Extended Subspace Projection Upon Sample Augmentation Based on Global Spatial and Local Spectral Similarity for Hyperspectral Imagery Classification

Jiaochan Hu, Xueji Shen, Haoyang Yu, *Member, IEEE*, Xiaodi Shang, *Student Member, IEEE*, Qiandong Guo, and Bing Zhang, *Fellow, IEEE*

***Abstract*—Band redundancy and limitation of labeled samples restrict the development of hyperspectral image classification (HSIC) greatly. To address the earlier issues, the classification models such as subspace-based support vector machines, which have gained a certain advance but mainly concentrate on the dimensionality reduction and ignore the augmentation of training samples. In fact, these two issues are equally important for improving the performance of classification, and should be addressed simultaneously. Therefore, this article proposes a novel method named extended subspace projection upon sample augmentation based on global spatial and local spectral similarity (GLSC) for HSIC, which takes both sample augmentation and dimensionality reduction into consideration. Specifically, it first exploits the GLSC to enlarge the original labeled sample set, which allows HSIC to obtain more prior information. Then, the augmented samples and the original labeled samples are combined to construct the extended subspace, which is more comprehensive to reflect the real situation of the ground objects. Finally, the original HSI is projected to the subspace and classified by the neighborhood activity degree-driven representation-based classifier. Experimental results on three real hyperspectral datasets demonstrate the practicality and effectiveness of the proposed method for HSIC tasks.**

***Index Terms*—Dimensionality reduction, hyperspectral remote sensing, labeled sample size, supervised classification.**

Manuscript received June 4, 2021; revised July 16, 2021; accepted August 12, 2021. Date of publication August 24, 2021; date of current version September 9, 2021. This work was supported in part by the National Science Foundation of China under Grant 42101350 and Grant 42001280, in part by the Chinese Postdoctoral Science Foundation under Grant 2021M690497 and Grant 2020M680925, and in part by the Fundamental Research Funds for Central Universities under Grant 3132021169. *(Corresponding author: Haoyang Yu.)*

Jiaochan Hu is with the College of Environmental Sciences and Engineering, Dalian Maritime University, Dalian 116026, China (e-mail: hujc@dlmu.edu.cn).

Xueji Shen, Haoyang Yu, and Xiaodi Shang are with the Center of Hyperspectral Imaging in Remote Sensing, Information Science and Technology College, Dalian Maritime University, Dalian 116026, China (e-mail: k17516201@163.com; yuhy@dlmu.edu.cn; shangxd329@dlmu.edu.cn).

Qiandong Guo is with the School of Geosciences, University of South Florida, Tampa, FL 33620 USA (e-mail: guo1@mail.usf.edu).

Bing Zhang is with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China, and also with the College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: zb@radi.ac.cn).

Digital Object Identifier 10.1109/JSTARS.2021.3107105

I. Introduction

REMOTE sensing is a comprehensive technology used for earth observation. Hyperspectral remote sensing is such a technique characterized by imaging land surface at numerous spectral bands [1]–[3]. With the high spectral resolution, hyperspectral remote sensing image (HSI) enables a complete spectral diagnosis of ground objects; thus, allowing a fine classification of land cover and land use classes [4], [5]. Some deep learning methods are also used for HSI classification (HSIC) in recent years [6], [7]. However, its accuracy of classification was usually limited by the spectral redundancy of HSI and the size of training samples.

Although the increase of spectral resolution in the HSI can capture more details, high correlations between adjacent bands induce high redundancy, meaning that the hyperspectral data can be condensed into a subspace with lower dimensions [8]. Two common approaches have been used to reduce the dimension of HSI, band selection, and feature extraction. Band selection aims to use the most discriminative bands to construct a simplified feature space [9], [10]. Maximum variance principal component analysis (PCA) [11], [12] and constrained band selection [13], [14] are two classic methods that have been proved effective in band selection. Feature extraction is another way of dimensionality reduction. It focuses on creating a new feature space which synthesizes information or enhances desired features through a mathematical transformation of original spectral space [15]. Some classical methods include PCA [16]–[18] and maximum noise fraction [19], [20], both of that aim to optimize the spectral features. Subspace projection algorithm has also been employed for spectral feature extraction [21], [22]. Its basic assumption is that high-dimensional samples of each class can be expressed by a set of vectors in a lower dimensional subspace. However, due to the existence of mixed pixels and intraclass spectral variation [23], the effectiveness of this algorithm is easily affected by the selection of training samples.

Another limitation is the vulnerability of supervised classification to training samples. With insufficient training samples, its accuracy could be improved by increasing spectral bands but to a limited degree as a result of the Hughes phenomenon. So, the abundance of training samples is required to achieve

a high accuracy in supervised classification [24]. However, it is challenging to acquire plentiful labeled samples in practice without consuming a lot of time and labor. Sometimes the limitation of objective factors makes it impossible to obtain enough training samples at all. In [25], based on the correlation of spectral bands, the original labeled samples were divided into several subsamples to alleviate this issue for HSIC. In [26], Cui *et al.* used the image segmentation algorithm to randomly select unlabeled samples in the same region where the original labeled samples are located to enlarge the sample set. These methods have a certain effect on training sample enlargement, but there is still a shortage of using global spatial information.

To improve the accuracy of HSIC, current methods mainly focused on either reducing spectral dimension or increasing the size of training sample. For instance, subspace-based support vector machine (SVM) [27]–[29] could effectively reduce spectral dimension but it does not consider automatically expanding the sample size used for classification. As dimensionality reduction and sample expansion are both critical in determining the accuracy of classification, balancing them could be an alternative way to improve the accuracy of supervised classification. However, seldom classification algorithms incorporated the two aspects altogether into their frameworks.

In this article, we proposed a new method named extended subspace projection upon sample augmentation based on global spatial and local spectral similarity (ESSA-GLSC) for HSIC, which takes both dimensionality reduction and sample augmentation into consideration. First, we applied a novel algorithm that exploited both global spatial and local spectral information to increase the number of training samples. Specifically, we take advantage of the spatial information of the region where the labeled pixel is located to find another non-neighborhood area that has similar structures. The central pixel of the found area will be marked as the same label. The basic assumption is that there may be some regions in the image with different distribution locations but have similar spatial characteristics, such as texture and material composition. By this way, the labeled sample size can be extended twice as much as before. Then, we make use of the spectral information to find a similar pixel in the neighborhood, the labeled sample is located. The found pixel is also marked as the corresponding class label to enlarge the labeled sample size. This method was inspired by the pixels adjacent to each other in space have a high probability of belonging to the same category. Since then, the training sample size has been tripled and can be effectively used to train the supervised classifier. Second, the extended subspace projection is introduced to reduce spectral dimension. Compared with the subspace projection, the proposed method exploits the augmented training samples and the original labeled samples together to construct the subspace, which is more comprehensive to reflect the real situation of the ground truth. Ultimately, we combine the above-mentioned implements with the neighborhood activity degree (NAD) driven representation-based classifier (NADRC) for HSIC. The NADRC is an improved version to the classic sparse representation-based classifier (SRC) [30]–[38], which has lower computational cost and takes spatial coherence into consideration. The main contributions of this article can be summarized as follows.

1) The proposed method takes both sample augmentation and dimensionality reduction into consideration, which further alleviates the effect of the Hughes phenomenon and enhances the performance of HSIC. The proposed method is more stable for the situation with limited training samples.
2) A concise and low-cost method that effectively expands the number of training samples through exploiting GLSC [39]–[46]. Then, the enlarged labeled samples and the original samples are used to construct the extended subspace, which is a more stable feature space and unties the concentration of spectral features caused by mixed pixels.
3) The projected image is classified by the spectral-spatial classifier NADRC, which was based on our previous work and gets a great performance in HSIC tasks.

The rest of this article is organized as follows. Section II introduces related subspace projection algorithm and sparse representation (SR) based models. Section III presents the proposed ESSA-GLSC in detail. Section IV evaluates the performance of the proposed method with three real hyperspectral datasets and compared with the results of other classic HSIC methods. Section V concludes this article with some remarks.

