Superpixel-Based Relaxed Collaborative Representation With Band Weighting for Hyperspectral Image Classification

Hongjun Su, Senior Member, IEEE, Yihan Gao, and Qian Du, Fellow, IEEE

Abstract—Representation learning methods, such as sparse representation (SR) and collaborative representation (CR), have been widely used in hyperspectral image classification. However, they merely considered the similarities between features. Due to the plentiful spatial and spectral information in hyperspectral images, the differences between features also need to be considered. Relaxed CR (RCR) is used in face recognition to accommodate the difference and similarity of features simultaneously. In this article, a novel method of RCR with band weighting based on superpixel segmentation is proposed for hyperspectral image classification. The $l_2$ norm on band coefficients and global average coefficients is exploited to ensure the similarity, and the variance determines the specific coefficient-related weight of each band. The training set is selected from each superpixel, which is considered as a subgraph rather than independent pixels. It is favorable for concentrating on the difference between similar bands since the samples in each superpixel are of high similarity. Furthermore, extended multilattribution profile (EMAP) features, Gabor features, and local binary pattern (LBP) features are employed to increase the diversity of features; thus, a method of multifeatures’ RCR based on superpixels is proposed. Three typical data are used to validate the related algorithms. The experiments demonstrate that the proposed algorithms can effectively improve classification accuracy compared to state-of-the-art classifiers.

Index Terms—Band weighting, hyperspectral classification, relaxed collaborative representation (RCR), superpixel segmentation.

I. INTRODUCTION

HYPER SPECTRAL remote sensing has been widely used in many aspects [1]–[4], such as mineral identification [5], water quality monitoring [6], and agricultural crop identification [7]. The classification of hyperspectral remote sensing images [8]–[12] is challenging mainly due to the Hughes phenomenon [13]. Therefore, the key problem to be solved is how to extract and utilize the various features [14]–[16] to obtain an ideal performance [17]–[21].

To solve these problems, representation learning, a parameter-free classification method that directly uses the original spectrum, has been proposed and successfully applied to various classification tasks, such as super-resolution reconstruction, data restoration, and image classification [22]. This type of method assigns the classes of testing samples based on the representation residual of a dictionary, and the key is to analyze the spectra according to certain similarity metrics. The original representative method is sparse representation classifier (SRC) [22], [23]. Li and Du [24] proposed an adaptive sparse representation method (ASRC). By applying an adaptive similarity measure between the testing samples and the labeled data, a more discriminative sparse code is then generated. Chen et al. [25] incorporated context information into the sparse representation (SR) model and proposed a joint SR classification (JSRC) based on synchronous orthogonal matching pursuit (SOMP). Peng et al. [15], [16] proposed a maximum likelihood estimation-based robust JSRC model that can effectively reduce the effect of inhomogeneous neighboring pixels or outliers and then further proposed a self-paced JSRC model that can adaptively select good pixels and effectively handle the noisy neighboring pixels or spectral bands. Zhang et al. [26] replaced the $l_1$ norm of SRC with the $l_2$ norm and proposed a collaborative representation classifier (CRC) with lower complexity but similar accuracy to SRC. Li et al. [27] calculated the similarity between the testing pixel and the atoms within the class by the distance-weighted Tikhonov regularization term, and a CRC-based nearest regular subspace (NRS) model is proposed, which considers the homogeneity of the local area. Li and Du [28] used local spatial average features to replace the spectral features in the NRS model and proposed a joint collaborative representation (JCRC) model. The representation-based classification method and its variants focus on how to emphasize the similarity between dictionaries and testing samples but ignore the differences between features, which is one of the natural characteristics and also has an impact on the classification results. Alternatively, the similarity and difference can be taken care of simultaneously in some recent methods, such as relaxed collaborative representation (RCR) [29] and region-based relaxed multiple kernel collaborative representation (R2MK) [30], which combines the RCR and kernel technique. According to RCR, a face image can be partitioned into several blocks that are represented by different features.
measure and a balance parameter are added to the objective function. Therefore, the distance between features will be regarded as the criterion to calculate the specific weight. In this method, each feature is coded on its subdictionary. To exploit the similarity among features, the variance of coding coefficients is supposed to be minimized. Due to the fact that adjacent bands are highly similar but still different in hyperspectral data, the similarity and difference can also be considered simultaneously, such as RCR, which takes both into consideration.

Several methods have been applied in exploring the difference between the original spectral features, such as band weighting. For instance, dimensionality reduction [31]–[33], such as the principal component analysis (PCA), can obtain part of the original spectrum, which has larger weights in expressing the image to reflect the hyperspectral data as fully as possible. In order to avoid discarding seemingly useless bands roughly, the band-weighting method has recently been studied [34]–[36]. Qi et al. [37] proposed two feature weighting methods: one is a support vector machine (SVM) with compactness and a separation coefficient (CSC-SVM), and the other is SVM with a similarity entropy (SE-SVM). In order to increase the separation of SVMs, the bands with lower uncertainty would be given comparatively higher weights. Imani and Ghassemian [38] proposed to measure the relative importance of each feature by using a supervised band weighting method to obtain weighted training samples. Nevertheless, these methods cannot provide flexible weight configuration, resulting in limited classification performance.

