

LUO, F., GUO, T., LIN, Z., REN, J. and ZHOU, X. 2020. Semisupervised hypergraph discriminant learning for dimensionality reduction of hyperspectral image. *IEEE journal of selected topics in applied earth observations and remote sensing* [online], 13, pages 4242-4256. Available from: <https://doi.org/10.1109/jstars.2020.3011431>





# Semisupervised hypergraph discriminant learning for dimensionality reduction of hyperspectral image.

LUO, F., GUO, T., LIN, Z., REN, J. and ZHOU, X.

2020

© 2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

# Semisupervised Hypergraph Discriminant Learning for Dimensionality Reduction of Hyperspectral Image

Fulin Luo , Member, IEEE, Tan Guo , Zhiping Lin , Senior Member, IEEE, Jinchang Ren , Senior Member, IEEE, and Xiaocheng Zhou

**Abstract**—Semisupervised learning is an effective technique to represent the intrinsic features of a hyperspectral image (HSI), which can reduce the cost to obtain the labeled information of samples. However, traditional semisupervised learning methods fail to consider multiple properties of an HSI, which has restricted the discriminant performance of feature representation. In this article, we introduce the hypergraph into semisupervised learning to reveal the complex multistuctures of an HSI, and construct a semisupervised discriminant hypergraph learning (SSDHL) method by designing an intraclass hypergraph and an interclass graph with the labeled samples. SSDHL constructs an unsupervised hypergraph with the unlabeled samples. In addition, a total scatter matrix is used to measure the distribution of the labeled and unlabeled samples. Then, a low-dimensional projection function is constructed to compact the properties of the intraclass hypergraph and the unsupervised hypergraph, and simultaneously separate the characteristics of the interclass graph and the total scatter matrix. Finally, according to the objective function, we can obtain the projection matrix and the low-dimensional features. Experiments on three HSI data sets (Botswana, KSC, and PaviaU) show that the proposed method can achieve better classification results compared with a few state-of-the-art methods. The result indicates that SSDHL can

simultaneously utilize the labeled and unlabeled samples to represent the homogeneous properties and restrain the heterogeneous characteristics of an HSI.

**Index Terms**—Dimensionality reduction (DR), graph learning, hyperspectral image (HSI) classification, locality-constrained linear coding, neighborhood margin.

## I. INTRODUCTION

A HYPERSPECTRAL image (HSI) is composed of tens or hundreds of consecutive narrow electromagnetic bands covering the visible-to-infrared spectrums [1], [2]. Due to the abundant spectral and spatial information of an HSI, it can be used to discriminate different types of land covers [3], [4]. Generally, an HSI covers a large region. Machine learning is an effective way to automatically discriminant the land cover types of HSI [5], [6]. For the high-dimensional structure of HSI data, traditional recognition methods face the challenge of the Hughes phenomenon [7], [8]. To address this challenge, the dimensionality reduction (DR) of HSI data becomes crucial.

In the past decades, DR has been widely used to process high-dimensional data, which reduces the dimensionality of data via transforming high-dimensional data into a low-dimensional space while preserving the useful information as much as possible [9], [10]. The classic DR methods include principal component analysis (PCA) and its variations [11], [12], linear discriminant analysis (LDA) [13], and maximum noise fraction (MNF) [14]. These methods utilize the variance property of the data to construct a DR model. Specifically, PCA maximizes the variance of the orthogonal projection of data, LDA considers the within-class and between-class variance, and MNF utilizes the noise variance to construct the projection model. However, these methods only consider the statistical properties, which neglect the intrinsic structures of the data. To better analyze the intrinsic properties of the data, manifold learning has been developed to reveal the geometry structures of the data [15], [16]. Some representative algorithms include isometric mapping (Isomap) [17], locally linear embedding (LLE) [18], and Laplacian eigenmaps (LE) [19]. These methods consider a certain local properties to describe the intrinsic manifold of the high-dimensional data in a low-dimensional space while the three methods belong to nonlinear projection that cannot obtain specific projection matrix to map the out-of-sample into the corresponding low-dimensional space. To address this out-of-sample extension problem, LLE and LE are linearized to neighborhood

Manuscript received May 20, 2020; revised July 6, 2020; accepted July 20, 2020. Date of publication July 23, 2020; date of current version August 7, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61801336, in part by the China Postdoctoral Science Foundation under Grant 2019M662717, in part by the Fundamental Research Funds for the Central Universities under Grant 2042020kf0013, in part by the Science and Technology Major Project of Hubei Province (Next-Generation AI Technologies) under Grant 2019AEA170, in part by the Science, and Technology Research Program of the Chongqing Municipal Education Commission under Grant KJQN201800632, in part by the Open Research Fund of State Key Laboratory of Integrated Services Networks under Grant ISN20-12, in part by Open Research Project of the Hubei Key Laboratory of Intelligent Geo-Information Processing under Grant KLIGIP-2018A06, and in part by the Open Research Fund of Key Laboratory of Spatial Data Mining, and Information Sharing of Ministry of Education, Fuzhou University, under Grant 2019LSDMIS06. (Corresponding author: Tan Guo.)

Fulin Luo is with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: luoflyn@163.com).

Tan Guo is with the School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: guot@cqupt.edu.cn).

Zhiping Lin is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: ezplin@ntu.edu.sg).

Jinchang Ren is with the Department of Electronic and Electrical Engineering, University of Strathclyde, G11XW Glasgow, U.K. (e-mail: jinchang.ren@strath.ac.uk).

Xiaocheng Zhou is with the Key Laboratory of Spatial Data Mining and Information Sharing of Ministry of Education, Fuzhou University, Fuzhou 350116, China (e-mail: zhouxc@fzu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2020.3011431

preserving embedding (NPE) [20] and locality preserving projections (LPP) [21], respectively. To generalize these methods, a unified graph framework was developed to represent these DR models [22]. This framework considers a certain statistic or geometry properties to construct a graph model. Thus, according to different statistic or geometry characteristics, different graph learning algorithms can be designed with various construction manners of similarity matrix and constraint matrix. Based on the view of graph framework, many advanced methods have been developed to better reveal the intrinsic properties of the data, such as marginal Fisher analysis (MFA) [22], local Fisher discriminant analysis (LFDA) [23], regularized local discriminant embedding (RLDE) [24], and local geometric structure Fisher analysis (LGSFA) [25].

The above-mentioned methods are unsupervised or supervised models. For these unsupervised methods, they do not use any prior information to construct the low-dimensional feature learning models. For these supervised methods, they need to know the label information of the training samples to design the DR models. In real-world applications, an unsupervised technique cannot achieve desired performance in general, whereas supervised algorithms usually require a high cost to obtain the labeled information for the training samples [26]. To enhance the discriminant performance or reduce the cost of labeling samples, semisupervised learning was proposed to simultaneously utilize the labeled and unlabeled samples to construct the DR models [27], [28]. Based on the margin criterion, a semisupervised maximum margin criterion (SSMMC) [29] was proposed to improve the margin discriminant power of different classes. According to the theory of graph learning, semisupervised graph learning (SEGL) [30] was designed to build a semisupervised graph with the labeled and unlabeled samples. Meanwhile, LDA was used to develop semisupervised discriminant analysis (SDA) [31] and semisupervised local discriminant analysis (SELD) [32]. In addition, two semisupervised methods, i.e., semisupervised submanifold preserving embedding (S3MPE) [33] and semisupervised sparse manifold discriminative analysis (S3MDA) [34], consider the manifold structures of the data to enhance the representation of intrinsic characteristics.

These traditional graph learning methods only consider the binary relationship between the sample points; hence, they show relatively unrealistic performance on an HSI due to the complex multiple structures it contains. Generally, the traditional graph is difficult to represent the intrinsic high-order relationships of HSIs. To reveal multiple properties of the data, hypergraph was introduced into the field of machine learning [35]–[37]. In hypergraph, each hyperedge contains more than two vertices, whereas the traditional graph just has two vertices on each edge [38], [39]. Subsequently, a series of hypergraph learning methods have been developed for DR of the data. According to LPP, a binary hypergraph (BH) [40] was proposed to extract the features of an HSI. In addition, a supervised hypergraph learning method was proposed to improve the representation performance of low-dimensional features, termed discriminant hyper-Laplacian projection (DHLP) [41]. For an HSI, the spatial-spectral information was used to construct different spatial-spectral hypergraph models, including spatial

hypergraph (SH) [40], hypergraph embedding based spatial-spectral joint features (SSHG) [42], and spatial-spectral hypergraph discriminant analysis (SSHGDA) [43]. However, these hypergraph methods cannot utilize the labeled and unlabeled samples to construct effective DR models, simultaneously.

In this article, we proposed a semisupervised discriminant hypergraph learning (SSDHL) method to obtain the effective low-dimensional features for HSI classification. In SSDHL, we construct an unsupervised hypergraph model with the  $K$  nearest neighbors (NNs) in the unlabeled samples and also design an intraclass hypergraph model with intraclass neighbors with the labeled samples. On the other hand, we utilize the interclass neighbors of the labeled samples to construct an interclass graph, which can enhance the separability of different classes. In order to represent the global structure, we develop a total scatter matrix for all the labeled and unlabeled samples. In the low-dimensional space, we also compact the similarity properties of the unsupervised hypergraph and the intraclass hypergraph while we should separate the different characteristics of the interclass graph and the total matrix as much as possible. According to this criterion, a DR model is designed to obtain the low-dimensional projection matrix. To demonstrate the effectiveness of the proposed method, we select three HSI data sets (i.e. the Botswana, KSC, and PaviaU data sets) to conduct several compared experiments with a few state-of-the-art methods.

The rest of this article is organized as follows. Section II briefly reviews some related works including graph embedding and hypergraph embedding. Section III details our proposed method. Experimental results are presented in Section IV to demonstrate the effectiveness of the proposed SSDHL method. Finally, Section V provides some concluding remarks and suggestions for future works.

## II. RELATED WORKS

For an HSI data set, the labeled samples are denoted as  $\mathbf{X}_l = [\mathbf{x}_{l,1}, \mathbf{x}_{l,2}, \dots, \mathbf{x}_{l,n_l}] \in \mathbf{R}^{B \times n_l}$  and the unlabeled samples are denoted as  $\mathbf{X}_u = [\mathbf{x}_{u,1}, \mathbf{x}_{u,2}, \dots, \mathbf{x}_{u,n_u}] \in \mathbf{R}^{B \times n_u}$ , where  $B$  is the number of bands and  $n_l$  and  $n_u$  are the numbers of labeled and unlabeled samples, respectively.  $\ell(\mathbf{x}_{l,i}) \in \{1, 2, \dots, c\}$  is the class label of  $\mathbf{x}_{l,i}$ , where  $c$  is the number of classes.  $\mathbf{X} = [\mathbf{X}_l, \mathbf{X}_u] \in \mathbf{R}^{B \times n}$  denotes the whole data set, where  $n = n_l + n_u$  is the total number of samples. The low-dimensional features are represented as  $\mathbf{Y} = [\mathbf{y}_{l,1}, \mathbf{y}_{l,2}, \dots, \mathbf{y}_{l,n_l}, \mathbf{y}_{u,1}, \mathbf{y}_{u,2}, \dots, \mathbf{y}_{u,n_u}] \in \mathbf{R}^{d \times n}$ , where  $d$  is the embedding dimension.  $\mathbf{X}$  can be transformed as  $\mathbf{Y} = \mathbf{V}^T \mathbf{X}$  with a projection matrix  $\mathbf{V} \in \mathbf{R}^{B \times d}$ .

### A. Graph Embedding

Graph has been widely used to reveal the relationships between different samples. For an undirected graph  $\mathbf{G}$ , it can be denoted as  $\mathbf{G} = \{\mathbf{X}, \mathbf{E}, \mathbf{W}\}$ , where  $\mathbf{X}$  is the vertex set,  $\mathbf{E}$  is the edge set, and  $\mathbf{W} = [w_{ij}]_{i,j=1}^n$  is the weight matrix of edges. To construct a graph, some similarity measure methods are adopted to define the connection of edges between two vertices and its corresponding weight. If vertices  $i$  and  $j$  are similar, an edge should be connected between vertices  $i$  and  $j$ , and a similarity weight of the edge should be defined at the same time.

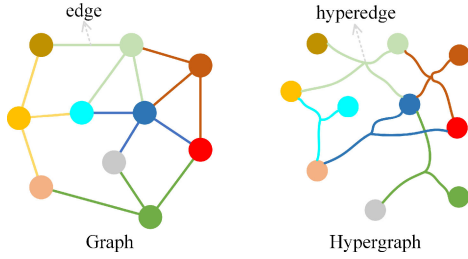


Fig. 1. Graph and hypergraph structures.