II. Related Model Description

A. SR-Based Classification

The original hyperspectral image is denoted as $\mathbf{X}$ with totally N pixels, B spectral bands and containing K thematic classes. SRC randomly selects the training samples from each class to construct the dictionary $\mathbf{D}$. Then, $\mathbf{D}$ is used to represent the testing pixel $\mathbf{x}_{i.j}$, as $\mathbf{x}_{i,j} \approx \mathbf{D}\alpha$, where α is a weighted vector. Theoretically, SRC adopts the l_0-norm to measure the sparsity of α, which calculates the number of nonzero values in α. However, because the optimization of the l_0-norm is an NP-hard problem, the l_1-norm minimization as the closest convex function is adopted in practice [47]. Therefore, SRC can be formulated as a Lagrange formulation as follows:

$$\begin{aligned} &\hat{\alpha}_{i,j} = \arg\min_{\alpha_{i,j}} \left\{ \|\mathbf{x}_{i,j} - \mathbf{D}\alpha_{i,j}\|_2^2 + \lambda \|\alpha_{i,j}\|_1 \right\} \\ &\text{s.t.} \parallel \mathbf{x}_{i,j} - \mathbf{D}\alpha_{i,j}\|_1 < \varepsilon \end{aligned} \tag{1}$$

where $\|\alpha_{i,j}\|_1 = \sum_{m=1}^{n} |\alpha_m|$ denotes the l_1-norm constraint, and λ is a scalar regularization parameter. The constant ε is used to balance the error of representation. Then, the class label of $\mathbf{x}_{i,j}$ is determined by evaluating the residual errors between the obtained approximation and the $\mathbf{x}_{i,j}$.

B. Subspace Projection

The linear mixture model assumption has been used to the subspace projection algorithm to reduce the dimension of HSI. For any pixel $\mathbf{x}_{i,j} \in \mathbf{X}$, it can be represented as follows:

$$\mathbf{x}_{i,j} = \sum_{k=1}^{K} \mathbf{U}^{(k)} \mathbf{z}_{i,j}{}^{(k)} + \mathbf{n}_{i,j} \tag{2}$$

where $\mathbf{U}^{(k)} = \{\mathbf{u}_1^{(k)}, \ldots, \mathbf{u}_{r^{(k)}}(k)\}$ is a set of $r^{(k)}$-dimensional orthonormal basis vectors for the subspace, associated with class $k \in [1, K]$. $n_{i,j}$ represents noise, and $\mathbf{z}_{i,j}^{(k)}$ denotes the

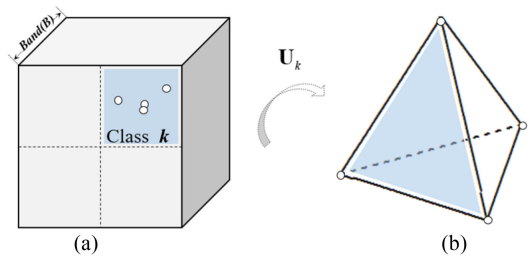

Fig. 1. Subspace spanned by the original labeled samples (○).

coordinates of $\mathbf{x}_{i,j}$ with respect to the basis $\mathbf{U}^{(k)}$. In order to get the appropriate $\mathbf{U}^{(k)}$, let $\mathbf{R}^{(k)}=E\{x_{l^{(k)}}^{(k)} x_{l^{(k)}}^{(k)^T}\}$ denotes the self-correlation matrix associated with class k, and $x_{l^{(k)}}(k)$ is the training set associated with class k containing $l^{(k)}$ samples. According to $\mathbf{R}^{(k)}=\mathbf{E}^{(k)}\mathbf{\Lambda}^{(k)}\mathbf{E}^{(k)^T}$, the eigenvalue matrix $\mathbf{\Lambda}=diag(\lambda_1^{(k)},\ldots,\lambda_B^{(k)})$ and eigenvector matrix $\mathbf{E}^{(k)}=\{\mathbf{e}_1^{(k)},\ldots\mathbf{e}_d^{(k)}\}$ can be calculated. The retention ratio $r^{(k)}$ of the original spectral information is controlled by the parameter τ

$$r^{(k)}=\min\left\{r^{(k)}:\sum_{i=1}^{r^{(k)}}\lambda_i^{(k)}\geq\sum_{i=1}^{d}\lambda_i^{(k)}\times\tau\right\}. \tag{3}$$

Then, $\mathbf{U}^{(k)}=\{\mathbf{e}_1^{(k)},\ldots\mathbf{e}_{r^{(k)}}^{(k)}\}$ can be obtained, where $r^{(k)}<B$. Therefore, the pixel $\mathbf{x}_{i,j}$ can be transformed as follows:

$$\phi(\mathbf{x}_{i,j})=\left\{\|\mathbf{x}_{i,j}\|^2,\left\|\mathbf{x}_{i,j}^T\mathbf{U}^{(1)}\right\|^2,\ldots,\left\|\mathbf{x}_{i,j}^T\mathbf{U}^{(K)}\right\|^2\right\}^T. \tag{4}$$

The dimension of $\phi(\mathbf{x}_{i,j})$ is $(K+1)$, which is unrelated with the number of the original bands and training sample size. Thus, the effect of the Hughes phenomenon was largely avoided. On this basis, the original image $\mathbf{X}$ can be projected as follows:

$$\phi(\mathbf{X})=[\phi(\mathbf{x}_1),\ldots,\phi(\mathbf{x}_N)]. \tag{5}$$

For example, the original labeled samples shown in Fig. 1(a) can be used to construct the $\mathbf{U}^{(k)}$, and the original HSI can be projected to obtain the lower dimensional image $\phi(\mathbf{X})$, which is shown in Fig. 1(b).

III. Proposed Methods

To alleviate the problem of insufficient labeled samples and band redundancy, this section first introduces a sample augmentation algorithm based on GLSC. Then, the ampliative labeled sample set is further applied to extended subspace projection which randomly selects some clusters from each class to construct a lower dimensional subspace. Finally, a new classifier named ESSA-GLSC is proposed, which applies the spectral-spatial classifier NADRC [48] to the projected image for the better classification performance.

A. GLSC-Based Sample Augmentation

1) Global Spatial Similarity-Based Sample Augmentation: Nonlocal self-similarity (NLSS) was first proposed in the field of image denoising [49]–[52]. NLSS was motivated by a fact that natural images contain many similar patches in different areas. Inspired by NLSS, the proposed sample augmentation method makes use of the texture information of the region where the labeled pixel is located to find another non-neighborhood area with similar structure. The central pixel of the area will be marked as the same class and be put into the dictionary to represent testing samples.

Let $\mathbf{x}_{i,j}=[x_{i,j}^1,\ldots,x_{i,j}^B]$ at the location (i,j) of $\mathbf{X}$ be a labeled pixel belonging to the original training sample set S_1. In order to search its global spatial similarity information in the image, $\mathbf{X}$ is first expanded to a $(r+w_1-1)\times(c+w_1-1)$ pixel-sized image, where r and c are the size of original rows and columns. Then, we denote a $\mathbf{x}_{i,j}$-centered $w_1\times w_1$ pixel-sized patch as $\mathbf{P}_{i,j}$ and extract the set of $w_1\times w_1$ pixel-sized patches centered on each pixel in $\mathbf{X}$, denoted as $\mathbf{P}$. To search the global spatial similarity of $\mathbf{x}_{i,j}$ is actually to find a patch $\hat{\mathbf{P}}_{a,b}$ in $\mathbf{P}$ with the highest similarity to $\mathbf{P}_{i,j}$. The similarity here is measured by the sum of Euclidean distances between pixels at corresponding positions in the block except for central pixel. The process of search can be expressed as follows:

$$\min_{a,b} dist(\mathbf{P}_{i,j},\hat{\mathbf{P}}_{a,b})=\min_{a,b}\left\{\sum_{m=1}^{w^2}d(\mathbf{x}_m,\hat{\mathbf{x}}_m)\right\} \quad s.t.\ \ m\neq\tfrac{1}{2}w^2 \tag{6}$$
$$\mathbf{x}\in\mathbf{P}_{i,j},\hat{\mathbf{x}}\in\hat{\mathbf{P}}_{a,b}$$

where $\mathbf{x}_m$ is the m_{th} pixel in $\mathbf{P}_{i,j}$ and $\hat{\mathbf{x}}_m$ is the m_{th} pixel in $\hat{\mathbf{P}}_{a,b}$, and $d(\mathbf{x}_m,\hat{\mathbf{x}}_m)$ represents the Euclidean distance between $\mathbf{x}_m$ and $\hat{\mathbf{x}}_m$. It should be noted that the search process should follow the principal that there is nonoverlapping between $\hat{\mathbf{P}}_{a,b}$ and $\mathbf{P}_{i,j}$.The purpose of the decentralized search is to reduce the influence of spectral variability and other objective factors, such as weather or imaging conditions. Considering the computational cost, the calculation is suggested to perform on the first p components of singular value decomposition (SVD), which maintains the relative distance of the original space under the orthogonal projection transformation. In the experiments, the original hyperspectral image has been processed by SVD to reduce the computational complexity and maintain the accuracy of the measurement of similarity. The execution steps of this method are summarized in Algorithm 1-1.

Through the above-mentioned implementation, the size of the labeled sample set S_2 has been extended twice as much as before in the nonlocal spatial level.

