# Tensor Discriminative Locality Alignment for Hyperspectral Image Spectral–Spatial Feature Extraction

Liangpei Zhang, *Senior Member, IEEE*, Lefei Zhang, *Student Member, IEEE*, Dacheng Tao, *Senior Member, IEEE*, and Xin Huang

***Abstract*—In this paper, we propose a method for the dimensionality reduction (DR) of spectral–spatial features in hyperspectral images (HSIs), under the umbrella of multilinear algebra, i.e., the algebra of tensors. The proposed approach is a tensor extension of conventional supervised manifold-learning-based DR. In particular, we define a tensor organization scheme for representing a pixel's spectral–spatial feature and develop tensor discriminative locality alignment (TDLA) for removing redundant information for subsequent classification. The optimal solution of TDLA is obtained by alternately optimizing each mode of the input tensors. The methods are tested on three public real HSI data sets collected by hyperspectral digital imagery collection experiment, reflective optics system imaging spectrometer, and airborne visible/infrared imaging spectrometer. The classification results show significant improvements in classification accuracies while using a small number of features.**

***Index Terms*—Classification, feature extraction, hyperspectral image (HSI), remote sensing, tensor.**

## I. Introduction

HYPERSPECTRAL imaging sensors, e.g., the hyperspectral digital imagery collection experiment (HYDICE) [1], can collect an image in which each pixel has contiguous bands of spectra [2]. This data product from a hyperspectral sensor is a "cube" data, which has two spatial dimensions (width and height) and a spectral dimension. For hyperspectral image (HSI) analysis and processing [3], previous researchers have demonstrated that the redundancy from interband correlation is very high and the data structure in the spectral dimension can be reduced without a significant loss of the useful information for subsequent utilizations [4]–[7]. Thus, there is a need for dimensionality reduction (DR) technologies that can: 1) reduce the redundancy among features; 2) preserve the discriminative information that is important for the subsequent classification; and 3) decrease the computational cost.

In general, DR for HSI can be achieved in essentially two ways: feature selection and feature extraction [8]. The problem of feature selection is defined as how to select a subset from the given candidate features, which performs optimally for the particular classifiers [9], [10]. Some representative feature selection techniques are the branch and bound method [11], genetic algorithm [12], clonal selection algorithm [13], and Fisher's score [14]. In this paper, we focus on feature extraction [14]–[16], which aims to find a transformation from a higher dimensional space to a lower dimensional subspace, with most of the desired information content preserved [17]. The most commonly used type of algorithm is principal component analysis (PCA) [18], a linear transformation to find principal components, in accordance with the maximum variance of a data matrix. Another popular transformation in HSI processing is maximum noise fraction [19], in which the transformed principal components are ranked by SNR. In recent years, manifold-learning-inspired feature-extraction methods have been widely applied in image processing areas [20]–[22] and on HSI data analysis in particular. Some studies have demonstrated the potential value of manifold learning for applications such as feature extraction [23], [24], classification [25], [26], and anomaly target detection [27], [28].

However, the aforementioned DR technologies only deal with a set of first-order data as inputs, i.e., the vector representation, which is commonly used to represent the spectral feature of a certain pixel in HSI [29]–[32]. Although the spectral feature vector contains a lot of information about the spectral properties of the pixel, the spectral-vector-based HSI analyses just process each pixel independently, without considering the spatial relationship of neighboring pixels. To overcome this point, some researchers have suggested using spectral information as well as spatial information to achieve further improvement in classification performance [33]–[37]. These studies have verified the enhanced performance obtained when combining spectral and spatial features, but there is an obvious shortcoming in that they still put the spectral–spatial feature into first-order data for analysis and neglect the spectral and spatial rearrangements of features. In fact, second-order data, such as gray-level images in computer vision and pattern recognition [38], third-order data, such as HSI in remote

Manuscript received September 13, 2011; revised February 21, 2012; accepted April 25, 2012. Date of publication June 6, 2012; date of current version December 19, 2012. This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2011CB707105, by the National Natural Science Foundation of China under Grants 40930532, 41101336, and 41061130553, and by the Australia Research Council Discovery Project under Grant ARC DP-120103730.

L. Zhang, L. Zhang, and X. Huang are with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: zlp62@public.wh.hb.cn; zhanglefei@whu.edu.cn; huang_whu@163.com).

D. Tao is with the Centre for Quantum Computation and Intelligent Systems and the Faculty of Engineering and Information Technology, University of Technology, Sydney, NSW 2007, Australia (e-mail: dacheng.tao@uts.edu.au).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2012.2197860

sensing [39]–[42], and high-order data, such as multifeature-represented local patches in HSI [43], can be unified into a tensor representation [44]. Under the umbrella of multilinear algebra [45], various tensor subspace learning algorithms have been developed [46]–[49], as pioneered by the work on "TensorFaces" [50]. Recent studies have also constructed some tensor-embedding frameworks for DR, which provide a multilinear projection with high-order tensor data that respects some kinds of local geometrical structure of the manifold [51]–[54].

Inspired by the successful work of the aforementioned tensor subspace learning algorithms and our previous patch alignment framework [55], [56], in this paper, for the first time, we propose a supervised manifold-learning algorithm, under the umbrella of multilinear algebra, i.e., tensor discriminative locality alignment (TDLA) for hyperspectral remote sensing image spectral–spatial feature representation and DR. The advantages of our method lie in the following three aspects.

1) The spectral–spatial information of the pixel is preserved. As pointed out by this paper, tensor representation preserves as many as possible the original spatial constraints of a certain pixel and its neighbors, which helps to better represent the pixel's spectral–spatial feature. Compared to the vector-based feature representation, such structural information in the tensor feature is a reasonable constraint to reduce the number of unknown parameters used in learning a feature DR model.
2) The discriminability of classes for classification is preserved. The neighboring samples of both the same classes and different classes are considered in the proposed TDLA optimization, so the discriminative information can be maintained. Apart from this key point, the proposed approach can also deal with the nonlinear nature of the sample distribution by taking into account the locality manifold of the samples, which can also help to achieve a better classification performance than the conventional linear DR methods.
3) A generalized DR framework for high-order data is provided. We propose a DR framework to accept high-order data, such as a collection of third-order data cubes or some high-order feature data inputs, as the inputs for DR. It will be shown that the previous DR [55] is a special case of our tensor method when the input data are first-order vectors.

The rest of this paper is structured as follows. In the next section, we give a brief introduction to tensor algebra. Section III describes the proposed tensor representation for the HSI spectral–spatial feature. In Section IV, we describe the TDLA algorithm in detail. Finally, the experiments are reported in Section V, followed by the conclusion in Section VI.

## II. Tensor Algebra

Tensors [38], [44] are multidimensional arrays of numbers that transform linearly under coordinate transformations, which can be represented as $X \in R^{L_1 \times L_2 \times \cdots \times L_M}$ with multilinear algebra [45] defined on them. Here, $M$ is the order of the tensor, the $i$th order of the tensor is of size $L_i$, and each order is also called the $i$th mode. An arbitrary element of $X$ is a scalar denoted by $X_{l_1,l_2,\ldots,l_M}$, where $1 \leqslant l_i \leqslant L_i$ and $1 \leqslant i \leqslant M$, and $l_i$ denotes the location of this element in mode $i$. Here, we briefly introduce the following relevant definitions in multilinear algebra.

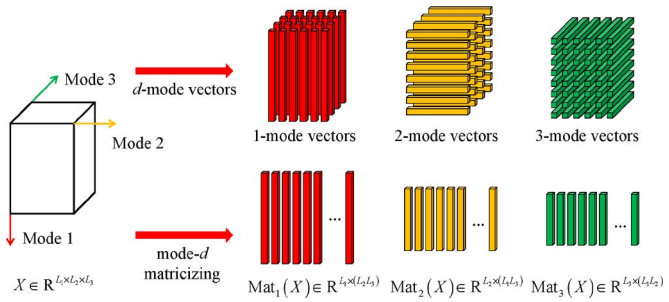



Fig. 1. Visual illustration of mode-$d$ matricizing and $d$-mode vectors of a third-order tensor.

1) Mode-$d$ matricizing ($d$-mode vectors): Defined as unfolding an $M$-order tensor $X \in R^{L_1 \times L_2 \times \cdots \times L_d \times \cdots \times L_M}$ to a matrix $Mat_d(X) \in R^{L_d \times \overline{L_d}}$ in which $\overline{L_d} = \prod_{i=1, i \neq d}^{M} L_i$, by keeping the index $l_d$ fixed and varying the other indices. The column vectors of the resulting matrix $Mat_d(X)$ is a set of $d$-mode vectors of size $L_d$, which can also be obtained by varying its index $l_d$ while keeping all the other indices fixed. A visual illustration of mode-$d$ matricizing and $d$-mode vectors on a third-order tensor (data cube) is shown in Fig. 1.
2) Mode-$d$ product ($_dU$): The mode-$d$ product $X \times {}_dU$ of tensor $X \in R^{L_1 \times L_2 \times \cdots \times L_d \times \cdots \times L_M}$ and matrix $U \in R^{L'_d \times L_d}$ is a tensor $(X \times {}_dU) \in R^{L_1 \times L_2 \times \cdots \times L_{d-1} \times L'_d \times L_{d+1} \times \cdots \times L_M}$ defined by

$$(X \times {}_dU)_{l_1,l_2,\ldots,l_{d-1},l'_d,l_{d+1},\ldots,l_M} = \sum_{l_d=1}^{L_d} \left( X_{l_1,l_2,\ldots,l_d,\ldots,l_M} U_{l'_d,l_d} \right). \tag{1}$$

To simplify the notation in this paper, when the $M$ times of mode-$d$ products are conducted on all modes, i.e., $X \times {}_1U_1^{\mathrm{T}} \times \cdots \times {}_MU_M^{\mathrm{T}}$, we denote this procedure as $X \prod_{i=1}^{M} \times_i U_i^{\mathrm{T}}$. When the $(M-1)$ times of mode-$d$ products are conducted on all modes except the $k$th mode, i.e., $X \times {}_1U_1^{\mathrm{T}} \times \cdots \times {}_{k-1}U_{k-1}^{\mathrm{T}} \times {}_{k+1}U_{k+1}^{\mathrm{T}} \times \cdots \times {}_MU_M^{\mathrm{T}}$, we denote this procedure as $X \prod_{i \neq k}^{M} \times_i U_i^{\mathrm{T}}$.
3) Tensor contraction: The contraction of tensors $X \in R^{L_1 \times L_2 \times \cdots \times L_M \times L'_1 \times L'_2 \times \cdots \times L'_{M'}}$ and $Y \in R^{L_1 \times L_2 \times \cdots \times L_M \times L''_1 \times L''_2 \times \cdots \times L''_{M''}}$ is defined as

$$[X \otimes Y; (1:M)(1:M)]_{l_1,l_2,\ldots,l_M} = \sum_{l_1=1}^{L_1} \cdots \sum_{l_M=1}^{L_M} (X)_{l_1,l_2,\ldots,l_M,l'_1,l'_2,\ldots,l'_{M'}} (Y)_{l_1,l_2,\ldots,l_M,l''_1,l''_2,\ldots,l''_{M''}}. \tag{2}$$

The condition for the contraction is that tensors $X$ and $Y$ are of the same size at the specific mode, which must

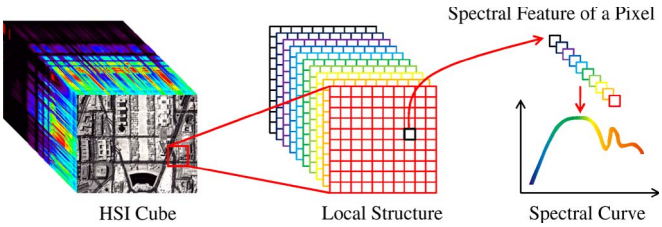



Fig. 2. Structure of the hyperspectral data cube.

be clearly indicated, e.g., in (2); $(1:M)$ suggests that the specific mode for contraction is from mode 1 to $M$, and this contraction reduces the tensor order by $2M$. To simplify the notation in this paper, when the contraction is conducted on all indices except for the index $k$ on the tensors $X, Y \in R^{L_1 \times L_2 \times \cdots \times L_M}$, we denote this procedure as $[X \otimes Y; (\overline{k})(\overline{k})]$. As a property of tensor contraction, we have [56]

$$\left[X \otimes Y; (\overline{k})(\overline{k})\right] = Mat_k(X) \cdot Mat_k^{\mathrm{T}}(Y). \tag{3}$$

In particular, the contraction of tensors $X, Y \in R^{L_1 \times L_2 \times \cdots \times L_M}$ from mode 1 to $M$ is scalar by definition (2), which is also called the inner product of tensors $X$ and $Y$.