In a low-dimensional space, the structure of the graph should be preserved and similar samples should be as compact as possible. Therefore, the objective function can be represented as the following:

$$\begin{aligned} \min_{\mathbf{Y}} \frac{1}{2} \sum_{i \neq j} \|\mathbf{y}_i - \mathbf{y}_j\|^2 w_{ij} &= \min_{\mathbf{Y}} \text{tr}(\mathbf{Y}\mathbf{L}\mathbf{Y}^T) \\ \text{s.t. } \text{tr}(\mathbf{Y}\mathbf{H}\mathbf{Y}^T) &= h \end{aligned} \quad (1)$$

where  $h$  is a constant,  $\mathbf{H}$  is a constraint matrix being used to avoid a trivial solution of the objective function,  $\mathbf{L} = \mathbf{D} - \mathbf{W}$  is the Laplacian matrix of the graph  $\mathbf{G}$ , and  $\mathbf{D} = \text{diag}([\sum_{j=1}^n w_{1j}, \sum_{j=1}^n w_{2j}, \dots, \sum_{j=1}^n w_{nj}])$  is a diagonal matrix. In general,  $\mathbf{H}$  is set to an identity matrix for scale normalization or a Laplacian matrix of penalty graph that is used to suppress some of the unwanted properties of data.

### B. Hypergraph Embedding

The main difference between the hypergraph and graph is the number of vertices for each edge. In a graph, each edge only has two vertices, whereas each edge can contain more than two vertices in a hypergraph as shown in Fig. 1.

For a hypergraph, it is defined as  $\mathbf{G}^H = \{\mathbf{X}, \mathbf{E}^H, \mathbf{W}^H\}$ , where  $\mathbf{X}$  is the vertex set,  $\mathbf{E}^H$  is the hyperedge set,  $\mathbf{W}^H$  is the weight matrix of the hyperedges, and each hyperedge  $e_i \in \mathbf{E}^H$  has a weight  $w_i^e$ . An incidence matrix  $\mathbf{H} = [h_{ij}^{ve}]_{i,j} \in \mathfrak{R}^{|\mathbf{V}^H| \times |\mathbf{E}^H|}$  is used to represent the relationship between the vertex  $v_i$  and the hyperedge  $e_j$ , and  $h_{ij}^{ve}$  is defined as follows:

$$h_{ij}^{ve} = \begin{cases} 1, & \text{if } v_i \in e_j \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

To reveal the properties of the hypergraph, the degree of vertex  $v_i$  is the summation of the weights of the hyperedges via vertex  $v_i$ , and the degree of hyperedge  $e_i$  is the number of vertices on the hyperedge  $e_i$ . Therefore, the degrees of vertex  $v_i$  and hyperedge  $e_j$  are defined as follows:

$$d_i^v = \sum_{j=1}^n w_j^e h_{ij}^{ve} \quad (3)$$

$$d_j^e = |e_j| = \sum_{i=1}^n h_{ij}^{ve} \quad (4)$$

In a low-dimensional space, the samples from the same hyperedge should be compact as close as possible, and the objective

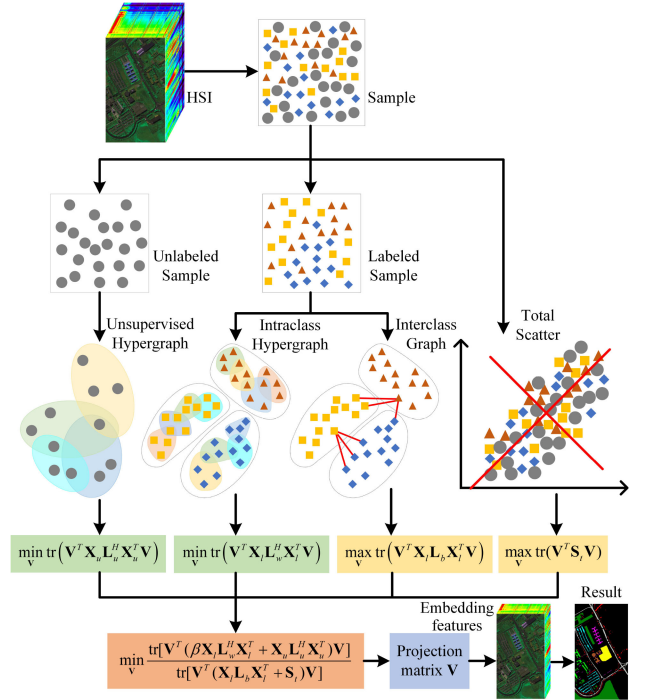


Fig. 2. Procedure of the proposed SSDHL method.

function can be represented as the following:

$$\begin{aligned} \min_{\mathbf{Y}} \frac{1}{2} \sum_{e_i \in \mathbf{E}^H} \frac{w_i^e}{d_i^e} \sum_{(\mathbf{x}_j, \mathbf{x}_k) \in e_i} \left\| \frac{\mathbf{y}_j}{\sqrt{d_j^v}} - \frac{\mathbf{y}_k}{\sqrt{d_k^v}} \right\|^2 &= \text{tr}(\mathbf{Y}\mathbf{L}^H\mathbf{Y}^T) \\ \text{s.t. } \text{tr}(\frac{1}{n}\mathbf{Y}\mathbf{Y}^T) &= 1 \end{aligned} \quad (5)$$

where  $\mathbf{L}^H = \mathbf{I} - (\mathbf{D}^v)^{-1/2} \mathbf{H}\mathbf{W}^H (\mathbf{D}^e)^{-1} (\mathbf{H})^T (\mathbf{D}^v)^{-1/2}$  is the hyper-Laplacian matrix.  $\mathbf{W}^H = \text{diag}([w_1^e, w_2^e, \dots, w_n^e])$  is the weight matrix.  $\mathbf{D}^e = \text{diag}([d_1^e, d_2^e, \dots, d_n^e])$  and  $\mathbf{D}^v = \text{diag}([d_1^v, d_2^v, \dots, d_n^v])$  are the diagonal matrices of the vertex and hyperedge degrees.

### III. SEMISUPERVISED DISCRIMINANT HYPERGRAPH LEARNING (SSDHL)

In this section, we propose a semisupervised hypergraph learning model to extract the low-dimensional features of an HSI with both the labeled and unlabeled samples. This model is termed as SSDHL. First, an unsupervised hypergraph is constructed to reveal the similarity of the unlabeled samples, and an intra-class hypergraph is designed to represent the relationships of the labeled samples from the same class. To enhance the discriminant performance of different classes, we also construct an interclass graph to separate the samples from different classes as much as possible. In addition, a total scatter matrix is used to disperse the distribution of all the labeled and unlabeled samples. A semisupervised DR model is constructed to compact the similar properties of the unsupervised hypergraph and the intra-class hypergraph and separate the difference characteristics of the interclass graph and the total scatter matrix. The workflow of the proposed SSDHL method is shown in Fig. 2.



### A. Unsupervised Hypergraph

For the unlabeled samples, an unsupervised hypergraph  $\mathbf{G}_u^H = \{\mathbf{X}_u, \mathbf{E}_u^H, \mathbf{W}_u^H\}$  is constructed to represent the multiple relationships of the unlabeled samples, where  $\mathbf{X}_u$  is the vertex set of the unlabeled samples,  $\mathbf{E}_u^H$  is the unsupervised hyperedge set, and  $\mathbf{W}_u^H$  is the unsupervised weight matrix of unsupervised hyperedges.  $K_u$  NNs are selected by the Euclidean distance to construct an unsupervised incidence matrix  $\mathbf{H}^u = [h_{ij}^u]_{i,j}$

$$h_{ij}^u = \begin{cases} 1, & \text{if } \mathbf{x}_{u,j} \in \mathcal{N}_{K_u}(\mathbf{x}_{u,i}) \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

where  $\mathcal{N}_{K_u}(\mathbf{x}_{u,i})$  denotes the  $K_u$  NNs of  $\mathbf{x}_{u,i}$  in the unlabeled set. An unsupervised hyperedge  $e_i^u \in \mathbf{E}_u^H$  possesses  $K_u$  vertices that are  $K_u$  NNs of  $\mathbf{x}_{u,i}$  from the unlabeled samples.

For an unsupervised hyperedge  $e_i^u$ , the weight is used to represent the similarity of these vertices on this hyperedge as defined in the following:

$$w_i^u = \sum_{\mathbf{x}_{u,j} \in \mathcal{N}_{K_u}(\mathbf{x}_{u,i})} \exp\left(-\frac{\|\mathbf{x}_{u,j} - \mathbf{x}_{u,i}\|^2}{2(t_i^u)^2}\right) \quad (7)$$

where  $t_i^u = (1/K_u) \sum_{\mathbf{x}_{u,j} \in \mathcal{N}_{K_u}(\mathbf{x}_{u,i})} \|\mathbf{x}_{u,j} - \mathbf{x}_{u,i}\|$  is the kernel parameter, and  $\mathbf{W}_u^H = \text{diag}([w_1^u, w_2^u, \dots, w_{n_u}^u])$  is the unsupervised hyperedge weight matrix.

According to the unsupervised incidence matrix and the unsupervised weight matrix, the degrees of vertex  $\mathbf{x}_{u,i}$  and hyperedge  $e_i^u$  are defined to reveal the intrinsic properties of unlabeled samples as the following:

$$d_{v,i}^u = \sum_{j=1}^{n_u} w_j^u h_{ij}^u \quad (8)$$

$$d_{e,i}^u = \sum_{i=1}^{n_u} h_{ij}^u. \quad (9)$$

In a low-dimensional space, we should preserve the structures of the unsupervised hypergraph and compact the similar samples as much as possible. Therefore, the objective function can be defined as follows:

$$\min_{\mathbf{V}} \frac{1}{2} \sum_{e_i^u \in \mathbf{E}_u^H} \frac{w_i^u}{d_{e,i}^u} \sum_{(\mathbf{x}_{u,j}, \mathbf{x}_{u,k}) \in e_i^u} \left\| \frac{\mathbf{V}^T \mathbf{x}_{u,j}}{\sqrt{d_{v,j}^u}} - \frac{\mathbf{V}^T \mathbf{x}_{u,k}}{\sqrt{d_{v,k}^u}} \right\|^2. \quad (10)$$

With some mathematical operations, we can simplify the objection function (10) with the following procedure in (11) shown at the bottom of this page, where  $\mathbf{L}_u^H = \mathbf{I} - (\mathbf{D}_v^u)^{-1/2} \mathbf{H}^u \mathbf{W}_u^H (\mathbf{D}_e^u)^{-1} (\mathbf{H}^u)^T (\mathbf{D}_v^u)^{-1/2}$  is the unsupervised hyper-Laplacian matrix, and  $\mathbf{D}_e^u = \text{diag}([d_{e,1}^u, d_{e,2}^u, \dots, d_{e,n_u}^u])$