2) Local Spectral Similarity-Based Sample Augmentation: The spectral information of adjacent pixels can also be used to find another pixel that belongs to the same class to enlarge the labeled sample size. Based on the basic assumption that the pixels adjacent to each other in space have a high probability of belonging to the same category. Therefore, the proposed algorithm, named local spectral similarity-based sample augmentation, aims to search the pixel $\mathbf{x}_{u,v}$ in $\mathbf{P}_{i,j}$, which is most like $\mathbf{x}_{i,j}$. The search process can be expressed as follows:

$$\min_{\mathbf{x}_{u,v}} dist(\mathbf{x}_{i,j},\mathbf{x}_{u,v})=\sum_{b=1}^{B}d(x_{i,j}^b,x_{u,v}^b) \tag{7}$$
$$\mathbf{x}_{u,v}\in\mathbf{P}_{i,j}$$

Algorithm 1-1: SA-GS.

Input: Original HSI $\mathbf{X}$, original labeled sample set S_1, the size of neighborhood w_1.

1. Pre-processing: obtain the dimension-reduced image $\mathbf{X'}_g$ by SVD.
2. expand $\mathbf{X'}_g$ to$\mathbf{X''}_g$ with size of$(r+w_1-1)\times(c+w_1-1)$
3. for $i=1$ to r, $j=1$ to c do patch $\leftarrow$ cube /* extract the set of patches $\mathbf{P}$ */
4. for $i=1$ to r, $j=1$ to c do $\mathbf{x}_{a,b}=f(\mathbf{x}_{i,j})$ /*search the most similar pixel $\mathbf{x}_{a,b}$ to $\mathbf{x}_{i,j}$ in S_1 via (6) */
5. $S_2=\{S_1,\mathbf{x}_{a,b}\}$ /* augment the original training sample set S_1 to S_2 */

Output: S_2.

Algorithm 1-2: SA-LS.

Input: Original HSI $\mathbf{X}$, the size of neighborhood w_2, the augmented labeled sample set S_2.

1. Pre-processing: obtain the dimension-reduced image $\mathbf{X'}_l$ by SVD.
2. expand $\mathbf{X'}_l$ to$\mathbf{X''}_l$ with size of$(r+w_2-1)\times(c+w_2-1)$
3. for $i=1$ to r, $j=1$ to c do patch $\leftarrow$ cube /* extract the set of patches $\mathbf{P}$ */
4. for $i=1$ to r, $j=1$ to c do $\mathbf{x}_{u,v}=f(\mathbf{x}_{i,j})$ /* search the most similar pixel $\mathbf{x}_{u,v}$ to $\mathbf{x}_{i,j}$in S_1 via (7) */
5. $S_3=\{S_2,\mathbf{x}_{u,v}\}$ /* extend S_2 to S_3 */

Output: S_3.

where $x^b_{i,j}$ and $x^b_{u,v}$ represent the pixel $\mathbf{x}_{i,j}$ and $\mathbf{x}_{u,v}$ at the band $b\in[1,B]$, respectively.$d(x^b_{i,j},x^b_{u,v})$ represents the Euclidean distance between $x^b_{i,j}$ and $x^b_{u,v}$. The pixel $\mathbf{x}_{a,b}$ and $\mathbf{x}_{u,v}$ maybe the same one in the experiments. In this case, only one of them is put into the labeled sample set. The pseudocode is summarized in Algorithm 1-2.

Now, the labeled sample set S_3 has been effectively tripled in the local spectral level, and can be further used as input to serve the subsequent subspace expansion.

B. Extended Subspace Projection

Based on the subspace projection algorithm, the extended subspace projection constructs the set of self-correlation matrixes in each class by randomly selecting several clusters from the appropriately labeled samples. With the selected clusters, the subspace can be spanned by more basis vectors, which is closer to the real situation of the ground truth. As defined before, let $\mathbf{R}^{(k,m)}=E\{x^{(k,m)}_{l^{(k)}}x^{(k,m)^T}_{l^{(k)}}\}$ represents the self-correlation matrix calculated by the labeled samples, which are randomly selected at m_{th} time, and $x^{(k,m)}_{l^{(k)}}$ denotes $l^{(k)}$ labeled samples chosen at m_{th} time from the class k. According to $\mathbf{R}^{(k,m)}=\mathbf{E}^{(k,m)}\mathbf{\Lambda}^{(k,m)}\mathbf{E}(k,m)^T$, $\mathbf{U}^{(k,m)}=\{e^{(k,m)}_1,\dots e^{(k,m)}_{r^{(k)}}\}$ can be obtained. Then, a more comprehensive basis for constructing the subspace of class k can be determined, denote as $\mathbf{U}^{(k)}=\{\mathbf{U}^{(k,m)}\}^M_{m=1}$, where M is the number of clusters. Therefore, the projected pixel is represented as follows (8) shown at the bottom of this page: where $\|\bullet\|_2$ denotes the l_2-norm. Thus, the projected image can be represented as follows:

$$\phi(\mathbf{X})=[\phi(\mathbf{x}_1),\dots,\phi(\mathbf{x}_N)]. \tag{9}$$

The $\phi(\mathbf{X})$ is an r-dimensional image where $r=K\times M$, independently of the size of labeled samples. As shown in Fig. 2, the extended subspace projection constructs a more comprehensive subspace to regard the distribution of the ground objects by the augmented labeled samples than the original labeled samples.

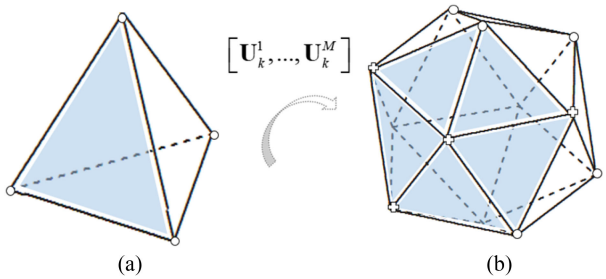

Fig. 2. Extended subspace spanned by the original labeled samples (○) and augmented labeled samples (+).

C. Neighborhood Activity Degree (NAD) Driven Representation-Based Classifier (NADRC)

As introduced in Section II, a testing pixel can be noted as $\mathbf{x}_{i,j}\approx\mathbf{D}\alpha$ by the classic SRC, the sparse coefficient α is processed in conjunction with the dictionary to calculate the approximate value of testing pixel, and the class label is determined by the class-dependent minimum residual error. However, there is latent discriminant information under the sparse coefficient, which can be further investigated. Specifically, α is sparsely constrained, most elements in it are zero. It means that only the labeled samples whose weight values are not zero participated in the representation, namely they are active in the current process. Moreover, the labeled samples with higher degree of activity play important roles in determining the class of the testing pixel.

$$\phi(\mathbf{x}_{i,j})=\left\{\|\mathbf{x}_{i,j}\|^2,\left\|\mathbf{x}_{i,j}{}^T\mathbf{U}^{(1,1)}\right\|^2_2,\dots,\left\|\mathbf{x}_{i,j}{}^T\mathbf{U}^{(1,M)}\right\|^2_2,\dots,\left\|\mathbf{x}_{i,j}{}^T\mathbf{U}^{(K,M)}\right\|^2_2\right\}^T \tag{8}$$

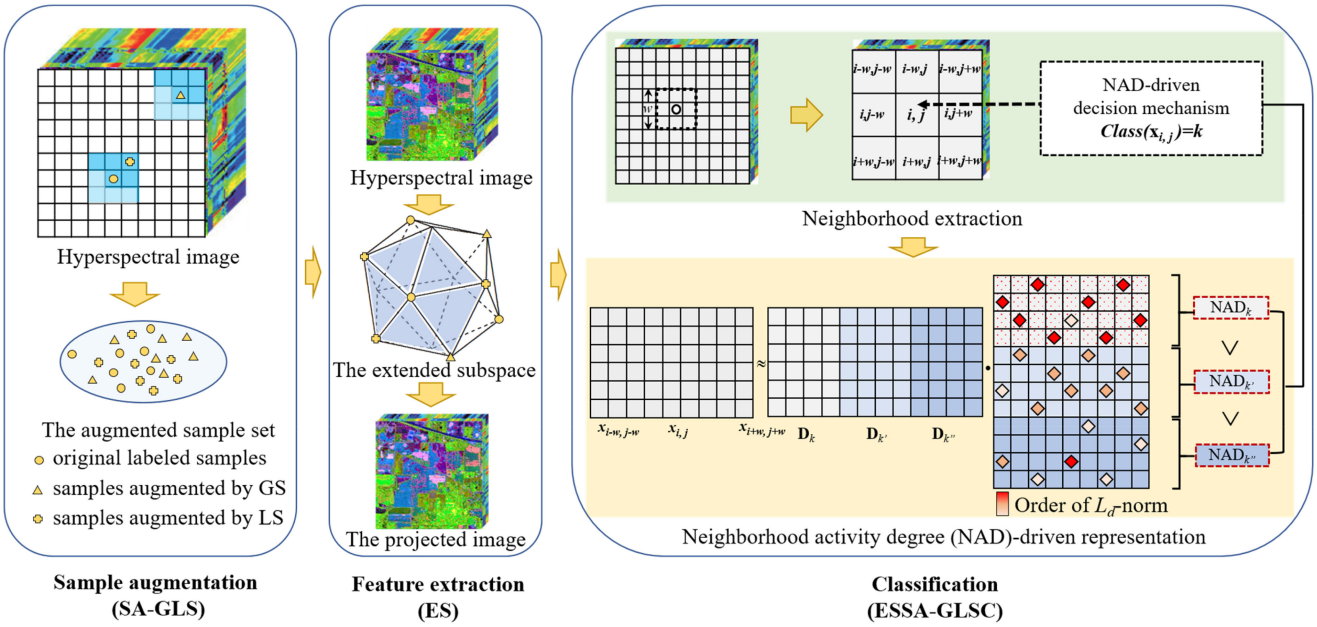

Fig. 3. Principal diagram of ESSA-GLSC.

