Abstract—Due to variations in imaging conditions, spectra of the same type of ground objects usually exhibit certain discrepancies, leading to intra-class spectral distance increase and inter-class distance decrease. As a result, classification accuracy is greatly affected, especially in cases with few labeled samples. For representation-based classifiers, the spectral variability within limited training samples is far from sufficient to represent diverse variations within testing ones. To handle this problem, a spectral variation augmented representation for hyperspectral imagery classification (SVARC) with few labeled samples is proposed in this article. First, a novel class-independent and class-dependent components-based linear representation model (CICD-LRM) is proposed to emphasize the representation of spectral variation. Second, depending on spatial and spectral correlation, the CICD-LRM-guided global and local spectral variation extraction schemes are designed, and a fused spectral variation dictionary is constructed by concatenation. Finally, a classifier for hyperspectral images based on the CICD-LRM and spectral variation dictionary is proposed, and specifically, three different spectral variation reconstruction strategies are designed. Similar to most representation-based classifiers, a residual-driven decision is also employed in the proposed classifier. Comparative experiments are conducted with eight classical and state-of-the-art methods using two benchmark datasets. The experimental results demonstrate that the proposed SVARC method significantly outperforms the compared ones in cases with few labeled samples.

Index Terms—Augmented representation, hyperspectral classification, limited training samples, spectral variation.

I. INTRODUCTION

HYPERSONTAL image (HSI) is acquired with a relatively new image-spectrum merging technology.

Compared with traditional remote sensing images, HSI possesses some unique spectral advantages. Specifically, it can record up to hundreds of narrow spectral bands. Its high spectral resolution can reveal very subtle spectral signatures and hence enhance the discrimination of different materials or objects, which makes HSI analysis an active topic in many applications [1], [2], [3].

As an important application, hyperspectral imagery classification (HIC) has attracted plenty of attention and has been applied in many practical scenarios. In early research, classical classifiers, such as k-nearest neighbor (KNN) [4] and support vector machines (SVMs) [5], are successfully generalized to HIC. With the development of deep learning, various artificial neural network-based classifiers have exhibited excellent performance in recent works [6], [7], [8]. When sufficient training samples are available, these classifiers can exploit abundant spectral characteristics or extract deep features very well, and thus generate satisfying classification results.

Since labeling pixels of HSIs is a highly labor-intensive and time-consuming task, the number of samples with accurate labels is usually limited, which greatly influences classification accuracy. The inevitable spectral variability further exacerbates this problem. Spectral variability primarily describes the widespread phenomena that the same objects with non-identical spectra and different objects with similar spectra. It is generally caused by internal (such as sensor parameters and material composition) and external factors (mainly imaging conditions) [9], [10]. Spectral variations saliently decrease the ground object discrimination. Therefore, HIC with few labeled training samples combined with inevitable spectral variability forms a challenging task.

In HIC, insufficient labeled samples can be addressed by spectral dimensionality reduction (DR) and training sample augmentation. Many spectral DR methods have been proposed, which are usually summarized into two categories: band selection (BS)-based methods and feature extraction (FE)-based methods. BS-based methods select a relevant subset from the original spectral bands, which retains the original information as much as possible and maintains the physical significance of true HSI data. Some clustering-based approaches, such as Affinity Propagation [12], improved...
sparse subspace clustering (ISSC) [13], and SpaBS [14] are the most commonly used BS-based methods.

Different from BS-based methods, FE-based ones are generally implemented by mapping the spectral vector into a new and lower-dimension space with a transformation matrix. FE-based methods are generally divided into spectral and spectral-spatial-based approaches. Principal component analysis (PCA) [15], maximum noise fraction (MNF) [16], and independent component analysis (ICA) [17] are classical spectral-based FE approaches. Considering the nonlinear and sparse characteristics in HSIs, other advanced methods, including Kernel PCA (KPCA) [18], sparse PCA [19], and sparse self-representation (SSR) [20], are developed. Similarly, linear discriminant analysis (LDA) and some nonlinear variations [21], [22], [23], [24] designed on the basis of prior information have also been widely used in HIC. On the other hand, some spectral–spatial FE works, containing spectral–spatial Kernel-based multivariate analysis (KMASSA) [25], spectral–spatial LDA [26], and spectral–spatial similarity measure-based DR [27], have also attracted wide attention.

BS- and FE-based methods can usually extract discriminant spectral features or bands to alleviate the Huges problem caused by insufficient labeled samples and reduce computational complexity, while some useful spectral bands or information may still be discarded, leading to poor classification performance. Thus, some researchers proposed training sample augmentation methods to handle this problem. Shahshahani and Landgrebe [28] pointed out that the unlabeled samples are potentially useful to alleviate the Hughes problem in HSI processing. Inspired by this thought, some training sample augmentation methods in semisupervised manners are proposed. In [29], SVM classifiers with multiple typical kernels are employed to predict the pseudo-labels of testing samples, and then some are selected via a voting method to augment the training set. In [30], a task-driven dictionary learning with Laplacian regularization-based semisupervised method is proposed to produce better classification results. Zheng et al. [31], utilized superpixel (SP) segmentation and distance-weighted linear regression to handle the small labeled training sample set problem. In [32], the classification-guided SP segmentation and the discriminative low-rank representation are iteratively conducted to suppress the spectral variations and increase intraclass similarity. In these methods, part of unlabeled samples is regarded as pseudo-training samples to enlarge the training sample set and then participate in classifier training. The classification performance is highly dependent on the selected pseudo-training samples.