4) Frobenius norm: The Frobenius norm of a tensor $X \in R^{L_1 \times L_2 \times \cdots \times L_M}$ is given by

$$\|X\| = \sqrt{[X \otimes X; (1:M)(1:M)]} = \sqrt{\sum_{l_1=1}^{L_1} \cdots \sum_{l_M=1}^{L_M} X_{l_1,\ldots,l_M}^2}. \tag{4}$$

5) Euclidean distance: The Euclidean distance between two tensors $X, Y \in R^{L_1 \times L_2 \times \cdots \times L_M}$ is given by

$$D(X, Y) = \|X - Y\|. \tag{5}$$

## III. HSI Spectral–Spatial Tensor Representation

Hyperspectral remote sensing images are composed of hundreds of spatially coregistered gray images, each of which is acquired in a particular spectral channel. Fig. 2 shows the spectral–spatial structure of a typical HSI data cube. For each pixel in the image, the spectral feature is obtained by scanning the digital numbers in all bands into a vector; this spectral feature is known as a spectral curve. For a routine pixel-based classification technique, each pixel is processed independently using such a spectral feature, without considering the spatial constraints.

In this paper, by introducing a tensor for feature representation, several kinds of $k$ nearest neighbor ($k$-NN) pixels are adopted to represent the local spatial information of the centric pixel, as shown in Fig. 3.

For an arbitrary pixel $a$, suppose its spectral vector is $x_a \in R^L$, where $L$ is the number of bands in HSI. The proposed tensor spectral–spatial feature representation for the analyzed pixel $a$ is constructed by

$$X_a^k = [x_a, x_{a1}, x_{a2}, \ldots, x_{ak}] \in R^{L \times (k+1)} \tag{6}$$

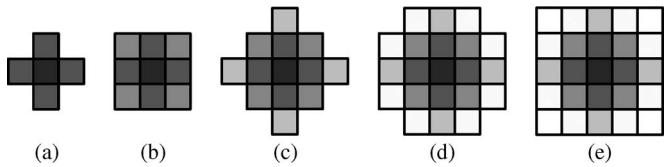



Fig. 3. Several kinds of $k$-NN local spatial structures of an analyzed pixel. (a) 4-NN. (b) 8-NN. (c) 12-NN. (d) 20-NN. (e) 24-NN.

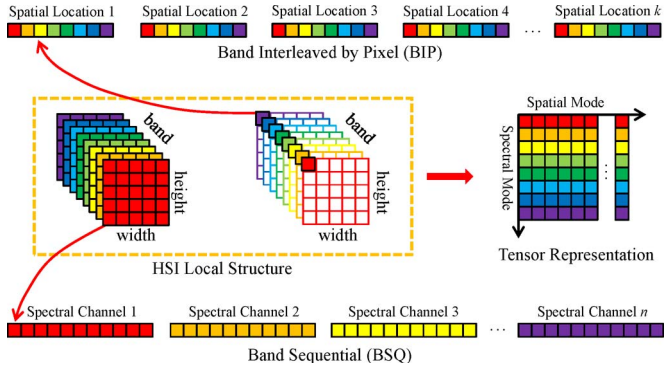



Fig. 4. Vector and tensor representations for a local patch in a hyperspectral data cube.

where $x_{ai}$ $(i = 1, 2, \ldots, k)$ is the spectral vector of pixels in the $k$-NNs. It is worth emphasizing that if the vector representation is adopted, we have to rearrange such a local patch of pixel $a$ to a first-order vector. There are two main approaches for such a rearrangement: band interleaved by pixel (BIP) and band sequential (BSQ). Fig. 4 shows the detailed data organizational structures of the vector and tensor representations. When we compare the tensor representation to BIP and BSQ, we observe that the detailed elements are identical; however, they have entirely different organizational structures. In the BIP representation, the spatially connected constraint among local pixels is lost, while in the BSQ representation, the spectrally connected constraint among spectral channels of a certain pixel is lost. However, the proposed tensor representation has two modes, in order to keep the original spatial structure connected in one mode and the spectral channels connected in the other mode, which helps to better represent the pixel's spectral–spatial feature.

By preserving as many as possible the original spectral–spatial constraints, tensor representation helps to reduce the number of unknown parameters used in learning a linear DR model. Here, we consider the tensor-based linear DR algorithm

$$Y = X \times_1 U_1^{\mathrm{T}} \times_2 U_2^{\mathrm{T}}. \tag{7}$$

Based on multilinear algebra, we have $U_1 \in R^{L_1 \times d_1}$ and $U_2 \in R^{L_2 \times d_2}$, in which $L_1$ and $L_2$ are the original tensor sizes in spectral and spatial modes and $d_1$ and $d_2$ are the reduced feature sizes, respectively. Thus, there are $\rho_1 = L_1 d_1 + L_2 d_2$ independent unknown parameters in the tensor-based DR model.

We now consider a vector-based linear DR algorithm

$$y = U^{\mathrm{T}} x. \tag{8}$$

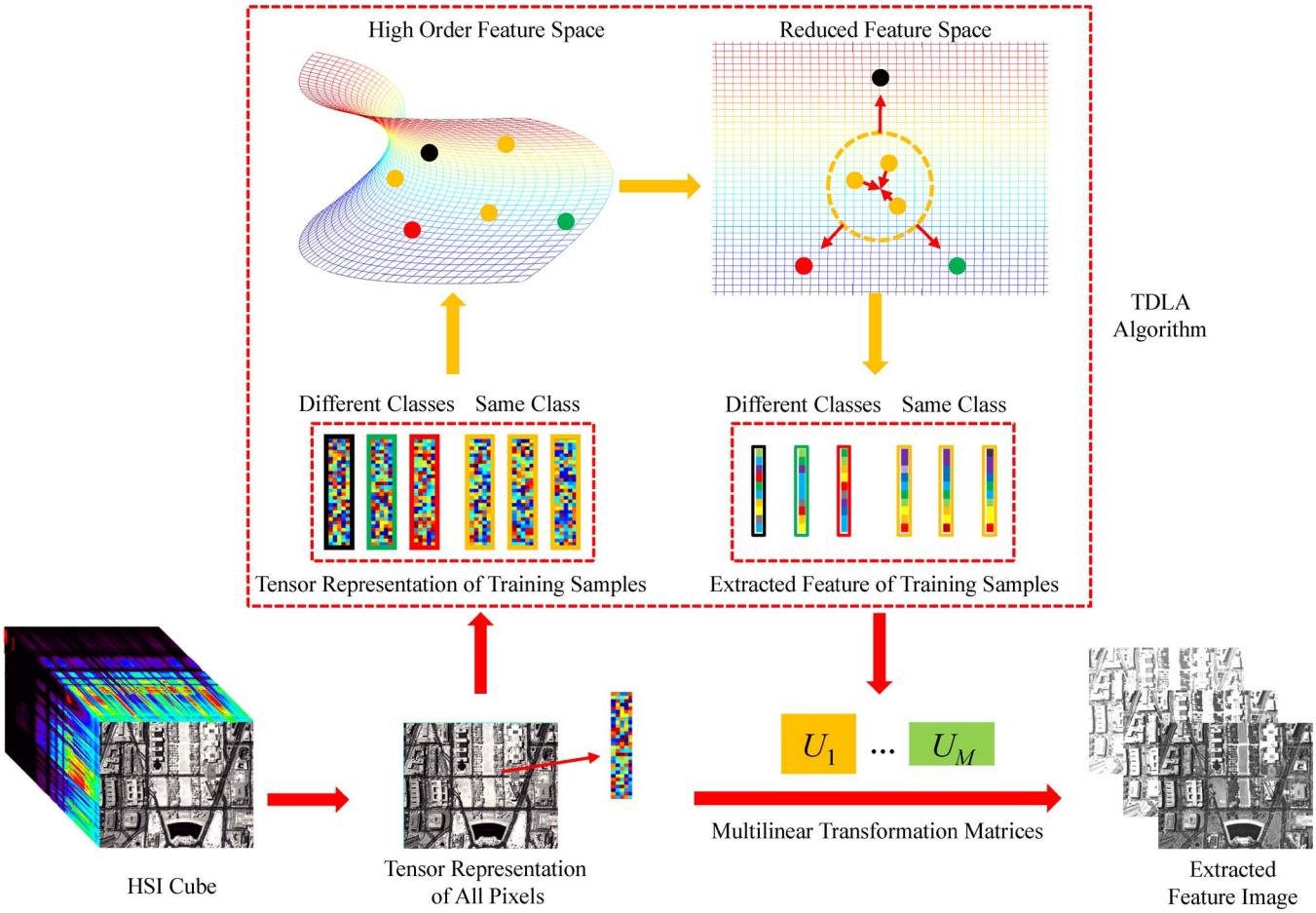



Fig. 5. Flowchart of the proposed approach.

Based on Fig. 4, we have $x \in R^{L_1 L_2}$, using either BIP or BSQ. According to [38], we know that if $U = U_1 \otimes U_2$ (here, we use $\otimes$ to denote the Kronecker product), then $y$ is equal to the vectorization of $Y$, i.e., the vector-based DR algorithm equals the tensor-based DR algorithm. However, the size of the projection matrix should be $U \in R^{L_1 L_2 \times d_1 d_2}$, which indicates that $\rho_2 = L_1 L_2 d_1 d_2$ unknown parameters are required in the vector-based DR model. Obviously, we can see that $\rho_1 \ll \rho_2$. Therefore, tensor representation helps reduce the number of parameters needed in the DR model. In statistical learning, we require a larger number of training samples to learn a reasonable solution when the model has a larger number of unknown parameters. So, tensor representation can learn a better DR model using limited training samples, which helps to improve the performance in the subsequent classification.

## IV. TDLA Algorithm

The proposed TDLA algorithm for HSI spectral–spatial feature extraction is a pixel-based approach, which can be divided into three main components, as shown in Fig. 5. In the first step, each pixel in HSI is processed using the spectral–spatial tensor representation proposed in Section III. Then, the TDLA algorithm, which preserves the discriminability of classes for the subsequent classification, is employed to obtain a multilinear transformation from the original high-order tensor space to the reduced feature space. Finally, the extracted feature representation in reduced feature space is achieved by the optimized multilinear transformation for each pixel of HSI.