Based on the abovementioned, a new concept is defined which calculates the contribution of the labeled samples in the dictionary to the representation of the testing pixel, called AD. The AD of a certain class k is calculated as follows:

$$\mathrm{AD}_k = \left\|\alpha_{i,j}^k\right\|_d \tag{10}$$

where $\|\bullet\|_d$ denotes a l_d-norm ($d = 1$ or 2) constraint. However, this decision mechanism only considers the spectral information of the testing pixel but without integration of spatial information, which may lead to misclassification in the HSIC tasks. To utilize the spatial information, for a testing pixel $\mathbf{x}_{i,j}$, we extract a $w \times w$ pixel-sized neighborhood from $\mathbf{X}$, denoted as $\mathbf{A}_{i,j}$, and define a new concept NAD to evaluate the contribution of different class-dependent subdictionaries to the representation of the testing neighborhood. The basic assumption is that spatial adjacent pixels are likely to belong to the same class. The NAD of a certain class is calculated by the sum of AD as follows:

$$\mathrm{NAD}_k = \sum_{v=1}^{\widehat{w}^2} \mathrm{AD}_k^{(v)} \tag{11}$$

where v represents the index of vector in the neighborhood. Therefore, the NADRC takes advantage of the spatial information and effectively corrects the latent misclassification.

D. Extended Subspace Projection Upon Sample Augmentation Based on Global Spatial and Local Spectral Similarity

In this context, the proposed approach can be implemented as three steps: First, augment the training samples by taking advantage of the GLSC so that the classifier can be better trained. Second, reduce the dimension by projecting the original HSI to a lower dimensional subspace, which was spanned by the extended labeled samples. Finally, integrate the above-mentioned implements with NADRC to bring better characterization of features and improvement of classification. Therefore, the objective function of ESSA-GLSC can be finally expressed as follows:

$$\begin{aligned}\hat{\alpha}_{i,j}^{ES} &= \arg\min_{\alpha_{i,j}^{ES}} \left\{\left\|\phi(\mathbf{x}_{i,j}) - \phi(\mathbf{D}^{ES})\alpha_{i,j}^{ES}\right\|_2^2 + \lambda\left\|\alpha_{i,j}^{ES}\right\|_1\right\} \\ \mathrm{class}(\phi(\mathbf{x}_{i,j})) &= \arg\min_k \left\{\| \phi(\mathbf{x}_{i,j}) - \phi(\mathbf{D}^{ES})\alpha_{i,j}^{ES} \|_2^2\right\}\end{aligned} \tag{12}$$

where $\phi(\mathbf{D}^{ES})$ denotes the dictionary, which has extended labeled samples and be projected into a lower dimensional subspace. The pseudocode for the proposed ESSA-GLSC is presented in Algorithm 2. Fig. 3 illustrates the schematic diagram of it.

Algorithm 2: ESSA-GLSC.

Input: Original HSI $\mathbf{X}$, the augmented labeled sample set S_3.

1. obtain $\mathbf{X}'$ by normalization.
2. construct the dictionary$\mathbf{D}^E$ by randomly selecting training samples from $S_3 = \{\mathbf{x}_{i,j}, \mathbf{x}_{a,b}, \ldots, \mathbf{x}_{u,v}\}$
3. $\phi(\mathbf{X}) \leftarrow \mathbf{X}$; $\phi(\mathbf{D}^{ES}) \leftarrow \mathbf{D}^E$
   /* exploit $\mathbf{D}^E$ to construct a lower-dimensional subspace and project the original scene to it. */
4. for $i = 1$ to r, $j = 1$ to c do
   $k = class(\phi(\mathbf{x}_{i,j}))k \in [1, K]$
   /* the image is classified by (12) */

Output: the identity of the testing pixel.

IV. Experimental Results and Analysis

In this section, ESSA-GLSC is evaluated via three real hyperspectral datasets from different sensors, of which the details are provided in Section IV-A. The related parameters of the sample augmentation stage, extended subspace projection, and NADRC are illustrated in Section IV-B. For comparison,

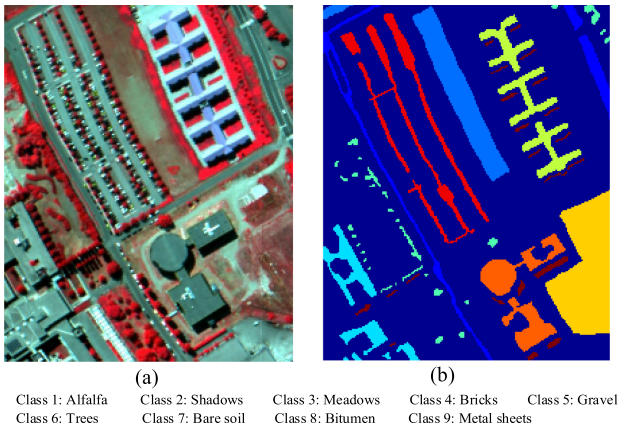

Fig. 4. ROSIS University of Pavia dataset. (a) False-color composite image. (b) Reference map.

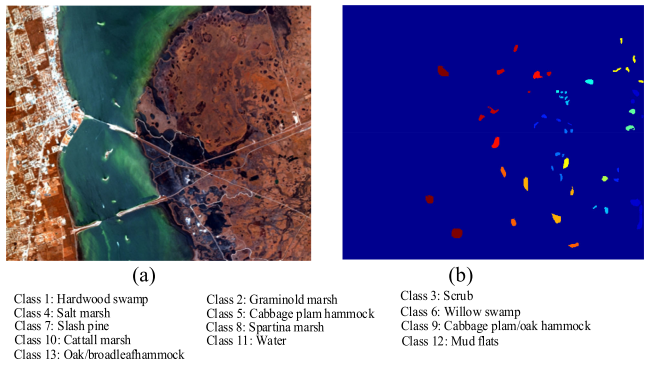

Fig. 5. AVIRIS Kennedy space center dataset. (a) False-color composite image. (b) Reference map.

the SVM, SRC, support vector machines-Markov random field (SVM-MRF), and joint SRC (JSRC) [53] are used to compare the performance of the proposed method. Moreover, labeled sample size and the number of dimensions is also used as variables to measure the effectiveness of the proposed method. From the perspective of the overall accuracy (OA) and the class-dependent accuracy (CA), Section IV-C presents the analysis of the classification results of different methods mentioned earlier.

A. Data Description

1) ROSIS University of Pavia Scene: The University of Pavia scene was acquired by the reflective optics system imaging spectrometer (ROSIS) sensor. The size of this scene is 250×200 pixels, and the spatial resolution is 1.3 m. The scene consists of 103 bands, with 12 bands are removed due to the high noise and water absorption, the spectral range is from 0.43 to 0.86 μm. Nine ground-truth classes, with a total of 12 889 labeled samples are provided in the reference data. Fig. 4(a) shows the false-color composite image, and Fig. 4(b) is the corresponding reference map.

2) AVIRIS Kennedy Space Center Scene: The Kennedy Space Center scene was collected by the airborne visible/infrared imaging spectrometer (AVIRIS) sensor. The scene contains 512×614 pixels with 176 bands after removing the high noise and water absorption bands. Its false-color image is shown in Fig. 5(a). Thirteen ground-truth classes, with a total of 5211 labeled samples are provided in the reference data shown in Fig. 5(b).

3) HYDICE Washington DC Mall Scene: The Washington DC Mall scene was collected by the hyperspectral digital image collection experiment (HYDICE) sensor over the Washington, DC, USA. The size of this scene is $280 \times 307 \times 191$, with a spatial resolution of 3 m. The reference data of this image contains six classes, with a total of 10 190 labeled samples. Fig. 6(a) shows the false-color composite image and the reference map of this scene is shown in Fig. 6(b).