Dictionary representation, including sparse representation (SR) [33], collaborative representation (CR) [34], and their varied versions [35], [36], [37], [38], has been proven to be an effective way to handle the spectral variation within HSI. SR-based classifier (SRC) and CR-based classifier (CRC) represent a testing sample by sparse or collaborative combination of training samples, respectively. Since CRC generally provides more competitive performance with lower complexity, many extended CRC methods have been developed. In [39], the CRC with Tikhonov regularization is proposed to influence the estimation of representation coefficients depending on the similarity level of the corresponding training samples and the testing sample. Similarly, structure-aware CR with Tikhonov regularization (SaCRT) is proposed in [40], which utilizes prior knowledge of training samples and spectral signatures of testing pixels to intervene in estimated coefficients and further improve the classification performance. For most representation-based classifiers, the diversity of spectral variation within the dictionary is totally sufficient to represent that within testing samples when plenty of training samples are available. However, the impact of spectral variation on classification cannot be ignored when only a few labeled training samples are available.

To simultaneously handle the insufficient labeled samples and spectral variation issues, a novel spectral variation augmented representation for hyperspectral imagery classification (SVARC) scheme is proposed for HSI. In the proposed method, the class-dependent and -independent components are innovatively defined, and a novel linear representation model is proposed. Specifically, the class-dependent component refers to the intrinsic spectral characteristics of a specific class, and the class-independent component is relatively implicit and distributed evenly throughout all classes. To further explore prior information on the class-independent components, both local and global spectral variations are extracted with the proposed class-independent and -dependent components-based linear representation model (CICD-LRM). Finally, three different restrictions are designed for reconstruction optimization, and a residual-driven prediction strategy is adopted to assign an explicit label for each testing sample.

The main contributions of this article can be summarized as follows.

1) In the proposed CICD-LRM, a unique class-dependent spectrum and several class-independent spectral variation atoms are employed to jointly represent the macroscopical variations in the same class, accounting for the intrinsic variability in physical properties caused by various environmental changes. The specialized spectral variation dictionary consists of more diverse spectral variations, leading to more accurate reconstruction.

2) Based on the spatial and spectral correlations, methods are designed to explore class-independent components in both local and global manners. The extracted spectral variation models different imaging conditions of gathering the ground objects belonging to the same class, within a limited spatial neighborhood and within the entire HSI scene, respectively. Their concatenation generates a fused spectral variation dictionary, which encloses more diverse variations and enhances representation capability.

3) A spectral variation augmented representation-based HIC scheme is proposed for cases with few labeled training samples, in which the signal reconstruction accuracy significantly influences the classification performance. Thus, three different constraints are adopted to search for the optimal reconstruction of testing samples with class-independent components. Furthermore, the residual-driven prediction converts reconstruction residuals into discriminant information, yielding three specific classification maps generated by
SVARC with three different constraints, that is, SVARC-SR, SVARC-CR, and SVARC-CRT.

The rest of this article is organized as follows. Section II describes three key problems of the proposed classification method, including CICD-LRM, spectral variation extraction (SVE), and spectral variation augmented classifier. Section III presents the experimental results and analysis. Finally, conclusions are drawn in Section IV.

II. PROPOSED METHOD

Fig. 1 depicts the flowchart of the proposed SVARC method. First, the spectral variability within the same class is represented. Second, local and global spectral variation characteristics are extracted and fused to represent the spectral variability. Finally, CICD-LRM is optimized by imposing constraints on representation coefficients corresponding to spectral variation atoms, and then the reconstruction coefficient vectors and residuals are estimated. By employing a residual-driven discriminator, the class label for each pixel is adaptively determined.

A. CICD-LRM

Most of the existing representation-based classifiers are developed based on a linear representation model, in which each pixel in HSIs is represented as a linear combination of training samples with the same label. Taking an $i$th class spectral signal $x_i \in \mathbb{R}^B$ with $B$ spectral bands for example, the linear representation model can be written as

$$x_i = E_i \alpha_i + n$$  \hspace{1cm} (1)

where $E \in \mathbb{R}^{B \times N}$ is the dictionary containing $N$ spectral signatures. Furthermore, $E$ can be decomposed into $w$ subsets according to the categories of samples, denoted as $E = [E_1, E_2, \ldots, E_w]$. $E_i \in \mathbb{R}^{B \times N_i}$ is the $i$th class dictionary with $N_i$ samples, $\alpha_i \in \mathbb{R}^{N_i}$ is the representation coefficient, and $n \in \mathbb{R}^B$ denotes the measurement error or noise vector.

Due to its simplicity and effectiveness, the linear representation model has been applied in many practical classification scenarios. However, when the number of atoms in a dictionary is very limited, the performance of these methods is greatly affected. One of the most important factors is spectral variation, which is generally caused by the inhomogeneity of imaging conditions (illumination, topographical changes, atmospheric effects, etc.). Since the spectral variation presented in an entire HSI is diverse and complicated, the diversity of spectral variation derived from limited training samples is apparently inadequate. Besides, from the perspective of signal reconstruction, the diversity of spectral variation within a training dictionary and testing samples may be unmatched. The insufficient dictionary atoms may increase the reconstruction residuals and decrease the differences between classes. In the classifier layer, these residuals can be further propagated, which greatly affects the final classification accuracy.

To handle this problem, a novel CICD-LRM is proposed, in which an observed spectrum can be written as the combination of class-independent and -dependent spectral components

$$x_i = \tilde{e}_i + \tilde{x}$$  \hspace{1cm} (2)

where $\tilde{x}$ presents the class-independent component accounting for the spectral variation, and $\tilde{e}_i$ represents the class-dependent spectral characteristics of the $i$th class obtained by

$$\tilde{e}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} e_j \quad \text{s.t.} \quad e_j \in E_i$$  \hspace{1cm} (3)

with $N_i$ being the number of training samples in the $i$th class. Thus, the class-independent component in (2) can
be rewritten as
\[ \tilde{x} = x_i - \tilde{e}_i \]  
which can be further expressed as the linear combination of a spectral variation dictionary
\[ \tilde{x} = x_i - \tilde{e}_i = V\beta_i \]  
where \( V \) and \( \beta_i \) are the spectral variation dictionary and corresponding representation coefficient vector, respectively. The signal \( x_i \) is explicitly divided into class-dependent and \( \beta_i \)-independent spectral components, in which the class-dependent component is derived from the intrinsic spectral characteristics set \( \{e_j|j \in [1,w]\} \). The class-independent spectral component is regarded as the subtle spectral differences between observed spectrum \( x_i \) and the intrinsic spectral characteristics \( \tilde{e}_i \), which can be linearly represented using spectral variation dictionary \( V \) with small residuals. The spectral differences between \( x_i \) and the intrinsic spectral characteristics of other classes \( \tilde{e}_j(j \neq i) \) would yield larger residuals. Therefore, a comprehensive spectral variation dictionary \( V \) is crucial for accurate representation and reconstruction.