Furthermore, different extracted features (e.g., texture and morphological features) can intensify the differences between similar surface materials, which have obvious differences merely in a certain range of the original spectrum. Applying them to the existing classifiers can offer the great capability for classification. Recently, some literature focuses on the effectiveness of extracted spatial–spectral features. It is noticed that extended multiattribute profiles (EMAPs) and Gabor are regarded as effective features in classification. Mirzapour and Ghassemian [39] extracted features from segmentation-based gray-level co-occurrence matrices and fused them with global Gabor features and morphological features for SVM classification [17]. Jia et al. [40] combined the EMAP features with 3-D Gabor features by convolution. Concerning the same coding patterns for different features, Su et al. [18] construct an updated dictionary that consists of multiple features by a dictionary learning model. Compared with spectral features, the generated dictionary is more diverse and has taken the difference of features into consideration. The methods are aimed at extracting the relevant spatial–spectral features rather than a single feature but do not combine them with the spatial map to utilize the spatial structure information fully. Alternatively, image segmentation used in image processing, pattern recognition, and artificial intelligence can help to obtain spatial structure information. Recently, superpixel segmentation, such as mean shift [41], normalized cuts (Ncut) [42], and entropy rate segmentation (ERS) [43], has led to great results in many fields. It segments an image into compact and uniform irregular blocks, which provides the condition of concentrating on the slight difference between similar features. Therefore, it is desirable to combine the multiple features with superpixels to achieve better classification performance.

As mentioned earlier, representation learning reduces computational complexity due to the training process without parameters. However, representation algorithms aim to strengthen the similarity between testing samples and atoms but ignore the fact that the features used for representation are naturally different. The method of band weighting and spatial–spectral feature extraction investigates the difference but ignores the constraints on similarity. Thus, how to balance the similarity and differences between features is still an open problem.

In this article, a new RCR with band weighting based on superpixel is proposed for hyperspectral image classification. A hyperspectral image with hundreds of bands is regarded as hundreds of single-band images. Since the training set is selected in accordance with the superpixels and is constant in each superpixel, the coefficient codes of each band can be calculated band by band. Next, the global average coefficient codes are calculated by taking the image as a whole with hundreds of bands. The similarity is ensured by the $l_2$ norm constraints on band coefficients and global average coefficients. Specifically, their variance represents the importance of each band and determines the related weights. In order to validate the weights, the EMAP features, local binary pattern (LBP) features, and Gabor features are extracted to enrich the diversity of features and weights, and then, a new improved multifeatures’ RCR with the band-weighting framework is further introduced. The appropriate weights would measure the differences between adjacent bands while ensuring similarity. The proposed methods are directly applied to hyperspectral image classification for the first time and performed better results than some current classical methods. The main contributions are given as follows.

1) Different from the traditional block-based RCR that assigns weights on subblocks of a whole image for face recognition, a band-weighting relaxed collaborative algorithm improves it as a more suitable method for hyperspectral data. It pays attention to the difference between bands without any dimensionality reduction and makes the classifier more flexible.

2) Considering the weights of different features, multiple features are extracted to enrich the diversity of the coefficient-related weights that reflect the differences between bands. Meanwhile, the $l_2$ norm on band coefficients and global average coefficients ensures the similarity. Taking the similarities and differences between various features into consideration can improve the effectiveness of diverse weights more obviously.

3) Because of the high complexity of updating the coefficients and coefficient-related weights, the superpixel algorithm has been employed as a preprocessing step. Moreover, the superpixel maps in various scales can characterize the local spatial information precisely, which provides favorable conditions for classification.

The organization of the rest is given as follows. Section II reviews representation learning and image segmentation methods. Section III introduces two methods of RCR with band weighting based on superpixels. The experimental results to evaluate the two proposed algorithms are given in Section IV.
Finally, some conclusions and future work are presented in Section V.

II. RELATED WORK

A. Superpixel Algorithm

As a typical method in clustering analysis, a superpixel algorithm has led to great results in many fields, such as feature extraction [44], target detection [45], image segmentation [46], [47], object tracking [48], and endmember detection [49]. Many superpixel algorithms, such as simple linear iterative extraction [44], target detection [45], and image segmentation [46], [47], have been proposed.

In this article, ERS is adopted, which uses the greedy algorithm [52] as a heuristic search strategy, the random walk model [53], [54], and the entropy rate to obtain homogeneous clusters. It can segment an image into adaptive superpixel blocks with multiple shapes and sizes according to its own distribution characteristics. The produced subgraphs have the following characteristics, including similarity, homogeneity, and adaptability. Fig. 1 shows the basic principle of the ERS method.