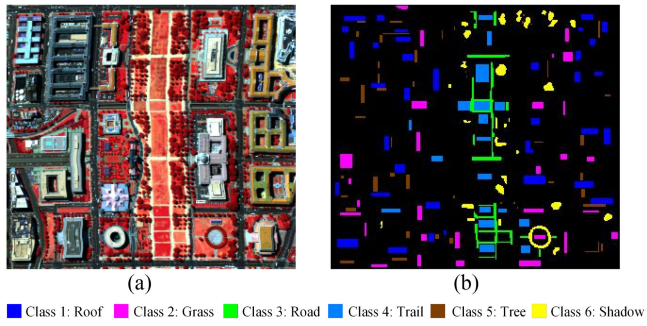

Fig. 6. HYDICE Washington DC Mall dataset. (a) False-color composite image. (b) Reference map.

B. Parameter Setting

There are some key parameters are investigated in detail in this section. First, at the sample augmentation stage, the size of the patches plays an important role in the classification. Considering the true ground texture and the spatial resolution synthetically, the size of the patch that used for global spatial similarity search w_1 and used for local spectral similarity search w_2 is set to 5 and 3, respectively, for all the datasets. And the number of components kept by SVD is set to 4. Second, when using the extended subspace projection to reduce the original dimension, the number of the clusters M that randomly selected from each class is also an important parameter to determine the performance of feature extraction. According to the cross validation, M is given to the values ranging from 1 to 6 for the University of Pavia scene and the Kennedy Space Center scene, for the Washington DC Mall scene, the M is denoted as 1 to 13. Moreover, the parameter τ which controls the retention ratio of the original spectral information is denoted to 99%. When the projected image is classified by the NADRC, the size of neighborhood w is set to 5 according to the cross validation.

Due to the limitation of objective factors, it is laborious to obtain sufficient training samples. Therefore, in practical application, the number of labeled samples is often insufficient, which leads to the poor performance of many classifiers. The method

TABLE I
OVERALL AND CLASS-DEPENDENT ACCURACIES (IN PERCENT) OBTAINED BY THE DIFFERENT TESTED METHODS FOR THE ROSIS UNIVERSITY OF PAVIA SCENE

Class	SVM	SRC	SRCsub	$SRC_{SA\text{-}GLS}$	$SRC_{ESSA\text{-}GLS}$	SVM-MRF	JSRC	ESSA-GLSC
1	64.89%	23.69%	48.54%	59.77%	65.67%	84.59%	23.10%	**93.37%**
2	67.93%	60.70%	49.69%	68.75%	89.04%	85.75%	82.37%	**97.38%**
3	66.30%	81.98%	80.25%	84.59%	85.26%	46.78%	80.83%	**95.76%**
4	92.40%	96.37%	96.58%	94.44%	96.58%	89.14%	96.15%	**97.86%**
5	96.16%	99.55%	99.41%	99.55%	99.93%	97.99%	99.85%	**100.00%**
6	67.74%	62.82%	59.47%	65.29%	62.85%	67.42%	61.05%	**72.65%**
7	70.65%	87.52%	78.65%	89.77%	86.02%	71.34%	85.56%	**95.41%**
8	58.75%	69.16%	67.58%	70.57%	62.73%	75.57%	**85.31%**	82.27%
9	**99.50%**	87.24%	93.83%	94.03%	96.91%	97.88 %	58.02%	99.18%
OA	71.50%	69.21%	68.25%	75.76%	77.97%	77.17%	73.03%	**89.14%**

The best results are highlighted in bold typeface (in all cases, **5** labeled samples per class were used).

TABLE II
OVERALL AND CLASS-DEPENDENT ACCURACIES (IN PERCENT) OBTAINED BY THE DIFFERENT TESTED METHODS FOR THE AVIRIS KENNEDY SPACE CENTER SCENE

Class	SVM	SRC	SRCsub	$SRC_{SA\text{-}GLS}$	$SRC_{ESSA\text{-}GLS}$	SVM-MRF	JSRC	ESSA-GLSC
1	75.81%	88.44%	85.55%	89.75%	95.01%	90.57%	90.67%	**97.90%**
2	73.74%	67.49%	75.72%	74.49%	77.37%	**82.69%**	75.72%	78.60%
3	58.29%	92.58%	67.19%	94.53%	90.63%	48.53%	81.25%	**96.88%**
4	24.17%	28.17%	40.08%	38.49%	48.81%	12.51%	**74.21%**	51.19%
5	48.40%	67.08%	69.57%	70.81%	73.29%	43.97%	75.78%	**93.17%**
6	44.29%	41.48%	38.43%	58.95%	47.60%	41.79%	38.86%	**63.32%**
7	83.60%	83.81%	94.29%	91.43%	99.05%	91.10%	**100.00%**	99.05%
8	45.26%	81.90%	78.42%	91.88%	77.26%	67.56%	77.26%	**94.43%**
9	64.76%	88.65%	89.04%	87.12%	94.62%	81.69%	90.00%	**99.81%**
10	69.10%	91.09%	82.43%	92.57%	92.33%	88.75%	94.55%	**97.28%**
11	87.73%	87.11%	87.35%	87.83%	83.77%	**93.94%**	89.26%	85.68%
12	80.36%	64.21%	70.58%	67.59%	78.93%	85.40%	71.57%	**91.45%**
13	97.91%	99.35%	99.68%	99.57%	99.68%	**100.00%**	**100.00%**	**100.00%**
OA	71.53%	81.12%	80.33%	84.48%	85.72%	79.41%	84.99%	**91.67%**

The best results are highlighted in bold typeface (in all cases, **5** labeled samples per class were used).

we proposed is committed to improving the classification accuracy in the case of small size of labeled samples, so we used 5 samples per class in the experiments to simulate the insufficient situation and illustrate the performance of this method.

C. Results Analysis and Discussion

In our experiments, we randomly select 5 labeled samples per class to construct the dictionary and train the classifier, where the remaining labeled samples are applied for validation. Tables I, II and III provide the results of OA and CA acquired by the comparative experiments, with their corresponding classification maps shown in Figs. 7–9. Fig. 10 shows the time complexity of the proposed method and other spatial-spectral SR-based classifiers. Based on these results, several conclusions can be drawn as follows.

1) Comparison of the Spectral Classifiers: Compared with SVM, SRC has nearly the same performance in the classification. The results are testaments to the effectiveness of representation-based framework in HSIC.

2) Comparison of Labeled Sample Size: The dictionary of SRC has 45 labeled samples in the University of Pavia scene, 65 labeled samples in the Kennedy Space Center scene and 30 labeled samples in the Washington DC Mall scene. After using the GLSC to sample augmentation ($SRC_{SA\text{-}GLS}$), the dictionary contains 127 labeled samples, 192 labeled samples, and 81 labeled samples, respectively. The classification accuracy is obviously improved (from 69.21% to 75.76% in the University of Pavia scene, from 81.12% to 84.48% in the Kennedy Space Center scene and from 80.40% to 81.17% in the Washington DC Mall scene) due to more training samples are used to train the supervised classifier.