$$\begin{aligned} & \frac{1}{2} \sum_{e_i^u \in \mathbf{E}_u^H} \frac{w_i^u}{d_{e,i}^u} \sum_{(\mathbf{x}_{u,j}, \mathbf{x}_{u,k}) \in e_i^u} \left\| \frac{\mathbf{V}^T \mathbf{x}_{u,j}}{\sqrt{d_{v,j}^u}} - \frac{\mathbf{V}^T \mathbf{x}_{u,k}}{\sqrt{d_{v,k}^u}} \right\|^2 \\ &= \sum_{e_i^u \in \mathbf{E}_u^H} \frac{w_i^u}{d_{e,i}^u} \sum_{(\mathbf{x}_{u,j}, \mathbf{x}_{u,k}) \in e_i^u} \text{tr} \left( \frac{\mathbf{V}^T \mathbf{x}_{u,j} \mathbf{x}_{u,j}^T \mathbf{V}}{d_{v,j}^u} - \frac{\mathbf{V}^T \mathbf{x}_{u,j} \mathbf{x}_{u,k}^T \mathbf{V}}{\sqrt{d_{v,j}^u} \sqrt{d_{v,k}^u}} \right) \\ &= \sum_{e_i^u \in \mathbf{E}_u^H} \frac{w_i^u}{d_{e,i}^u} \sum_{\mathbf{x}_{u,j}, \mathbf{x}_{u,k} \in \mathbf{X}_u} h_{ij}^u h_{ik}^u \text{tr} \left( \frac{\mathbf{V}^T \mathbf{x}_{u,j} \mathbf{x}_{u,j}^T \mathbf{V}}{d_{v,j}^u} - \frac{\mathbf{V}^T \mathbf{x}_{u,j} \mathbf{x}_{u,k}^T \mathbf{V}}{\sqrt{d_{v,j}^u} \sqrt{d_{v,k}^u}} \right) \\ &= \sum_{e_i^u \in \mathbf{E}_u^H} \sum_{\mathbf{x}_{u,j}, \mathbf{x}_{u,k} \in \mathbf{X}_u} \left( \frac{h_{ik}^u}{d_{e,i}^u} \times \frac{w_i^u h_{ij}^u \text{tr}(\mathbf{V}^T \mathbf{x}_{u,j} \mathbf{x}_{u,j}^T \mathbf{V})}{d_{v,j}^u} - \frac{\text{tr}(\mathbf{V}^T \mathbf{x}_{u,j} h_{ij}^u w_i^u h_{ik}^u \mathbf{x}_{u,k}^T \mathbf{V})}{d_{e,i}^u \sqrt{d_{v,j}^u} \sqrt{d_{v,k}^u}} \right) \\ &= \sum_{\mathbf{x}_{u,j} \in \mathbf{X}_u} \left( \text{tr}(\mathbf{V}^T \mathbf{x}_{u,j} \mathbf{x}_{u,j}^T \mathbf{V}) \frac{1}{d_{v,j}^u} \sum_{e_i^u \in \mathbf{E}_u^H} \left( w_i^u h_{ij}^u \frac{\sum_{\mathbf{x}_{u,k} \in \mathbf{X}_u} h_{ik}^u}{d_{e,i}^u} \right) \right) \\ &\quad - \sum_{e_i^u \in \mathbf{E}_u^H} \sum_{\mathbf{x}_{u,j}, \mathbf{x}_{u,k} \in \mathbf{X}_u} \text{tr} \left( \mathbf{V}^T \mathbf{x}_{u,j} \frac{1}{\sqrt{d_{v,j}^u}} h_{ij}^u w_i^u \frac{1}{d_{e,i}^u} h_{ik}^u \frac{1}{\sqrt{d_{v,k}^u}} \mathbf{x}_{u,k}^T \mathbf{V} \right) \\ &= \text{tr}(\mathbf{V}^T \mathbf{X}_u \mathbf{X}_u^T \mathbf{V}) \\ &\quad - \text{tr} \left( \mathbf{V}^T \mathbf{X}_u (\mathbf{D}_v^u)^{-1/2} \mathbf{H}^u \mathbf{W}_u^H (\mathbf{D}_e^u)^{-1} (\mathbf{H}^u)^T (\mathbf{D}_v^u)^{-1/2} \mathbf{X}_u^T \mathbf{V} \right) \\ &= \text{tr} \left\{ \mathbf{V}^T \mathbf{X}_u \left[ \mathbf{I} - (\mathbf{D}_v^u)^{-1/2} \mathbf{H}^u \mathbf{W}_u^H (\mathbf{D}_e^u)^{-1} (\mathbf{H}^u)^T (\mathbf{D}_v^u)^{-1/2} \right] \mathbf{X}_u^T \mathbf{V} \right\} \\ &= \text{tr}(\mathbf{V}^T \mathbf{X}_u \mathbf{L}_u^H \mathbf{X}_u^T \mathbf{V}). \end{aligned} \quad (11)$$

and  $\mathbf{D}_v^u = \text{diag}([d_{v,1}^u, d_{v,2}^u, \dots, d_{v,n_u}^u])$  are the degree matrices of the unsupervised hyperedges and the vertices.

According to the transformation procedure of (17), for the unlabeled samples, the low-dimension embedding function can be denoted by

$$\min_{\mathbf{V}} \text{tr}(\mathbf{V}^T \mathbf{X}_u \mathbf{L}_u^H \mathbf{X}_u^T \mathbf{V}). \quad (12)$$

### B. Intraclass Hypergraph

To reveal the multiple properties of the samples from the same class, we construct an intraclass hypergraph  $\mathbf{G}_w^H = \{\mathbf{X}_l, \mathbf{E}_w^H, \mathbf{W}_w^H\}$ , where  $\mathbf{X}_l$  is the vertex set of the labeled samples,  $\mathbf{E}_w^H$  is the intraclass hyperedge set, and  $\mathbf{W}_w^H$  is the intraclass weight matrix of intraclass hyperedges. An intraclass incidence matrix can be defined as  $\mathbf{H}^w = [h_{ij}^w]_{i,j}$  via selecting  $K_w$  NNs from the same class with the Euclidean distance in the labeled set

$$h_{ij}^w = \begin{cases} 1, & \text{if } \mathbf{x}_{l,j} \in \mathcal{N}_{K_w}(\mathbf{x}_{l,i}) \text{ and } \ell(\mathbf{x}_{l,i}) = \ell(\mathbf{x}_{l,j}) \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

where  $\mathcal{N}_{K_w}(\mathbf{x}_{l,i})$  denotes the  $K_w$  NNs of  $\mathbf{x}_{l,i}$  from the same class.

For an intraclass hyperedge  $e_i^w$ , the hyperedge weight, which reflects the similarity of the vertices on the hyperedge  $e_i^w$ , is set as the following:

$$w_i^w = \sum_{\mathbf{x}_{l,j} \in \mathcal{N}_{K_w}(\mathbf{x}_{l,i})} \exp\left(-\frac{\|\mathbf{x}_{l,j} - \mathbf{x}_{l,i}\|^2}{2(t_i^w)^2}\right) \quad (14)$$

where  $t_i^w = (1/K_w) \sum_{\mathbf{x}_{l,j} \in \mathcal{N}_{K_w}(\mathbf{x}_{l,i})} \|\mathbf{x}_{l,j} - \mathbf{x}_{l,i}\|$  is the kernel parameter, and  $\mathbf{W}_w^H = \text{diag}([w_1^w, w_2^w, \dots, w_{n_l}^w])$  is the intra-class hyperedge weight matrix.

According to the intraclass incidence matrix, we can obtain the degrees of intraclass hyperedge  $e_i^w$  and vertex  $\mathbf{x}_{l,i}$  to represent the intrinsic properties of hypergraph, i.e.,

$$d_{v,i}^w = \sum_{j=1}^{n_l} w_j^w h_{ij}^w \quad (15)$$

$$d_{e,j}^w = \sum_{i=1}^{n_l} h_{ij}^w. \quad (16)$$

In a low-dimensional space, the intraclass hypergraph structure is preserved to reveal the intrinsic properties of the labeled samples and the intraclass similar samples are compacted as much as possible. With the derived  $h_{ij}^w$ ,  $d_{v,i}^w$ ,  $w_i^w$ , and  $d_{e,j}^w$ , we can obtain  $h_{ij}^{wr}$ ,  $d_{v,i}^{wr}$ ,  $w_i^{wr}$ , and  $d_{e,j}^{wr}$  of the  $r$ th class, and the low-dimensional projection function can be determined by

$$\min_{\mathbf{V}} \frac{1}{2} \sum_{r=1}^c \sum_{e_i^{wr} \in \mathbf{E}_w^H} \frac{w_i^{wr}}{d_{e,i}^{wr}} \sum_{(\mathbf{x}_{r,j}, \mathbf{x}_{r,k}) \in e_i^{wr}} \left\| \frac{\mathbf{V}^T \mathbf{x}_{r,j}}{\sqrt{d_{v,j}^{wr}}} - \frac{\mathbf{V}^T \mathbf{x}_{r,k}}{\sqrt{d_{v,k}^{wr}}} \right\|^2 \quad (17)$$

where  $e_i^{wr}$  is the  $i$ th hyperedge in the  $r$ th class,  $\mathbf{x}_{r,j}$  is the  $j$ th sample in the  $r$ th class,  $d_{e,i}^{wr}$  is the degree of intraclass hyperedge  $e_i^{wr}$ , and  $d_{v,j}^{wr}$  is the degree of vertex  $\mathbf{x}_{r,j}$ .

With the same transformation procedure of (11), the low-dimensional embedding function of (17) can be simplified as

$$\min_{\mathbf{V}} \sum_{r=1}^c \text{tr}(\mathbf{V}^T \mathbf{X}_r \mathbf{L}_{wr}^H \mathbf{X}_r^T \mathbf{V}) = \min_{\mathbf{V}} \text{tr}(\mathbf{V}^T \mathbf{X}_l \mathbf{L}_w^H \mathbf{X}_l^T \mathbf{V}) \quad (18)$$

where  $\mathbf{X}_r = [\mathbf{x}_{r,j}]_j$  is the  $r$ th class samples,  $\mathbf{X}_l = [\mathbf{X}_r]_{r=1}^c$ ,  $\mathbf{L}_{wr}^H = \mathbf{I} - (\mathbf{D}_v^{wr})^{-1/2} \mathbf{H}^{wr} \mathbf{W}_w^H (\mathbf{D}_e^{wr})^{-1} (\mathbf{H}^{wr})^T (\mathbf{D}_v^{wr})^{-1/2}$  is the intraclass hyper-Laplacian matrix of the  $r$ th class,  $\mathbf{D}_e^{wr} = \text{diag}([d_{e,1}^{wr}, d_{e,2}^{wr}, \dots, d_{e,n_r}^{wr}])$  and  $\mathbf{D}_v^{wr} = \text{diag}([d_{v,1}^{wr}, d_{v,2}^{wr}, \dots, d_{v,n_r}^{wr}])$  are the degree matrices of the intraclass hyperedges and the vertices of the  $r$ th class,  $\mathbf{H}^{wr} = [h_{ij}^{wr}]_{i,j=1}^{n_r}$  is the incidence matrix of the  $r$ th class,  $n_r$  is the number of the  $r$ th class, and  $\mathbf{L}_w^H = \text{diag}([\mathbf{L}_{wr}^H]_{r=1}^c)$  is the intraclass hyper-Laplacian matrix.

### C. Interclass Graph

To further enhance the difference of the samples from different classes, we construct a graph  $\mathbf{G}_b = \{\mathbf{X}_l, \mathbf{E}_b, \mathbf{W}_b\}$  to reveal the relationships of interclass samples that should be restrained in low-dimensional space. In a graph  $\mathbf{G}_b$ ,  $\mathbf{X}_l$  is the vertex set consisting of the labeled samples,  $\mathbf{E}_b$  denotes the interclass edge set, and  $\mathbf{W}_b$  is the interclass weight matrix of the interclass edges. To construct the graph  $\mathbf{G}_b$ , we adopt the Euclidean distance to select  $K_b$  interclass NNs and connect each neighbor with an edge. For each edge, we set a weight  $w_{ij}^b$  to reveal the similarity of vertices  $i$  and  $j$ . If two vertices are unconnected, the weight is set to zero. Therefore, the weight can be defined as follows:

$$w_{ij}^b = \begin{cases} \exp\left(-\frac{\|\mathbf{x}_{l,j} - \mathbf{x}_{l,i}\|^2}{2(t_i^b)^2}\right), & \text{if } \mathbf{x}_{l,j} \in \mathcal{N}_{K_b}(\mathbf{x}_{l,i}) \\ & \text{and } \ell(\mathbf{x}_{l,i}) \neq \ell(\mathbf{x}_{l,j}) \\ 0, & \text{otherwise} \end{cases} \quad (19)$$

where  $t_i^b = (1/K_b) \sum_{\mathbf{x}_{l,j} \in \mathcal{N}_{K_b}(\mathbf{x}_{l,i})} \|\mathbf{x}_{l,j} - \mathbf{x}_{l,i}\|$  is the kernel parameter.  $\mathcal{N}_{K_b}(\mathbf{x}_{l,i})$  denotes the  $K_b$  NNs of  $\mathbf{x}_{l,i}$  from different classes.  $\mathbf{W}_b = [w_{ij}^b]_{i,j=1}^{n_l}$  is the interclass weight matrix. Since the graph is undirected, the weight matrix should be symmetric and it should be symmetrized to  $\mathbf{W}_b = \max(\mathbf{W}_b, \mathbf{W}_b^T)$ .

The degree of vertex  $\mathbf{x}_{l,i}$  is denoted as

$$d_{v,i}^b = \sum_{j=1}^{n_l} w_{i,j}^b. \quad (20)$$

In a low-dimensional space, the relationships of  $\mathbf{G}_b$  should be separated as much as possible to restrain the similarity of the interclass NNs. According to  $w_{ij}^b$  and  $d_{v,i}^b$ , we can obtain  $d_{v,i}^{bp}$  of the  $p$ th class and  $w_i^{bpq}$  between the  $p$ th class and the  $q$ th class, and an objective function can be defined as follows:

$$\max_{\mathbf{V}} \frac{1}{2} \sum_{p=1}^c \sum_{q=1, q \neq p}^c \sum_{i=1}^{n_p} \sum_{j=1}^{n_q} w_{ij}^{bpq} \left\| \frac{\mathbf{V}^T \mathbf{x}_{p,i}}{\sqrt{d_{v,i}^{bp}}} - \frac{\mathbf{V}^T \mathbf{x}_{q,j}}{\sqrt{d_{v,j}^{bq}}} \right\|^2 \quad (21)$$

where  $\mathbf{x}_{p,i}$  is the  $i$ th sample from the  $p$ th class and  $d_{v,i}^{bp}$  is the degree of vertex  $\mathbf{x}_{p,i}$ .