The original image is regarded as a collection of vertices and edges $G = (V, E)$, and the segmented image is $G = (V, A)$. Each edge should be given a certain weight $\omega_{i,j}$, which can be defined as

$$\omega_{i,j} = \exp\left\{ -\frac{d(v_i, v_j)^2}{2\sigma^2} \right\}$$

(1)

$$d(v_i, v_j) = d_1 \ast \arccos\left( \frac{v_i^T v_j}{||v_i||_2 ||v_j||_2} \right)$$

(2)

where $d(v_i, v_j)$ is the product of the Euclidean distance over spatial coordinates and the spectral angular distance that represents the intensity difference. Assume that $X = \{X_t | t \in \mathcal{Z}, X_t \in \mathcal{U}\}$ is a random walk on the graph, and the transition probability is $p_{i,j} = P_r(X_{t+1} = v_j | X_t = v_i) = (\omega_{i,j})/\omega_i$, where $\omega_i = \sum_{q,i,q \in E} \omega_{i,q}$. The expression of a stationary distribution is

$$\rho = (\rho_1, \rho_2, \ldots, \rho_n) = \left( \frac{\omega_1}{\omega_z}, \frac{\omega_2}{\omega_z}, \ldots, \frac{\omega_n}{\omega_z} \right)$$

(3)

where $\omega_z = \sum_{i=1}^{n} \omega_i$ is the normalization constant. The entropy rate of a random walk can be calculated as

$$H(X) = H(X|X_1) = \sum_i \rho_i H(X|X_1 = v_i) = -\sum_i \rho_i \sum_j p_{i,j} \log p_{i,j}$$

and correspondsingly, the entropy rate of the graph $G = (V, A)$ is shown as

$$H(A) = -\sum_i \rho_i \sum_j p_{i,j}(A) \log p_{i,j}(A).$$

(5)

Transition probabilities can be computed by

$$p_{i,j} = (A) = \begin{cases} e_j, & i \neq j, e_{i,j} \in A \\ 0, & i \neq j, e_{i,j} \in A \\ 1 - \frac{\sum_{j \neq i \in A} e_{i,j}}{e_i}, & i = 1. \end{cases}$$

(6)

Finally, the balance term is introduced as

$$B(A) = H(Q_A) - L_A = -\sum_i p_{Q_A}(i) \log (p_{Q_A}(i)) - L_A$$

(7)

where $L_A$ is the number of elements selected in the graph and $Z_A$ is the distribution of cluster members. Thus, the objective function is written as

$$\max_A = H(A) + \eta B(A)$$

(8)

where $\eta$ is the weight parameter.

B. Feature Extraction

Extended attribute profiles (EAPs) can effectively preserve the structure and shape information of hyperspectral images [58], especially in areas with more regular distribution of ground objects. LBP [62] can eliminate the illumination change to a certain extent. It achieves dimensionality reduction and speeds up the calculation. Due to its invariance of gray level and rotation, it has led to great results in hyperspectral image classification [63]. The principle is shown in Fig. 2. Besides, Gabor filtering has led to great performance in feature weighted classifiers [68] and collaborative representation (CR) using the 3-D Gabor feature [69]. Due to the obvious advantages of these features in the hyperspectral field recently, we select them to represent texture features, gray-scale features, and frequency features, respectively.
1) EMAP: Multiattribute profiles (MPs) use a structuring element (SE) [55] to perform repeated closing and opening operations on the image. For an image \( g \), the MP operation can be written as

\[
\text{MP}(g) = \left\{ \begin{array}{ll}
\prod_{i} = \prod_{0}, & v = n + i \ \forall v \in [0, n] \\
\prod_{i} = \prod_{1}, & v = i \ \forall v \in [n + 1, 2n + 1]
\end{array} \right.
\]  

(9)

where \( \prod_{0} \) is a concatenation of a closing profile and \( \prod_{1} \) is a concatenation of an opening profile. Benediktsson et al. [56] and Dalla Mura et al. [57] proposed an EMAP method of applying MP to hyperspectral images. The dimensions are reduced to several PCs first with a PCA, and the calculation is performed on each PC, which is expressed as

\[
\text{EMP} = \{\text{MP}(\text{PC}_1), \text{MP}(\text{PC}_2), \ldots, \text{MP}(\text{PC}_n)\}.
\]  

(10)

The generation of APs depends on thickening and thinning AP, which are, respectively, expressed as

\[
\text{AP}(g) = \left\{ \begin{array}{ll}
\prod_{i} = \prod_{0}^{M'}, & v = n + i \ \forall v \in [0, n] \\
\prod_{i} = \prod_{1}^{M'}, & v = i \ \forall v \in [n + 1, 2n + 1]
\end{array} \right.
\]

(11)

where \( M' = \{M_1, M_2, \ldots, M_n\} \) is an ordered set of criteria. Similarly, EAP calculated based on APs is expressed as

\[
\text{EAP} = \{\text{AP}(\text{PC}_1), \text{AP}(\text{PC}_2), \ldots, \text{AP}(\text{PC}_n)\}.
\]

(12)

EMAPs are extended by EAPs [58]–[60]

\[
\text{EMAP} = \{\text{EAP}_{a_1}; \text{EAP}_{a_2}; \ldots; \text{EAP}_{a_n}\}.
\]

(13)

Due to the multiattribute information extraction of APs, EMAPs have more advantages in extracting spatial information. EMAPs have been applied in many fields and lead to great results [58], [61].