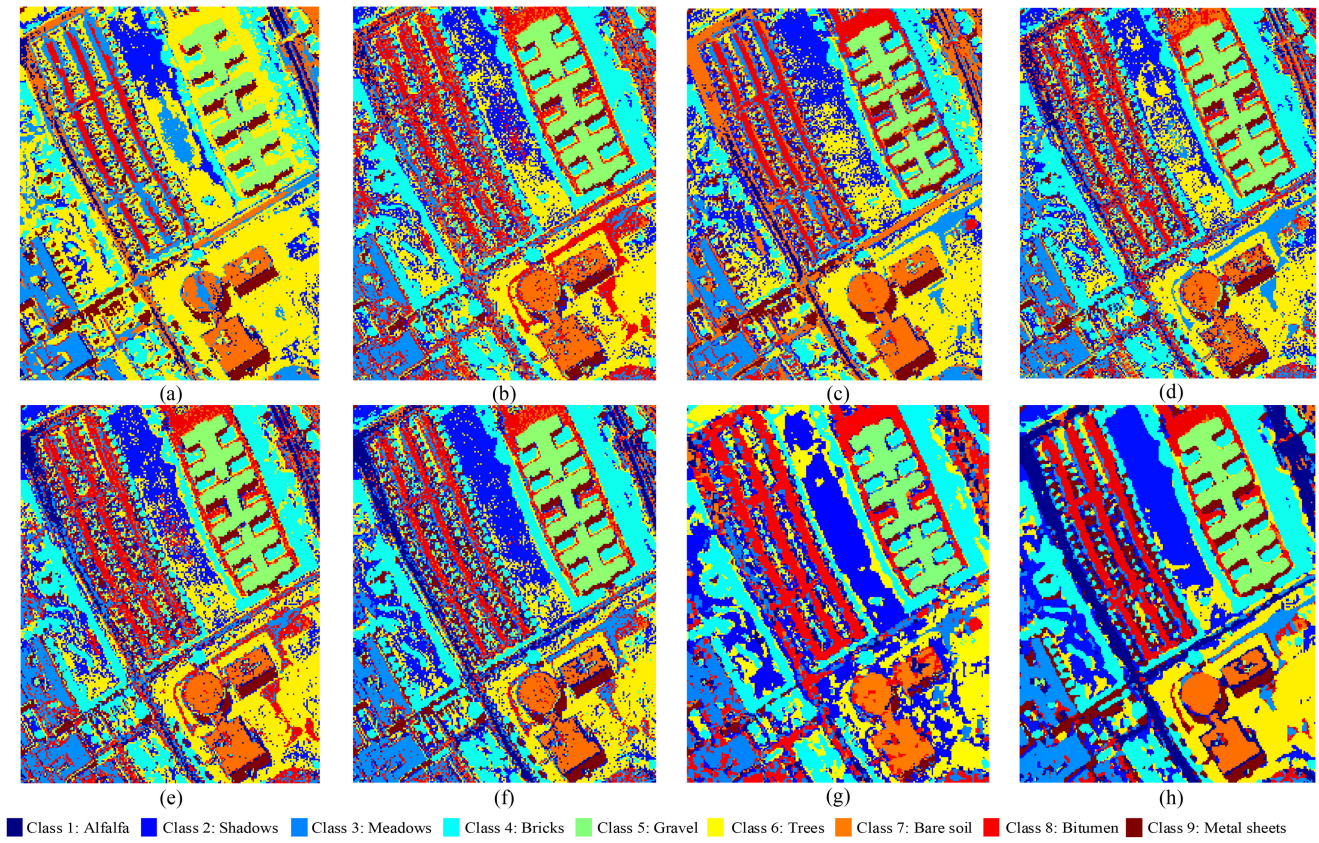

Fig. 7. Classification maps obtained by the different tested methods for the ROSIS University of Pavia dataset. The overall accuracies are given in the parentheses. (a) SVM (71.50%). (b) SRC (69.21%). (c) SRCsub (68.25%). (d) $\text{SRC}_{\text{SA-GLS}}$ (75.76%). (e) $\text{SRC}_{\text{ESSA-GLS}}$ (77.97%). (f) SVM-MRF (77.17%). (g) JSRC (73.03%). (h) ESSA-GLSC (89.14%).

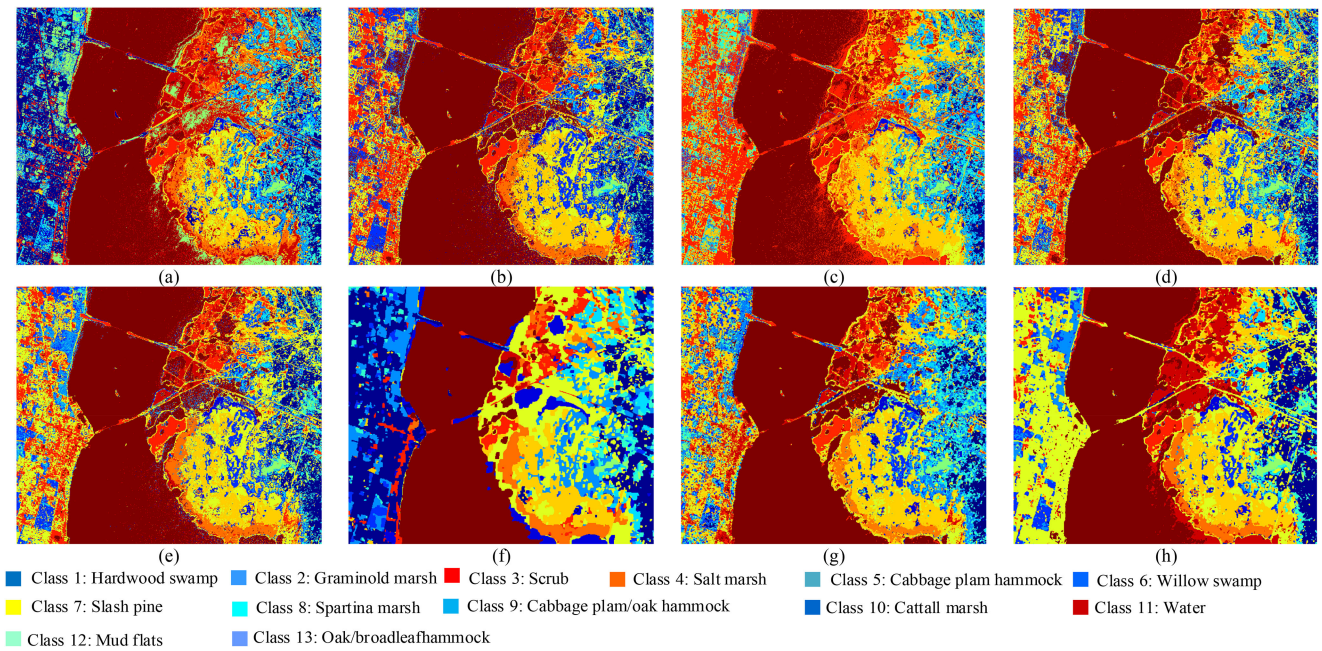

Fig. 8. Classification maps obtained by the different tested methods for the AVIRIS Kennedy space center dataset. The overall accuracies are given in the parentheses. (a) SVM (71.53%). (b) SRC (81.12%). (c) SRCsub (80.33%). (d) $\text{SRC}_{\text{SA-GLS}}$ (84.48%). (e) $\text{SRC}_{\text{ESSA-GLS}}$ (85.72%). (f) SVM-MRF (79.41%). (g) JSRC (84.99%). (h) ESSA-GLSC (91.67%).

TABLE III
OVERALL AND CLASS-DEPENDENT ACCURACIES (IN PERCENT) OBTAINED BY THE DIFFERENT TESTED METHODS FOR THE HYDICE WASHINGTON DC MALL SCENE

Class	SVM	SRC	SRCsub	$SRC_{SA\text{-}GLS}$	$SRC_{ESSA\text{-}GLS}$	SVM-MRF	JSRC	ESSA-GLSC
1	78.94%	60.07%	61.93%	58.28%	76.23%	**86.37%**	54.63%	80.57%
2	77.34%	70.39%	48.24%	72.69%	65.09%	77.00%	**80.04%**	66.61%
3	75.60%	81.28%	70.93%	81.87%	78.85%	69.18%	**92.17%**	74.64%
4	88.38%	80.12%	17.41%	82.59%	84.87%	**92.34%**	90.06%	84.39%
5	**98.39%**	76.86%	55.96%	80.03%	83.97%	91.92%	85.18%	80.52%
6	88.90%	77.23%	27.92%	76.71%	78.10%	92.18%	**95.32%**	80.84%
OA	83.28%	80.40%	52.08%	81.17%	86.73%	84.95%	78.52%	**87.59%**

The best results are highlighted in bold typeface (in all cases, **5** labeled samples per class were used).