For (21), we can transform it into another form with the total labeled samples by

$$\max_{\mathbf{V}} \frac{1}{2} \sum_{i=1}^{n_l} \sum_{j=1}^{n_l} w_{ij}^b \left\| \frac{\mathbf{V}^T \mathbf{x}_{l,i}}{\sqrt{d_{v,i}^b}} - \frac{\mathbf{V}^T \mathbf{x}_{l,j}}{\sqrt{d_{v,j}^b}} \right\|^2. \quad (22)$$

With some mathematical operations, (22) can be transformed as a matrix operation as follows:

$$\begin{aligned} & \frac{1}{2} \sum_{i=1}^{n_l} \sum_{j=1}^{n_l} w_{ij}^b \left\| \frac{\mathbf{V}^T \mathbf{x}_{l,i}}{\sqrt{d_{v,i}^b}} - \frac{\mathbf{V}^T \mathbf{x}_{l,j}}{\sqrt{d_{v,j}^b}} \right\|^2 \\ &= \sum_{i=1}^{n_l} \sum_{j=1}^{n_l} \text{tr} \left( \mathbf{V}^T \mathbf{x}_{l,i} \frac{w_{ij}^b}{d_{v,i}^b} \mathbf{x}_{l,i}^T \mathbf{V} - \mathbf{V}^T \mathbf{x}_{l,i} \frac{w_{ij}^b}{\sqrt{d_{v,i}^b} \sqrt{d_{v,j}^b}} \mathbf{x}_{l,j}^T \mathbf{V} \right) \\ &= \text{tr} \left( \mathbf{V}^T \left\{ \sum_{i=1}^{n_l} \left[ \mathbf{x}_{l,i} \left( \frac{\sum_{j=1}^{n_l} w_{ij}^b}{d_{v,i}^b} \right) \mathbf{x}_{l,i}^T \right] \right. \right. \\ & \quad \left. \left. - \sum_{i,j=1}^{n_l} \left( \mathbf{x}_{l,i} \frac{w_{ij}^b}{\sqrt{d_{v,i}^b} \sqrt{d_{v,j}^b}} \mathbf{x}_{l,i}^T \right) \right\} \mathbf{V} \right) \\ &= \text{tr} \left( \mathbf{V}^T \left[ \sum_{i=1}^{n_l} (\mathbf{x}_{l,i} \mathbf{x}_{l,i}^T) \right. \right. \\ & \quad \left. \left. - \sum_{i,j=1}^{n_l} \left( \mathbf{x}_{l,i} (d_{v,i}^b)^{-1/2} w_{ij}^b (d_{v,j}^b)^{-1/2} \mathbf{x}_{l,i}^T \right) \right] \mathbf{V} \right) \\ &= \text{tr}(\mathbf{V}^T \mathbf{X}_l [\mathbf{I} - (\mathbf{D}_v^b)^{-1/2} \mathbf{W}_b (\mathbf{D}_v^b)^{-1/2}] \mathbf{X}_l^T \mathbf{V}) \\ &= \text{tr}(\mathbf{V}^T \mathbf{X}_l \mathbf{L}_b \mathbf{X}_l^T \mathbf{V}) \end{aligned} \quad (23)$$

where  $\mathbf{L}_b = \mathbf{I} - (\mathbf{D}_v^b)^{-1/2} \mathbf{W}_b (\mathbf{D}_v^b)^{-1/2}$  denotes the interclass Laplacian matrix, and  $\mathbf{D}_v^b = \text{diag}(d_{v,1}^b, d_{v,2}^b, \dots, d_{v,n_l}^b)$  is the diagonal matrix of the vertex degree.

According to (23), the interclass graph embedding function can be simplified as

$$\max_{\mathbf{V}} \text{tr}(\mathbf{V}^T \mathbf{X}_l \mathbf{L}_b \mathbf{X}_l^T \mathbf{V}). \quad (24)$$

#### D. Total Scatter

To better combine the labeled and unlabeled samples, we construct a total scatter matrix to control the distribution of all the samples. The total scatter matrix is defined as the following:

$$\mathbf{S}_t = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{X}})(\mathbf{x}_i - \bar{\mathbf{X}})^T \quad (25)$$

where  $\bar{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$  is the mean of  $\mathbf{X}$ .

In the low-dimensional space, the total scatter should be dispersed as much as possible to better discriminate these samples from different classes. Therefore, the objective function can be defined as follows:

$$\max_{\mathbf{V}} \text{tr}(\mathbf{V}^T \mathbf{S}_t \mathbf{V}). \quad (26)$$

#### E. Feature Embedding

To reduce the heterogeneity of samples from the same class and enhance the difference of samples from different classes, the intraclass samples should be compacted and the interclass samples should be separated as much as possible in the low-dimensional space. According to the intraclass hypergraph, the interclass graph, the unsupervised hypergraph, and the total scatter matrix, we design a combined projection function to obtain the low-dimensional discriminant features and reveal the complex intrinsic properties of an HSI. The projection matrix can be computed by the following optimization problem:

$$\min_{\mathbf{V}} \frac{\text{tr}[\mathbf{V}^T (\beta \mathbf{X}_l \mathbf{L}_w^H \mathbf{X}_l^T + \mathbf{X}_u \mathbf{L}_u^H \mathbf{X}_u^T) \mathbf{V}]}{\text{tr}[\mathbf{V}^T (\mathbf{X}_l \mathbf{L}_b \mathbf{X}_l^T + \mathbf{S}_t) \mathbf{V}]} = \min_{\mathbf{V}} \frac{\text{tr}(\mathbf{V}^T \mathbf{P}^l \mathbf{V})}{\text{tr}(\mathbf{V}^T \mathbf{P}^g \mathbf{V})} \quad (27)$$

where  $\beta > 1$  is a label weight parameter to enhance the contribution of the labeled samples.  $\mathbf{P}^l = \beta \mathbf{X}_l \mathbf{L}_w^H \mathbf{X}_l^T + \mathbf{X}_u \mathbf{L}_u^H \mathbf{X}_u^T$  is the local related matrix and  $\mathbf{P}^g = \mathbf{X}_l \mathbf{L}_b \mathbf{X}_l^T + \mathbf{S}_t$  denotes the global related matrix. The local related matrix reveals the homogeneity of the similar samples that may be from the same class, whereas the global related matrix represents the heterogeneity of the differential samples that may be from different classes.

According to the principle of the generalized Rayleigh quotient, the optimization function of (27) is equal to the following problem:

$$\begin{aligned} & \min_{\mathbf{V}} \mathbf{V}^T \mathbf{P}^l \mathbf{V} \\ & \text{s.t. } \mathbf{V}^T \mathbf{P}^g \mathbf{V} = \mathbf{I}. \end{aligned} \quad (28)$$

To solve the optimization problem (28), we use the method of Lagrangian multipliers [44], [45] to obtain the optimal projection matrix, and the Lagrangian function can be defined as follows:

$$L(\mathbf{V}, \lambda) = \mathbf{V}^T \mathbf{P}^l \mathbf{V} - \lambda (\mathbf{V}^T \mathbf{P}^g \mathbf{V} - \mathbf{I}) \quad (29)$$

where  $\lambda$  is the Lagrangian multiplier.

To obtain the optimal projection, we set the derivative of  $L(\mathbf{V}, \lambda)$  to zero, and the optimization problem can be transformed to a generalized eigenvalue problem

$$\mathbf{P}^l \mathbf{v}_i = \lambda_i \mathbf{P}^g \mathbf{v}_i \quad (30)$$

where  $\lambda_i$  is an eigenvalue and its corresponding eigenvector is  $\mathbf{v}_i$ . The optimization solution is composed of the  $d$  largest eigenvalues corresponding eigenvectors, and the projection matrix can be denoted as the following:

$$\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d] \in \mathbb{R}^{B \times d}. \quad (31)$$

According to the projection matrix, we can derive the low-dimensional embedding features of  $\mathbf{x}_i$  as

$$\mathbf{y}_i = \mathbf{V}^T \mathbf{x}_i. \quad (32)$$

#### F. Illustration of SSDHL

In summary, the proposed SSDHL method constructs an intraclass hypergraph and an interclass graph with the labeled samples to represent the intrinsic properties of the data. SSDHL also generates an unsupervised hypergraph to reveal the

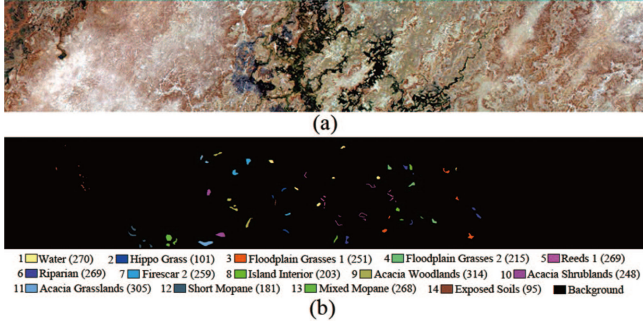


Fig. 3. Botswana HSI. (a) False-color image. (b) Ground truth. (Note that the number of samples for each class is shown in brackets.)

similarity of the unlabeled samples. In addition, a total scatter matrix is designed to control the distribution of all the samples, which can better combine the labeled and unlabeled samples. For SSDHL, the local information is represented with the intraclass and unsupervised hypergraphs, whereas the global information is described with the interclass graph and the total scatter matrix. The proposed method can use a few labeled samples and a large number of unlabeled samples to effectively reveal the intrinsic structures of the HSI data, and it can better extract the low-dimensional embedding features to improve the discriminant performance of an HSI. The detailed procedure of SSDHL is shown in Algorithm 1.

#### IV. EXPERIMENTAL RESULTS

To demonstrate the effectiveness of the proposed SSDHL, we conduct comprehensive experiments on three publicly available HSI data sets and benchmarked with a few state-of-the-art methods for comparison and performance evaluation.

##### A. Data Sets

1) *Botswana Data Set*: This data set was captured by the NASA Earth Observing-1 satellite over the Okavango Delta, Botswana on May 31, 2001. The spatial resolution is 30 m with a size of  $1476 \times 256$  pixels. This image contains 242 bands with the electromagnetic spectrums in the range of 400–2500 nm. After removing noise and water absorption bands, the retained 145 bands were used for scientific research. This data set contains 3428 labeled samples in 14 types of land cover. The false-colored image and the ground truth are shown in Fig. 3.

2) *KSC Data Set*: This data set was acquired by the NASA Airborne Visible Infrared Imaging Spectrometer sensor over the Kennedy Space Center (KSC) on March 23, 1996. The image has a spatial resolution of 18 m with 224 electromagnetic spectrum bands. Due to the influence of water absorption and low signal-to-noise ratio, only 176 bands were retained for experimental analysis. This data set has a size of  $512 \times 614$  pixels and a total of 5211 pixels were labeled in 13 types of land cover. Fig. 4 displays the false color image and its corresponding ground truth.

3) *PaviaU Data Set*: This image was captured by the Reflective Optics System Imaging Spectrometer sensor in 2001. The

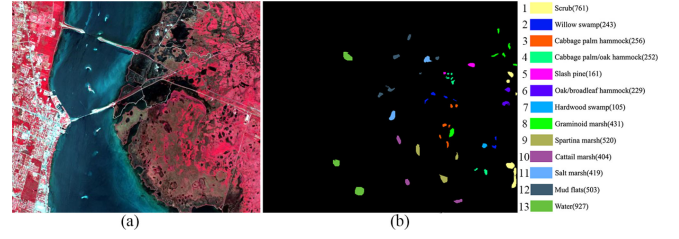


Fig. 4. KSC HSI. (a) False-color image. (b) Ground truth. (Note that the number of samples for each class is shown in brackets.)

---

#### Algorithm 1: SSDHL.

---

**Input:** Labeled samples

$\mathbf{X}_l = [\mathbf{x}_{l,1}, \mathbf{x}_{l,2}, \dots, \mathbf{x}_{l,n_l}] \in \mathbb{R}^{B \times n_l}$  and their corresponding class labels

$\{\ell(\mathbf{x}_{l,1}), \ell(\mathbf{x}_{l,2}), \dots, \ell(\mathbf{x}_{l,n_l})\} \in \{1, 2, \dots, c\}$ , unlabeled samples  $\mathbf{X}_u = [\mathbf{x}_{u,1}, \mathbf{x}_{u,2}, \dots, \mathbf{x}_{u,n_u}] \in \mathbb{R}^{B \times n_u}$ ,

$\mathbf{X} = [\mathbf{X}_l, \mathbf{X}_u] \in \mathbb{R}^{B \times n}$ , embedding dimension  $d$  ( $d < B$ ), neighbor size  $K_u, K_w, K_b$ , and weight parameter  $\beta > 1$ .

**Output:** Low-dimensional embedding features

$\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n] \in \mathbb{R}^{d \times n}$ .

1: Construct the intraclass hypergraph  $\mathbf{G}_w^H$  with the  $K_w$  nearest neighbors from the same class according to Euclidean distances in the labeled samples.

