Extended Collaborative Representation-Based Hyperspectral Imagery Classification

Bobo Xie, Graduate Student Member, IEEE, Shaohui Mei†, Senior Member, IEEE, Ge Zhang†, Graduate Student Member, IEEE, Yifan Zhang†, Member, IEEE, Yan Feng†, Member, IEEE, and Qian Du‡, Fellow, IEEE

Abstract—Collaborative representation (CR) has been demonstrated to be very effective for hyperspectral image classification. However, insufficient diversity of training samples often results in limited classification accuracy under small-training-sample conditions, especially when diverse spectral variation is presented in testing samples. In order to alleviate such a problem, a spectral variation augmented-based linear mixed model (SV-LMM) is proposed, in which the spectral variation is extracted by conducting singular value decomposition (SVD) over training samples. Such spectral variation is further utilized to extend the CR for hyperspectral classification. Experiments over two benchmark datasets, i.e., the Pavia Center dataset and the University of Houston dataset, demonstrate that the proposed extended CR-based classifier (ECRC) clearly improves the performance of conventional CRC for hyperspectral classification and outperforms several state-of-the-art algorithms.

Index Terms—Extended collaborative representation (CR), hyperspectral classification, limited training samples, spectral variation.

I. INTRODUCTION

HYPERSPECTRAL remote sensing is an image-spectrum merging technology, in which the spectral dimension of each pixel is an approximately continuous curve to reflect abundant spectral information of the objects. Such an advantage over conventional remote sensing images can provide richer and more discriminative information over different materials. Over the past decades, hyperspectral remote sensing image processing has developed into a current research hotspot in the field of information acquisition and processing. It has actively been researched for many fields of applications, such as object classification, environmental monitoring, resource exploration, and disaster prevention and mitigation industries [1]–[3].

Affected by the inhomogeneity of imaging conditions (illumination, topographical changes, atmospheric effects, and so on) [4], even pixels composed of exactly the same materials and acquired by the same sensors may present certain differences among their spectral characteristics. Such spectral variation has caused the situation that the same object with nonidentical spectra and different objects with similar spectrum appeared in a hyperspectral image. From the perspective of spectral characteristic-based classification, these two phenomena will increase within-class diversity and between-class similarity. Apparently, it will seriously affect the accuracy of ground object classification and recognition [5].

Dictionary-based representation has shown satisfying performance in reducing the negative impact of spectral variation. Many dictionary-based classifiers have been proposed in the past few decades, such as sparse representation-based classification (SRC) [6], collaborative representation (CR)-based classification (CRC) [7], and their varied versions [8]–[10]. Furthermore, with integrated spatial and spectral information, many novel joint or structured sparse representation (SR) or CR models have been also proposed [11]–[15]. For example, Li et al. [15] proposed joint sparse representation for superpixel-level hyperspectral image classification. Ma et al. [16] proposed a similarity-based block sparse subset selection (SB2S3) model for video summarization and gained superior performance. Yu et al. [17] proposed global spatial and local spectral similarity-based manifold learning group sparse representation for hyperspectral imagery classification and gained good performance. Li et al. [18] proposed a collaborative-representation-based nearest neighbor classifier, which is based on CR computed by an $\ell_2$-norm minimization with a Tikhonov regularization matrix. Besides, the kernel trick is also applied to CR-based classifier to solve some physical nonlinear problems [19]–[22].

Though the influence of spectral variation has been decreased to some extent by using the spectral dictionary to accommodate variation among spectra, the classification performance may also degrade when the number of atoms in the dictionary becomes small. This is because the spectral variation distributed throughout the image scene is usually complicated. When training samples are limited, the diversity of within-class spectra may be insufficient to represent variations in testing samples. Therefore, how to increase the within-class variation is a key issue to improve the performance of dictionary-based classification under the small-training-sample situation.

In order to address the aforementioned issue, the within-class diversity is further increased using spectral variation within other classes. Specifically, a singular value decomposition (SVD)-based spectral variation separation strategy is designed to extract spectral variation atoms from an end-member dictionary. Such spectral variation is further used
III. PROPOSED ECR FOR HSI CLASSIFICATION

Fig. 1 presents the whole framework of the proposed extended CR-based classifier (ECRC) for hyperspectral classification, which is mainly composed of an improved version of the representation model, the extraction of spectral variation, and the extended version of the classifier.