In this DR framework, the proposed TDLA algorithm finds a multilinear transformation from the original high-order feature space $X \in R^{I_1 \times I_2 \times \cdots \times I_M}$ to the reduced feature space $X^* \in R^{P_1 \times P_2 \times \cdots \times P_M}$

$$X^* = X \prod_{i=1}^{M} \times_i U_i^{\mathrm{T}} \tag{9}$$

where $U_i \in R^{I_i \times P_i}$ $(i = 1, 2, \ldots, M)$ and $P_i \leqslant I_i$ are a set of projection matrices in the multilinear transformation. It is worth emphasizing that if $P_i = 1$, then the $i$th projection matrix degenerates to a projection vector $U_i \in R^{I_i}$; therefore, the resulting output tensor $X^*$ reduces its order by one.

The input data of the TDLA algorithm is a set of training samples $X_i$ $(i = 1, 2, \ldots, N)$ and their class labels $y_i$ $(y_i \in [1, 2, \ldots, C])$, where $N$ is the number of samples and $C$ is the number of classes. For each sample $X_i$, we divide the other $N - 1$ samples $X_j$ $(i = 1, 2, \ldots, i-1, i+1, \ldots, N)$ into two groups by their class labels: the samples of the same class and the samples of different classes. Then, we sort the two groups of samples by the Euclidean distance $D(X_i, X_j)$. The $n_1$ nearest samples of the group "same class" and $n_2$ nearest samples of

the group "different classes" are selected to build a patch of sample $X_i$

$$Patch(X_i) = \left\{X_i, S_{i(1)}, \ldots, S_{i(n_1)}, D_{i(1)}, \ldots, D_{i(n_2)}\right\} \in R^{I_1 \times I_2 \times \cdots \times I_M \times (1+n)} \quad 10$$

where $n = n_1 + n_2$, $S_{i(j)}$ $(j = 1, 2, \ldots, n_1)$ is the $j$th sample in the same class, and $D_{i(j)}(j = 1, 2, \ldots, n_2)$ is the $j$th sample in a different class.

For each $Patch(X_i)$, the corresponding representation in the output-reduced feature space is denoted by

$$Patch\left(X_i^*\right) = \left\{X_i^*, S_{i(1)}^*, \ldots, S_{i(n_1)}^*, D_{i(1)}^*, \ldots, D_{i(n_2)}^*\right\} \in R^{P_1 \times P_2 \times \cdots \times P_M \times (1+n)}. \quad (11)$$

In the output-reduced feature space, in order to preserve the discriminability of classes for classification, TDLA suggests that distances between $X_i$ and samples in the same class are as small as possible, while the distances between $X_i$ and samples in a different class are as large as possible, as shown in Fig. 5. Based on this point, we have the following optimizations on patch $X_i$:

$$\arg \min_{X_i^*} \sum_{j=1}^{n_1} D^2\left(X_i^* - S_{i(j)}^*\right) \quad (12)$$

$$\arg \max_{X_i^*} \sum_{j=1}^{n_2} D^2\left(X_i^* - D_{i(j)}^*\right). \quad (13)$$

We unify (12) and (13) by introducing a combination factor $\alpha$

$$\arg \min_{X_i^*} \sum_{j=1}^{n_1} D^2\left(X_i^* - S_{i(j)}^*\right) - \alpha \sum_{j=1}^{n_2} D^2\left(X_i^* - D_{i(j)}^*\right). \quad (14)$$

In order to simplify the following derivation, we set

$$\beta = \left[\overbrace{1, \ldots, 1}^{n_1}, \overbrace{-\alpha, \ldots - \alpha}^{n_2}\right]^{\mathrm{T}}. \quad (15)$$

Then, (14) reduces to

$$\begin{aligned}
&\arg \min_{X_i^*} \sum_{j=1}^{n_1} \beta_j \cdot \left\|X_i^* - S_{i(j)}^*\right\|^2 + \sum_{j=1}^{n_2} \beta_{n_1+j} \cdot \left\|X_i^* - S_{i(j)}^*\right\|^2 \\
&= \arg \min_{X_i^*} \sum_{j=1}^{n} \beta_j \cdot \left\|X_{P_i(1)}^* - X_{P_i(j+1)}^*\right\|^2 \\
&= \arg \min_{X_i^*} \sum_{g=1}^{n+1} \sum_{h=1}^{n+1} \left( \begin{bmatrix} \sum_{j=1}^{n} \beta_j & -\beta^{\mathrm{T}} \\ -\beta & \mathrm{diag}(\beta) \end{bmatrix}_{g,h} \cdot \left[X_{P_i(g)}^* \otimes X_{P_i(h)}^*; (1:M)(1:M)\right] \right) \quad (16)
\end{aligned}$$

where $X_{P_i(k)}$ and $X_{P_i(k)}^*$ denote the $k$th tensor sample in (10) and (11), respectively. If we define

$$Q = \begin{bmatrix} \sum_{j=1}^{n} \beta_j & -\beta^{\mathrm{T}} \\ -\beta & \mathrm{diag}(\beta) \end{bmatrix} \in R^{(1+n)\times(1+n)} \quad (17)$$

for each sample $X_i$, we have a final representation of the patch optimization

$$\arg \min_{X_i^*} \sum_{g=1}^{n+1} \sum_{h=1}^{n+1} \left(Q_{g,h} \cdot \left[X_{P_i(g)}^* \otimes X_{P_i(h)}^*; (1:M)(1:M)\right]\right). \quad (18)$$

Then, the whole alignment of TDLA is obtained by summing over all the patch optimizations of the training samples from one to $N$. Because the local patch provided by (10) is unique for each sample, we must unite the samples in each patch to a unified system by assuming that the $(1 + n)$ samples in (10) are selected from the training samples

$$Whole(X) = \{X_1, \ldots, X_N\} \in R^{I_1 \times I_2 \times \cdots \times I_M \times N}. \quad (19)$$

The unification can be achieved by using a selection matrix $S_i \in R^{N \times (1+n)}$ defined by

$$S_{i(a,b)} = \begin{cases} 1, & \text{if } a = F_i\{b\} \\ 0, & \text{else} \end{cases} \quad (20)$$

where $F_i \in \{i, i_1, \ldots, i_n\}$ denotes the set of global indices of samples in $Patch(X_i)$. Then, the sum of all the patch optimizations described in (18) can be written as

$$\arg \min_{X_i^*, \ldots, X_N^*} \sum_{g=1}^{N} \sum_{h=1}^{N} \left(\Omega_{g,h} \cdot \left[X_g^* \otimes X_h^*; (1:M)(1:M)\right]\right) \quad (21)$$

in which

$$\Omega = \sum_{i=1}^{N} S_i Q_i S_i^{\mathrm{T}} \in R^{N \times N}. \quad (22)$$

The full optimization (21) aims to obtain an optimal subspace from the original high-order feature space for the subsequent classification. It is worth emphasizing that the feature mapping $X \in R^{I_1 \times I_2 \times \cdots \times I_M} \to X^* \in R^{P_1 \times P_2 \times \cdots \times P_M}$ from the original feature space to the reduced subspace can be implicit. For linearization, we simply put (9) into (21)

$$\arg \min_{U_1, \ldots, U_M} \sum_{g=1}^{N} \sum_{h=1}^{N} \Omega_{g,h} \cdot \left[\left(X_g \prod_{i=1}^{M} \times_i U_i^{\mathrm{T}}\right) \otimes \left(X_h \prod_{i=1}^{M} \times_i U_i^{\mathrm{T}}\right); (1:M)(1:M)\right]. \quad (23)$$

To avoid trivial solutions, we impose the following constraints to uniquely determine the projection matrices:

$$U_i^{\mathrm{T}} U_i = I, \qquad (i = 1, \ldots, M). \quad (24)$$

The objective function of the final optimization of TDLA (23) and (24) is a multiparameter and nonconvex problem, and there is no known optimal solution which allows for the simultaneous optimization of all the projection matrices. To overcome this problem, inspired by alternating optimization, we compute $U_i$ $(i = 1, 2, \ldots, M)$ as follows. We iteratively optimize $U_k$ from $k = 1$ to $k = M$ by fixing other $(M-1)$ matrices rather than compute the optimal $U_i$ $(i = 1, 2, \ldots, M)$ simultaneously. Here, we assume that only the projection matrix $U_k$ is unknown. Then, according to the tensor contraction, the objective function (23) and (24) can be rewritten as (25). Note that (25) is derived from the fact that for an arbitrary matrix $A$, we have $\|A\|^2 = \mathrm{tr}(A^{\mathrm{T}}A)$

$$\begin{aligned}
&\arg \min_{U_k^{\mathrm{T}}U_k = I} \sum_{g=1}^{N}\sum_{h=1}^{N} \Omega_{g,h} \\
&\quad \cdot \left\{ \left[ \left( X_g \prod_{I \neq k}^{M} \times_i U_i^{\mathrm{T}} \right) \times_k U_k^{\mathrm{T}} \right] \right. \\
&\qquad \left. \otimes \left[ \left( X_h \prod_{I \neq k}^{M} \times_i U_i^{\mathrm{T}} \right) \times_k U_k^{\mathrm{T}} \right] ; (1:M)(1:M) \right\} \\
&= \arg \min_{U_k^{\mathrm{T}}U_k = I} \sum_{g=1}^{N}\sum_{h=1}^{N} \Omega_{g,h} \\
&\quad \cdot \mathrm{tr} \left\{ U_k^{\mathrm{T}} \left[ \left( X_g \prod_{I \neq k}^{M} \times_i U_i^{\mathrm{T}} \right) \right. \right. \\
&\qquad \left. \left. \otimes \left( X_h \prod_{I \neq k}^{M} \times_i U_i^{\mathrm{T}} \right) ; (\overline{k})(\overline{k}) \right] U_k \right\}. \quad (25)
\end{aligned}$$

Based on the tensor contraction property given in (3), we can further rewrite (25) as

$$\begin{aligned}
\arg \min_{U_k^{\mathrm{T}}U_k = I} \mathrm{tr} \left[ U_k^{\mathrm{T}} \left( \sum_{g=1}^{N}\sum_{h=1}^{N} \Omega_{g,h} \cdot Mat_k \left( X_g \prod_{I \neq k}^{M} \times_i U_i^{\mathrm{T}} \right) \right. \right. \\
\left. \left. \cdot Mat_k^{\mathrm{T}} \left( X_h \prod_{I \neq k}^{M} \times_i U_i^{\mathrm{T}} \right) U_k \right] . \quad (26)
\end{aligned}$$

Since the center part of optimization (26) can be computed by known $\Omega$ and fixed $U_i$ $(i = 1, 2, \ldots, M, i \neq k)$, we simply denote

$$\begin{aligned}
F(k) = \sum_{g=1}^{N}\sum_{h=1}^{N} \Omega_{g,h} \cdot Mat_k \left( X_g \prod_{I \neq k}^{M} \times_i U_i^{\mathrm{T}} \right) \\
\cdot Mat_k^{\mathrm{T}} \left( X_h \prod_{I \neq k}^{M} \times_i U_i^{\mathrm{T}} \right) \in \mathrm{R}^{I_k \times I_k}. \quad (27)
\end{aligned}$$

Then, we put (27) into (26) to further reduce the optimization of $U_k$ to

$$\arg \min_{U_k^{\mathrm{T}}U_k = I} \mathrm{tr} \left( U_k^{\mathrm{T}} F(k) U_k \right). \quad (28)$$

The solution of (28) is acquired by combining the eigenvectors associated with the smallest $P_k$ eigenvalues of matrix $F(k)$ [57].

