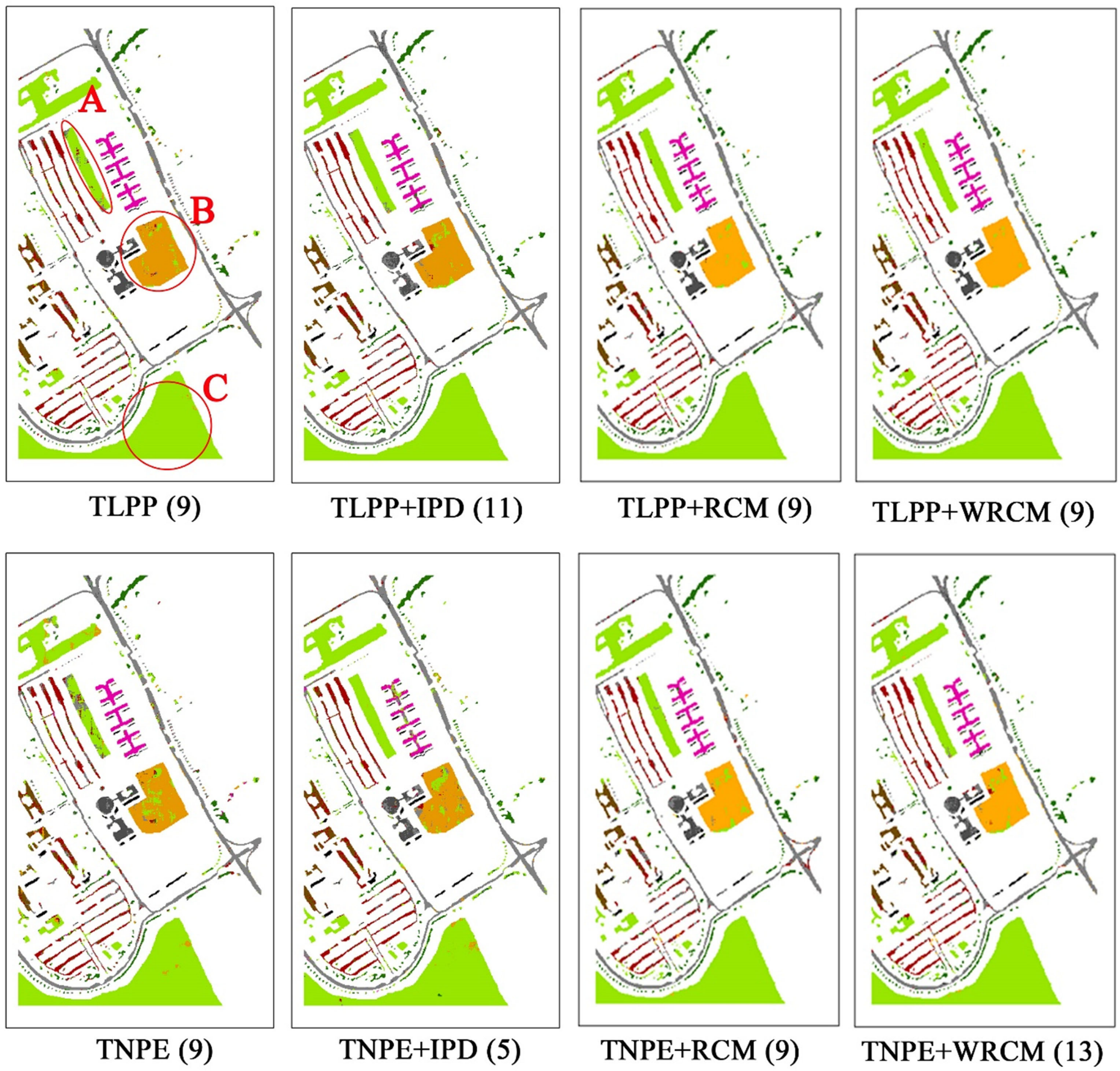

Fig. 15. Classification maps for tensor-patch-based DR results with highest overall accuracy for University of Pavia, ROSIS image (numbers in parentheses are the corresponding window sizes for the highest overall accuracy).