B. Spectral Variation Extraction

Assume that \( X \in \mathbb{R}^{L \times H \times B} \) is an HSI with \( B \) spectral bands and \( L \times H \) pixels that can be classified into \( w \) categories except for the background. In this article, \( U \) pixels are randomly selected for each class as the training sample set \( E \in \mathbb{R}^{B \times N} \) with \( N = U \times w \). The rest of the pixels form the testing sample set, which is denoted as \( Y \in \mathbb{R}^{B \times M} \) with \( M \gg N \).

Generally, all the spectral variations among spectral signatures from the same class are caused by the inhomogeneity of imaging conditions and presented as subtle intraclass spectral differences. To extract the intraclass spectral variations and exclude interclass spectral differences, a new SVE method, that is, fused spectral variation extraction (FSVE) method, is proposed, which is implemented by fusing the local spectral variation with the global one.

1) Local Spectral Variation Extraction: A typical characteristic of HSIs is spatial correlation. Pixels in a small spatial neighborhood are likely to belong to the same class, that is, their spectra are similar. This characteristic has been widely utilized in many spatial–spectral classifiers. A local spectral variation extraction method (LSVE) is designed based on the hypothesis that all pixels in a small spatial neighborhood generally share the same class label. The scheme of the LSVE method is shown in Fig. 2.

Guided by CICD-LRM, the LSVE is designed with two key steps. First, a local class-dependent component within a small spatial neighborhood is calculated. Specifically, for each training sample \( e_i \), its spatial neighborhood \( \Omega_{\text{spatial}}(e_i) \) is defined as a \( q \times q \) square window centered at \( e_i \):
\[ \Omega_{\text{spatial}}(e_i) = \left\{ e(s, t) \mid s \in [s_i - h, s_i + h], t \in [t_i - h, t_i + h] \right\} \]  
where \((s, t)\) is the coordinates of \( e_i \) and \( h = (q - 1)/2 \). An average operation conducted in this square window generates a local class-dependent component
\[ \Omega_{\text{spatial}}(e_i) = \frac{1}{q^2} \sum_{s = s_i - h}^{s_i + h} \sum_{t = t_i - h}^{t_i + h} e(s, t). \]  

The synthetic spectrum \( \Omega_{\text{spatial}}(e_i) \) is the representation of intrinsic spectral characteristics and can provide more useful information by introducing spatially adjacent pixels. Since pixels in a small window most likely belong to the same class, their spectra exhibit high similarity. The subtle difference between \( \Omega_{\text{spatial}} \) and \( e_i \) is regarded as the intraclass spectral variation
\[ v_i^L = e_i - \Omega_{\text{spatial}}(e_i). \]  

It is notable that (8) is a special case of (5), in which the weight corresponding to \( v_i^L \) is 1. It is an approximate description of the subtle spectral difference within a limited spatial region, which models the local imaging condition differences. Thus, the local spectral variations extracted from all training samples \( E \) and the corresponding spatial neighborhood are gathered to build local spectral variation dictionary \( V^L = \{v_i^j | j \in [1, N]\} \). Note that the class labels of training samples are ignored in the construction of \( V^L \), and therefore, the proposed LSVE method is actually an unsupervised one.

2) Global Spectral Variation Extraction: With available training samples, another extraction method of spectral variation, the global spectral variation extraction (GSVE) method is designed to model the global imaging condition differences in a supervised manner, whose scheme is shown in Fig. 3.

In GSVE, with randomly selected training samples, spectral variation is extracted from those belonging to the same class,
which reflects the difference in their acquisition conditions over the entire scene. Similar to LSVE, the class-dependent component is first calculated, while the averaging operation is executed for each subset in a spectral neighborhood rather than a spatial one employed in LSVE. That is, the global class-dependent component of the $i$th class can be calculated as
\[
\mathbf{E}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \mathbf{e}_j^i
\]
where $\mathbf{E}_i = \{\mathbf{e}_1^i, \mathbf{e}_2^i, \ldots, \mathbf{e}_{N_i}^i\}$ is the subset of the $i$th class. The global spectral variation is then derived with CICD-LRM. For the $j$th sample in the $i$th class, its corresponding spectral variation vector is
\[
\mathbf{v}_j^i = \mathbf{e}_j^i - \mathbf{E}_i.
\]

Different from the LSVE method, it is an approximate description of the subtle spectral difference among the same class objects within the entire scene, which models the global imaging condition differences. Afterward, the global spectral variation dictionary is constructed as
\[
\mathbf{V}^G = \{\mathbf{v}_j | j \in [1, w], i \in [1, N_i]\}. \tag{11}
\]

3) Local and Global Spectral Variation Fusion: As mentioned above, local and global spectral variations are extracted based on spatial and spectral correlation within HSI, respectively. To fully utilize the prior spatial and spectral knowledge of HSI, the obtained global and local spectral variation dictionary, denoted as $\mathbf{V}^G$ and $\mathbf{V}^L$, are fused by concatenation to form
\[
\mathbf{V}^F = [\mathbf{V}^L \ \mathbf{V}^G].
\]

It is obvious that the fused spectral variation dictionary $\mathbf{V}^F$ encloses much more diverse variations, leading to enhanced representation capability.