Algorithm I summarizes the aforementioned procedure for optimization. First, we initialize $U_i$ $(i = 1, 2, \ldots, M)$ as unit matrices, i.e., $U_i = I_{I_i}$. Then, for each iteration, we iteratively optimize $U_k$ from $k = 1$ to $k = M$ by fixing other $(M-1)$ matrices by (28) and replace the original $U_k$ with the latest optimized value. It has been theoretically demonstrated that such an alternating optimization procedure converges to a local optimum [38], [56]. In the next section, we will empirically show that the alternating optimization procedure converges within a few iterations. The following index $\Lambda$ is computed at the end of each iteration round:

$$\begin{aligned}
\Lambda = \sum_{g=1}^{N}\sum_{h=1}^{N} \left\{ \Omega_{g,h} \cdot \left[ \left( X_g \prod_{i=1}^{M} \times_i U_i^{\mathrm{T}} \right) \right. \right. \\
\left. \left. \otimes \left( X_h \prod_{i=1}^{M} \times_i U_i^{\mathrm{T}} \right) ; (1:M)(1:M) \right] \right\}. \quad (29)
\end{aligned}$$

In each round of optimization, the value of $\Lambda$ is decreased by the proposed procedure. If the error of $\Lambda$ values between the $t$th iteration and $(t-1)$th iteration is small enough, it indicates that the algorithm has converged.

---

Algorithm I: Procedure of the TDLA Algorithm

---

**Inputs**: Training samples $X \in R^{I_1 \times I_2 \times \cdots \times I_M}$ $(i = 1, 2, \ldots, N)$ and their class labels $y_i$ $(y_i \in [1, 2, \ldots, C])$. Size of the samples in reduced feature space $X^* \in R^{P_1 \times P_2 \times \cdots \times P_M}$.

---

**Step 1**: Construct the alignment matrix $\Omega$ by (22);
**Step 2**: Initialization $U_i = I_{I_i}$ $(i = 1, 2, \ldots, M)$;
**Step 3**:
for $t = 1$ to $T$
  for $k = 1$ to $M$
    Optimize $U_k$ by (28);
  end
  Compute $\Lambda$ by (29) to check if convergence is reached.
end

---

**Outputs**: Set of projection matrices $U_i \in R^{I_i \times P_i}$ $(i = 1, 2, \ldots, M)$ for multilinear transformation.

---

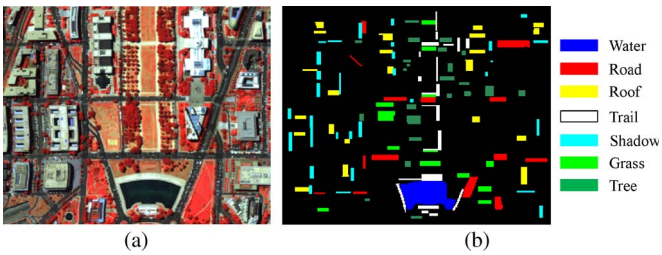



Fig. 6. (a) RGB composites of Data Set 1 (bands 65, 52, and 36 for red, green, and blue, respectively). (b) Reference data of Data Set 1.

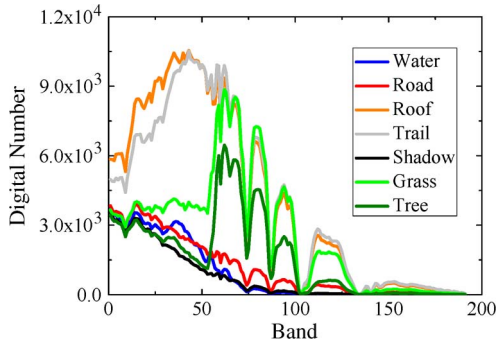



Fig. 7. Representative spectral curves in Data Set 1.

## V. Experimental Results

The experimental analysis was conducted on hyperspectral remote sensing images from the following three public data sets.

The first data set is an urban site of the airborne HYDICE from the mall in Washington, DC, which has an original size of 1280 × 307 pixels. In this study, we used a size of 250 × 307. A total of 210 bands were collected in the 0.4–2.4-$\mu$m region of the visible and infrared spectrums. The water absorption bands were then removed, resulting in 191 channels [58].

The second experiment was performed on an airborne HSI data set, which was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) at the urban test area of Pavia, northern Italy. The whole data set size is 1400 × 512 pixels, and we used a size of 400 × 400 in this study. Some channels were removed due to noise, and the remaining 102 spectral dimensions from 0.43–0.83 $\mu$m were processed. This data set was provided by the Data Fusion Technical Committee of the IEEE Geoscience and Remote Sensing Society [59].

The third data set is from a mixed forest/agricultural site at the Indian Pine test site in northwest Indiana, taken on June 12, 1992, which was gathered by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor. This data set was obtained from an aircraft flown at 19 812-m altitude and operated by the National Aeronautics and Space Administration Jet Propulsion Laboratory. The data set has a size of 145 × 145 pixels and 220 spectral bands, measuring approximately 20 m across the ground [58].

### A. Experiment 1: HYDICE Data Set

Data Set 1 and the reference data are shown in Fig. 6(a) and (b). There are seven classes of pixels to analyze: water, road, roof, trail, shadow, grass, and tree. It is challenging to classify in the hyperspectral remote sensing area, mainly because road, roof, and trail are spectrally similar, in that they may be made of similar materials.

The representative spectral curves of each class are shown in Fig. 7. It is worth emphasizing that there is no single spectral response representative of class "roof" for this data set; the detailed analysis of this point is given in Table I.

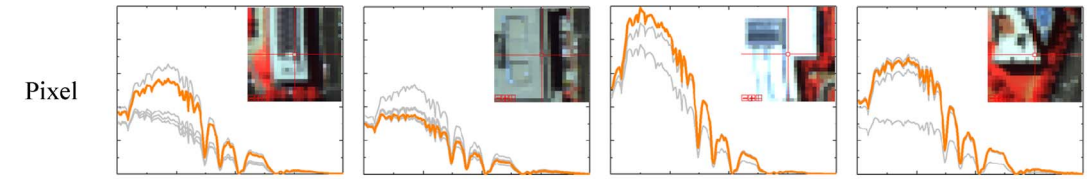
We first show four pixels of the class "roof" and their corresponding spectral curves in the first row of Table I. In the experiments, as a case study, we used the 4-NN local spatial structure of the analyzed pixel for feature extraction. The analyzed pixels are located at the center of each subimage, and the curves with a gray color represent the spectral curves of the neighboring pixels of the center pixel. The six subtables under the first row show the normalized distances among the four pixels in the following extracted feature spaces: original feature space, reduced feature space using PCA [18], locality preserving projections (LPP) [60], linear discriminant analysis (LDA) [14], DLA [55], and TDLA. In the vector-based feature-extraction approaches, we used the BIP representation as the input and reduced the size of feature to ten, while in the proposed method, we also experimentally set the reduced feature to a first-order tensor of size ten. There are 20 samples per class selected as the input training samples in the TDLA algorithm.

From Table I, we can see that the pixel distances of the same class were obviously reduced in the proposed TDLA-extracted feature space; the sum of the distances decreased from 2.38 in the original feature space to 1.956 in the reduced feature space, which is also the minimum value among the five feature-extraction approaches. To further investigate the ability of the proposed method to preserve the discriminability of classes for classification, we also compared the normalized distances in the extracted feature spaces of four pixels in different classes. The results of this comparison are shown in Table II.

In Table II, we chose pixels from the classes of roof, road, trail, and water. The classes of road, roof, and trail are spectrally similar in this data set, which can be observed in the first row of Table II. Thus, the distances between them in the original feature space and the two unsupervised feature-extracted feature spaces are very close, as shown in Table II. The sum values of the distances in the original feature space, PCA feature, and LPP feature space are 1.987, 2.621, and 2.990, respectively. However, the sum value of the distances in the proposed TDLA feature space reaches 6.231, which is the largest index in Table II. In order to further compare many more pixels' distribution in the aforementioned six feature spaces, Fig. 8 shows all the pixels of the reference data in different feature spaces, under a 2-D condition. We use the analyzed pixel features of bands 32 and 65 for Fig. 8(a), while we use the top-two significant components of the extracted features by different algorithms for Fig. 8(b)–(f), respectively. The number of pixels in the reference data is given in the "test data" of Table III. The results shown in Fig. 8(a)–(e) demonstrate that the existing feature-extraction algorithms merge different classes of pixels in the low-dimensional space. In contrast, the proposed TDLA can effectively separate different classes

TABLE I

SAMPLES OF THE SAME CLASS AND THEIR NORMALIZED DISTANCES IN EXTRACTED FEATURE SPACE




NORMALIZED DISTANCES IN EXTRACTED FEATURE SPACE

ORIGINAL FEATURE SPACE: 2.387

| | (a) | (b) | (c) | (d) |
|---|---|---|---|---|
| (a) | 0 | 0.128 | 0.238 | 0.090 |
| (b) | 0.128 | 0 | 0.361 | 0.217 |
| (c) | 0.238 | 0.361 | 0 | 0.156 |
| (d) | 0.090 | 0.217 | 0.156 | 0 |

PCA FEATURE SPACE: 2.950

| | (a) | (b) | (c) | (d) |
|---|---|---|---|---|
| (a) | 0 | 0.084 | 0.366 | 0.168 |
| (b) | 0.084 | 0 | 0.399 | 0.206 |
| (c) | 0.366 | 0.399 | 0 | 0.249 |
| (d) | 0.168 | 0.206 | 0.249 | 0 |

LPP FEATURE SPACE: 2.020

| | (a) | (b) | (c) | (d) |
|---|---|---|---|---|
| (a) | 0 | 0.193 | 0.161 | 0.121 |
| (b) | 0.193 | 0 | 0.179 | 0.192 |
| (c) | 0.161 | 0.179 | 0 | 0.162 |
| (d) | 0.121 | 0.192 | 0.162 | 0 |

LDA FEATURE SPACE: 2.546

| | (a) | (b) | (c) | (d) |
|---|---|---|---|---|
| (a) | 0 | 0.266 | 0.209 | 0.206 |
| (b) | 0.266 | 0 | 0.146 | 0.265 |
| (c) | 0.209 | 0.146 | 0 | 0.178 |
| (d) | 0.206 | 0.265 | 0.178 | 0 |

DLA FEATURE SPACE: 1.978

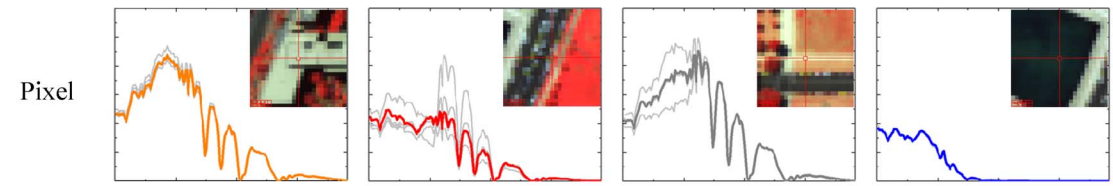
| | (a) | (b) | (c) | (d) |
|---|---|---|---|---|
| (a) | 0 | 0.157 | 0.152 | 0.115 |
| (b) | 0.157 | 0 | 0.155 | 0.244 |
| (c) | 0.152 | 0.155 | 0 | 0.165 |
| (d) | 0.115 | 0.244 | 0.165 | 0 |

TDLA FEATURE SPACE: 1.956

| | (a) | (b) | (c) | (d) |
|---|---|---|---|---|
| (a) | 0 | 0.097 | 0.071 | 0.196 |
| (b) | 0.097 | 0 | 0.079 | 0.292 |
| (c) | 0.071 | 0.079 | 0 | 0.240 |
| (d) | 0.196 | 0.292 | 0.240 | 0 |