2) LBP: For the center point \( h_c \) of each window, the pixels on the circle with the radius \( r \) starting from the clockwise direction are denoted as \( (h_0, h_1, \ldots, h_{p-1}) \). An image is expressed as

\[
F = f(h_c, h_0, h_1, \ldots, h_{p-1}).
\]

(14)

Assuming that the coordinates of the center pixel on the map are \((0, 0)\), then the coordinates of each neighboring pixel are \((-r \sin(2\pi p/p), r \cos(2\pi p/p))\). The gray value can be converted into a certain binary label by comparing the value of each neighbor pixel with the center and using the sign function

\[
F = f(s(h_0 - h_c), s(h_1 - h_c), \ldots, s(h_{p-1} - h_c))
\]

\[
s(x) = \begin{cases} 
1, & x \geq 0 \\
0, & x < 0.
\end{cases}
\]

(15)

(16)

Then, it can be written as the corresponding decimal number

\[
\text{LBP}_{p,r} = \sum_{p=0}^{p-1} s(h_p - h_0)2^p.
\]

(17)

3) Gabor Filter: The Gabor function was originally proposed to imitate the effect of the human eyes [64], so it is often used for texture recognition and has achieved great results [65], [66]. When the window function in the short-time Fourier transform is taken as a Gaussian function, it is called the Gabor transform.

In 2-D space, a Gabor filter [67] can be obtained by superimposing a Gaussian function and a sine function. Its mathematical expression is given as follows:

\[
G(x, y, \xi, \phi, \delta, \epsilon) = \exp\left(-\frac{x'^2 + \epsilon^2 y'^2}{2\delta^2}\right) \exp\left(i\left(2\pi \frac{x'}{\xi} + \phi\right)\right).
\]

(18)

The real part and the imaginary part are written as

\[
G(x, y, \xi, \phi, \delta, \epsilon) = \exp\left(-\frac{x'^2 + \epsilon^2 y'^2}{2\delta^2}\right) \cos\left(2\pi \frac{x'}{\xi} + \phi\right)
\]

(19)

\[
G(x, y, \xi, \phi, \delta, \epsilon) = \exp\left(-\frac{x'^2 + \epsilon^2 y'^2}{2\delta^2}\right) \sin\left(2\pi \frac{x'}{\xi} + \phi\right)
\]

(20)

respectively, where \( \xi \) is the wavelength, \( \phi \) represents the direction, \( \delta \) is the phase shift, \( \epsilon \) refers to the standard deviation of the Gaussian factor, \( \delta \) is the aspect ratio, and \( wb \) is the bandwidth. The coordinates are \( x' = x \cos \theta + y \sin \theta, y' = -x \sin \theta + y \cos \theta \). \( \delta \) is computed by

\[
\frac{\delta}{\xi} = \frac{1}{\pi} \sqrt{\frac{\ln 2}{2} \frac{2w+1}{2w-1}}.
\]

(21)

The analysis of texture features can be realized by using the real part, and six parameters need to be set at first.

C. Relaxed Collaborative Representation

SRC and CRC can be regarded as a linear regression problem

\[
\arg\min_{a} ||y - Da||_2^2 + \lambda ||a||_p.
\]

(22)

When \( p = 1 \), it is SR [22]

\[
\arg\min_{a} ||y - Da||_2^2 + \lambda ||a||_1.
\]

(23)

When \( p = 2 \), it is CR [24]

\[
\arg\min_{a} ||y - Da||_2^2 + \lambda ||a||_2^2.
\]

(24)

RCR [29], [30] considers the similarity and differences between features compared with CRC and SRC. It uses feature-related subdictionaries for representation and weight optimization to ensure the diversity of encodings of different features in face recognition. Meanwhile, weighted regularization is introduced to reduce variance. The objective function can be written as

\[
\arg\min_{a_k, w_k} \sum_{k=1}^{K} (||\epsilon_k - D_k a_k||_2^2 + \lambda ||a_k||_2^2)
\]

\[
+ \tau \omega_k ||a_k - \overline{a}||_2^2
\]

(25)

where \( K \) is the number of blocks, \( \tau \) is the balance parameter, and \( \overline{a} \) is the average value of \( a_k \). \( \omega_k \) is the weight of each block.
There are three situations for its weight.
1) For strong prior, (25) can have a closed-form solution.
2) For moderate prior, the prior of weights can be set as
   \[
   \begin{cases}
   -K_k \ln \omega_k > \sigma \\
   \mu_1 \leq \omega \leq \mu_c \\
   0 \leq \omega.
   \end{cases}
   \] (26)
3) For weak prior, the weights will not be measured except that the entropy satisfies (27), and \( \varepsilon \) is the threshold to consider whether it is a weak prior.