C. Spectral Variation Augmented Classifier

As suggested in (4), the average spectrum of each class is subtracted from the observed samples $\mathbf{y}_p (p = 1, 2, \ldots, M)$ to generate multiclass spectral components
\[
\mathbf{y}_p^i = \mathbf{y}_p - \bar{\mathbf{v}}_i, \quad i = 1, 2, \ldots, w
\]
where $M$ denotes the number of testing samples. The multiclass spectral components $\mathbf{y}_p^i (i = 1, 2, \ldots, w)$ represent the differences between the testing sample $\mathbf{y}_p$ and the average spectrum of the $i$th class. The generated variation set $\mathbf{y}_p^i (i = 1, 2, \ldots, w)$ includes an intraclass spectral variation vector and $w-1$ interclass spectral variation vectors. Since the extracted spectral variation in both global and local manners is essentially spectral variation within the same class, the fused dictionary is capable of representing the spectral variation vector within the same class while incapable of representing the rest $w-1$ interclass spectral variations. Therefore, the spectral variation augmented representation is an effective way to separate the intraclass spectral variation from interclass spectral variations.

The major challenge is to estimate representation and reconstruction coefficients for the class-independent component of the observed signal. Referring to the relevant research in signal reconstruction, including CR [34], SR [33], and CRT [39], constraints can be imposed on coefficient vectors to acquire an approximation of spectral variation. Specifically, the proposed classifiers based on these three reconstruction strategies are termed as SVARC-SR, SVARC-CR, and SVARC-CRT, respectively. SVARC-SR adopts an $\ell_0$ or $\ell_1$-norm regularization, while SVARC-CR and SVARC-CRT adopt an $\ell_2$-norm regularization. SVARC-CRT can be regarded as a variant of SVARC-CR by adding Tikhonov’s regularization.

1) SVARC-SR: In the proposed SVARC-SR, the multiclass variation component $\mathbf{y}_p^i$ is regarded as a linear combination of several atoms in a spectral variation dictionary. In practice, sparse coefficient $\beta_p^i$ can be estimated by imposing an $\ell_0$-norm constraint with the following relaxed equation:
\[
\hat{\beta}_p^i = \arg \min_{\beta_p^i} \|\mathbf{y}_p^i - \mathbf{V} \beta_p^i\|_2^2 \text{ s.t. } \|\beta_p^i\|_0 \leq K \tag{13}
\]
where $\mathbf{V}$ is the spectral variation dictionary, and it can be the local version $\mathbf{V}^L$, global version $\mathbf{V}^G$, or their fused version $\mathbf{V}^F$. Here, $K$ is the sparsity level, and $\beta_p^i$ is approximately estimated with the orthogonal matching pursuit (OMP) algorithm [41] in our work.

2) SVARC-CR: SVARC-CR is designed based on the hypothesis that CR leads to more accurate reconstruction rather than competitive and SR. Hence, the $\ell_2$-norm regularization is introduced to ensure that all atoms contribute to representation, which is formed as
\[
\hat{\beta}_p^i = \arg \min_{\beta_p^i} \|\mathbf{y}_p^i - \mathbf{V} \beta_p^i\|_2^2 + \lambda_1 \|\beta_p^i\|_2^2 \tag{14}
\]
where $\lambda_1$ is the regularization parameter. $\beta_p^i$ is estimated as
\[
\hat{\beta}_p^i = (\mathbf{V}^T \mathbf{V} + \lambda_1 I)^{-1} \mathbf{V}^T \mathbf{y}_p^i. \tag{15}
\]

3) SVARC-CRT: SVARC-CRT employs Tikhonov regularization to estimate representation coefficients as
\[
\hat{\beta}_p^i = \arg \min_{\beta_p^i} \|\mathbf{y}_p^i - \mathbf{V} \beta_p^i\|_2^2 + \lambda_2 \|\mathbf{y}_p^i - \mathbf{V} \beta_p^i\|_2^2 \tag{16}
\]
where $\lambda_2$ is the regularization parameter balancing residual and regularization terms, and Tikhonov matrix $\Gamma_{\mathbf{y}, \mathbf{y}_p}$ is calculated by
\[
\Gamma_{\mathbf{y}, \mathbf{y}_p} = \begin{bmatrix} \|\mathbf{y}_p^i - \mathbf{v}_1\|_2 & 0 & 0 \\
0 & \ddots & 0 \\
0 & 0 & \|\mathbf{y}_p^i - \mathbf{v}_o\|_2 \end{bmatrix} \tag{17}
\]
with $\mathbf{v}_j (j = 1, 2, \ldots, o)$ being the $j$th atom in $\mathbf{V}$. The similarity between variation component $\mathbf{y}_p^i$ and each atom in the variation dictionary is measured, and more similar to the variation component are assigned larger weights. $\beta_p^i$ is then estimated by
\[
\hat{\beta}_p^i = (\mathbf{V}^T \mathbf{V} + \lambda_2 \left(\Gamma_{\mathbf{y}, \mathbf{y}_p}\right)^T \Gamma_{\mathbf{y}, \mathbf{y}_p})^{-1} \mathbf{V}^T \mathbf{y}_p^i. \tag{18}
\]
The label of sample \( y_p \), denoted as \( l_p \), is then determined by comparing multiclass residuals

\[
l_p = \arg\min_{i=1,2,...,N} \| y_p - V \hat{\beta}_p \|.
\]

The detailed SVARC method is depicted in Algorithm 1.

**Algorithm 1 SVARC**

**Input:** Hyperspectral image \( X \in \mathbb{R}^{L \times H \times B} \), Labeled training sample set \( E \in \mathbb{R}^{B \times N} \), Testing sample set \( Y \in \mathbb{R}^{B \times M} \).