A. Spectral Variation Augmented Representation Model

Due to simplicity and flexibility, the linear mixture model (LMM) has become a widely used tool to represent spectral signatures. For a given hyperspectral image $X \in \mathbb{R}^{L \times W \times B}$ with $L \times W$ pixels over $B$ bands, the observed spectrum vector denoted as $x \in \mathbb{R}^B$ is a linear combination of endmembers, which can be represented as

$$x = E\alpha + n$$

where $E \in \mathbb{R}^{B \times N}$ is the endmember dictionary containing $N$ spectral signatures, $\alpha \in \mathbb{R}^N$ is the representation coefficient, and $n \in \mathbb{R}^B$ denotes the measurement error.

Generally, the LMM is an excellent representation model, which can explain the spectral response very well in HSI. Some linear classifiers based on LMM have been widely used in hyperspectral classification, such as SRC and CRC. However, under the small-training-sample condition, the diversity of training samples is inadequate to well match one of the testing samples. The immediate consequence of this incident is that the representation error increases significantly, which seriously influences the classification performance.

To increase the diversity of training samples, the simplest solution is to add more training samples. However, it is impractical in many situations. The spectral variation, caused by the external atmosphere or internal sensors and so on, maybe evenly distributed throughout the scene and share similarities between different classes. Therefore, by further considering such additional spectral variation in the LMM, we proposed a novel spectral variation augmented LMM (SV-LMM) to resolve possible spectral drift between spectral dictionary and testing samples. Specifically, in the proposed representation model, each pixel can be written as

$$x = E\alpha + V\beta + n$$

where $V \in \mathbb{R}^{B \times M}$ denotes the spectral variation dictionary and $\beta \in \mathbb{R}^M$ represents its corresponding weighting coefficients.

B. Extraction of Spectral Variation

In remote sensing, the spectra of ground objects are inevitably affected by the imaging environment. As a result, spectral variation is usually presented in the spectra even with the same label. When SVD is applied to spectral signatures, Boardman [23] proposed the hypothesis that the smaller singular value and the corresponding singular vector mainly reflect the spectral variation and the measurement error, while the larger singular values and their singular vectors can explain the intrinsic spectra of different ground objects. Such a hypothesis has also been demonstrated to be very effective in determining the purity of spatial neighborhood in endmember extraction [24]. Inspired by this theory, we proposed a spectral variation extraction method using SVD, as shown in Fig. 2.

Let the $i$th class dictionary be denoted by $E_i = [e_{i1} e_{i2} \ldots e_{iN}] (i = 1, 2, \ldots, w)$, in which $N_i$ and $w$ represent the numbers of training samples and classes, respectively, and $e_{ij} \in \mathbb{R}^B (j = 1, 2, \ldots, N_i)$ represents the $j$th spectral characteristics. As shown in Fig. 2, there are two simple steps to extract spectral variation for each class, including SVD-based reconstruction and spectral variation estimation.

Fig. 1. Framework of the proposed algorithm.

Fig. 2. Framework of extracting spectral variation dictionary for each class using SVD.
1) SVD-Based Reconstruction: The dictionary of a certain class can be decomposed by SVD as

$$E_i = U_i \sum_i V_i^T, \quad U_i \in \mathbb{R}^{B \times B}, \quad V_i \in \mathbb{R}^{N_i \times N_i}$$

where $U_i$ and $V_i$ are left and right singular matrices, whose column vectors are eigenvectors of $E_iE_i^T$ and $E_i^TE_i$, respectively. As the column index increases, the corresponding eigenvalue will reduce. $\sum_i \in \mathbb{R}^{B \times B}$ is a square matrix with only the diagonal elements being nonzero, which can be described as

$$\sum_i (j, j) = \lambda_i^j \quad (j = 1, 2, \ldots, N_i)$$

where the singular value $\lambda_i^j$ is sorted in a descending order by computing the square of the $j$th larger eigenvalue of $E_i^TE_i$.