Here, \( \Phi \) is the relative relation matrix of \( \omega_k \) (e.g., \( \omega_2 \geq \omega_1 \geq 0, \Phi = [-1, 1], \mu_c = 0, \mu_1 = -\infty \))
\[
- \sum_{k=1}^{K} \omega_k \ln \omega_k > \varepsilon .
\] (27)

We choose the corresponding weighted method according to different datasets.

III. PROPOSED ALGORITHM

A. SRCR-BW

It is necessary to take the balance between similarity and difference, which characterize different features into consideration. In this section, a method of band-weighting RCR classifier based on superpixel is presented to improve the accuracy and stability with optimizing weights and coefficients simultaneously. Subgraphs \( \mathcal{N}_s(l = 1, 2, \ldots, L) \) obtained from ERS are used to form training sets to ensure uniform distribution. For each subgraph, we extract training samples \( \mathcal{D}_l \in \mathbb{R}^{B \times n_l} \), where \( n_l \) is the number of training samples and \( N = \sum_l n_l \) so that a global dictionary is \( \mathcal{D} = \{ \mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_L \} \in \mathbb{R}^{B \times N} \). Then, the proposed RCR-BW classifier will be applied in each subgraph for classification. Although the features in a subgraph merely retain the spectral information for arraying as vectors, the superpixels have retained the original spatial information. The objective function of RCR-BW is shown as
\[
\arg\min_{\alpha_l, \omega_b} \sum_{b=1}^{B} \left( ||y_l,b - D_{l,b} \alpha_{l,b}||^2 + \lambda ||\alpha_{l,b}||^2_2 \right) + \tau \omega_b ||\alpha_{l,b} - \alpha_0||^2_2 \quad (l = 1, \ldots, L). \tag{28}
\]
A closed-form solution can be calculated as
\[
\alpha_{l,b} = \alpha^0_{l,b} + \frac{\omega_b}{\sum_{b=1}^{B} \omega_b} M_{l,b} N_l \sum_{b=1}^{B} \omega_b \alpha^0_{l,b}. \tag{29}
\]

After traversing all subgraphs, it produces a reconstructed image. Algorithm 1 provides more details.

1) Generate Subgraphs and Training Set: According to the characteristic of the entropy rate, ERS can obtain compact and uniform subgroups. As the number of pixels increases, its undersegmentation error (UE) will decrease; meanwhile, the boundary recall (BR) and achieve segmentation accuracy (ASA) will increase. Therefore, an appropriate number of superpixels needs to be preset, which is the prerequisite for improving accuracy. Set the number of superpixels as: \( N_s = [50 100 200 300 400] \). ERS has two input parameters: \( \lambda \) and \( \sigma \). According to previous experiments, the segmentation will be better when \( \lambda = 0.5 \), but the influence of \( \sigma \) is relatively less...
affected [43], which can be set as a constant. The segmented results are shown in Fig. 3

\[
M_{l,b} = (D_{l,b}^TD_{l,b} + I(\lambda + \tau \omega_b))^{-1}
\]

\[
\omega^0_b = M_b D_{l,b}^T y_{l,b}
\]

\[
N_l = \left( I - \sum_{b=1}^{B} \tau M_{l,b}/B \right)^{-1}
\]

\[
\omega_l = \frac{1}{\sum_{b=1}^{B} \omega_b \ln \omega_b}
\]
Algorithm 1 SRCR-BW Algorithm

Input:
1. A hyperspectral image with spectral features
2. Set parameters: regulation parameter $\lambda$, balance parameter $\tau$, Langrange multiplier $\gamma$, number of superpixels $N^S$

Segmentation: Generate the blocks using ERS superpixel

Main Loop:
For each block
1. input: Training set $D_l$ and testing set $Y_l \in R^B \times K_l$, where $k_l$ is the number of testing samples, $Y_l \in R^B \times K$ and $K = \sum_k k_i$, maximum iteration times $\text{iters}$
2. initialization: $\omega_b^0 = 1$
3. RCR-BW
   for each testing vector in $y_l$
      while $||\omega_b^{t+1} - \omega_b^t||_2/||\omega_b^t||_2 > \zeta$ (t represents $t^{th}$ iteration)
         Update coefficient vectors with (29)
         Update weights via (35)
      end while
      Identity via (37)
   end for
End for
Output: class1 ($Y_l$)