TABLE II

SAMPLES OF DIFFERENT CLASSES AND THEIR NORMALIZED DISTANCES IN EXTRACTED FEATURE SPACE




NORMALIZED DISTANCES IN EXTRACTED FEATURE SPACE

ORIGINAL FEATURE SPACE: 1.987

| | (a) | (b) | (c) | (d) |
|---|---|---|---|---|
| (a) | 0 | 0.148 | 0.056 | 0.249 |
| (b) | 0.148 | 0 | 0.153 | 0.119 |
| (c) | 0.056 | 0.153 | 0 | 0.265 |
| (d) | 0.249 | 0.119 | 0.265 | 0 |

PCA FEATURE SPACE: 2.621

| | (a) | (b) | (c) | (d) |
|---|---|---|---|---|
| (a) | 0 | 0.194 | 0.092 | 0.320 |
| (b) | 0.194 | 0 | 0.196 | 0.172 |
| (c) | 0.092 | 0.196 | 0 | 0.334 |
| (d) | 0.320 | 0.172 | 0.334 | 0 |

LPP FEATURE SPACE: 2.990

| | (a) | (b) | (c) | (d) |
|---|---|---|---|---|
| (a) | 0 | 0.339 | 0.275 | 0.286 |
| (b) | 0.339 | 0 | 0.203 | 0.198 |
| (c) | 0.275 | 0.203 | 0 | 0.191 |
| (d) | 0.286 | 0.198 | 0.191 | 0 |

LDA FEATURE SPACE: 3.048

| | (a) | (b) | (c) | (d) |
|---|---|---|---|---|
| (a) | 0 | 0.288 | 0.277 | 0.250 |
| (b) | 0.288 | 0 | 0.213 | 0.196 |
| (c) | 0.277 | 0.213 | 0 | 0.297 |
| (d) | 0.250 | 0.196 | 0.297 | 0 |

DLA FEATURE SPACE: 4.808

| | (a) | (b) | (c) | (d) |
|---|---|---|---|---|
| (a) | 0 | 0.226 | 0.158 | 0.563 |
| (b) | 0.226 | 0 | 0.247 | 0.558 |
| (c) | 0.158 | 0.247 | 0 | 0.650 |
| (d) | 0.563 | 0.558 | 0.650 | 0 |

TDLA FEATURE SPACE: 6.231

| | (a) | (b) | (c) | (d) |
|---|---|---|---|---|
| (a) | 0 | 0.287 | 0.259 | 0.655 |
| (b) | 0.287 | 0 | 0.296 | 0.738 |
| (c) | 0.259 | 0.296 | 0 | 0.878 |
| (d) | 0.655 | 0.738 | 0.878 | 0 |

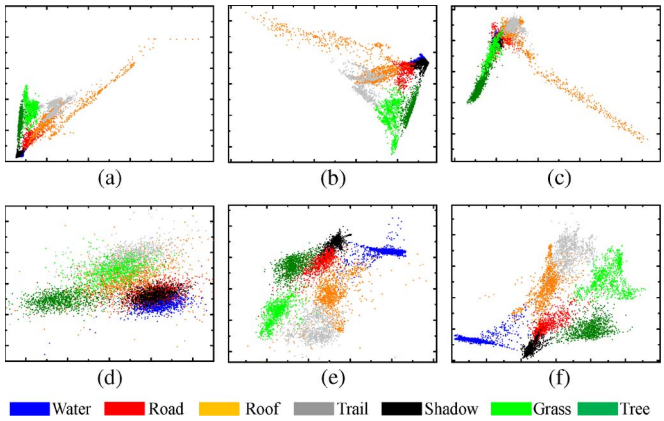



Fig. 8. Two-dimensional representation of features for the different algorithms. (a) Original. (b) PCA. (c) LPP. (d) LDA. (e) DLA. (f) TDLA.

TABLE III
TRAINING AND TEST SAMPLES FOR CLASSIFICATION IN DATA SET 1

| Class | Water | Road | Roof | Trail | Shadow | Grass | Tree | Total |
|---|---|---|---|---|---|---|---|---|
| Training | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 350 |
| Test | 1433 | 1463 | 1342 | 1033 | 1059 | 1026 | 1068 | 8424 |

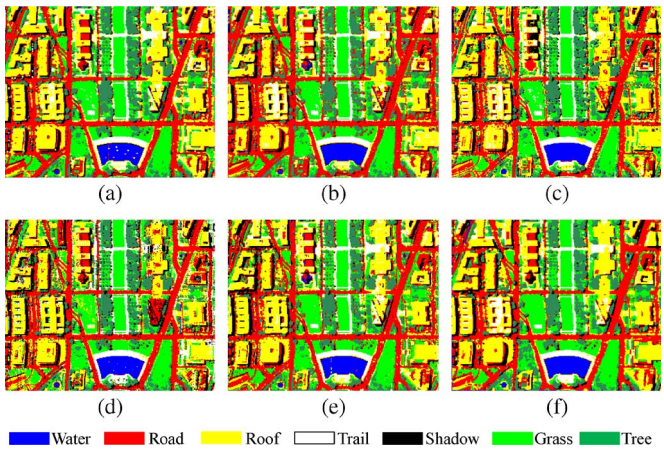



Fig. 9. Classification maps of all the methods in Data Set 1, based on SVM. (a) Original. (b) PCA. (c) LPP. (d) LDA. (e) DLA. (f) TDLA.

because the tensor representation preserves the spectral–spatial information of the pixels and the optimization of TDLA preserves the discriminability of classes for classification.

The classification maps for HSI using different features are shown in Fig. 9, based on support vector machine (SVM). In classification, training samples are selected from the reference data randomly while using all of the reference data for testing. The number of training samples and test samples is given in Table III.

As shown in Fig. 9 and Table IV, the proposed TDLA-based classification achieved the best results in both accuracy and visual interpretation. In these maps, it can be seen that roof pixels exist in the road pixels in all the classification maps. There are also some roof pixels existing in the water pixels in Fig. 9(a) and (d), because these pixel pairs have similar spectra. Although the spatial information is addressed in all the methods, only a little improvement can be observed in Fig. 9(a)–(e), because the spectral–spatial structure information is completely lost in the vector representation. However, Fig. 9(f) shows fewer misclassifications than the rest of the classification maps, particularly for the class pairs of roof–road, road–shadow, and trail–roof.

In order to thoroughly evaluate the different feature representations, the averaged classification rates in 20 independent experiments are compared in Table IV. The classification rates are reported by using two common classifiers, i.e., SVM and NN. In each experiment, the training samples were selected from the reference data randomly. Classification rates observed in Table IV indicate that TDLA obtained several of the top classification rates in the individual classes and achieved the highest overall accuracy (OA) and kappa coefficient, both by SVM and NN.

The classification OAs, with respect to reduced dimensionality for all the feature-extraction methods, are shown in Fig. 10(a) and (b). From this, we can see that the proposed method performed better than the other methods when the dimensionality was larger than five and achieved the best classification OA for both SVM and NN. The classification OA of all the algorithms stabilized when the reduced dimensionality increased to 20 and 5, for SVM and NN, respectively.

### B. Parameter Analysis

*1) Size of Extracted Feature:* In the experiment, the input tensor $X_i \in R^{191\times 5}$ was reduced to an output tensor of size $P_1 \times P_2$, where the parameters $P_1$ and $P_2$ determined the size of the multilinear transformation matrices, i.e., the solution of $U_i$ was acquired by combining the eigenvectors associated with the smallest $P_k$ eigenvalues of matrix $F(k)$. Here, we experimentally show the eigenvectors in the first and second orders, sorted in ascending order (Fig. 11).

For the first order, 191 eigenvalues were distributed as follows. From indices 1 to 7, the eigenvalues are negative and with large values. From indices 8 to 140, the eigenvalues are very close to zero and very slowly increase to nonnegative. Then, starting from index 150, the eigenvalues increase quickly to large values. Therefore, the best choice of $P_1$ is around seven. While for the second order, there are five eigenvalues, but only the first one is negative, and the others are nonnegative, so the best choice of $P_2$ is around one. The corresponding classification OA, with respect to variations of $P_1$ and $P_2$, is shown in Fig. 12. When $P_2 > 1$, the output feature is a second-order tensor, so we compute Fig. 12 using NN as the unique classifier. The surface in Fig. 12 also suggests that the optimal values of $P_1$ and $P_2$ are seven and one, respectively. Based on the aforementioned analysis, the optimal size of $P_k$ for the TDLA algorithm ought to be set experimentally by considering the eigenvalues of matrix $F(k)$ in the $k$th mode optimization.

*2) Parameters for Building a Patch:* There are three parameters, i.e., $n_1$, $n_2$, and $\alpha$, used in building a patch in the TDLA algorithm. Of these parameters, $n_1$ and $n_2$ are the numbers of nearest samples for building a patch, and $\alpha$ is a tradeoff weight of discrimination. In order to avoid cross-validation, we usually set these parameters according to their physical meanings. That is, $n_1$ and $n_2$ are decided by the size of the training samples, and $\alpha$ is usually more than one in the classification tasks. In

TABLE IV
CLASS-SPECIFIC RATES IN PERCENTAGE FOR VARIOUS FEATURES IN DATA SET 1

| DR+Classifier | | Water | Road | Roof | Trail | Shadow | Grass | Tree | OA | Kappa |
|---|---|---|---|---|---|---|---|---|---|---|
| Original | *SVM* | 90.88 | 92.33 | 88.91 | 97.88 | 90.12 | 98.14 | 98.05 | 93.71 | 0.9265 |
| | *NN* | 81.20 | 81.83 | 38.65 | 79.64 | 88.68 | 96.00 | 97.79 | 80.27 | 0.7694 |
| PCA | *SVM* | 94.58 | 93.99 | 78.67 | 81.94 | 95.09 | 98.15 | 97.50 | 91.33 | 0.8986 |
| | *NN* | 84.58 | 79.45 | 41.96 | 79.30 | 80.71 | 95.57 | 97.62 | 79.46 | 0.7601 |
| LPP | *SVM* | 94.72 | 94.64 | 84.68 | 84.42 | 96.66 | 97.78 | 98.94 | 93.07 | 0.9189 |
| | *NN* | 71.79 | 96.75 | 69.40 | 86.01 | 90.10 | 91.38 | 98.38 | 86.87 | 0.8461 |
| LDA | *SVM* | 96.38 | 90.42 | 91.41 | 95.38 | 93.64 | 96.08 | 98.54 | 94.32 | 0.9336 |
| | *NN* | 85.30 | 83.74 | 73.10 | 84.68 | 90.77 | 93.68 | 90.44 | 85.79 | 0.8338 |
| DLA | *SVM* | 94.46 | 94.54 | 91.74 | 93.91 | 94.41 | 97.94 | 98.71 | 95.06 | 0.9422 |
| | *NN* | 90.90 | 82.32 | 74.76 | 83.29 | 90.25 | 88.29 | 93.77 | 85.90 | 0.8349 |
| TDLA | *SVM* | 95.38 | 96.79 | 97.54 | 98.32 | 97.70 | 99.12 | 98.68 | 97.66 | 0.9726 |
| | *NN* | 89.90 | 79.32 | 85.82 | 91.41 | 92.62 | 96.26 | 97.19 | 89.90 | 0.8819 |

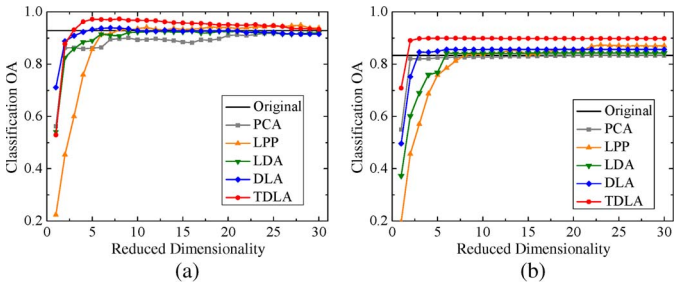



Fig. 10. Classification OAs, with respect to reduced dimensionality in Data Set 1. (a) SVM. (b) NN.