is greatly alleviated (circles A and B). The misclassification of painted asphalt is alleviated in TLPP, TNPE, and their IPD versions, and is completely avoided in RCM- and WRCM-based TLPP and TNPE methods (circle C). The misclassifications of concrete and dark shingle roofs (circles D and E) and grass (circle F) are also eliminated in the tensor-patch-based methods compared to patch-based methods. In Fig. 14, the small misclassified areas within most of the croplands are avoided. In circles A and D, TNPE-based results show better accuracy than the TLPP-based method. In Fig. 15, the misclassification happening in circles A, B, and C have been greatly improved from the patch-based results in Fig. 11.

IV. Discussion

From the classification results, we can deduce that both patch- and tensor-patch-based DR methods increase the OA, AA, and Kappa coefficient from the traditional and up-to-date DR methods provided in 3.2 Classification results for comparison. The highest overall accuracies among different window sizes are given in Tables III–V for the patch-based methods and in Tables VI–VIII for tensor-patch-based methods. For the three tested images the tensor-patch-based methods achieve better classification results than the patch-based methods. The appropriate window sizes for the patch- and tensor-patch-based methods vary regarding the specific method and the target image. We found that in general, the IPD-based methods usually achieve better overall accuracy with smaller window sizes (<11), and the RCM- and WRCM-based methods usually require larger window sizes to achieve better overall accuracy (>11). The appropriate window sizes for the patch-based methods are usually larger than the appropriate window sizes for the tensor-patch-based methods. In the end, the use of the first ten principal components in the RCM/WRCM calculation provides a practical

TABLE III
OVERALL CLASSIFICATION ACCURACY AMONG PATCH-BASED DR METHODS (SURREY, BC, CASI)

		OA	AA	Kappa
Existing methods	LPP	79.6%	76.4%	0.73
	NPE	86.3%	83.6%	0.81
Best roposed methods	IPD-LPP (3)	80.2%	75.2%	0.73
	RCM-LPP (7)	87.3%	87.6%	0.82
	WRCM-LPP (7)	**87.3%**	**88.0%**	**0.83**
	IPD-NPE (9)	85.3%	85.0%	0.80
	RCM-NPE (11)	86.8%	86.1%	0.82
	WRCM-NPE (13)	87.1%	86.9%	0.84

TABLE IV
OVERALL CLASSIFICATION ACCURACY AMONG PATCH-BASED DR METHODS (INDIAN PINES, AVIRIS)

		OA	AA	Kappa
Existing methods	LPP	**79.7%**	**87.6%**	**0.77**
	NPE	73.6%	83.8%	0.70
Proposed methods	IPD-LPP (3)	77.8%	86.1%	0.75
	RCM-LPP (13)	85.6%	86.8%	0.84
	WRCM-LPP (13)	85.7%	91.3%	0.84
	IPD-NPE (3)	80.0%	88.5%	0.77
	RCM-NPE (11)	86.6%	92.5%	0.84
	WRCM-NPE (11)	**86.7%**	**92.9%**	**0.85**

TABLE V
OVERALL CLASSIFICATION ACCURACY AMONG PATCH-BASED DR METHODS (UNIVERSITY OF PAVIA, ROSIS)

		OA	AA	Kappa
Existing methods	LPP	69.6%	75.2%	0.62
	NPE	79.6%	83.0%	0.79
Proposed methods	IPD-LPP (3)	70.7%	77.9%	0.69
	RCM-LPP (13)	77.2%	85.6%	0.75
	WRCM-LPP (11)	77.8%	86.9.%	0.78
	IPD-NPE (5)	82.9%	86.1%	0.80
	RCM-NPE (13)	88.6%	90.1%	0.88
	WRCM-NPE (13)	**89.5%**	**92.6%**	**0.89**