Algorithm 2 MSRCR-BW Algorithm

Input:
1. A hyperspectral image with multi-features
2. Set parameters: regulation parameter $\lambda$, balance parameter $\tau$, Langrange multiplier $\gamma$, number of superpixels: $N^S$

Segmentation: Generate the blocks using ERS superpixel

Main Loop:
For each block
1. input: Training set $D_m$ and testing set $y_m \in R^B \times k_l$, where $k_l$ is the number of testing samples, $B^m$ is the dimension of the $m^{th}$ feature, $y_m \in R^B \times K$ and $K = \sum_k k_i$, maximum iteration times $\text{iters}$
2. initialization: $\omega_b^0 = 1$
3. RCR-BW
   for each testing vector in $y_m$
      while $||\omega_b^{t+1,m} - \omega_b^{t,m}||_2/||\omega_b^{t,m}||_2 > \zeta$ (t and m represents $t^{th}$ iteration and $m^{th}$ feature respectively)
         Update coefficient vectors with (29)
         Update weights via (38)
      end while
      Identity via (40)
   end for
End for
Output: class2 ($Y_l$)

1) Multifeature Extraction: There are three types of features being employed. EMAP can extract spatial information with multiple attributes. LBP contains invariant rotation and gray levels. The Gabor filter realizes the analysis of texture features in the frequency domain. The extraction results are shown in Fig. 8. For EMAP, four attributes were selected: “a” refers to the area, “d” refers to the length of the diagonal, “s” refers to the standard deviation, and “i” refers to the moment of inertia. (i.e., generating a stack of 123 features, $41 \times 3$ PCs). A multiscale fusion is used to integrate the extracted information of LBP. Besides, the results of the Gabor filter in four directions and the scales were calculated. The extraction of multiple features not only improves the diversity of features but also preserves spatial information to a certain extent.

2) Update Multiweights: Since weighting on bands plays a significant role in the proposed methods, the diversity of features is supposed to be considered to calculate the multiweights, and (35) should be changed into (38). Features in an HSI are inherently different, and then, take the difference between bands and the fusion of plentiful feature information into consideration. The weight results are shown in Fig. 9, where $m$ refers to the $m^{th}$ feature

$$\omega_b^m = \exp\left\{-1 - \tau ||a_b^m - \overline{a}_b||_2/\gamma\right\}. \tag{38}$$

3) Classification Criteria: Since multifeatures were calculated, the classification error for each class becomes an overall multidimensional error superposition, which can be

weight is according to the variance of coding coefficients between each band and all other bands. Bands with higher residuals are given lower weights. Considering the balance of the difference and similarity between the bands, the value of the weight is close to 1 although some bands should have small or zero weights according to recent band weighting methods. Besides, there should have no significant difference in weights between adjacent bands. Fig. 7 described the change of weights in a subdictionary.

3) Classification Criteria: The classification is based on the superposition of all errors in spectral dimension for each class. The criteria can be computed as

$$r_{i,l} = \sum_{b=1}^{B} o_b ||y_{l,b} - D_{l,b}^i o_{l,b}||_2^2 \tag{36}$$

where $D_{l,b}^i$ and $o_{l,b}$ are the subset of the subdictionary $D_{l,b}$ and the coefficient vector from $o_{l,b}$ associated with the $i^{th}$ class, respectively. The class label is assigned according to

$$\text{Class1}(y_l) = \arg\min_i \{r_{i,l}\} \ (l = 1, 2, \ldots, L). \tag{37}$$

B. MSRCR-BW

SRCR-BW takes advantage of differences based on spectral information. Generally, multiple features are used to describe the hyperspectral data from multiple views. In this section, extracted multifeatures are combined to increase the diversity and intensify feature information so that the classifier can be more flexible and diverse. Algorithm 2 elaborates the specific steps.
Fig. 8. Four kinds of features from the AVIRIS data. (a) Spectral feature. (b) EMAP. (c) LBP. (d) Gabor (four orientations: 0, π/4, π/2, and 3π/4 are employed).

Fig. 9. Comparison of different weights from multifeatures before and after updating. (a) Residuals of each band on EMAP features extracted from ROSIS data. (b) Weights of different bands on EMAP features extracted from ROSIS data. (c) Residuals of each band on LBP features extracted from ROSIS data. (d) Weights of different bands on LBP features extracted from ROSIS data. (e) Residuals of each band on Gabor features extracted from ROSIS data. (f) Weights of different bands on Gabor features extracted from ROSIS data.

expressed as

\[
\begin{align*}
    r_{ij}^m &= \sum_{b=1}^{B} \omega^m_b ||y_{ij}^m - D_{ij,b}^m ||^2_2. \\
\end{align*}
\]  

(39)

The class label is assigned as

\[
\text{Class2}(y_l) = \arg \min \left\{ r_{ij}^m \right\} \quad (l = 1, 2, \ldots, L). 
\]  

(40)

IV. EXPERIMENTS

A. Dataset

1) Indian Pines: The first hyperspectral data are from the Indian Pines Northwestern Indiana. It was collected by AVIRIS sensor in 1992, which consists of 145 × 145 pixels and 224 bands, and the wavelength range is 400–2500 nm. The available data includes 200 bands after removing the bands of water absorption and 16 classes. Table I shows the labeled pixels.