**Output:** Predicted class label vector: \( L \).

1. Extract the local spectral variation by Eq. (6)-(8), and construct local dictionary \( V^L = \{ v^L_j \mid j \in [1, N] \} \).
2. Extract the global spectral variation by Eq. (9)-(11), and construct global dictionary \( V^G = \{ v^G_i \mid i \in [1, w], j \in [1, N] \} \).
3. Fuse the local and global spectral variation, and construct the fused variation dictionary \( V^F \).
4. Calculate the average spectrum of each class by Eq. (3).
5. For all \( y_p (p = 1, 2, \ldots, M) \in Y \) do
   6. Generate multi-class components \( y^F_p (i = 1, 2, \ldots, w) \) by Eq. (12).
   7. Calculate the multi-class coefficients \( \hat{\beta}^F_p (i = 1, 2, \ldots, w) \) with respect to \( V = V^F \) for \( y^F_p \) reconstruction. Specifically, coefficients acquired with SVARC-SR, SVARC-CR or SVARC-CRT are estimated by solving Eq. (13), (15), and (18), respectively.
   8. Predict the class label \( l_p \) by Eq. (19).
9. End for
10. Return \( L = \{ l_p | p = 1, 2, \ldots, M \} \).

### III. Experimental Results

**A. Datasets**

The proposed SVARC is performed on two benchmark HSI datasets frequently used for classification: Pavia Center (PaviaC) and University of Houston (UH) [42], [43].

1. **PaviaC**: PaviaC is captured by the reflective optics system imaging spectrometer (ROSIS) sensor in Italy. It contains 102 spectral bands and 1096 × 1096 pixels in each band with a spatial resolution of 1.3 meters and spectral resolution of 4 nm. Part of the scene containing no information is normally discarded in practical analysis, producing 1096 × 715 remained pixels. The corresponding ground truth with nine land cover classes of interest and false color image is shown in Fig. 4.

   ![Fig. 4. PaviaC. (a) False-color image (R-G-B bands: 59-29-14). (b) Ground-truth map. (c) Classes by colors.](image)

2. **UH**: UH is collected by the ITRES CASI-1500 sensor covering UH campus and the neighboring region. It consists of 144 bands ranging from 380 nm to 1050 nm and 349 × 1905 pixels in each band with a geometric resolution of 2.5 m. Different from other datasets, the disjoint training and testing samples of UH are directly provided to users, containing 2834 and 12197 pixels, respectively, which are divided into 15 categories. The false-color image of UH and the ground truth of disjoint training and testing samples are shown in Fig. 5.

   ![Fig. 5. UH. (a) False-color image (R-G-B bands: 55-35-14). (b) Disjoint training sample set. (c) Ground truth of the testing sample set. (d) Classes by colors.](image)

**B. Parameter Tuning**

In the proposed SVARC, there are some parameters impacting SVE and reconstruction performance and furthermore affecting classification performance. Specifically, the parameters including the spatial window size \( q \) in LSVE, sparsity level \( K \) in SVARC-SR, and tradeoff regularization \( \lambda_1 \) and \( \lambda_2 \) in SVARC-CR and SVARC-CRT are discussed with the following experiments.

Experiments are conducted to study classification performance with different values for parameters \( q \) and \( K/\lambda_1/\lambda_2 \), employing the overall accuracy (OA), average accuracy (AA), and kappa coefficient (Kappa) as evaluation measurements. Five pixels are randomly selected as training samples for each class, and the rest of the pixels serve as testing samples. In the experiments, the window size \( q \) is set to be \( \{3, 5, 7, 9, 11\} \) and \( K \) is set to be \( \{5, 7, 9, 11, 13\} \). The tradeoff regularization parameters \( \lambda_1 \) and \( \lambda_2 \) are set to be \( \{0.001, 0.01, 0.1, 1, 10\} \).
The regularization parameters \( \lambda \) of Figs. 7 and 8 depict the variations of OA over different values of \( \lambda \). Hence, it becomes more difficult to match the spectral variations within testing samples distributed through the entire HSI, which exhibit much more diversity. While a too-large \( q \) also causes problems. Pixels other than the central pixel generate not only intraclass, but also interclass spectral variations, which will saliently decrease the discriminative ability of the proposed SVARC. The experiments show consistent results. With fixed parameters, that is, \( K = 9 \) for SVARC-SR, \( \lambda_1 = 0.01 \) for SVARC-CR, and \( \lambda_2 = 0.1 \) for SVARC-CRT, OA is improved as \( q \) increases first and then degraded significantly. The peaks are presented at \( q = 3 \), \( q = 7 \), and \( q = 9 \) in SVARC-SR, SVARC-CR, and SVARC-CRT, respectively, for the PaviaC dataset. While for the UH dataset, the best results are achieved when \( q = 5 \), \( q = 3 \), and \( q = 9 \) for these three versions, respectively. In the following experiments, \( q \) is set to be 5 for all the cases without losing generality.

Fig. 6 shows the performance of SVARC-SR using a fused dictionary with different parameters. It reveals the trends of OA with \( q \) and \( K/\lambda_1/\lambda_2 \). The variations of OA with \( q \) are exhibited in Figs. 6–8 with denoted optimal results. Figs. 7 and 8 depict the variations of OA over different values of \( \lambda_1 \) and \( \lambda_2 \). With a fixed \( q \), there is basically an “inverse u” relation between OA and \( \lambda_1 \) and \( \lambda_2 \). That is, the OA significantly increases and then decreases as \( \lambda_1 \) and \( \lambda_2 \) increases. The peaks appear around \( \lambda_1 = 0.01 \) and \( \lambda_2 = 0.001 \) for the PaviaC and UH datasets, respectively. As for \( \lambda_2 \), the optimal OA is attained with \( \lambda_2 = 0.1 \) on both datasets. Therefore, in the following experiments, \( \lambda_1 \) is set to be 0.01 in SVARC-CR, and \( \lambda_2 \) is set to be 0.1 in SVARC-CRT for both datasets.