TABLE VI
OVERALL CLASSIFICATION ACCURACY AMONG TENSOR-PATCH-BASED DR METHODS (SURREY, BC, CASI)

		OA	AA	Kappa
Existing methods	TLPP (5)	86.8%	85.1%	0.82
	IPD-TLPP (3)	89.8%	86.9%	0.86
	RCM-TLPP (12)	91.1%	90.3%	0.87
	TNPE (3)	84.0%	78.6%	0.78
Proposed methods	**WRCM-TLPP (12)**	**91.3%**	**90.2%**	**0.88**
	IPD-TNPE (5)	89.2%	86.6%	0.85
	RCM-TNPE (7)	90.6%	88.7%	0.87
	WRCM-TNPE (9)	90.6%	89.4%	0.87

solution to the high computation associated with the huge hyperspectral datasets and removes redundant information in spectral bands. Furthermore, it can be observed that WRCM-based patch/tensor methods often achieve the best overall accuracy among the three different patch-based adjacency map/weight matrix calculation methods.

TABLE VII
OVERALL CLASSIFICATION ACCURACY AMONG TENSOR-PATCH-BASED DR METHODS (INDIAN PINES, AVIRIS)

		OA	AA	Kappa
Existing methods	TLPP (5)	93.3%	96.1%	0.92
	IPD-TLPP (13)	94.7%	96.9%	0.94
	RCM-TLPP (7)	98.4%	98.8%	0.98
	TNPE (5)	95.1%	96.7%	0.94
Proposed methods	WRCM-TLPP (9)	98.5%	99.2%	0.99
	IPD-TNPE (3)	95.0%	97.0%	0.94
	RCM-TNPE (11)	98.3%	98.8%	0.98
	WRCM-TNPE (11)	**99.4%**	**99.8%**	**0.99**

TABLE VIII
OVERALL CLASSIFICATION ACCURACY AMONG TENSOR-PATCH-BASED DR METHODS (UNIVERSITY OF PAVIA, ROSIS)

		OA	AA	Kappa
Existing methods	TLPP (9)	88.9%	87.9%	0.88
	IPD-TLPP (11)	89.3%	90.2%	0.89
	RCM-TLPP (9)	93.7%	95.6%	0.91
	TNPE (9)	84.8%	85.3%	0.86
Proposed methods	**WRCM-TLPP (9)**	**94.3%**	**92.6%**	**0.93**
	IPD-TNPE (5)	87.0%	88.7%	0.89
	RCM-TNPE (9)	89.5%	90.6%	0.87
	WRCM-TNPE (13)	90.1%	91.6%	0.90

*The number in each parenthesis after the method is the corresponding window size for the highest overall accuracy among different window sizes.

V. CONCLUSION

This article proposed weight modified graph-based DR methods: WMTLPP and WMTNPE. Experiments were conducted on two airborne hyperspectral images with different spatial resolutions separately in urban and agricultural scenes. The tensor-patch-based representation employs the spatial information in the processes of generating adjacency map/weight matrix and solving the three target eigenproblems along each dimension. Three major findings were observed from the results. First, we find that the proposed weight modified graph-based DR methods: WMTLPP and WMTNPE by preserving more spatial information outperform the PCA and the two up-to-date methods. Second, in the experiments in this article, LPP- and TLPP-based methods outperformed the NPE- and TNPE-based methods in the Indian Pines image; and the NPE- and TNPE-based methods outperformed the LPP and TLPP-based methods in the Surrey, BC image. Third, the proposed use of the principle components in the RCM and WRCM calculation provided reasonable results, while computing time and data redundancy were reduced.

ACKNOWLEDGMENT

We would like to thank the City of Surrey for providing the airborne hyperspectral images, which are captured by the CASI-1500 sensor during May 2013.