2) Pavia University: The second hyperspectral data is over the Pavia in Northern Italy from Pavia University with nine classes. It was acquired by ROSIS sensor and includes 103 bands and 610 × 340 pixels with a resolution of 1.3 m. The wavelength range is 430–860 nm. Labeled pixels are shown in Table II.

3) Mall of Washington DC: The third hyperspectral data come from the Mall of Washington DC, which is acquired by the HYDICE sensor. It includes 210 bands ranging from 400 to 2400 nm. Due to the complex atmospheric environment, only 191 of the bands are used for experiments. The spatial coverage size is 266 × 304 pixels. There are six classes, as shown in Table III.

B. Classification

1) Experimental Design: To illustrate the superiority of the proposed algorithms, three kinds of features (EMAP, LBP,
and Gabor) are combined with CRC, respectively. Specifically, CRC on EMAP (CRC-EMAP), CRC on LBP (CRC-LBP), and CRC on Gabor (CRC-Gabor) are involved. Moreover, MCRC-DL [18] is also considered for its combination of three kinds of features by dictionary learning. In particular, R2MK that has considered the RCR method is taken into consideration. Besides, two popular classifiers, SVM and JCRC, are also examined for comparison. Because of the different numbers of samples between three datasets and superpixels, 1%–5% of samples are used as the training set for the Pavia University dataset and 3%–7% percent for the others. The remaining labels are considered as testing samples. The regularization parameter $\lambda$ of all algorithms is set to $[10^{-4}, 10^{-5}, \ldots, 10^{-1}]$. The balance parameter $\tau$ is set to $[10^{-3}, 10^{-2}, \ldots, 10^{-1}]$ for Indian Pines and Pavia University data, which is set as $[10^{-3}, 10^{-2}, \ldots, 10^{-1}]$ for DC data. The Langrange multiplier is set as $[10^{-3}, 10^{-2}, \ldots, 10^{-1}]$. Set the iterations as 5 according to the speed of convergence. The experimental training set for quantitative evaluation is selected from each superpixel.

In this article, the overall accuracy (OA), the average accuracy (AA), and the kappa coefficient are calculated for quantitative evaluation. The first metric refers to the global accuracy and is calculated by dividing the correctly classified samples by the number of all testing samples. The second one is the average value of the correlative accuracy for each class. The third metric reflects the classification performance from the perspective based on the confusion matrix. A higher value of each metric means better performance.

2) Performance: In order to demonstrate the effectiveness of band weighting, the unweighted SRCR is also used for classification. In Fig. 10, it can be seen that SRCR-BW obtains higher accuracy than SRCR without band weighting. Compared with SRCR, SRCR-BW can improve the accuracy rate by 5%.

First, Table IV shows the optimal parameters of all datasets. Due to the various classes of Indian Pines data, accurate classification is challenging. Table V shows the classification results in the case of optimal parameters. The diagram is shown in Fig. 17. The proposed MSRCR-BW method can produce the best performance with 99% classification accuracy. Compared with other classifiers, SRCR-BW can significantly improve the accuracy, especially in small sample datasets that are more obvious. After spatial feature extraction, MSRCR-BW can improve the classification accuracy further. Among the different features, the combination of EMAPs and spectral information has the best accuracy, and the superposition of multifeatures may slightly reduce the classification accuracy. In general, SRCR-BW and MSRCR-BW can reach a better performance than other classifiers so that the effectiveness of the algorithm can be proved.

For the second dataset, the classification performance is shown in Fig. 18 and Table VI. Similar to the first data, SRCR-BW can effectively improve classification accuracy. The three kinds of multifeature superimposed spectral information are all effective, but MSRCR-EMAP has the best effect. Compared with SVM, CRC, JCRC, and MCRC-DL that use multifeatures and R2MK that uses spatial information, the OA of MSRCR-BW is increased by 5.97%, 24.71%, 0.40%, 10.96%, and 1.62%, respectively. Meanwhile, the best OA can reach more than 99%.

For Washington DC data with scattered labeled samples, the classification performance is quantitatively analyzed, as shown in Fig. 19. The proposed SRCR-BW can improve the accuracy rate by 3.78% compared with CRC similar to the Indian Pines and Pavia University data, and MSRCR-BW can increase it further. The superposition of three extracted features on these data is better than a single feature.

It can be seen that MSRCR-EMAP can obtain better performance than the other two in most situations and is
almost equal to MSRCR-BW, which uses the three types of features occasionally. In general, MSRCR-BW can obtain higher accuracy than SRCR-BW. SRCR-BW and MSRCR-BW perform better according to the three evaluation indicators.