2: Compute the intraclass hyper-Laplacian matrix of the  $r$ th class:

$$3: \quad \mathbf{L}_{wr}^H = \mathbf{I} - (\mathbf{D}_v^{wr})^{-1/2} \mathbf{H}^{wr} \mathbf{W}_{wr}^H (\mathbf{D}_e^{wr})^{-1} (\mathbf{H}^{wr})^T (\mathbf{D}_v^{wr})^{-1/2}$$

4: Achieve the intraclass hyper-Laplacian matrix:

$$5: \quad \mathbf{L}_w^H = \text{diag}([\mathbf{L}_{wr}^H]_{r=1}^c)$$

6: Construct the unsupervised hypergraph  $\mathbf{G}_u^H$  with the  $K_u$  nearest neighbors according to Euclidean distances in the unlabeled samples.

7: Compute the unsupervised hyper-Laplacian matrix:

$$8: \quad \mathbf{L}_u^H = \mathbf{I} - (\mathbf{D}_v^u)^{-1/2} \mathbf{H}^u \mathbf{W}_u^H (\mathbf{D}_e^u)^{-1} (\mathbf{H}^u)^T (\mathbf{D}_v^u)^{-1/2}$$

9: Construct the interclass graph  $\mathbf{G}_b$  with the  $K_b$  nearest neighbors from different classes according to Euclidean distances in the labeled samples.

$$10: \quad \mathbf{L}_b = \mathbf{I} - (\mathbf{D}_v^b)^{-1/2} \mathbf{W}_b (\mathbf{D}_v^b)^{-1/2}$$

11: Construct the total scatter matrix with all the samples as:

$$12: \quad \mathbf{S}_t = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{X}})(\mathbf{x}_i - \bar{\mathbf{X}})^T$$

13: Compute the local related matrix:

$$14: \quad \mathbf{P}^l = \beta \mathbf{X}_l \mathbf{L}_w^H \mathbf{X}_l^T + \mathbf{X}_u \mathbf{L}_u^H \mathbf{X}_u^T$$

15: Compute the global related matrix:

$$16: \quad \mathbf{P}^g = \mathbf{X}_l \mathbf{L}_b \mathbf{X}_l^T + \mathbf{S}_t$$

17: Solve the generalized eigenvalue problem:

$$18: \quad \mathbf{P}^l \mathbf{v}_i = \lambda_i \mathbf{P}^g \mathbf{v}_i$$

19: Obtain the projection matrix with  $d$  largest eigenvalues corresponding eigenvectors:

$$20: \quad \mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d] \in \mathbb{R}^{B \times d}$$

21: Calculate the low-dimensional embedding features:

$$22: \quad \mathbf{y}_i = \mathbf{V}^T \mathbf{x}_i$$


---



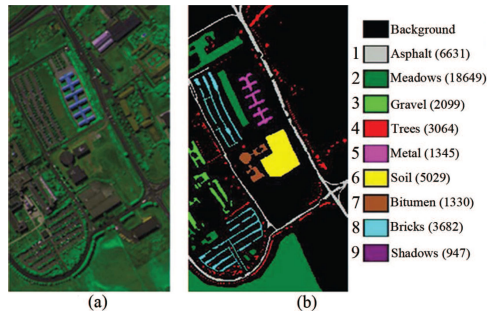


Fig. 5. PaviaU HSI. (a) False color image. (b) Ground truth. (Note that the number of samples for each class is shown in brackets.)

image scene covers the University of Pavia (PaviaU) with a size of  $610 \times 340$  pixels. Since 12 bands are influenced by noise and water absorption, the remaining 103 bands are used for scientific research. This data set has 9 land cover types and 42776 labeled samples. The false color image and its ground truth are shown in Fig. 5.

### B. Experimental Setup

In the experiments, we randomly selected a few samples from a data set as the labeled set and a large number of samples from the remaining samples as the unlabeled set. The training set consists of the labeled and unlabeled sets, and the other samples are composed of the test set. First, we use the training set to train each DR model, where the unsupervised and semisupervised methods use all the labeled and unlabeled samples for training and the supervised methods just use the labeled samples for training. Then, we can obtain the low-dimensional embedding features of all the samples with these DR models. Finally, a classifier is used to discriminate the class type of each test sample; meanwhile, we apply the classification accuracy of each class, the average accuracy of all the classes (AA), the overall classification accuracy (OA), and the kappa coefficient (KC) to quantitatively evaluate the classification results of each method.

As we proposed a semisupervised DR method in this article, some similar DR methods are used for comparison. To demonstrate the efficacy of the proposed method, we selected several semisupervised methods for comparison, i.e., SDA, SSMC, S3MPE, and S3MDA. In addition, we compared two hypergraph learning methods, termed BH and DHLP. After extracting the embedding features, three nonparameter classifiers were used for classification, including the NN classifier, the spectral angle mapping (SAM) classifier, and the sparse representation classifier (SRC) [46]. Meanwhile, we also took the classification results of the raw spectral feature with NN, SAM, and SRC as the baseline (BL) for comparison.

In the experiments, to compare the classification result of each method, we randomly selected 20 labeled samples from each class and 1000 unlabeled samples from the remaining samples to form the training set. All other samples were used as the test set. For the neighbor size  $K_w$ ,  $K_u$ , and  $K_b$ , for convenience, we set  $K_u = K_w$  and  $K_b = \alpha K_w$ . This is because  $K_w$  and  $K_u$  are both used to reveal the similar relationship of the homogeneous samples, whereas  $K_b$  is used to represent the heterogeneous

structure of the interclass samples; generally,  $K_b$  is larger than  $K_w$ . Therefore, for the neighbor size, we just need to set the values of  $K_w$  and  $\alpha$  in the experiments, and  $K_w$  and  $\alpha$  are set to 7 and 5 for the Botswana data set, 9 and 9 for the KSC data set, and 9 and 5 for the PaviaU data set. For the weight parameter  $\beta$ , we set it to 3 for the Botswana data set and 5 for the KSC and PaviaU data sets. For all the methods, the embedding dimension is set to 30. To robustly analyze each algorithm, we repeated each experiment ten times in each condition and obtained the average OAs with standard deviations (STDs) and the average KCs. All the experiments were conducted on a personal computer with E5-2620 v3 central processing unit, 16-GB memory, and 64-bit Windows 10 using MATLAB 2017b.

### C. Performance Comparison

To compare the classification results of each method, we used the training set containing 20 labeled samples from each class and 1000 unlabeled samples to learn the projection model of each method. For the supervised methods, only the labeled samples were used for training. According to the training models, we can obtain the low-dimensional features of each method. Then, we adopted the NN classifier to discriminate the labels of the test set. The classification results on the three data sets are shown in Tables I–III, where the bold font denotes the best result in each row.

According to Tables I–III, the proposed method obtains better results than the other semisupervised methods (i.e., SDA, SSMC, S3MPE, and S3MDA) in most classes and generates the best AA, OA, and KC, which indicates that the proposed method can more effectively reveal the complex intrinsic properties of the HSI and enhance the compactness of the homogeneous samples and the separability of the heterogeneous samples. Compared with two hypergraph learning methods (i.e., BH and DHLP), the proposed SSDHL method achieves better classification results, because SSDHL can simultaneously utilize the labeled and unlabeled samples to train the projection model, whereas DHLP just uses the labeled samples and BH does not apply the label information, which, both, cannot obtain good discriminant features to improve the classification performance of an HSI.

To visualize the classification results, Figs. 6–8 display the classification maps of the Botswana, KSC, and PaviaU data sets. In the figures, SSDHL achieves more homogeneous areas than the other methods on three data sets, especially in the regions of *Water*, *Hippo Grass*, *Floodplain Grasses 1*, *Floodplain Grasses 2*, *Reeds 1*, *Riparian*, *Island Interior*, *Acacia Woodlands*, *Acacia Shrublands*, and *Short Mopane* for the Botswana data set, *Cabbage Palm Hammock*, *Slash Pine*, *Oak/Broadleaf Hammock*, *Graminoid Marsh*, and *Salt Marsh* for the KSC data set, and *Meadows* and *Gravel* for the PaviaU data set. The reason is that SSDHL constructs an intraclass hypergraph and an interclass graph to enhance the discrimination of labeled samples and a total scatter matrix to maximize the variance distribution of the labeled and unlabeled samples. The results indicate that the proposed SSDHL method can learn a good semisupervised model with the labeled and unlabeled samples to reveal the intrinsic

TABLE I  
CLASSIFICATION RESULTS OF EACH CLASS ON THE BOTSWANA DATA SET

Class	BL	BH	DHLP	SDA	SSMMC	S3MPE	S3MDA	SSDHL
1	69.25	71.48	63.47	68.59	69.65	77.79	66.10	<b>78.58</b>
2	83.54	82.30	82.72	81.89	83.54	85.60	88.48	<b>90.12</b>
3	50.00	62.50	54.69	45.31	49.61	71.88	67.58	<b>75.00</b>
4	50.79	59.13	50.40	46.43	51.19	48.81	44.05	<b>69.05</b>
5	57.76	63.98	59.63	52.17	55.28	66.46	69.57	<b>71.43</b>
6	35.37	42.36	41.05	33.19	33.62	43.67	48.91	<b>68.56</b>
7	82.86	83.81	89.52	81.90	83.81	<b>91.43</b>	85.71	85.71
8	58.93	65.20	61.95	56.15	58.47	71.69	72.85	<b>77.49</b>
9	84.42	92.88	87.31	80.96	83.27	95.38	94.42	<b>96.15</b>
10	75.99	89.36	86.88	73.02	75.74	89.36	91.83	<b>95.79</b>
11	92.84	95.23	94.51	93.56	92.84	<b>95.70</b>	94.75	94.99
12	72.17	78.73	75.55	70.97	73.96	79.32	75.75	<b>84.69</b>
13	97.84	98.06	<b>99.68</b>	99.24	97.84	98.06	97.95	98.17
AA	70.14	75.77	72.87	67.95	69.91	78.09	76.76	<b>83.52</b>
OA	74.96	80.02	76.90	73.44	74.88	82.23	80.18	<b>86.36</b>
KC	72.26	77.87	74.41	70.56	72.18	80.29	78.06	<b>84.86</b>

TABLE II  
CLASSIFICATION RESULTS OF EACH CLASS ON THE KSC DATA SET

Class	BL	BH	DHLP	SDA	SSMMC	S3MPE	S3MDA	SSDHL
1	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>
2	98.02	<b>100.00</b>	99.01	99.01	98.02	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>
3	96.41	99.20	94.42	97.61	94.82	96.41	96.02	<b>100.00</b>
4	95.81	99.07	97.67	99.07	95.81	95.81	<b>99.53</b>	98.14
5	75.09	79.93	78.81	80.30	76.21	78.44	81.04	<b>89.22</b>
6	65.43	69.14	66.54	64.68	65.06	74.72	51.67	<b>76.21</b>
7	96.91	96.91	96.91	96.91	96.91	95.75	<b>98.07</b>	<b>98.07</b>
8	94.58	97.04	96.06	97.04	94.09	90.15	99.01	<b>100.00</b>
9	73.89	84.39	77.71	79.30	74.20	83.44	<b>94.59</b>	83.76
10	76.21	85.48	81.45	85.89	75.00	<b>93.95</b>	91.53	93.55
11	86.89	90.49	88.85	88.20	86.23	92.46	89.84	<b>93.11</b>
12	94.48	95.03	95.58	96.13	94.48	95.58	<b>98.34</b>	<b>98.34</b>
13	86.57	91.04	84.70	89.55	85.82	89.55	89.18	<b>94.40</b>
14	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	98.95	<b>100.00</b>	<b>100.00</b>
AA	88.59	91.98	89.84	90.98	88.33	91.80	92.06	<b>94.63</b>
OA	86.88	90.70	88.24	89.47	86.61	90.70	90.76	<b>93.60</b>
KC	85.80	89.93	87.26	88.60	85.50	89.93	90.00	<b>93.06</b>

TABLE III  
CLASSIFICATION RESULTS OF EACH CLASS ON THE PAVIAU DATA SET

Class	BL	BH	DHLP	SDA	SSMMC	S3MPE	S3MDA	SSDHL
1	65.98	<b>68.12</b>	56.04	72.04	66.58	62.87	57.29	64.65
2	61.41	60.56	61.28	62.08	60.94	59.99	70.08	<b>79.75</b>
3	53.26	54.98	65.32	73.46	55.07	62.17	65.41	<b>77.13</b>
4	82.44	83.39	84.69	83.84	84.07	89.33	<b>89.98</b>	88.64
5	99.03	99.26	98.29	99.26	99.03	98.14	<b>99.55</b>	<b>99.55</b>
6	57.29	62.68	59.77	64.70	57.65	56.93	75.28	<b>75.86</b>
7	92.48	94.36	87.07	93.46	93.61	93.83	<b>95.86</b>	91.50
8	68.22	<b>68.33</b>	55.87	71.08	66.62	57.90	65.15	51.77
9	<b>100.00</b>	<b>100.00</b>	97.04	99.89	<b>100.00</b>	99.79	<b>100.00</b>	99.47
AA	75.57	76.85	73.93	79.98	75.95	75.66	79.85	<b>80.93</b>
OA	66.33	67.15	64.46	69.81	66.36	65.24	71.87	<b>76.48</b>
KC	57.97	59.11	55.77	62.33	58.07	57.00	64.85	<b>69.99</b>

multiple properties of an HSI and improve the discriminant power of low-dimensional features.