REFERENCES

[1] J. Xia, N. Falco, J. A. Benediktsson, P. Du, and J. Chanussot, "Hyperspectral image classification with rotation random forest via KPCA," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 4, pp. 1601–1609, Apr. 2017.

[2] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1923–1938, Apr. 2019.

[3] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "Cospace: Common subspace learning from hyperspectral-multispectral correspondences," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4349–4359, Jul. 2019.

[4] L. He *et al.*, "Pansharpening via detail injection based convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 4, pp. 1188–1204, Apr. 2019.

[5] J. Khodr and R. Younes, "Dimensionality reduction on hyperspectral images: A comparative review based on artificial datas," in *Proc. 4th Int. Congr. Image Signal Process.*, 2011, pp. 1875–1883.

[6] M. Cui, S. Prasad, W. Li, and L. M. Bruce, "Locality preserving genetic algorithms for spatial-spectral hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 3, pp. 1688–1697, Jun. 2013.

[7] Y. Zhou, J. Peng, and C. P. Chen, "Dimension reduction using spatial and spectral regularized local discriminant embedding for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 1082–1095, Feb. 2014.

[8] H. Pu, Z. Chen, B. Wang, and G.-M. Jiang, "A novel spatial–spectral similarity measure for dimensionality reduction and classification of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7008–7022, Nov. 2014.

[9] Y. Zhou, J. Peng, and C. P. Chen, "Dimension reduction using spatial and spectral regularized local discriminant embedding for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 1082–1095, Feb. 2015.

[10] W. Zhao and S. Du, "Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544–4554, Aug. 2016.

[11] Y. Gao, X. Wang, Y. Cheng, and Z. J. 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.

[12] 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, pp. 452–471, 2017.