From Figs. 6–8, it can be observed that it is quite challenging to set a specific \( q \) so that all the proposed reconstruction strategy-based classifiers (including SVARC-SR, SVARC-CR, and SVARC-CRT) can achieve the best performance for both datasets. When \( q \) is too small, the enclosed pixels exhibit limited diversity and complexity in spectral variations because of highly similar imaging conditions. Therefore, it becomes more difficult to match the

\[ \text{TABLE I} \]

<table>
<thead>
<tr>
<th></th>
<th>SVARC-SR</th>
<th>SVARC-CR</th>
<th>SVARC-CRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>PaviaC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fused dictionary</td>
<td>91.15</td>
<td>94.53</td>
<td>94.88</td>
</tr>
<tr>
<td>Local dictionary</td>
<td>90.47</td>
<td>94.3</td>
<td>91.7</td>
</tr>
<tr>
<td>Global dictionary</td>
<td>92.1</td>
<td>92.68</td>
<td>94.69</td>
</tr>
<tr>
<td>UH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fused dictionary</td>
<td>64.53</td>
<td>67.51</td>
<td>67.06</td>
</tr>
<tr>
<td>Local dictionary</td>
<td>66.05</td>
<td>64.68</td>
<td>63.23</td>
</tr>
<tr>
<td>Global dictionary</td>
<td>61.74</td>
<td>66.71</td>
<td>66.36</td>
</tr>
</tbody>
</table>

In the following experiments, the parameters \( q, K, \lambda_1 \), and \( \lambda_2 \) are set to be 5, 9, 0.01, and 0.1, respectively. It is notable that these parameters are selected because they are moderate ones for the purposes of generality.

C. Component Analysis

As mentioned above, there are two important components in the proposed SVARC methods: spectral variation dictionary and spectral variation reconstruction strategy. In this section, an ablation study is employed to investigate their contributions. The experimental results of the ablation study are listed in Table I.

It can be observed that in SVARC-SR, the fused spectral variation dictionary not always outperforms the local or global one. This can be attributed to the fixed sparsity level \( K \) that determines how many spectral variation atoms are allowed to participate in the reconstruction. Even more diverse spectral
variation atoms are contained in the fused dictionary, only \( K \) competitive ones contribute to the reconstruction. Under this circumstance, the poor utilization of spectral variations in SVARC-SR may lead to limited improvement in classification accuracy. When the spectral variation reconstruction strategy is fixed, the best performance is achieved by using a fused spectral variation dictionary in most cases (for SVARC-CR and SVARC-CRT), which illustrates the advantages of the fused spectral variation dictionary over the local and global ones.

The classification performance of CRCs (SVARC-CR and SVARC-CRT) and SR-based one (SVARC-SR) employing the fused spectral variation dictionary \( V^F \) with different numbers of training samples are further investigated. The results are depicted in Fig. 9. Theoretically, the CRCs (SVARC-CR and SVARC-CRT) are designed to utilize all spectral variations in \( V^F \) to reconstruct the class-independent spectral components. Thus, the information enclosed in the fused spectral variation dictionary is completely explored and utilized, leading to better reconstruction and classification accuracy. While the performance of SRC (SVARC-SR) is influenced by the limitation of information exploitation. It can also be observed in Fig. 9 that the CRCs always outperform the SRC with the same fused spectral variation dictionary. The experimental results again demonstrate that the competitive manner is more appropriate for the representation of class-independent spectral components using the fused spectral variation atoms.

**D. Comparison Experiment**

In this section, the proposed SVARC method is compared with some state-of-the-art methods, including SRC, CRC, CRT [39], SVM, discriminative marginalized least-squares regression (DMLSR) [44], interclass sparsity based discriminative least-squares regression (ICS-DLSR) [45], SaCRT [40], and extended CR-based classifier (ECRC) [46]. SRC and CRC are baselines for comparison in many dictionary representation-based HIC approaches, and CRT, DMLSR, ICS-DLSR, and SaCRT are their recently proposed variants. Especially, ECRC is the latest work of our research team, which is also a residual-based classification method proposed for the purpose of reducing the impact of spectral variations on classification accuracy. Three main differences lie between the newly proposed SVARC and ECRC, including the employed representation model, SVE method, and constraint imposed on reconstruction coefficients. To illustrate the advancement of SVARC, ECRC is also included as one of the comparison methods. For the compared methods, parameters are set basically following the suggestions in the original article. However, there are also some exceptions. To make a fair comparison, the sparsity \( K \) in both SRC and SVARC-SR are set to 9, and the tradeoff parameters \( \lambda \) in CRC and CRT are also set to 0.01 and 0.1, respectively, as the ones employed in SVARC-CR and SVARC-CRT.

1) **Quantitative Comparison:** The classification accuracy of different methods with the two datasets is reported in Tables II
TABLE III