#### D. Labeled Samples Analysis

To analyze the classification accuracies with respect to different numbers of labeled samples, we selected 5–40 (with

an interval of 5) labeled samples from each class and 1000 unlabeled samples from the remaining samples of each data set, and the other samples were used for testing. After obtaining the embedding features, we adopt the NN classifier to discriminate the labels of the test samples. Fig. 9 shows the classification results with different numbers of labeled samples on the three data sets.



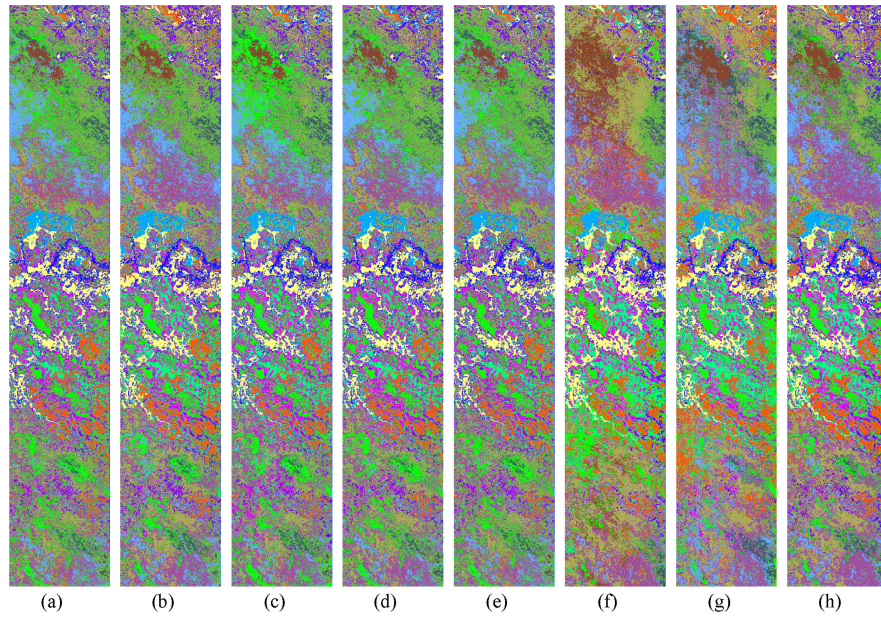


Fig. 6. Classification maps of each method on the Botswana data set. (a) BL. (b) BH. (c) DHLP. (d) SDA. (e) SSMMC. (f) S3MPE. (g) S3MDA. (h) SSDHL.

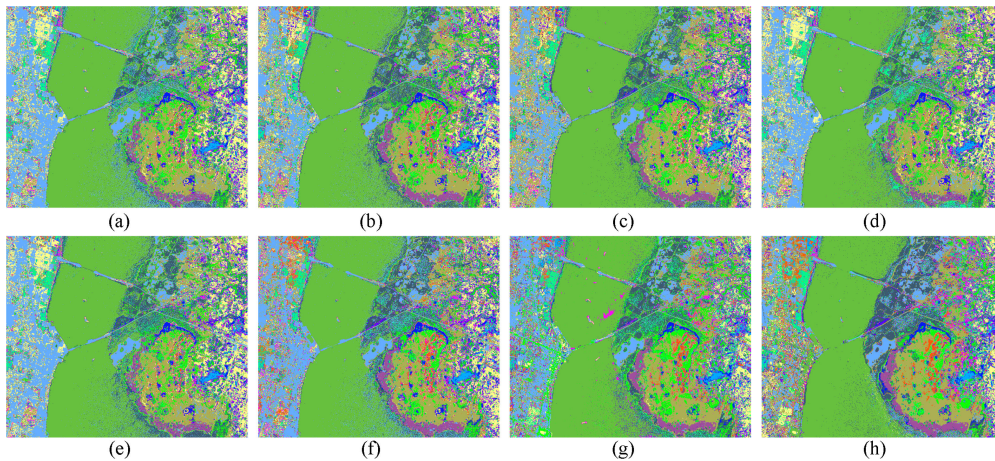


Fig. 7. Classification maps of each method on the KSC data set. (a) BL. (b) BH. (c) DHLP. (d) SDA. (e) SSMMC. (f) S3MPE. (g) S3MDA. (h) SSDHL.

According to Fig. 9, we can obtain the same conclusion about the change in the number of labeled samples on the three data sets. With the increase in the number of labeled samples, the classification results first increase and then reach to a peak value. The reason is that the more priori information can be used to improve the classification accuracies with the increased number of labeled samples. When the number of labeled samples exceeds a certain value, there is adequate information to represent the intrinsic properties of an HSI, which will make the classification accuracy maintain a stable value, hence no further improvement from increasing the number of labeled samples.

#### E. Unlabeled Samples Analysis

To analyze the classification results with different numbers of unlabeled samples, we fixed the number of labeled samples to 20 in each class, and we set the number of unlabeled samples

to  $\{50, 100, 300, 500, 1000, 1500, \text{ and } 2000\}$ . The classification results with the NN classifier with different numbers of unlabeled samples are shown in Fig. 10.

As shown in Fig. 10, with the increasing number of unlabeled samples, the OAs and KCs improve and then reach to a stable value, because a large number of unlabeled samples can reduce the influence of outlier samples and provide more available information. The results indicate that the unlabeled samples are beneficial to promote the extraction of the intrinsic structures of the HSI and enhance the representation power of the embedding features.

#### F. Dimensionality Analysis

For DR, the embedding dimension will influence the classification result. Thus, we conduct further experiments to analyze the OAs with respect to the embedding dimension. The results



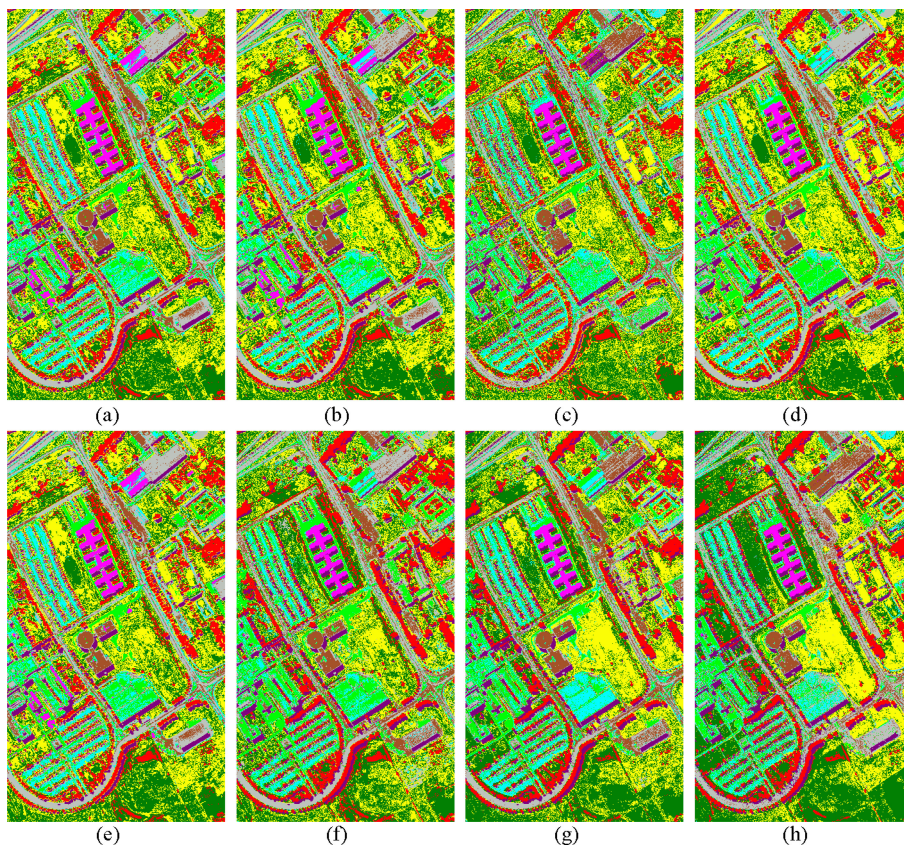


Fig. 8. Classification maps of each method on the PaviaU data sets. (a) BL. (b) BH. (c) DHLP. (d) SDA. (e) SSMMC. (f) S3MPE. (g) S3MDA. (h) SSDHL.

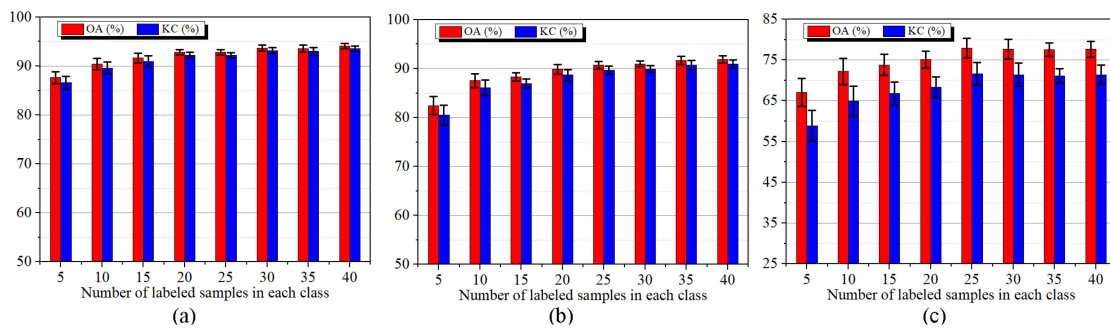


Fig. 9. Classification results with different numbers of labeled samples on (a) Botswana, (b) KSC, and (c) PaviaU data sets.

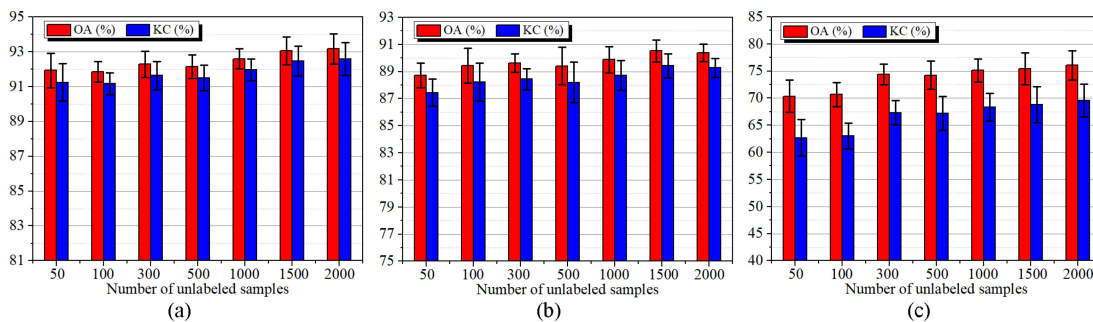


Fig. 10. Classification results with different numbers of unlabeled samples on (a) Botswana, (b) KSC, and (c) PaviaU data sets.



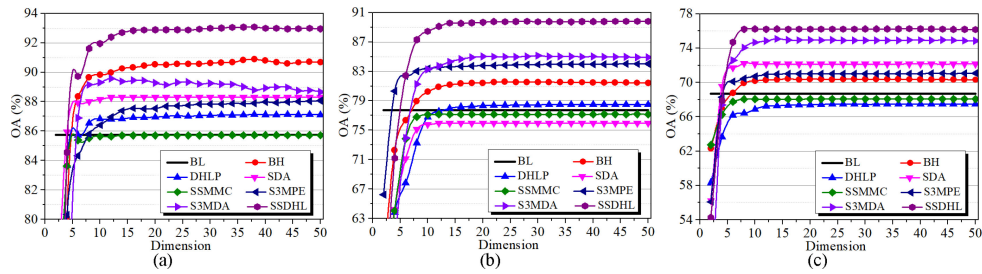


Fig. 11. Classification results with different embedding dimensions on (a) Botswana, (b) KSC, and (c) PaviaU data sets.