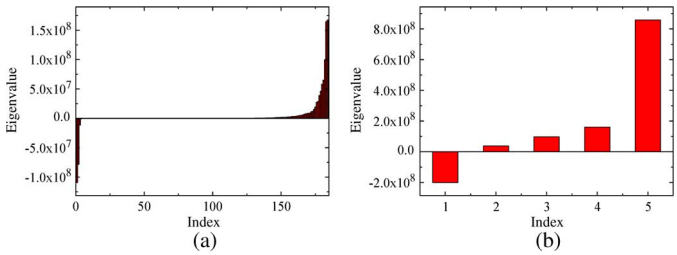



Fig. 11. Eigenvectors of Data Set 1 in the first and second orders. (a) First order. (b) Second order.

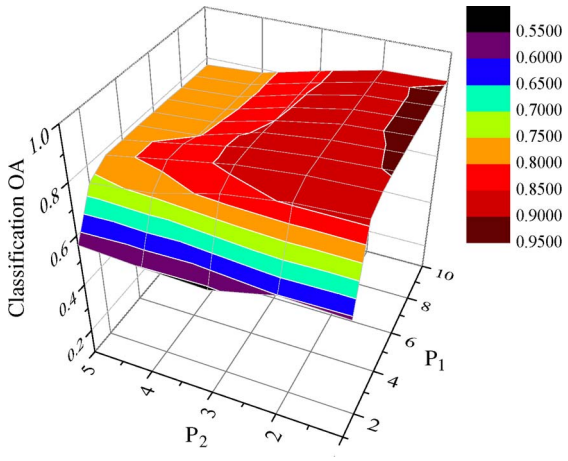



Fig. 12. Classification OAs, with respect to the size of extracted feature of Data Set 1.

our experiments, we fix $\alpha$ at two. Fig. 13(a) and (b) shows the effect of $n_1$ and $n_2$ on the classification OA in Data Set 1, based on SVM and NN. The classification OA was obtained by

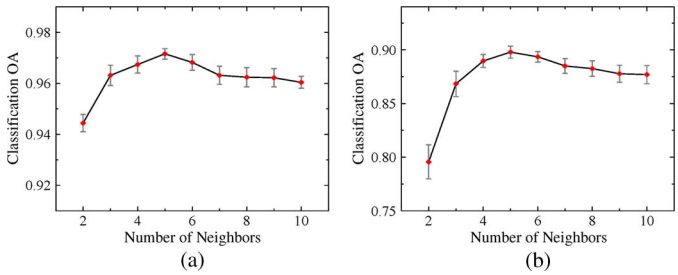



Fig. 13. Classification OAs, with respect to the number of samples for building a patch of Data Set 1. (a) SVM. (b) NN.

reporting the mean and standard deviation values of 20 groups of classification results, using independent training samples. Here, we experimentally set $n_1 = n_2$. We considered that there are 20 samples per class selected for training in the TDLA algorithm. The classification curves in the two figures suggest that a small size for the building patch helps to represent the local geometry property in the TDLA algorithm. There is a peak in the curves when the number of neighbors reaches five.

*3) Convergence Analysis:* Fig. 14(a)–(e) shows the error of $\Lambda$ values between the $t$th iteration and $(t-1)$th iteration in the alternating optimization procedure of various local spatial structures in Data Set 1. It is clear that the TDLA algorithm often converges at stable values in about three iterations. We experimentally found that, when the iteration number is at five, the error value reduces to no more than 0.001% of the $\Lambda$ value. We have also observed the same trend in the other two data sets, as well as with different parameters. Experimental results suggest that, in practice, we could fix the number of iterations at five to guarantee that the TDLA algorithm has converged.

## C. Experiment 2: ROSIS Data Set

Data Set 2 and the reference data are shown in Fig. 15(a) and (b). The six classes of pixels that were analyzed are water, roof, road, grass, tree, and shadow (Table V). The representative spectral curves of each class are shown in Fig. 16.

In this data set, we still selected 20 samples per class as input training tensors of TDLA. The classification result maps are shown in Fig. 17, while the detailed averaged classification rates are reported in Table VI. The classification results are

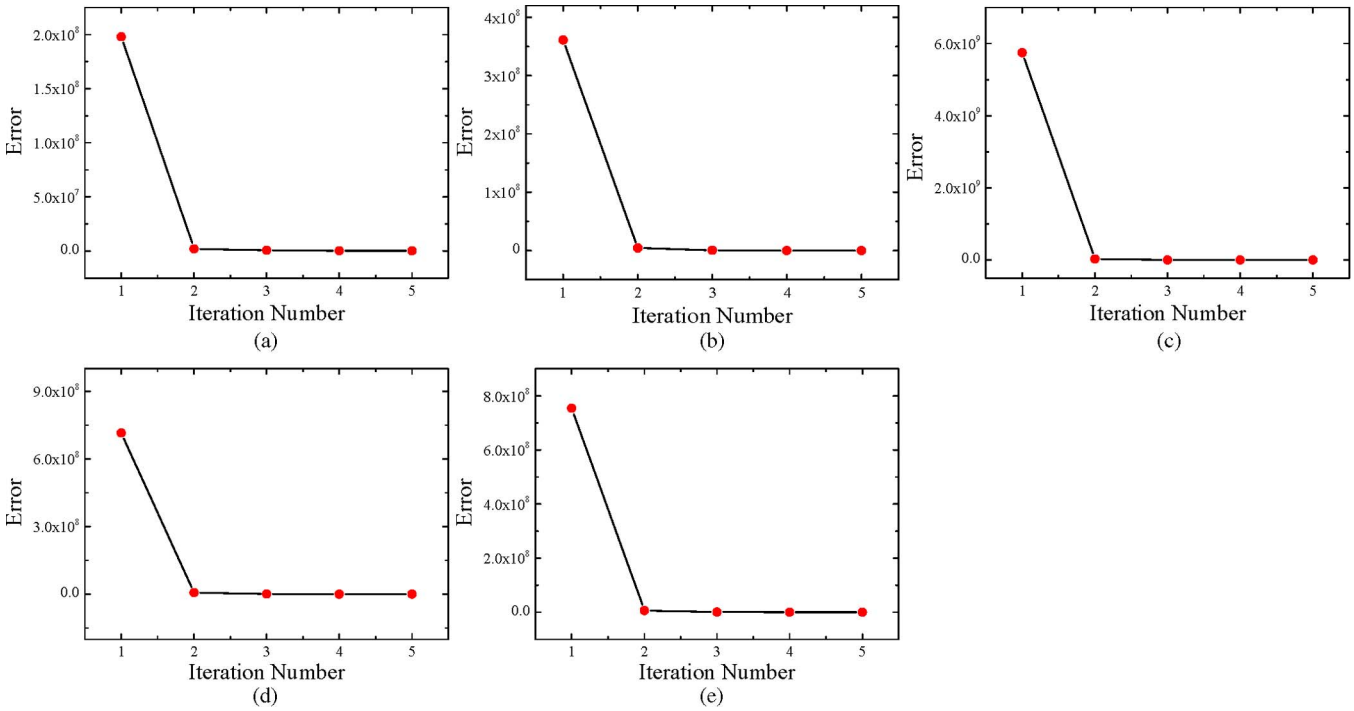



Fig. 14. Convergence report of the proposed algorithm with Data Set 1. (a) 4-NN. (b) 8-NN. (c) 12-NN. (d) 20-NN. (e) 24-NN.

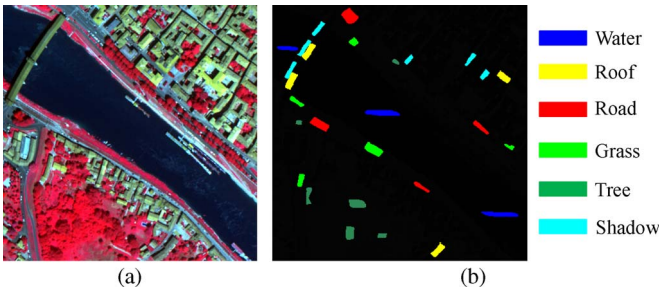



Fig. 15. (a) RGB composites of Data Set 2 (bands 102, 56, and 31 for red, green, and blue, respectively). (b) Reference data of Data Set 2.

TABLE V
TRAINING AND TEST SAMPLES FOR CLASSIFICATION IN DATA SET 2

| Class | Water | Roof | Road | Grass | Tree | Shadow | Total |
|---|---|---|---|---|---|---|---|
| Training | 30 | 30 | 30 | 30 | 30 | 30 | 180 |
| Test | 950 | 892 | 1009 | 807 | 879 | 806 | 5343 |

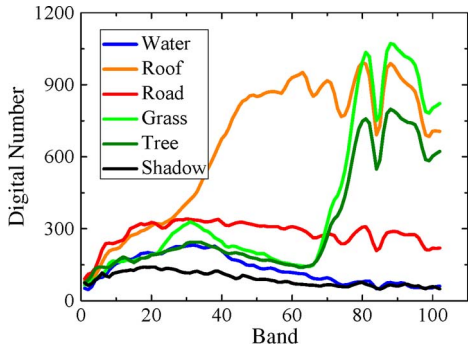



Fig. 16. Representative spectral curves of classes in Data Set 2.

similar to the reports described earlier with Data Set 1. The proposed TDLA-based classification achieved the best performance, particularly at the bridge over the river (distinguishing

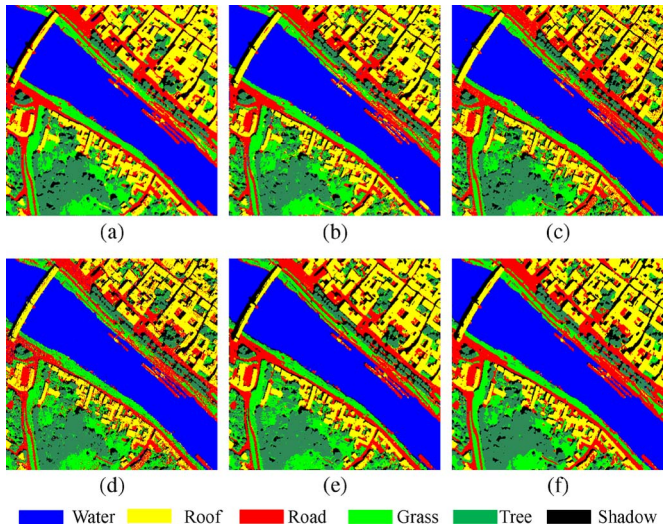



Fig. 17. Classification maps of all the methods in Data Set 2, based on SVM. (a) Original. (b) PCA. (c) LPP. (d) LDA. (e) DLA. (f) TDLA.

the pair of roof–shadow) and along the river (distinguishing the pairs of grass–tree and grass–road). On the detailed classification rate, the proposed algorithm achieved several of the top classification rates of the individual classes and gave the best OA and kappa, for both SVM and NN.