CLASSIFICATION ACCURACIES (%) OF DIFFERENT METHODS OVER THE UH DATASET

<table>
<thead>
<tr>
<th>class</th>
<th>Train/Test</th>
<th>SRC</th>
<th>CRC</th>
<th>CRT</th>
<th>SVM</th>
<th>DMLSR</th>
<th>ICS-DMLSR</th>
<th>SaCRT</th>
<th>ECRC</th>
<th>SVARC-SR</th>
<th>SVARC-CR</th>
<th>SVARC-CRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grass Healthy</td>
<td>5/1053</td>
<td>77.21</td>
<td>74.07</td>
<td>80.72</td>
<td>81.67</td>
<td>80.15</td>
<td>77.40</td>
<td>79.68</td>
<td>92.97</td>
<td>78.25</td>
<td>75.26</td>
<td>82.24</td>
</tr>
<tr>
<td>Grass Stressed</td>
<td>5/1064</td>
<td>85.06</td>
<td>91.54</td>
<td>84.40</td>
<td>81.58</td>
<td>82.99</td>
<td>76.69</td>
<td>74.81</td>
<td>85.90</td>
<td>83.27</td>
<td>83.65</td>
<td>83.18</td>
</tr>
<tr>
<td>Grass Synthetic</td>
<td>5/1055</td>
<td>84.55</td>
<td>99.80</td>
<td>99.01</td>
<td>98.81</td>
<td>99.41</td>
<td>99.60</td>
<td>99.21</td>
<td>99.80</td>
<td>100.00</td>
<td>100.00</td>
<td>99.60</td>
</tr>
<tr>
<td>Tree</td>
<td>5/1050</td>
<td>34.56</td>
<td>71.69</td>
<td>95.27</td>
<td>88.54</td>
<td>66.95</td>
<td>93.66</td>
<td>68.94</td>
<td>76.04</td>
<td>79.45</td>
<td>82.39</td>
<td>88.07</td>
</tr>
<tr>
<td>Soil</td>
<td>5/1056</td>
<td>88.45</td>
<td>99.91</td>
<td>99.43</td>
<td>97.16</td>
<td>98.30</td>
<td>97.35</td>
<td>91.10</td>
<td>98.48</td>
<td>85.51</td>
<td>89.77</td>
<td>83.90</td>
</tr>
<tr>
<td>Water</td>
<td>5/143</td>
<td>80.42</td>
<td>79.02</td>
<td>99.30</td>
<td>76.22</td>
<td>81.82</td>
<td>82.52</td>
<td>86.71</td>
<td>83.92</td>
<td>92.31</td>
<td>81.12</td>
<td>95.10</td>
</tr>
<tr>
<td>Residential</td>
<td>5/1076</td>
<td>7.93</td>
<td>11.66</td>
<td>52.33</td>
<td>29.85</td>
<td>31.72</td>
<td>56.34</td>
<td>33.02</td>
<td>41.98</td>
<td>71.92</td>
<td>48.60</td>
<td>67.63</td>
</tr>
<tr>
<td>Commercial</td>
<td>5/1053</td>
<td><strong>78.92</strong></td>
<td>39.41</td>
<td>27.54</td>
<td>29.72</td>
<td>29.06</td>
<td>16.90</td>
<td>55.75</td>
<td>49.10</td>
<td>41.03</td>
<td>59.45</td>
<td>26.69</td>
</tr>
<tr>
<td>Road</td>
<td>5/1059</td>
<td>8.88</td>
<td>17.66</td>
<td>48.06</td>
<td>67.89</td>
<td>60.06</td>
<td>37.58</td>
<td>42.78</td>
<td>57.98</td>
<td>56.66</td>
<td>44.00</td>
<td><strong>71.95</strong></td>
</tr>
<tr>
<td>Highway</td>
<td>5/1036</td>
<td>10.91</td>
<td>30.50</td>
<td>39.48</td>
<td><strong>44.98</strong></td>
<td>34.65</td>
<td>30.41</td>
<td>32.82</td>
<td>29.73</td>
<td>28.76</td>
<td>28.19</td>
<td>28.57</td>
</tr>
<tr>
<td>Railway</td>
<td>5/1054</td>
<td>1.90</td>
<td>9.30</td>
<td>64.14</td>
<td>59.77</td>
<td>55.22</td>
<td>50.95</td>
<td>45.64</td>
<td>51.23</td>
<td>55.31</td>
<td><strong>77.13</strong></td>
<td>76.09</td>
</tr>
<tr>
<td>Parking Lot1</td>
<td>5/1041</td>
<td>0.19</td>
<td>57.16</td>
<td>24.59</td>
<td>22.48</td>
<td>52.45</td>
<td>39.87</td>
<td><strong>66.47</strong></td>
<td>45.34</td>
<td>33.14</td>
<td>62.06</td>
<td>34.77</td>
</tr>
<tr>
<td>Tennis Court</td>
<td>5/247</td>
<td>75.30</td>
<td>98.38</td>
<td><strong>99.60</strong></td>
<td>99.19</td>
<td>93.12</td>
<td>98.79</td>
<td>91.09</td>
<td>93.52</td>
<td>98.95</td>
<td>97.76</td>
<td>95.95</td>
</tr>
</tbody>
</table>

OA - 44.41 55.67 64.97 63.35 62.18 61.19 62.55 65.87 64.83 67.51 67.06
AA - 50.35 62.36 68.99 66.05 65.03 66.03 66.36 68.87 68.70 69.16 69.79
Kappa - 40.13 52.30 62.09 60.55 59.10 58.43 59.49 62.96 61.84 64.70 64.29

Fig. 10. Classification maps of different methods over the PC dataset. (a) SRC (56.88 %). (b) CRC (82.45 %). (c) CRT (89.32 %). (d) SVM (87.73 %). (e) DMLSR (87.01 %). (f) ICS-DMLSR (88.41 %). (g) SaCRT (56.48 %). (h) ECRC (89.86 %). (i) SVARC-SR (91.15 %). (j) SVARC-CR (94.53 %). (k) SVARC-CRT (94.88 %). (l) Ground truth.

and III, in which the optimal and suboptimal results are highlighted in bold and italics, respectively.

The proposed SVARC methods (including SVARC-SR, SVARC-CR, and SVARC-CRT) usually achieve the best OA, AA, and Kappa over the PaviaC and UH datasets (except SVARC-SR over the UH dataset). Specifically, the OA of SVARC-CRT over the PaviaC dataset is improved by 5.02% and 38.4%, respectively, than the second-best (obtained by ECRC method) and worst (obtained by SaCRT method) results, respectively. For the UH dataset, the OA of SVARC-CR boosts up to 11.84% than the CRC method and 1.64% than the second-best method (ECRC). It demonstrates that the proposed SVARC method is capable of producing much more competitive HSI classification results.