 TABLE IV  
 CLASSIFICATION RESULTS WITH DIFFERENT CLASSIFIERS ON THE BOTSWANA DATA SET

Method	NN		SAM		SRC	
	OA $\pm$ STD(%)	KC $\pm$ STD(%)	OA $\pm$ STD(%)	KC $\pm$ STD(%)	OA $\pm$ STD(%)	KC $\pm$ STD(%)
BL	85.76 $\pm$ 0.85	84.56 $\pm$ 0.92	87.46 $\pm$ 0.87	86.40 $\pm$ 0.95	87.46 $\pm$ 0.87	86.40 $\pm$ 0.95
BH	90.62 $\pm$ 1.07	89.83 $\pm$ 1.17	90.75 $\pm$ 1.14	89.97 $\pm$ 1.24	90.75 $\pm$ 1.14	89.97 $\pm$ 1.24
DHLP	87.06 $\pm$ 1.01	85.97 $\pm$ 1.10	87.83 $\pm$ 1.24	86.80 $\pm$ 1.35	87.83 $\pm$ 1.24	86.80 $\pm$ 1.35
SDA	88.30 $\pm$ 0.91	87.32 $\pm$ 0.98	89.35 $\pm$ 0.71	88.46 $\pm$ 0.77	89.35 $\pm$ 0.71	88.46 $\pm$ 0.77
SSMMC	85.72 $\pm$ 0.76	84.51 $\pm$ 0.82	86.46 $\pm$ 1.08	85.32 $\pm$ 1.17	86.46 $\pm$ 1.08	85.33 $\pm$ 1.17
S3MPE	87.84 $\pm$ 1.84	86.81 $\pm$ 2.00	88.16 $\pm$ 1.47	87.16 $\pm$ 1.60	88.17 $\pm$ 1.47	87.17 $\pm$ 1.60
S3MDA	89.25 $\pm$ 1.01	88.35 $\pm$ 1.10	89.24 $\pm$ 0.86	88.34 $\pm$ 0.93	89.24 $\pm$ 0.86	88.34 $\pm$ 0.93
SSDHL	<b>92.98<math>\pm</math>0.82</b>	<b>92.39<math>\pm</math>0.89</b>	<b>93.13<math>\pm</math>0.81</b>	<b>92.54<math>\pm</math>0.88</b>	<b>93.13<math>\pm</math>0.81</b>	<b>92.54<math>\pm</math>0.88</b>

 TABLE V  
 CLASSIFICATION RESULTS WITH DIFFERENT CLASSIFIERS ON THE KSC DATA SET

Method	NN		SAM		SRC	
	OA $\pm$ STD(%)	KC $\pm$ STD(%)	OA $\pm$ STD(%)	KC $\pm$ STD(%)	OA $\pm$ STD(%)	KC $\pm$ STD(%)
BL	77.80 $\pm$ 1.59	75.31 $\pm$ 1.73	78.53 $\pm$ 1.01	76.14 $\pm$ 1.11	78.53 $\pm$ 1.01	76.14 $\pm$ 1.11
BH	81.91 $\pm$ 1.31	79.87 $\pm$ 1.43	80.57 $\pm$ 1.33	78.40 $\pm$ 1.46	81.88 $\pm$ 1.47	79.84 $\pm$ 1.61
DHLP	79.09 $\pm$ 1.05	76.75 $\pm$ 1.15	78.47 $\pm$ 1.02	76.07 $\pm$ 1.11	78.92 $\pm$ 1.04	76.57 $\pm$ 1.14
SDA	74.86 $\pm$ 1.90	72.06 $\pm$ 2.07	75.41 $\pm$ 1.48	72.72 $\pm$ 1.61	74.97 $\pm$ 2.11	72.18 $\pm$ 2.30
SSMMC	76.96 $\pm$ 1.71	74.38 $\pm$ 1.87	76.43 $\pm$ 1.04	73.83 $\pm$ 1.14	77.20 $\pm$ 1.72	74.64 $\pm$ 1.88
S3MPE	84.29 $\pm$ 1.72	82.50 $\pm$ 1.90	81.74 $\pm$ 1.52	79.69 $\pm$ 1.66	84.25 $\pm$ 1.79	82.47 $\pm$ 1.98
S3MDA	85.65 $\pm$ 1.17	84.01 $\pm$ 1.29	86.66 $\pm$ 1.10	85.14 $\pm$ 1.22	85.62 $\pm$ 1.24	83.98 $\pm$ 1.37
SSDHL	<b>89.88<math>\pm</math>0.93</b>	<b>88.72<math>\pm</math>1.03</b>	<b>89.91<math>\pm</math>0.90</b>	<b>88.76<math>\pm</math>0.99</b>	<b>89.73<math>\pm</math>1.05</b>	<b>88.55<math>\pm</math>1.15</b>

with the NN classifier under different embedding dimensions are shown in Fig. 11.

In Fig. 11, the OAs improve with the increase in the embedding dimension and then reach a peak value for each method, because a large embedding dimension contains more information to enhance the discriminant power of low-dimensional features. Most methods generate better OAs than BL on the three data sets, which indicates that the classification results will improve after using the process of DR. This is because these DR methods reduce the redundant information and enhance the discriminating power of features. The proposed method achieves the best classification accuracies than the other methods under different embedding dimensions on the Botswana, KSC, and PaviaU data sets. The reason is that the proposed method can reveal the multiple structures of an HSI and enhance the compactness of the intraclass samples and the separability of the interclass samples. To obtain better classification accuracy, we set the embedding dimension to 30 for all the methods on the three data sets.

### G. Results on Different Classifiers

To demonstrate the effectiveness of the proposed method for different classifiers, we selected NN, SAM, and SRC to discriminate the labels of test samples after extracting the

low-dimensional features. In this experiments, again 20 labeled samples from each class and 1000 unlabeled samples were used to train the DR models and the classification results with the three classifiers are given in Tables IV–VI.

According to Tables IV–VI, SSDHL achieves the best classification accuracies in terms of OA and KC than the other methods on the three data sets, because the proposed method considers the multiple structures to represent the homogenous properties of an HSI, the binary relationships to reveal the heterogeneous information of the HSI, and the total scatter to control the distribution of the labeled and unlabeled samples in the HSI. SSDHL can compact the homogenous samples and separate the heterogeneous samples to improve the discriminant performance of low-dimensional features with the labeled and unlabeled samples.

### H. Parameter Analysis

In the proposed method, it has two neighbor size parameters  $K_w$  and  $\alpha$ , and a label weight parameter  $\beta$ . Thus, we conduct two experiments to analyze the neighbor size and the label weight for the influence of classification accuracy.

1) *Neighbor Size*: For the neighbor size parameters  $K_w$  and  $\alpha$ , we randomly selected 20 labeled samples from each class

TABLE VI  
CLASSIFICATION RESULTS WITH DIFFERENT CLASSIFIERS ON THE PAVIAU DATA SET

Method	NN		SAM		SRC	
	OA $\pm$ STD(%)	KC $\pm$ STD(%)	OA $\pm$ STD(%)	KC $\pm$ STD(%)	OA $\pm$ STD(%)	KC $\pm$ STD(%)
BL	67.60 $\pm$ 2.57	59.27 $\pm$ 2.99	67.28 $\pm$ 2.66	58.78 $\pm$ 2.90	67.28 $\pm$ 2.66	58.78 $\pm$ 2.90
BH	70.08 $\pm$ 1.95	62.22 $\pm$ 2.31	71.52 $\pm$ 2.73	63.82 $\pm$ 3.04	71.52 $\pm$ 2.73	63.82 $\pm$ 3.04
DHLP	66.63 $\pm$ 2.56	58.17 $\pm$ 2.65	67.50 $\pm$ 2.82	58.88 $\pm$ 3.00	67.51 $\pm$ 2.82	58.89 $\pm$ 3.00
SDA	71.29 $\pm$ 2.93	63.76 $\pm$ 3.24	69.16 $\pm$ 3.06	61.18 $\pm$ 3.32	69.16 $\pm$ 3.06	61.18 $\pm$ 3.32
SSMMC	67.13 $\pm$ 2.94	58.75 $\pm$ 3.35	66.52 $\pm$ 2.83	58.03 $\pm$ 2.99	66.53 $\pm$ 2.83	58.03 $\pm$ 2.99
S3MPE	69.87 $\pm$ 3.23	62.17 $\pm$ 3.54	68.01 $\pm$ 4.58	59.78 $\pm$ 5.18	68.02 $\pm$ 4.57	59.78 $\pm$ 5.18
S3MDA	75.34 $\pm$ 2.71	68.62 $\pm$ 3.12	72.88 $\pm$ 1.82	65.30 $\pm$ 2.13	72.88 $\pm$ 1.82	65.30 $\pm$ 2.13
SSDHL	<b>75.99<math>\pm</math>2.99</b>	<b>69.33<math>\pm</math>3.39</b>	<b>75.89<math>\pm</math>3.06</b>	<b>69.21<math>\pm</math>3.47</b>	<b>75.94<math>\pm</math>3.05</b>	<b>69.26<math>\pm</math>3.46</b>

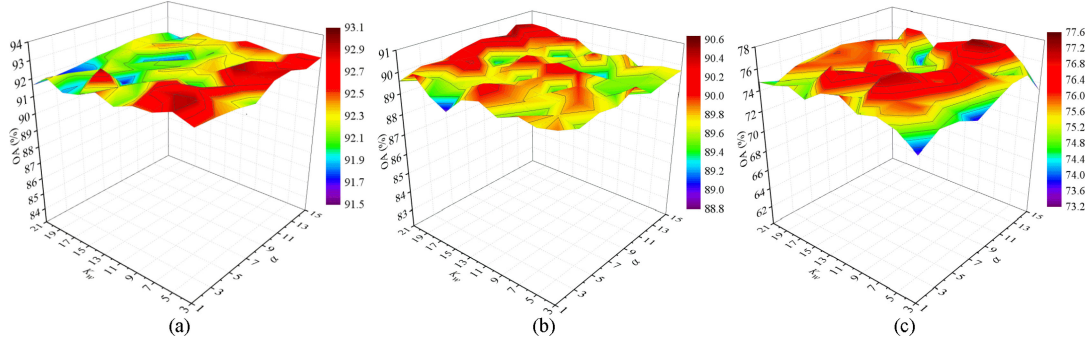


Fig. 12. Classification results with respect to the intraclass neighbor  $K_w$  and the interclass neighbor  $K_b = \alpha K_w$  on (a) Botswana, (b) KSC, and (c) PaviaU data sets.

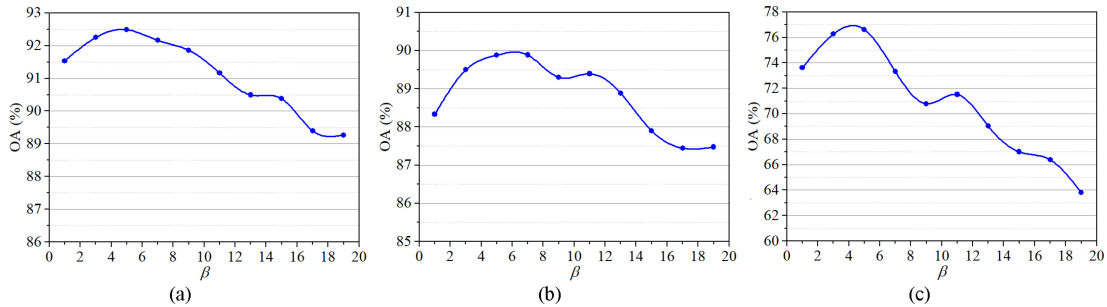


Fig. 13. Classification results with respect to the label weight  $\beta$  on (a) Botswana, (b) KSC, and (c) PaviaU data sets.

and 1000 unlabeled samples to conduct this experiment, and then we used the NN classifier to discriminate the labels of the other samples. Fig. 12 shows the average classification results about the change of  $K_w$  and  $\alpha$  with ten repeated experiments.

According to Fig. 12, the classification accuracy is fluctuating under different  $K_w$  and  $\alpha$  on the three data sets, and the fluctuation is very small in an appropriate range. For the Botswana data set, the results with small  $K_w$  and  $\alpha$  are better than those with large values, where  $K_w$  and  $\alpha$  are set to 7 and 5. For the KSC data set, the parameters  $K_w$  and  $\alpha$  have a small influence in terms of the classification accuracy, and we set them to 9 and 9. For the PaviaU data set, the OAs improve with the increase in  $K_w$  and  $\alpha$  and then decrease when  $K_w$  and  $\alpha$  exceed a value, thus we set  $K_w$  and  $\alpha$  to 9 and 5.

2) *Label Weight*: To enhance the contribution of the labeled samples, we analyzed the classification accuracy with respect to the label weight  $\beta$  and used the training set including 20 labeled samples from each class and 1000 unlabeled samples. Fig. 13

displays the classification results with different  $\beta$  on the three data sets.

In Fig. 13, the OAs first increase and then decrease with the increase in  $\beta$ . The reason is that a large weight will enhance the contribution of the labeled samples to better represent the intrinsic properties of the HSI. When  $\beta$  exceeds a certain value and continues to increase, it will degrade the contribution of the unlabeled samples, which cannot effectively reflect the distribution of the HSI. Therefore, we set  $\beta$  to 5 for the Botswana, KSC, and PaviaU data sets.