### D. Experiment 3: AVIRIS Data Set

The AVIRIS hyperspectral data set used in this experiment is shown in Fig. 18(a). The image shows a typical agricultural site with many kinds of crops. The advantage in using this data set is the availability of a reference map prepared from the field surveys conducted at the time of image acquisition. The ten major classes of reference data are shown in Fig. 18(b), and

TABLE VI
CLASS-SPECIFIC RATES IN PERCENTAGE FOR VARIOUS FEATURES IN DATA SET 2.

| DR+Classifier | | Water | Roof | Road | Grass | Tree | Shadow | OA | Kappa |
|---|---|---|---|---|---|---|---|---|---|
| Original | *SVM* | 99.87 | 98.75 | 91.99 | 93.96 | 92.96 | 90.60 | 94.48 | 0.9366 |
| | *NN* | 99.87 | 67.38 | 59.59 | 79.40 | 66.92 | 74.36 | 74.16 | 0.6900 |
| PCA | *SVM* | 100 | 96.99 | 92.11 | 97.80 | 88.05 | 93.16 | 94.38 | 0.9325 |
| | *NN* | 99.87 | 67.38 | 59.47 | 77.61 | 65.96 | 73.75 | 73.59 | 0.6831 |
| LPP | *SVM* | 99.49 | 90.09 | 95.51 | 91.07 | 92.74 | 94.59 | 93.94 | 0.9271 |
| | *NN* | 95.04 | 77.04 | 64.02 | 79.95 | 76.31 | 85.50 | 79.67 | 0.7554 |
| LDA | *SVM* | 99.49 | 89.84 | 89.44 | 88.60 | 87.30 | 89.38 | 90.54 | 0.8863 |
| | *NN* | 90.20 | 82.94 | 71.72 | 73.49 | 69.48 | 80.69 | 77.94 | 0.7345 |
| DLA | *SVM* | 99.75 | 95.23 | 91.38 | 92.72 | 92.53 | 91.22 | 93.66 | 0.9238 |
| | *NN* | 97.71 | 68.38 | 55.70 | 85.16 | 85.59 | 91.42 | 80.59 | 0.7708 |
| TDLA | *SVM* | 100 | 97.87 | 94.17 | 96.98 | 95.62 | 94.59 | 96.42 | 0.9569 |
| | *NN* | 96.18 | 78.80 | 68.08 | 86.40 | 88.26 | 90.70 | 84.91 | 0.8181 |

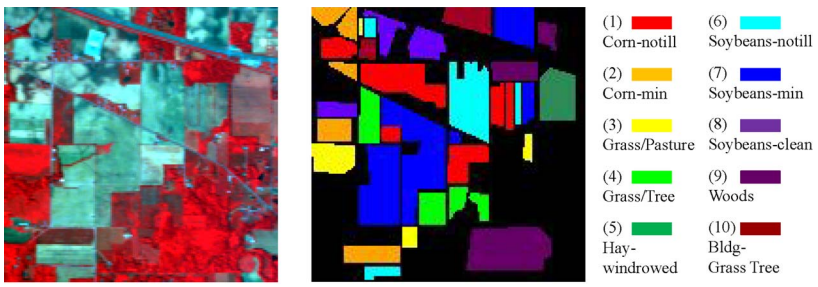



Fig. 18. (a) RGB composites of Data Set 3 (bands 57, 27, and 17 for red, green, and blue, respectively). (b) Reference data of Data Set 3.

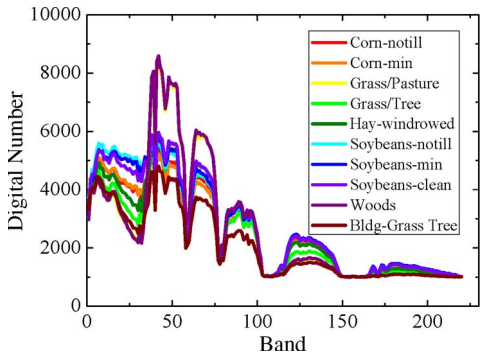



Fig. 19. Representative spectral curves of the classes in Data Set 3.

TABLE VII
TRAINING AND TEST SAMPLES FOR CLASSIFICATION IN DATA SET 3

| Class | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Training | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 300 |
| Test | 984 | 549 | 326 | 494 | 357 | 664 | 1759 | 240 | 882 | 254 | 6509 |

the representative spectral curves of all classes are shown in Fig. 19. This data set has been extensively used to test various HSI analysis algorithms. The training and test samples were generated from reference data randomly, the number of which is listed in Table VII.

A similar classification performance can be observed from Fig. 20 and Table VIII, in that the proposed TDLA-based classification achieved the best performance from both visual interpretation and accuracy. The analysis based on the three HSI data sets demonstrated that the proposed TDLA is an effective and robust feature-extraction algorithm.

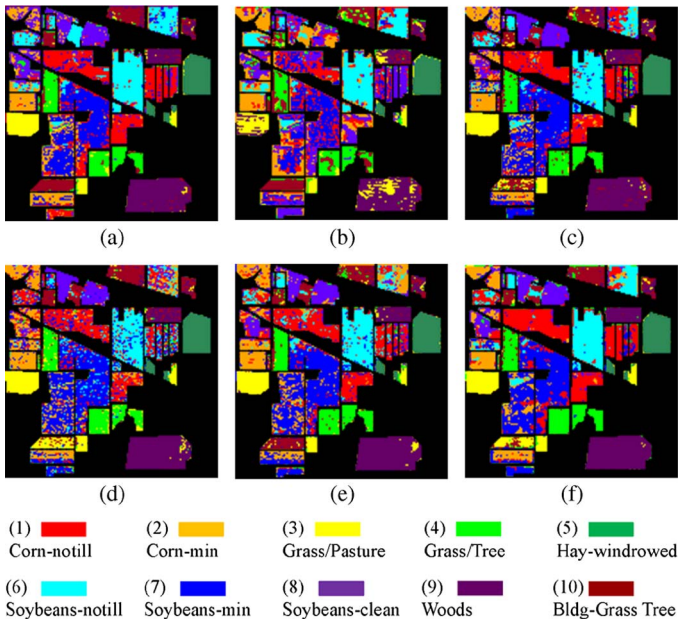



Fig. 20. Classification maps of all the methods in Data Set 3, based on SVM. (a) Original. (b) PCA. (c) LPP. (d) LDA. (e) DLA. (f) TDLA.

## VI. CONCLUSION

In this paper, we have proposed a new scheme for HSI spectral–spatial feature extraction. First, the spectral–spatial feature of a pixel in HSI is represented as a second-order tensor. Then, the TDLA algorithm is used to preserve the discriminability of the classes for classification by considering the discriminative locality information in the optimization. Finally,

TABLE VIII
CLASS-SPECIFIC RATES IN PERCENTAGE FOR VARIOUS FEATURES IN DATA SET 3

| DR+Classifier | | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | OA | Kappa |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Original | *SVM* | 71.85 | 70.13 | 88.65 | 96.96 | 100 | 90.36 | 64.92 | 87.50 | 92.52 | 77.17 | 79.60 | 0.7647 |
| | *NN* | 8.94 | 35.88 | 41.72 | 50.20 | 100 | 64.46 | 24.05 | 16.67 | 70.52 | 63.39 | 41.48 | 0.3475 |
| PCA | *SVM* | 39.33 | 55.37 | 76.07 | 72.27 | 100 | 90.66 | 52.13 | 47.50 | 74.04 | 62.20 | 62.90 | 0.5711 |
| | *NN* | 7.83 | 34.61 | 38.04 | 40.89 | 100 | 64.31 | 21.66 | 15.42 | 69.73 | 59.45 | 39.35 | 0.3241 |
| LPP | *SVM* | 79.98 | 78.51 | 96.63 | 94.33 | 100 | 91.57 | 75.21 | 89.17 | 90.14 | 65.75 | 83.93 | 0.8135 |
| | *NN* | 54.17 | 34.06 | 67.18 | 79.15 | 98.32 | 63.40 | 52.30 | 52.50 | 81.97 | 56.69 | 61.68 | 0.5604 |
| LDA | *SVM* | 66.16 | 76.32 | 95.40 | 90.28 | 100 | 78.01 | 78.62 | 82.92 | 89.57 | 90.55 | 81.49 | 0.7843 |
| | *NN* | 57.93 | 57.92 | 84.66 | 80.77 | 100 | 54.07 | 53.38 | 73.75 | 85.94 | 81.10 | 66.97 | 0.6191 |
| DLA | *SVM* | 76.12 | 72.31 | 94.17 | 96.76 | 100 | 92.32 | 78.57 | 87.50 | 95.58 | 83.07 | 85.22 | 0.8281 |
| | *NN* | 59.35 | 77.23 | 58.28 | 95.95 | 100 | 62.80 | 45.48 | 70.00 | 60.32 | 67.72 | 63.27 | 0.5827 |
| TDLA | *SVM* | 89.33 | 81.06 | 96.32 | 98.38 | 100 | 98.19 | 79.25 | 91.25 | 96.26 | 91.73 | 89.54 | 0.8786 |
| | *NN* | 71.56 | 82.70 | 58.28 | 89.07 | 100 | 61.30 | 47.70 | 85.42 | 94.78 | 79.92 | 71.22 | 0.6709 |

the extracted feature is obtained by multilinear transformation, under the definition of tensor algebra. Some advantages of this work are the following: 1) The proposed tensor representation can preserve as many as possible the original spatial constraints of a certain pixel, and 2) the proposed TDLA is a generalized DR framework for high-order data, which indicates that a more comprehensive high-order feature could be processed by the proposed framework directly. A range of experiments based on several kinds of HSI data demonstrated that the proposed method significantly improved the classification accuracies. Nevertheless, there may still be room for improvement of the input feature representations in TDLA, addressing more discriminative features, e.g., the morphological feature [61], texture feature [62], and polarimetric feature [63], which may further improve the HSI classification performance. This will be explored in our future work.

## ACKNOWLEDGMENT

The authors would like to thank Prof. D. Landgrebe from Purdue University, West Lafayette, IN, for providing the free downloads of the hyperspectral digital imagery collection experiment hyperspectral data set and AVIRIS data set, Prof. Gamba from the Data Fusion Technical Committee of the IEEE Geoscience and Remote Sensing Society for providing the reflective optics system imaging spectrometer data set, and the handling editor and anonymous reviewers for their careful reading and helpful comments.

## REFERENCES

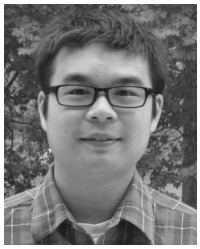
[1] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.
[2] A. F. H. Goetz, G. Vane, J. E. Solomon, and B. N. Rock, "Imaging spectrometry for earth remote sensing," *Science*, vol. 228, no. 4704, pp. 1147–1153, Jun. 1985.
[3] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, no. 1, pp. 110–122, Sep. 2009.
[4] L. O. Jimenez and D. A. Landgrebe, "Supervised classification in high-dimensional space: Geometrical, statistical, and asymptotical properties of multivariate data," *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 28, no. 1, pp. 39–54, Feb. 1998.
[5] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis: An Introduction*. Berlin, Germany: Springer-Verlag, 2006.
[6] C.-I. Chang and H. Safavi, "Progressive dimensionality reduction by transform for hyperspectral imagery," *Pattern Recognit.*, vol. 44, no. 10, pp. 2760–2773, Oct. 2011.
[7] H. Shen and L. Zhang, "A MAP-based algorithm for destriping and inpainting of remotely sensed images," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 5, pp. 1492–1502, May 2009.
[8] A. R. Webb, *Statistical Pattern Recognition*. West Sussex, U.K.: Wiley, 2002.
[9] A. Jain and D. Zongker, "Feature selection: Evaluation, application, and small sample performance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 2, pp. 153–158, Feb. 1997.
[10] C.-I. Chang and S. Wang, "Constrained band selection for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1575–1585, Jun. 2006.
[11] P. M. Narendra and K. Fukunaga, "A branch and bound algorithm for feature subset selection," *IEEE Trans. Comput.*, vol. C-26, no. 9, pp. 917–922, Sep. 1977.
[12] M. L. Raymer, W. F. Punch, E. D. Goodman, L. A. Kuhn, and A. K. Jain, "Dimensionality reduction using genetic algorithms," *IEEE Trans. Evol. Comput.*, vol. 4, no. 2, pp. 164–171, Jul. 2000.
[13] L. Zhang, Y. Zhong, B. Huang, J. Gong, and P. Li, "Dimensionality reduction based on clonal selection for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 4172–4186, Dec. 2007.
[14] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. San Diego, CA: Academic, 1990.
[15] D. Tao, X. Li, X. Wu, and S. J. Maybank, "Geometric mean for subspace selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 260–274, Feb. 2009.
[16] W. Bian and D. Tao, "Max–min distance analysis by using sequential SDP relaxation for dimension reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 1037–1050, May 2011.
[17] C. Lee and D. A. Landgrebe, "Feature extraction based on decision boundaries," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 4, pp. 388–400, Apr. 1993.
[18] I. T. Jolliffe, *Principal Component Analysis*. New York: Springer-Verlag, 2002.
[19] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 26, no. 1, pp. 65–74, Jan. 1988.
[20] X. Jiang, "Linear subspace learning-based dimensionality reduction," *IEEE Signal Process Mag.*, vol. 28, no. 2, pp. 16–26, Mar. 2011.
[21] M. H. C. Law and A. K. Jain, "Incremental nonlinear dimensionality reduction by manifold learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 3, pp. 377–391, Mar. 2006.
[22] D. Song and D. Tao, "Biologically inspired feature manifold for scene classification," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 174–184, Jan. 2010.
[23] C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina, "Exploiting manifold geometry in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 441–454, Mar. 2005.
[24] J. He, L. Zhang, Q. Wang, and Z. Li, "Using diffusion geometric coordinates for hyperspectral imagery representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 767–771, Oct. 2009.