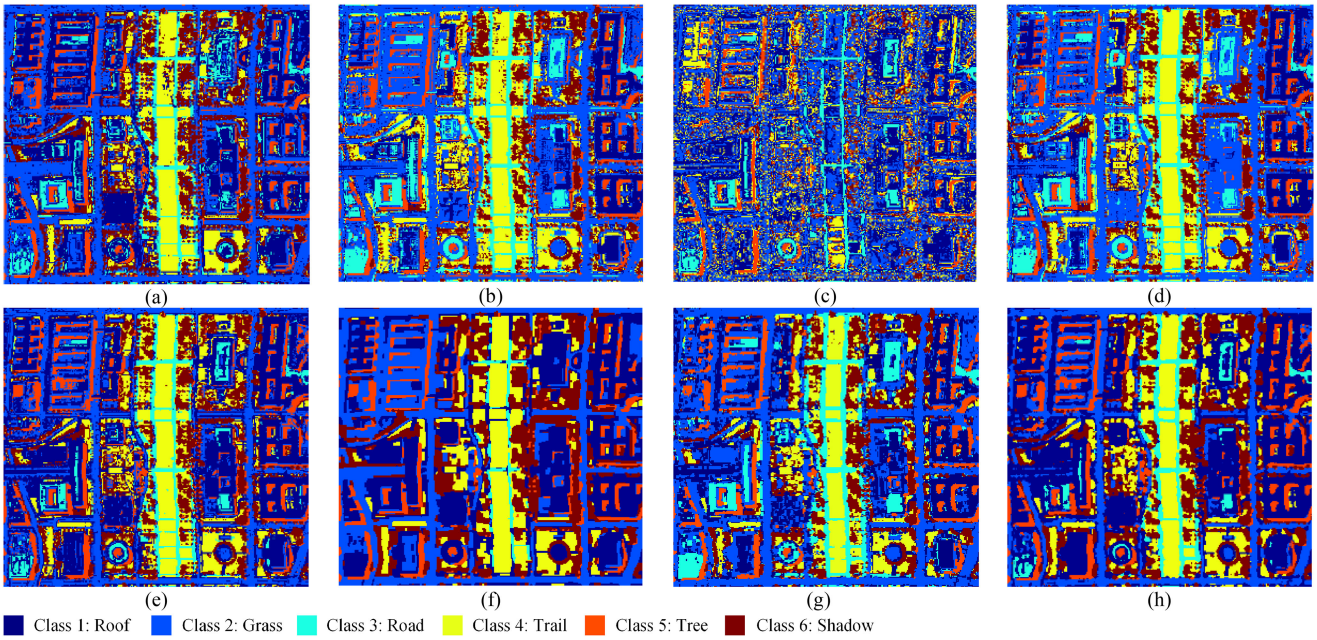

Fig. 9. Classification maps obtained by the different tested methods for the HYDICE Washington DC Mall dataset. The overall accuracies are given in the parentheses. (a) SVM (83.28%). (b) SRC (80.40%). (c) SRCsub (52.08%). (d) $SRC_{SA\text{-}GLS}$ (81.17%). (e) $SRC_{ESSA\text{-}GLS}$ (86.73%). (f) SVM-MRF (84.95%). (g) JSRC (78.52%). (h) ESSA-GLSC (87.59%).

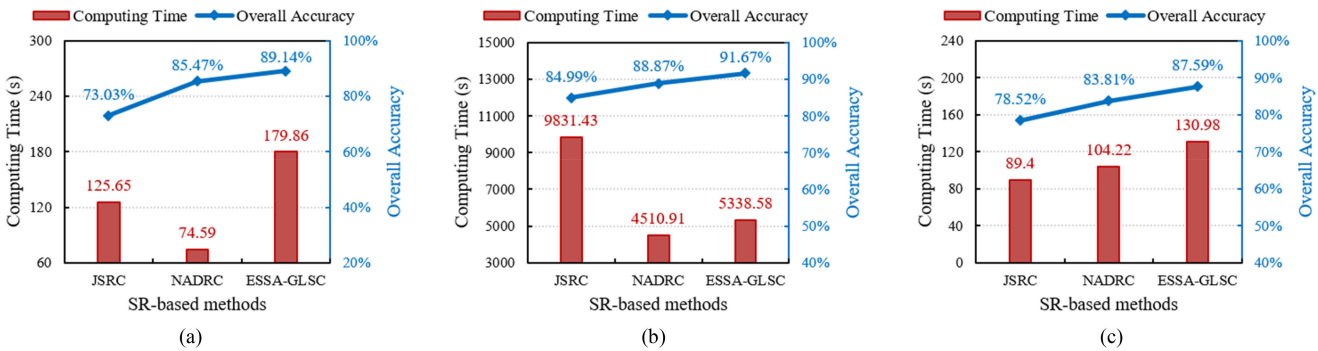

Fig. 10. Comparison of computing time and OA for three SR-based methods for three datasets. (a) University of Pavia scene. (b) Kennedy Space Center scene. (c) Washington DC Mall scene.

3) Comparison of Dimensionality Reduction Effects: The subspace-based SRC (SRCsub) has a lower OA compared with SRC for all datasets (from 69.21% to 68.25%, from 81.12% to 80.33%, and from 80.40% to 52.08%). Based on the abovementioned, there are some limitations in the subspace projection that constraint its application.

Compared with $\mathrm{SRC}_{\mathrm{SA\text{-}GLS}}$, the $\mathrm{SRC}_{\mathrm{ESSA\text{-}GLS}}$ which exploits extended subspace projection to $\mathrm{SRC}_{\mathrm{SA\text{-}GLS}}$, achieves suitable improvements for all the experimental datasets (from 75.76% to 77.97%, from 84.48% to 85.72%, and from 81.17% to 86.73%), which demonstrates the effectiveness of the extended subspace projection as a feature extraction method.

4) Comparison of Other Spectral-Spatial Classifiers: The ESSA-GLSC we proposed as a comparative test with other spectral-spatial classifiers such as SVM-MRF and JSRC to illustrate the effectiveness of our method, and obtains the best performance for all datasets, with the overall accuracies of 89.14%, 91.67%, and 87.59%, respectively. Considering the time complexity, the spatial-spectral SR-based classifiers JSRC and NADRC are compared to illustrate the effectiveness of the proposed method. For the University of Pavia scene, the time cost of ESSA-GLSC (179.86 s) is slightly higher than JSRC (125.65 s) and NADRC (74.59 s), but the accuracy can be improved by about 10%. For Kennedy Space Center scene, the time complexity of ESSA-GLSC is similar to that of NADRC, but much lower than that of JSRC, and its classification accuracy has reached the highest than other classifiers (84.99% and 88.87%). For the Washington DC Mall scene, ESSA-GLSC achieves the highest OA in the time close to the other two classifiers. The results demonstrate that the time complexity of these methods is close and acceptable. Meanwhile, the proposed method can achieve a better classification performance. The above-mentioned results are shown in Fig. 10 in a more intuitive form. It needs to be reiterated that the number of training samples in all experiments was 5.

V. Conclusion

This article proposed a method which takes both insufficient labeled samples and redundant bands into account in HSIC. Specifically, a sample augmentation method based on the GLSC is put forward to enlarge the original labeled sample set, which guarantees that the HSIC can obtain more prior information than before. Then, the extended labeled sample set is used to construct a lower dimensional subspace, which is more comprehensive to reflect the distribution of ground objects. Finally, the original HSI is projected to the subspace and classified by the spectral-spatial classifier NADRC. Experimental results on three real hyperspectral datasets from different sensors demonstrate the availability and effectiveness of the proposed method.

References

[1] A. Plaza *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. S110–S122, Sep. 2009.

[2] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[3] C.-I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. Norwell, MA, USA: Kluwer, 2003.

[4] M. Fauvel *et al.*, "Advances in spectral-spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, Mar. 2013.

[5] M. Song, X. Shang, and C.-I. Chang, "3-D receiver operating characteristic analysis for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 8093–8115, Nov. 2020.

[6] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2021.

[7] D. Hong *et al.*, "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, May 2021.

[8] D. Hong *et al.*, "Learning to propagate labels on graphs: An iterative multitask regression framework for semi-supervised hyperspectral dimensionality reduction," *ISPRS J. Photogrammetry*, vol. 158, pp. 35–49, Dec. 2019.

[9] H. Su, B. Yong, and Q. Du, "Hyperspectral band selection using improved firefly algorithm," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 68–72, Jan. 2016.

[10] Y. Wang *et al.*, "Constrained-target band selection for multiple-target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 6079–6103, Aug. 2019.

[11] C.-I. Chang *et al.*, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 6, pp. 2631–2641, Nov. 1999.

[12] 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.

[13] L. Wang, H. Li, B. Xue, and C. Chang, "Constrained band subset selection for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 14, no. 11, pp. 2032–2036, Nov. 2017.

[14] A. Sellami, M. Farah, I. R. Farah, and B. Solaiman, "Hyperspectral imagery semantic interpretation based on adaptive constrained band selection and knowledge extraction techniques," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1337–1347, Apr. 2018.

[15] M. Zhang *et al.*, "Unsupervised feature extraction in hyperspectral images based on Wasserstein generative adversarial network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 5, pp. 2669–2688, May 2019.

[16] A. L. Price *et al.*, "Principal components analysis corrects for stratification in genome-wide association studies," *Nature Genetics*, vol. 38, no. 8, pp. 904–909, Aug. 2006.

[17] T. Chin and D. Suter, "Incremental Kernel principal component analysis," *IEEE Trans. Image Process.*, vol. 16, no. 6, pp. 1662–1674, Jun. 2007.

[18] M. Imani and H. Ghassemian, "Principal component discriminant analysis for feature extraction and classification of hyperspectral images," in *Proc. Iranian Conf. Intell. Syst.*, Bam, Iran, 2014, pp. 1–5.

[19] B. Zhao, L. Gao, and B. Zhang, "An optimized method of Kernel minimum noise fraction for dimensionality reduction of hyperspectral imagery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2016, pp. 48–51.

[20] C. Gordon, "A generalization of the maximum noise fraction transform," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 1, pp. 608–610, Jan. 2000.