[13] J. An, Y. Song, Y. Guo, X. Ma, and X. Zhang, "Tensor Discriminant Analysis via compact feature representation for hyperspectral images dimensionality reduction," *Remote Sens.*, vol. 11, pp. 1822–1840, 2019.
[14] 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.
[15] D. Xu, S. Yan, L. Zhang, S. Lin, H.-J. Zhang, and T. S. Huang, "Reconstruction and recognition of tensor-based objects with concurrent subspaces analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 1, pp. 36–47, Jan. 2008.
[16] Y. Luo, D. Tao, K. Ramamohanarao, C. Xu, and Y. Wen, "Tensor canonical correlation analysis for multi-view dimension reduction," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 11, pp. 3111–3124, Nov. 2015.
[17] 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, Oct. 2010.
[18] N. Renard, S. Bourennane, and J. Blanc-Talon, "Denoising and dimensionality reduction using multilinear tools for hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 138–142, Apr. 2008.
[19] J. An, J. Lei, Y. Song, X. Zhang, and J. Guo, "Tensor based multiscale low rank decomposition for hyperspectral images dimensionality reduction," *Remote Sens.*, vol. 11, pp. 1485–1502, 2019.
[20] M. Jouni, M. Dalla Mura, and P. Comon, "Classification of hyperspectral images as tensors using nonnegative CP decomposition," in *Proc. Int. Symp. Math. Morphology Appl. Signal Image Process.*, 2019, pp. 189–201.
[21] R. Yan, J. Peng, D. Ma, and D. Wen, "Spectral tensor synthesis analysis for hyperspectral image spectral–spatial feature extraction," *J. Indian Soc. Remote Sens.*, vol. 47, pp. 91–100, 2019.
[22] C. Dantas, J. Cohen, and R. Gribonval, "Learning tensor-structured dictionaries with application to hyperspectral image denoising," in *Proc. 27th Eur. Signal Process. Conf.*, 2019, pp. 1–5.
[23] T. Zhang, D. Tao, X. Li, and J. Yang, "Patch alignment for dimensionality reduction," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1299–1313, Sep. 2009.
[24] 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.
[25] S. Velasco-Forero and J. Angulo, "Classification of hyperspectral images by tensor modeling and additive morphological decomposition," *Pattern Recognit.*, vol. 46, pp. 566–577, 2013.
[26] B. Du, M. Zhang, L. Zhang, R. Hu, and D. Tao, "PLTD: Patch-based low-rank tensor decomposition for hyperspectral images," *IEEE Trans. Multimedia*, vol. 19, no. 1, pp. 67–79, Jan. 2017.
[27] J. A. Lee and M. Verleysen, "Graph-based dimensionality reduction," 2012.
[28] 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.
[29] S.-E. Qian and G. Chen, "A new nonlinear dimensionality reduction method with application to hyperspectral image analysis," in *Proc. IEEE Int. Geosci. Rem. Sens. Symp.*, 2007, pp. 270–273.
[30] J. A. Lee and M. Verleysen, *Nonlinear Dimensionality Reduction*. Berlin, Germany: Springer, 2007.
[31] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, 2000.
[32] X. He, D. Cai, S. Yan, and H.-J. Zhang, "Neighborhood preserving embedding," in *Proc. 10th IEEE Int. Conf. Comput. Vis.*, 2005, pp. 1208–1213.
[33] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, pp. 1373–1396, 2003.
[34] X. He and P. Niyogi, "Locality preserving projections," in *Proc. Advances Neural Inf. Process. Syst.*, 2004, pp. 153–160.
[35] A. Mohan, G. Sapiro, and E. Bosch, "Spatially coherent nonlinear dimensionality reduction and segmentation of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 2, pp. 206–210, Apr. 2007.
[36] D. Hong, N. Yokoya, and X. X. Zhu, "Learning a robust local manifold representation for hyperspectral dimensionality reduction," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 6, pp. 2960–2975, Jun. 2017.
[37] Z. Zhong *et al.*, "Discriminant tensor spectral–spatial feature extraction for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, pp. 1028–1032, 2015.
[38] Y. Gao, X. Wang, Y. Cheng, and Z. J. 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. 5, pp. 1582–1593, May 2015.
[39] G. Dai and D.-Y. Yeung, "Tensor embedding methods," in *Proc. Assoc. Adv. Artif. Intell.*, 2006, pp. 330–335.
[40] X. He, D. Cai, and P. Niyogi, "Tensor subspace analysis," in *Proc. Advances Neural Inf. Process. Syst.*, 2006, pp. 499–506.
[41] O. Tuzel, F. Porikli, and P. Meer, "Region covariance: A fast descriptor for detection and classification," in *Proc. Eur. Conf. Computer Vision*, 2006, pp. 589–600.
[42] H. Qin, L. Qin, L. Xue, and Y. Li, "A kernel Gabor-based weighted region covariance matrix for face recognition," *Sensors*, vol. 12, pp. 7410–7422, 2012.

Boyu Feng (Member, IEEE) received the B.E. degree in geomatics from the China University of Geosciences, Wuhan, China, in 2012, the M.S. degree. From University of Kansas, Lawrence, KS, USA, and the Ph.D. degree from the Department of Geography, University of Western Ontario, London, ON, Canada, with a focus on tensor-based image processing.

From 2013 to 2014, she was with the Center for Remote Sensing of Ice Sheet, University of Kansas. Her current research interests lie in advanced image processing algorithms for urban and forest environments.

Jinfei Wang (Member, IEEE) received the B.S. and M.Sc. degrees from Peking University, Beijing, China, in 1982 and 1984, respectively, and the Ph.D. degree in Geography from University of Waterloo, Waterloo, ON, Canada, in 1989.

She is currently a Professor with the Department of Geography, the University of Western Ontario, London, ON, Canada. Her current research interests include methods for information extraction from high resolution remotely sensed imagery, land use and land cover monitoring in urban environments and agricultural crop monitoring using multiplatform multispectral, hyperspectral, lidar and radar data, and unmanned aerial vehicle data.