In order to extract the intrinsic spectra within a specific class, an information ratio threshold $\eta$ is preset to control the degree of approximation as

$$\sum_{i=1}^{K_i} \lambda_i^1 / \sum_{i=1}^{N_i} \lambda_i^1 \geq \eta. \quad (5)$$

Once the proportion of the first $K_i$ largest singular values is no less than $\eta$, the local diagonal matrix can be composed of the selected singular values

$$\sum_{i=1}^{K_i} = \text{Diag}(\lambda_i^1, \lambda_i^2, \ldots, \lambda_i^{K_i}). \quad (6)$$

Correspondingly, the first $K_i$ column vectors in the left and right singular matrices are also used for the subset $U_i^{K_i} \in \mathbb{R}^{B \times K_i}$ and $V_i^{K_i} \in \mathbb{R}^{N_i \times K_i}$. Afterward, the intrinsic spectral dictionary of a certain class can be approximately estimated by

$$\hat{E}_i = U_i^{K_i} \sum_{i=1}^{K_i} (V_i^{K_i})^T. \quad (7)$$

2) Spectral Variation Estimation: The observed spectra are composed of the intrinsic components of the characteristic spectra and the spectral variations. In this letter, the reconstructed dictionary in (7) is regarded as the intrinsic components, and consequently, the spectral variation within each class can be calculated by

$$V_i = E_i - \hat{E}_i. \quad (8)$$

The spectral variation dictionary mentioned later is actually a set of the spectral variation within all of the classes, which is denoted as $V = \{V_1, V_2, \ldots, V_w\}$.

C. ECRC for Hyperspectral Classification

In the proposed ECRC, the CR is imposed on both endmember representation and spectral variation representation, which is realized by imposing $\ell_2$-norm-based constraint over their corresponding representation coefficients as follows:

$$\varphi_i = \arg \min \varphi_i \| x - G_i\varphi_i \|_2^2 + \lambda \| \varphi_i \|_2^2 \quad (9)$$

where $\lambda$ is regularization parameter, $G_i = [E_i, V]$ is the concatenation of the spectral dictionary of $i$th class and variation dictionary of all classes, whose corresponding weight vector is $\varphi_i$, and it can be calculated by

$$\varphi_i = (G_i^TG_i + \lambda I)^{-1}(G_i)^T x. \quad (10)$$

Finally, the category labels are assigned to the signal by solving reconstruction residuals as

$$\text{class}(x) = \arg \min_{i} r_i(x)$$

s.t. $r_i(x) = \| x - G_i\varphi_i \|_2 \quad (i = 1, 2, \ldots, w). \quad (11)$

III. EXPERIMENTAL RESULTS

A. Experiment Settings

Two benchmark hyperspectral classification datasets, i.e., the Pavia Center (PaviaC) and the University of Houston (UH), are selected. There are nine and 15 land cover classes of interests in PaviaC and UH datasets, which totally provided 7456 and 15031 (2834 training pixels and 12197 testing pixels) labeled pixels, respectively. We first fixed the training samples for each class as 6 in these two datasets to analyze the influence of regularization parameter $\lambda$ and information ratio threshold $\eta$ over the classification performance. In practice, the parameter $\lambda$ is respectively set as 0.1, 0.01, and 0.001, while $\eta$ ranges from 0.85 to 0.99 with the step of 0.02. The corresponding classification performance with these parameter settings is plotted in Fig. 3, in which the performance is evaluated by three general criteria, including the overall accuracy (OA), the average accuracy (AA), and the kappa coefficient.

Observed from Fig. 3, it can be easily found that the evaluation values in both PaviaC and UH datasets increase when $\eta$ increases at first. After reaching a peak, the values of OA, AA, and kappa gradually decrease. It is worth noting that the values are basically at the peak when $\eta$ is fixed as 0.95 and 0.97, respectively, in PaviaC and UH datasets. Generally speaking, the influence trends of the $\eta$ on the class performance over PaviaC and UH are consistent. However, the situation of $\lambda$
is quite different. As shown in the left column of Fig. 3, the performance is proportionate to the value of $\lambda$ over the PaviaC dataset, whereas they are inversely proportionate over the UH dataset. It is obvious that there are no parameters $\lambda$ and $\eta$ to optimize the performances of both datasets. By compromising the performance over these two datasets, we set $\eta = 0.95$ and $\lambda = 0.01$ for both datasets in the subsequent experiments.