## V. CONCLUSION AND DISCUSSION

In this article, we proposed an SSDHL method for DR in an HSI. In SSDHL, we utilize the labeled samples to construct an intraclass hypergraph and an interclass graph; meanwhile, we use the unlabeled samples to design an unsupervised hypergraph. In addition, a total scatter matrix is adopted to control the

distribution of the labeled and unlabeled samples. Then, we construct an objective function to obtain the projection matrix, which can compact the intra-class hypergraph and the unsupervised hypergraph and separate the inter-class graph and the total scatter matrix as much as possible. Finally, we use three data sets (Botswana, KSC, and PaviaU) to demonstrate the effectiveness of the proposed method. The proposed method achieves better classification accuracy compared with a few state-of-the-art algorithms. SSDHL can compact the homogenous information in the intra-class hypergraph and the unsupervised hypergraph to represent the intrinsic properties of the samples from the same class while it can also separate the heterogeneous information in the inter-class graph and the total scatter matrix to enhance the discriminant performance of the samples from different classes.

In the classification maps, there still exist salt and pepper noise in the homogeneous areas for these methods does not consider the spatial relationship of HSIs. However, the spatial distribution is very important in HSI. Therefore, we can apply the spatial information to develop spatial-spectral semisupervised method in the future.

In addition, the proposed method needs to set two important neighbor parameters, which is very difficult to select a proper value; because the neighbor structures may be different for each class land cover in an HSI. If using the stable neighbor sizes to represent different land-cover types, we may not achieve satisfactory classification results for an HSI. Thus, in the future, self-adaption semisupervised learning can be designed based on SSDHL to represent the intrinsic information of an HSI.

#### ACKNOWLEDGMENT

The authors would like to thank the handling editor and the anonymous reviewers for their detailed and constructive comments and suggestions, which indeed helped to improve the quality of this article.

#### REFERENCES

- [1] W. Sun, G. Yang, J. Peng, and Q. Du, "Hyperspectral band selection using weighted kernel regularization," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 9, pp. 3665–3676, Sep. 2019.
- [2] L. Zhang, L. Zhang, B. Du, J. You, and D. Tao, "Hyperspectral image unsupervised classification by robust manifold matrix factorization," *Inf. Sci.*, vol. 485, pp. 154–169, Feb. 2019.
- [3] J. Peng and Q. Du, "Robust joint sparse representation based on maximum correntropy criterion for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7152–7164, Dec. 2017.
- [4] J. Zabalza *et al.*, "Novel two-dimensional singular spectrum analysis for effective feature extraction and data classification in hyperspectral imaging," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4418–4433, Aug. 2015.
- [5] W. Sun, J. Peng, G. Yang, and Q. Du, "Correntropy-based sparse spectral clustering for hyperspectral band selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 3, pp. 484–488, Mar. 2020.
- [6] Z. Lv, T. Liu, P. Zhang, J. Benediktsson, T. Lei, and X. Zhang, "Novel adaptive histogram trend similarity approach for land cover change detection by using bitemporal very-high-resolution remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9554–9574, Dec. 2019.
- [7] X. Zheng, Y. Yuan, and X. Lu, "Dimensionality reduction by spatial-spectral preservation in selected bands," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5185–5197, Sep. 2017.
- [8] G. Jin *et al.*, "An advanced phase synchronization scheme for LT-1," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 1735–1746, Mar. 2020.
- [9] W. Sun and Q. Du, "Hyperspectral band selection: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 118–139, Jun. 2019.
- [10] F. Luo, L. Zhang, B. Du, and L. Zhang, "Dimensionality reduction with enhanced hybrid-graph discriminant learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5336–5353, Aug. 2020.
- [11] J. Zabalza *et al.*, "Novel folded-PCA for improved feature extraction and data reduction with hyperspectral imaging and SAR in remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 93, pp. 112–122, 2014.
- [12] W. Sun, G. Yang, J. Peng, and Q. Du, "Lateral-slice sparse tensor robust principal component analysis for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 1, pp. 107–111, Jan. 2020.
- [13] H. Wang, X. Lu, Z. Hu, and W. Zheng, "Fisher discriminant analysis with L1-norm," *IEEE Trans. Cybern.*, vol. 44, no. 6, pp. 828–842, Jun. 2014.
- [14] L. Guan, W. Xie, and J. Pei, "Segmented minimum noise fraction transformation for efficient feature extraction of hyperspectral images," *Pattern Recognit.*, vol. 48, no. 10, pp. 3216–3226, Oct. 2015.
- [15] 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.
- [16] W. Wu, Y. Zhang, Q. Wang, F. Liu, P. Chen, and H. Yu, "Low-dose spectral CT reconstruction using image gradient  $\ell_0$ -norm and tensor dictionary," *Appl. Math. Model.*, vol. 63, pp. 538–557, Jul. 2018.
- [17] J. Tenenbaum, V. de Silva, and J. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.
- [18] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.
- [19] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, Jun. 2003.
- [20] X. He, D. Cai, S. Yan, and H. Zhang, "Neighborhood preserving embedding," in *Proc. IEEE Int. Conf. Comput. Vision*, Dec. 2005, pp. 1208–1213.
- [21] R. Wang, F. Nie, R. Hong, X. Chang, X. Yang, and W. Yu, "Fast and orthogonal locality preserving projections for dimensionality reduction," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 5019–5030, Oct. 2017.
- [22] S. Yan, D. Xu, B. Zhang, H. 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.
- [23] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis," *J. Mach. Learn. Res.*, vol. 8, pp. 1027–1061, May 2007.
- [24] Y. Zhou, J. Peng, and C. L. 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, Sep. 2015.
- [25] F. Luo, H. Hong, Y. Duan, J. Liu, and Y. Liao, "Local geometric structure feature for dimensionality reduction of hyperspectral imagery," *Remote Sens.*, vol. 9, no. 8, Aug. 2017, Art. no. 790.
- [26] Q. Shi, L. Zhang, and B. Du, "Semisupervised discriminative locally enhanced alignment for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4800–4815, Sep. 2013.
- [27] K. Tan, J. Hu, J. Li, and P. Du, "A novel semi-supervised hyperspectral image classification approach based on spatial neighborhood information and classifier combination," *ISPRS J. Photogramm. Remote Sens.*, vol. 105, pp. 19–29, Jul. 2015.
- [28] S. Yang, P. Jin, B. Li, L. Yang, W. Xu, and L. Jiao, "Semisupervised dual-geometric subspace projection for dimensionality reduction of hyperspectral image data," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3587–3593, Jun. 2014.
- [29] Y. Song, F. Nie, C. Zhang, and S. Xiang, "A unified framework for semi-supervised dimensionality reduction," *Pattern Recognit.*, vol. 41, no. 9, pp. 2789–2799, Sep. 2008.
- [30] R. Luo, W. Liao, X. Huang, Y. Pi, and W. Philips, "Feature extraction of hyperspectral images with semisupervised graph learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4389–4399, Sep. 2016.
- [31] D. Cai, X. He, and J. Han, "Semi-supervised discriminant analysis," in *Proc. IEEE Int. Conf. Comput. Vision*, Dec. 2007, pp. 1–7.
- [32] W. Liao, A. Pižurica, P. Scheunders, W. Philips, and Y. Pi, "Semisupervised local discriminant analysis for feature extraction in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 184–198, Jan. 2013.
- [33] Y. Song, F. Nie, and C. Zhang, "Semi-supervised sub-manifold discriminant analysis," *Pattern Recognit. Lett.*, vol. 29, no. 13, pp. 1806–1813, Oct. 2008.



- [34] F. Luo, H. Huang, Z. Ma, and J. Liu, "Semisupervised sparse manifold discriminative analysis for feature extraction of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6197–6211, Oct. 2016.
- [35] D. Zhou, J. Huang, and B. Schölkopf, "Learning with hypergraphs: Clustering, classification, and embedding," in *Proc. Neural Inf. Process. Syst.*, Dec. 2006, vol. 19, pp. 1601–1608.
- [36] J. Yu, D. Tao, and M. Wang, "Adaptive hypergraph learning and its application in image classification," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3262–3272, Jun. 2012.
- [37] F. Luo, L. Zhang, X. Zhou, T. Guo, Y. Cheng, and T. Yin, "Sparse-adaptive hypergraph discriminant analysis for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 6, pp. 1082–1086, Jun. 2020.
- [38] S. Gao, I. W. Tsang, and L. Chia, "Laplacian sparse coding, hypergraph Laplacian sparse coding, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 92–104, Jun. 2013.
- [39] Z. Zhang, L. Bai, Y. Liang, and E. Hancock, "Joint hypergraph learning and sparse regression for feature selection," *Pattern Recognit.*, vol. 63, pp. 291–309, Mar. 2017.
- [40] H. Yuan and Y. Y. Tang, "Learning with hypergraph for hyperspectral image feature extraction," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 8, pp. 1695–1699, Aug. 2015.
- [41] S. Huang, D. Yang, Y. Ge, and X. Zhang, "Discriminant hyper-Laplacian projections and its scalable extension for dimensionality reduction," *Neurocomputing*, vol. 173, pp. 145–153, Jan. 2016.
- [42] Y. Sun, S. Wang, Q. Liu, R. Hang, and G. Liu, "Hypergraph embedding for spatial-spectral joint feature extraction in hyperspectral images," *Remote Sens.*, vol. 9, no. 5, May 2017, Art. no. 506.
- [43] F. Luo, B. Du, L. Zhang, L. Zhang, and D. Tao, "Feature learning using spatial-spectral hypergraph discriminant analysis for hyperspectral image," *IEEE Trans. Cybern.*, vol. 49, no. 7, pp. 2406–2419, Jun. 2019.
- [44] M. Xu, D. Hu, F. Luo, F. Liu, S. Wang, and W. Wu, "Limited angle X ray CT reconstruction using image gradient  $\ell_0$  norm with dictionary learning," *IEEE Trans. Radiat. Plasma Med. Sci.*, to be published, doi: [10.1109/TRPMS.2020.2991887](https://doi.org/10.1109/TRPMS.2020.2991887).
- [45] W. Wu, P. Chen, S. Wang, V. Vardhanabuthi, F. Liu, and H. Yu, "Image-domain material decomposition for spectral CT using a generalized dictionary learning," *IEEE Trans. Radiat. Plasma Med. Sci.*, to be published, doi: [10.1109/TRPMS.2020.2997880](https://doi.org/10.1109/TRPMS.2020.2997880).
- [46] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.



**Fulin Luo** (Member, IEEE) received the B.S. degree in mechanical engineering and automation from Southwest Petroleum University, Chengdu, China, in 2011, and the M.S. and Ph.D. degrees in instrument science and technology from Chongqing University, Chongqing, China, in 2013 and 2016, respectively.

He is currently an Associate Researcher with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan, China. He was a Postdoctoral Researcher in LIESMARS from 2017 to 2019. His research interests include hyperspectral image classification, image processing, sparse representation, and manifold learning in general.



**Tan Guo** received the B.S. degree from the Henan University of Science and Technology, Luoyang, China, in 2011, and the M.S. and Ph.D. degrees from Chongqing University, Chongqing, China, in 2014 and 2017, respectively.

He is currently a Lecturer with the School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing, China. His current research interests include biometrics, pattern recognition, and machine learning.



**Zhiping Lin** (Senior Member, IEEE) received the B.Eng. degree in control engineering from the South China Institute of Technology, Canton, China, in 1982, and the Ph.D. degree in information engineering from the University of Cambridge, Cambridge, U.K., in 1987.

Since 1999, he has been with Nanyang Technological University, Singapore. His current research interests include multidimensional systems and signal processing, statistical and biomedical signal processing, and machine learning.

Dr. Lin was the Editor-in-Chief for *Multidimensional Systems and Signal Processing* from 2011 to 2015. He was an Associate Editor for the *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II* from 2010 to 2011. He was a Reviewer of *Mathematical Reviews* from 2011 to 2013. He is currently a Subject Editor for the *Journal of the Franklin Institute*. He was the recipient (as a coauthor) of the 2007 Young Author best paper award from the IEEE Signal Processing Society, the Distinguished Lecturer of the IEEE Circuits and Systems Society from 2007 to 2008, and received the best paper awards from International Conference of Extreme Learning Machine (ELM2015 and ELM2017).



**Jinchang Ren** (Senior Member, IEEE) received the Ph.D. degree in electronic imaging and media communication from Bradford University, Bradford, U.K., in 2009.

He is currently a Senior Lecturer with the Centre for Excellence for Signal and Image Processing and the Deputy Director of the Strathclyde Hyperspectral Imaging Centre, University of Strathclyde, Glasgow, U.K. His current research interests include visual computing and multimedia signal processing, especially on semantic content extraction for video

analysis and understanding and more recently hyperspectral imaging.

Xiaocheng Zhou, photograph and biography not available at the time of publication.