[21] J. Li, J. Bioucas-Dias, and A. Plaza, "Spectral–spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 809–823, Mar. 2012.

[22] H. Yu, L. Gao, W. Liao, B. Zhang, A. Pižurica, and W. Philips, "Multiscale superpixel-level subspace-based support vector machines for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 2142–2146, Dec. 2017.

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

[24] X. Shang, M. Song, and C. Chang, "An iterative random training sample selection approach to constrained energy minimization for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 9, pp. 1625–1629, Sep. 2021.

[25] L. Yu, J. Xie, S. Chen, and L. Zhu, "Generating labeled samples for hyperspectral image classification using correlation of spectral bands," *Frontiers Comput. Sci.*, vol. 10, no. 2, pp. 292–301, Apr. 2016.

[26] B. Cui, X. Ma, F. Zhao, and Y. Wu, "A novel hyperspectral image classification approach based on multiresolution segmentation with a few labeled samples," *Int. J. Adv. Robot. Syst.*, vol. 14, no. 3, pp. 1–10, 2017.

[27] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson, "SVM- and MRF-based method for accurate classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 736–740, Oct. 2010.
[28] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.
[29] H. Yu *et al.*, "Spectral-spatial hyperspectral image classification using subspace-based support vector machines and adaptive Markov random fields," *Remote Sens.*, vol. 8, no. 4, Apr. 2016, Art. no. 355.
[30] S. Xu, J. Li, M. Khodadadzadeh, A. Marinoni, P. Gamba, and B. Li, "Abundance-indicated subspace for hyperspectral classification with limited training samples," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 4, pp. 1265–1278, Apr. 2019.
[31] L. Gao *et al.*, "Subspace-based support vector machines for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 12, no. 2, pp. 349–353, Feb. 2015.
[32] Y. Tang, H. Yuan, and L. Li, "Manifold-based sparse representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7606–7618, Dec. 2014.
[33] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification via Kernel sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 217–231, Jan. 2013.
[34] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 471–478.
[35] X. Sun, Q. Qu, N. M. Nasrabadi, and T. D. Tran, "Structured priors for sparse-representation-based hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 7, pp. 1235–1239, Dec. 2013.
[36] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.
[37] H. Yuan and Y. Tang, "Sparse representation based on set-to-set distance for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2464–2472, Jun. 2015.
[38] C. Chen, N. Chen, and J. Peng, "Nearest regularized joint sparse representation for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 424–428, Mar. 2016.
[39] H. Yu *et al.*, "Global spatial and local spectral similarity-based manifold learning group sparse representation for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3043–3056, May 2020.
[40] H. Yu, L. Gao, W. Li, Q. Du, and B. Zhang, "Locality sensitive discriminant analysis for group sparse representation-based hyperspectral imagery classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 8, pp. 1358–1362, Aug. 2017.
[41] L. Zhuang and J. M. Bioucas-Dias, "Hyperspectral image denoising based on global and non-local low-rank factorizations," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2017, pp. 1900–1904.
[42] X. Zhang, Z. Gao, L. Jiao, and H. Zhou, "Multifeature hyperspectral image classification with local and nonlocal spatial information via Markov random field in semantic space," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1409–1424, Mar. 2018.
[43] H. Yu, L. Gao, W. Liao, and B. Zhang, "Group sparse representation based on nonlocal spatial and local spectral similarity for hyperspectral imagery classification," *Sensors.*, vol. 18, no. 6, May 2018, Art. no. 1695.
[44] J. Zou, W. Li, C. Chen, and Q. Du, "Scene classification using local and global features with collaborative representation fusion," *Inf. Sci.*, vol. 348, pp. 209–226, Jun. 2016.
[45] L. Fang, N. He, S. Li, A. J. Plaza, and J. Plaza, "A new spatial–spectral feature extraction method for hyperspectral images using local covariance matrix representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3534–3546, Jun. 2018.
[46] W. Zhao and S. Du, "Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544–4554, Aug. 2016.
[47] D. Donoho, "For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution," *Commun. Pure Appl. Math.*, vol. 59, no. 6, pp. 797–829, Mar. 2006.
[48] H. Yu, X. Zhang, C. Yu, L. Gao, and B. Zhang, "Neighborhood activity-driven representation for hyperspectral imagery classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4506–4517, Aug. 2020.
[49] Z. Zha, X. Yuan, B. Wen, J. Zhang, J. Zhou, and C. Zhu, "Simultaneous nonlocal self-similarity prior for image denoising," in *Proc. IEEE Int. Conf. Image Process.*, Dec. 2019, pp. 1119–1123.
[50] R. Wang and H. Li, "Nonlocal similarity regularization for sparse hyperspectral unmixing," in *Proc. IEEE Geosci. Remote Sens. Symp.*, 2014, pp. 2926–2929.
[51] L. Zhuang and J. M. Bioucas-Dias, "Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 730–742, Mar. 2018.
[52] S. Wang, Z. Liu, W. Dong, L. Jiao, and Q. Tang, "Total variation-based image deblurring with nonlocal self-similarity constraint," *Electron. Lett.*, vol. 47, no. 16, pp. 916–918, Aug. 2011.
[53] W. Li and Q. Du, "Joint within-class collaborative representation for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2200–2208, Jun. 2014.

Jiaochan Hu received the B.S. degree in remote sensing science and technology from Wuhan University, Wuhan, China, in 2012, the M.S. degree in surveying and mapping engineering from the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences (CAS), Beijing, China, in 2015, and the Ph.D. degree in cartography and geographic information system from the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, CAS, Beijing, China, in 2019.

She is currently a Lecture with the College of Environmental Sciences and Engineering, Dalian Maritime University, Dalian, China. Her research interests focus on hyperspectral image processing, analysis and applications.

Xueji Shen received the B.S. degree in light-chemical engineering from the Shaanxi University of Science and Technology, Shaanxi, China, in 2018. She is currently working toward the M.S. degree in computer technology with the Center of Hyperspectral Imaging in Remote Sensing, Information Science and Technology College, Dalian Maritime University, Dalian, China.

Her research interests include hyperspectral image processing and pattern recognition.

Haoyang Yu (Member, IEEE) received the B.S. degree in information and computing science from Northeastern University, Shenyang, China, in 2013, and the Ph.D. degree in cartography and geographic information system from the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China, in 2019.

He is currently a Xing Hai Associate Professor with the Center of Hyperspectral Imaging in Remote Sensing, Information Science and Technology College, Dalian Maritime University, Dalian, China. His research interests include models and algorithms for hyperspectral image processing, analysis and applications.

Xiaodi Shang (Student Member, IEEE) received the B.S. degree in software engineering from Qingdao University, Qingdao, China, in 2016. She is currently working toward the Ph.D. degree in computer application technology with Dalian Maritime University, Dalian, China.

Her research interests include hyperspectral image classification, band selection and applications.

Qiandong Guo received the B.E. degree in remote sensing science and technology from Wuhan University, Wuhan, China, in 2011, the M.S. degree in cartography and geographic information system from the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China, in 2014, and the Ph.D. degree in geography from the University of South Florida, Tampa, FL, USA, in 2018.

His research interests include hyperspectral image processing, target detection, and land use and cover changes.

Bing Zhang (Fellow, IEEE) received the B.S. degree in geography from Peking University, Beijing, China, in 1991, and the M.S. and Ph.D. degrees in remote sensing from the Institute of Remote Sensing Applications, Chinese Academy of Sciences (CAS), Beijing, China, in 1994 and 2003, respectively.

He is a Full Professor and the Deputy Director of the Aerospace Information Research Institute, CAS, where he has been leading lots of key scientific projects in the area of hyperspectral remote sensing for more than 25 years. He has developed five software systems in image processing and applications. His creative achievements were rewarded ten important prizes from the Chinese government, and special government allowances of the Chinese State Council. He has authored more than 300 publications, including more than 170 journal articles. He has edited six books/contributed book chapters on hyperspectral image processing and subsequent applications. His research interests include the development of mathematical and physical models and image processing software for the analysis of hyperspectral remote sensing data in many different areas.

Dr. Zhang has been serving as a Technical Committee Member of IEEE Workshop on Hyperspectral Image and Signal Processing, since 2011, and has been the President of Hyperspectral Remote Sensing Committee of China National Committee of International Society for Digital Earth since 2012, and has been the Standing Director of Chinese Society of Space Research (CSSR) since 2016. He is the Student Paper Competition Committee Member in International Geoscience and Remote Sensing Symposium (IGARSS) from 2015 to 2020. He received the National Science Foundation for Distinguished Young Scholars of China in 2013, and the 2016 Outstanding Science and Technology Achievement Prize of the Chinese Academy of Sciences, the highest level of Awards for the CAS Scholars. He is serving as an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.