The classification accuracy of some specific classes (including Water, Bitumen, etc. in PaviaC, and Grass Synthetic, Railway, etc. in the UH dataset) produced by the proposed SVARC method saliently outperforms other methods. And SVARC also generates competitive classification accuracy in other classes. As a result, the proposed SVARC method can produce the optimal AA in most cases, indicating its excellent robustness in different classes.
The performance of SVARC on the PaviaC dataset is better than that on the UH dataset, which is mainly due to the characteristics of datasets, especially the spatial distribution and spectral similarity. The comparatively smaller interclass differences in the UH dataset rise challenges for HIC. Meanwhile, in the UH dataset, ground objects are generally sparsely distributed in different regions. That is, diverse ground objects may appear in a limited spatial neighborhood, which conflicts with the application assumption of the proposed LSVE method. These two factors affect the classification performance of the proposed SVARC. Therefore, less classification performance improvement is observed in experiments on the UH dataset than that on the PaviaC dataset.

It can also be observed that the CR-based SVARC methods (SVARC-CR and SVARC-CRT) always outperform the SR-based (SVARC-SR) one. It can be attributed that the extracted spectral variation atoms are more likely distributed in the same feature subspace, and thus the collaborative rather than competitive manner is more appropriate for spectral variation representation and reconstruction.

2) Qualitative Comparison: The corresponding classification maps are depicted in Figs. 10 and 11. The three specifically designed modules enclosed in the proposed SVARC (the CICD-LRM model, the SVE method, and the spectral variation augmented classifier) jointly improve discrimination of different materials or objects, yielding the least misclassification, which is consistent with the quantitative analysis.

IV. CONCLUSION

In this article, a novel SVARC for HSIs with few labeled training samples is proposed. The SVARC consists of a CICD-LRM model, specifically designed SVE, and a spectral variation augmented classifier. It takes the presentation of subtle intraclass spectral variations into consideration. Guided by this model, two SVE methods, for the purpose of extracting intraclass and excluding interclass spectral differences, are proposed to construct a fused spectral variation dictionary with more diversity. Furthermore, three spectral variation augmented classifiers are designed, which are capable of discriminating and representing exclusive intraclass spectral variation within each testing sample, so that the label of the testing sample can be accurately decided. Apparently, they are beneficial to and cooperative with each other, and jointly improve discrimination of different materials or objects. The experimental results conducted on the PaviaC and UH datasets demonstrate the effectiveness and superiority of the proposed SVARC methods (especially SVARC-CR and SVARC-CRT) in the classification task with few labeled training samples.

REFERENCES


Yifan Zhang (Member, IEEE) received the B.S. degree in electronics and information technology and the M.S. and Ph.D. degrees in signal and information processing from Northwestern Polytechnical University, Xi’an, China, in 2001, 2004, and 2007, respectively.

From 2007 to 2010, she worked as a Post-Doctoral Researcher with the Vision Laboratory, Department of Physics, University of Antwerp, Antwerp, Belgium. She is currently an Associate Professor with the School of Electronics and Information, Northwestern Polytechnical University of China. Her research interests include hyperspectral image analysis, image fusion, and image restoration.

Shaohui Mei (Senior Member, IEEE) received the B.S. degree in electronics and information engineering and the Ph.D. degree in signal and information processing from Northwestern Polytechnical University, Xi’an, China, in 2005 and 2011, respectively.

He was a Visiting Student with the University of Sydney, Camperdown, NSW, Australia, from October 2007 to October 2008. He is an Associated Professor with the School of Electronics and Information, Northwestern Polytechnical University. His research interests include hyperspectral remote sensing image processing and applications, intelligent signal and information acquisition and processing, video processing, and pattern recognition.

Dr. Mei was a recipient of the Excellent Doctoral Dissertation Award of Shaanxi Province in 2014 and the Best Paper Award of the IEEE International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS) 2017. He served as the Registration Chair of the IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP) 2014. He is currently serving as the Topical Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (IEEE TGRS), an Associated Editor of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (IEEE JSTARS), the reviewer of more than 20 international famous academic journals, and was awarded the Best Reviewer of the IEEE JSTARS in 2019.

Ge Zhang (Graduate Student Member, IEEE) received the B.S. degree in electronics and information engineering from Northwestern Polytechnical University, Xi’an, China, in 2018, where is currently pursuing the Ph.D. degree in communication engineering with the School of Electronics and Information.

Her research interests include hyperspectral image unmixing and signal processing.

Yan Feng (Member, IEEE) received the B.S. degree in electronics engineering and the M.S. and Ph.D. degrees in signal and information processing from Northwestern Polytechnical University, Xi’an, China, in 1984, 1989, and 2007, respectively.

She is currently a Professor with the School of Electronics and Information, Northwestern Polytechnical University. Her research interests include image processing, compression and classification of hyperspectral images, compressed sensing theory, and intelligent information processing.

Qian Du (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Maryland at Baltimore County, Baltimore, MD, USA, in 2000.

She is a Bobby Shackouls Professor with the Department of Electrical and Computer Engineering, Mississippi State University, MS, USA. Her research interests include hyperspectral remote sensing image analysis and applications, pattern classification, data compression, and neural networks.

Dr. Du is a Fellow of the International Society for Optics and Photonics (SPIE). She was a recipient of the 2010 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society (GRSS). She was a Co-Chair for the Data Fusion Technical Committee of the IEEE Geoscience and Remote Sensing Society from 2009 to 2013 and the Chair for Remote Sensing and Mapping Technical Committee of the International Association for Pattern Recognition from 2010 to 2014. She was the General Chair for the Fourth IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing in Shanghai, China, in 2012. She served as an Associate Editor of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, the Journal of Applied Remote Sensing, and the IEEE SIGNAL PROCESSING LETTERS. Since 2016, she has been the Editor-in-Chief of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.