Two strategies to construct a spectral dictionary are also discussed. For the PaviaC dataset, the OA of ECRC using endmember dictionary $E$ is 89.65%, whereas the OA of ECRC using the reconstructed dictionary $\hat{E}$ is 89.62%. For the UH dataset, the OAs using these two dictionary are 64.42% and 63.34%, respectively. It is obvious that the classification performance using endmember dictionary $E$ as a dictionary is better than that using the reconstruction dictionary $\hat{E}$. Therefore, the endmember dictionary $E$ is selected as a spectral dictionary in the follow-up experiments.

### B. Experiment Results

In this subsection, we carried out several experiments to discuss the influence of the number of training samples. Specifically, we randomly selected three, six, nine, 12, and 15 labeled samples for each class in these two datasets. In the PaviaC dataset, the training samples account for about 0.36%–1.8% of total samples, and the rest pixels are used for testing. As for the UH dataset, the training samples are picked out from the provided training samples, which do not overlap with the testing samples. Thus, no matter how many training samples are selected from the UH dataset, the number of testing samples is always 12 197. Apparently, the evaluation is a typical small-training-sample situation. Note that all experimental results presented in this subsection are the average numerical accuracy of five repeated experiments.

The experimental results with different numbers of training samples over these two datasets are summarized in Tables I and II, respectively. It is observed that, for the PaviaC dataset, the performance of ECRC and CRC becomes better particularly increased by 8.4% than CRC when setting the number of training samples as 15. These obvious performance improvements of our proposed ECRC over CRC demonstrate the superiority of considering spectral variation in the LMM-based representation model.

Several classical and state-of-the-art methods are selected for further comparison, including SRC, CRC, CRT [20], SVM, DMLSR [25], ICS-DLSR [26], and SaCRT [27], to further evaluate the performance of our method. Specifically, five training samples per class are employed for classification. For SRC and CRC, the parameter $\lambda$ is set to 0.01, which is consistent with our proposed ECRC. It is different that $\lambda$ is set as $10^{-3}$ for CRT. According to [25], the parameters $\lambda_1$ and $\lambda_2$ of DMLSR are fixed as $10^{-3}$ and $10^{-2}$. For ICS-DLSR, the parameters $\lambda_1$, $\lambda_2$, and $\lambda_3$ are $10^{-1}$, $10^{-1}$, and $10^{-2}$. As suggested in [27], $\lambda_1$, $\lambda_2$, and $\lambda_3$ are set to $10^{-1}$, $10^{-2}$, and $10^{-2}$ for SaCRT.

The overall classification accuracy of these methods is reported in Fig. 4. It is observed that the proposed ECRC obviously outperforms other algorithms over these two datasets under the small-training-sample situation. This is due to the extraction of spectral variation, and the proposed extended signal representation model, including spectral variation, clearly increases the diversity of within-class training samples, while the SV-LMM alleviates the mismatch between testing and training samples caused by an inconsistent acquisition environment.
In order to further verify the performance of our proposed algorithm, we have also conducted classification experiments by further adding synthetic noises to these two datasets. Specifically, each pixel in PaviaC and UH datasets is added a random Gaussian noise with a random SNR ranging from 10 to 50 dB. The proposed ECRC algorithm, together with the seven compared algorithms, is applied to these noisy images. The classification performance in terms of OA for these methods is plotted in Fig. 5. It is observed that the proposed ECRC obviously outperforms other algorithms over the two noisy datasets, demonstrating that ECRC is capable of addressing the noisy hyperspectral images.

IV. CONCLUSION

In this letter, an ECRC method is proposed to improve the performance of the original CRC in hyperspectral image classification. Specifically, we proposed an SVD-based strategy to extract spectral variation, a spectral variation augmented LMM to increase the diversity of dictionaries, and a corresponding ECRC to improve the classification performance. Experimental results of CRC and ECRC over two benchmark datasets and their noisy datasets indicate that the spectral variation augmented dictionary is capable to represent more complicated variation among the spectra of testing and training samples and, consequently, assign more accurate labels to signatures under the small-training-sample condition. Moreover, the comparison with several state-of-the-art methods has further verified the effectiveness of the proposed ECRC.

REFERENCES