[25] T. Han and D. G. Goodenough, "Investigation of nonlinearity in hyperspectral imagery using surrogate data methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 10, pp. 2840–2847, Oct. 2008.
[26] L. Zhang, L. Zhang, D. Tao, and X. Huang, "On combining multiple features for hyperspectral remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 879–893, Mar. 2012.
[27] M. M. Crawford, L. Ma, and W. Kim, "Exploring nonlinear manifold learning for classification of hyperspectral data," in *Opt. Remote Sens.*, vol. 3, S. Prasad, L. M. Bruce, and J. Chanussot, Eds. Berlin, Germany: Springer-Verlag, 2011, pp. 207–234.
[28] B. Du and L. Zhang, "Random-selection-based anomaly detector for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 5, pp. 1578–1589, May 2011.
[29] C.-I. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 608–619, Mar. 2004.
[30] A. Plaza, P. Martínez, J. Plaza, and R. Pérez, "Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 466–479, Mar. 2005.
[31] F. Li, M. K. Ng, and R. J. Plemmons, "Coupled segmentation and denoising/deblurring models for hyperspectral material identification," *Numer. Linear Algebra Appl.*, vol. 19, no. 1, pp. 153–173, Jan. 2012.
[32] H. Li and L. Zhang, "A hybrid automatic endmember extraction algorithm based on a local window," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4223–4238, Nov. 2011.
[33] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.
[34] L. Zhang, X. Huang, B. Huang, and P. Li, "A pixel shape index coupled with spectral information for classification of high spatial resolution remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2950–2961, Oct. 2006.
[35] Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Multiple spectral–spatial classification approach for hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4122–4132, Nov. 2010.
[36] Q. Zhang, R. Plemmons, D. Kittle, D. Brady, and S. Prasad, "Joint segmentation and reconstruction of hyperspectral data with compressed measurements," *Appl. Opt.*, vol. 50, no. 22, pp. 4417–4435, Aug. 2011.
[37] Q. Yuan, L. Zhang, and H. Shen, "Multi-frame super-resolution employing a spatially weighted total variation model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 3, pp. 379–392, Mar. 2012.
[38] D. Tao, X. Li, X. Wu, W. Hu, and S. J. Maybank, "Supervised tensor learning," *Knowl. Inf. Syst.*, vol. 13, no. 1, pp. 1–42, Sep. 2007.
[39] S. Bourennane, C. Fossati, and A. Cailly, "Improvement of target-detection algorithms based on adaptive three-dimensional filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 4, pp. 1383–1395, Apr. 2011.
[40] D. Letexier and S. Bourennane, "Estimation of $N$-mode ranks of hyperspectral images for tensor denoising," in *Proc. Eur. Signal Process. Conf.*, Glasgow, Scotland, 2009, pp. 2594–2597.
[41] Q. Zhang, H. Wang, R. J. Plemmonsz, and V. P. Pauca, "Tensor methods for hyperspectral data processing: A space object identification study," *J. Opt. Soc. Amer. A*, vol. 25, no. 12, pp. 3001–3012, Dec. 2008.
[42] N. Renard and S. Bourennane, "Improvement of target detection methods by multiway filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2407–2417, Aug. 2008.
[43] L. Zhang, L. Zhang, D. Tao, and X. Huang, "A multifeature tensor for remote-sensing target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 2, pp. 374–378, Mar. 2011.
[44] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "A survey of multilinear subspace learning for tensor data," *Pattern Recognit.*, vol. 44, no. 7, pp. 1540–1551, Jul. 2011.
[45] L. D. Lathauwer, "Signal processing based on multilinear algebra," Ph.D. dissertation, Katholike Univ. Leuven, Leuven, Belgium, 1997.
[46] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H.-J. Zhang, "Multilinear discriminant analysis for face recognition," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 212–220, Jan. 2007.
[47] 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.
[48] 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.
[49] D. Muti and S. Bourennane, "Multidimensional filtering based on a tensor approach," *Signal Process.*, vol. 85, no. 12, pp. 2338–2353, Dec. 2005.
[50] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear analysis of image ensembles: TensorFaces," in *Proc. Eur. Conf. Comput. Vis.*, Copenhagen, Denmark, 2002, pp. 447–460.
[51] X. Li, S. Lin, S. Yan, and D. Xu, "Discriminant locally linear embedding with high-order tensor data," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 38, no. 2, pp. 342–352, Feb. 2008.
[52] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, Jan. 2007.
[53] Y. Liu, Y. Liu, and K. C. C. Chan, "Tensor distance based multilinear locality-preserved maximum information embedding," *IEEE Trans. Neural Netw.*, vol. 21, no. 11, pp. 1848–1854, Nov. 2010.
[54] Y. Mu, D. Tao, X. Li, and F. Murtagh, "Biologically inspired tensor features," *Cogn. Comput.*, vol. 1, no. 4, pp. 327–341, Dec. 2009.
[55] 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.
[56] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and Gabor features for gait recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1700–1715, Oct. 2007.
[57] R. Bhatia, *Matrix Analysis*. New York: Springer-Verlag, 1997.
[58] D. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*. New Jersey: Wiley, 2003.
[59] G. Licciardi, F. Pacifici, D. Tuia, S. Prasad, T. West, F. Giacco, C. Thiel, J. Inglada, E. Christophe, J. Chanussot, and P. Gamba, "Decision fusion for the classification of hyperspectral data: Outcome of the 2008 GRS-S data fusion contest," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3857–3865, Nov. 2009.
[60] X. He and P. Niyogi, "Locality preserving projections," in *Proc. Adv. Neural Inf. Process. Syst.*, 2004, pp. 153–160.
[61] M. D. Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3747–3762, Oct. 2010.
[62] X. Huang and L. Zhang, "An adaptive mean-shift analysis approach for object extraction and classification from urban hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 12, pp. 4173–4185, Dec. 2008.
[63] S. R. Cloude and E. Pottier, "An entropy based classification scheme for land applications of polarimetric SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 1, pp. 68–78, Jan. 1997.

**Liangpei Zhang** (M'06–SM'08) received the B.S. degree in physics from Hunan Normal University, Changsha, China, in 1982, the M.S. degree in optics from the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China, in 1988, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1998.

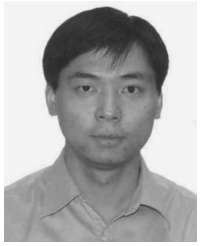
He is currently the Head of the Remote Sensing Division, State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University. He is also a "Changjiang Scholar" Chair Professor appointed by the Ministry of Education of China. He is currently a Principal Scientist with China State Key Basic Research Project (2011–2016) appointed by the Ministry of National Science and Technology of China to lead the remote sensing program in China. He has more than 240 research papers. He is the holder of five patents. He also serves as an Associate Editor of the *International Journal of Ambient Computing and Intelligence*, *International Journal of Image and Graphics*, *International Journal of Digital Multimedia Broadcasting*, *Journal of Geo-spatial Information Science*, and *Journal of Remote Sensing*. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence.

Dr. Zhang is a Fellow of the Institution of Electrical Engineers, Executive Member (Board of Governor) of China National Committee of International Geosphere-Biosphere Programme, Executive Member of China Society of Image and Graphics, etc. He regularly serves as a Cochair of the series International Society for Optics and Photonics (SPIE) Conferences on Multispectral Image Processing and Pattern Recognition, Conference on Asia Remote Sensing, and many other conferences. He edits several conference proceedings, issues, and geoinformatics symposiums.



**Lefei Zhang** (S'11) received the B.S. degree in sciences and techniques of remote sensing from Wuhan University, Wuhan, China, in 2008, where he is currently working toward the Ph.D. degree in the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing.

His research interests include hyperspectral data analysis, high-resolution image processing, and pattern recognition in remote sensing images.



**Dacheng Tao** (M'07–SM'12) received the B.Eng. degree from the University of Science and Technology of China, Hefei, China, the M.Phil. degree from The Chinese University of Hong Kong, Shatin, Hong Kong, and the Ph.D. degree from the University of London, London, U.K.

He is a Professor of computer science with the Centre for Quantum Computation and Intelligent Systems and the Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia. He mainly applies statistics and mathematics for data analysis problems in data mining, computer vision, machine learning, multimedia, and video surveillance. He has authored and coauthored more than 100 scientific articles at top venues, including IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, IEEE TRANSACTIONS ON IMAGE PROCESSING, *Neural Information Processing Systems*, International Conference on Machine Learning, The Conference on Uncertainty in Artificial Intelligence (UAI), International Conference on Artificial Intelligence and Statistics (AISTATS), ICDM, The International Joint Conference on Artificial Intelligence (IJCAI), Association for the Advancement of Artificial Intelligence (AAAI), IEEE International Conference on *Computer Vision and Pattern Recognition* (CVPR), European Conference on Computer Vision (ECCV), *ACM Transactions on Knowledge Discovery from Data* (ACM TKDD), Multimedia, and International Conference on Knowledge Discovery and Data Mining (KDD).

Dr. Tao was the recipient of the Best Theory/Algorithm Paper Runner-up Award in IEEE 2007 International Conference on Data Mining (ICDM).

**Xin Huang** received the Ph.D. degree in Photogrammetry and Remote Sensing at the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan, China, in 2009.

He is currently an Associate Professor at the LIESMARS, Wuhan University. His research interests include hyperspectral data analysis, high resolution image processing, pattern recognition and remote sensing applications. He has published more than 25 peer-reviewed articles in international journals such as IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, *Photogrammetric Engineering and Remote Sensing* and *International Journal of Remote Sensing*.

Dr. Huang has served as a Reviewer for most of the international journals for remote sensing. He was the recipient of the Top-Ten Academic Star of Wuhan University in 2009. In 2010, he received the Boeing Award for the best paper in image analysis and interpretation from the American Society for Photogrammetry and Remote Sensing. In 2011, he was the recipient of the New Century Excellent Talents in University from the Ministry of Education of China. In 2011, he was recognized by the IEEE Geoscience and Remote Sensing Society as the Best Reviewer of IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.