Fig. 10. OA of different methods. (SRCR is without band weighting, and SRCR-BW is the proposed band-weighting method). (a) Indian Pines. (b) Pavia University. (c) Washington DC.
Fig. 11. Classification accuracy versus $\lambda$. (a) Indian Pines. (b) Pavia University. (c) Washington DC.

Fig. 12. Classification accuracy versus $\tau$. (a) Indian Pines. (b) Pavia University. (c) Washington DC.

Fig. 13. Classification accuracy versus $\gamma$. (a) Indian Pines. (b) Pavia University. (c) Washington DC.

Fig. 14. Classification accuracy versus $N_s$ and training samples for Indian Pines data. (a) SRCR-BW. (b) MSRCR-EMAP. (c) MSRCR-LBP. (d) MSRCR-Gabor. (e) MSRCR-BW.

The two methods can obtain higher OA, AA, and the kappa coefficient.

3) Parameters Analysis: Fig. 10 shows the changes in classification accuracy as the number of training samples
increases. It can provide and maintains a high accuracy but will slightly decrease with increased training samples occasionally.

The effects of the two proposed methods on the regularization parameter $\lambda$ are evaluated, as shown in Fig. 11. The classification accuracy first increases as the value of $\lambda$ decreases and then maintains at a certain level. In general, the accuracy is insensitive to $\lambda$. For Washington DC data, it is more sensitive for the MSRCR-BW method to the regularization parameter $\lambda$ compared with the other two datasets. The optimal range is from $1e-5$ to $1e-2$. For the effect of balance parameter $\tau$ that is illustrated in Fig. 12, the classification accuracy will reach its best when $\tau = 1e+3$.

Fig. 13 shows the sensitivity of the Lagrangian factor to the proposed methods. As $\gamma$ deceases, the classification accuracy presents a trend of first being stable and then decreasing, and it changes stably between $1e-2$ and 1 with the optimal parameter $\gamma = 1$. In particular, the trend of MSRCR-EMAP is to remain unchanged and then decrease but increases when $\gamma = 1e-4$.

Due to the multiscale image segmentation, the effects of the parameter $Ns$ are evaluated. As shown in
Fig. 17. Classification results with the Indian Pines dataset. (a) False color image. (b) Ground truth. (c) SVM. (d) CRC. (e) CRC-EMAP. (f) CRC-LBP. (g) CRC-Gabor. (h) MCRC-DL. (i) JCRC. (j) R2MK (using superpixel). (k) SRCR-BW. (l) MSRCR-EMAP. (m) MSRCR-LBP. (n) MSRCR-Gabor. 

**TABLE VII**

CLASSIFICATION ACCURACY FOR THE WASHINGTON DC DATASET

<table>
<thead>
<tr>
<th>Class</th>
<th>Train</th>
<th>SVM-CK</th>
<th>CRC</th>
<th>CRC-EMAP</th>
<th>CRC-LBP</th>
<th>CRC-Gabor</th>
<th>MCR-DL</th>
<th>JCRC</th>
<th>R2MK</th>
<th>SRCR-BW</th>
<th>MSRCR-EMAP</th>
<th>MSRCR-LBP</th>
<th>MSRCR-Gabor</th>
<th>MSRCR-BW</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>32</td>
<td>96.61</td>
<td>94.10</td>
<td>98.58</td>
<td>96.07</td>
<td>93.33</td>
<td>98.47</td>
<td>99.23</td>
<td>98.91</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>C2</td>
<td>31</td>
<td>90.56</td>
<td>98.61</td>
<td>96.57</td>
<td>96.89</td>
<td>97.42</td>
<td>96.46</td>
<td>98.28</td>
<td>96.57</td>
<td>98.71</td>
<td>99.88</td>
<td>96.56</td>
<td>99.57</td>
<td>99.78</td>
</tr>
<tr>
<td>C3</td>
<td>24</td>
<td>67.61</td>
<td>97.88</td>
<td>99.47</td>
<td>99.12</td>
<td>83.19</td>
<td>99.82</td>
<td>86.90</td>
<td>98.41</td>
<td>99.65</td>
<td>92.31</td>
<td>99.82</td>
<td>97.70</td>
<td>99.64</td>
</tr>
<tr>
<td>C4</td>
<td>16</td>
<td>98.36</td>
<td>100</td>
<td>99.51</td>
<td>100</td>
<td>98.85</td>
<td>99.94</td>
<td>96.71</td>
<td>98.85</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>C5</td>
<td>32</td>
<td>91.81</td>
<td>99.38</td>
<td>100</td>
<td>99.38</td>
<td>96.60</td>
<td>100</td>
<td>99.53</td>
<td>99.69</td>
<td>98.76</td>
<td>99.83</td>
<td>100</td>
<td>100</td>
<td>98.60</td>
</tr>
<tr>
<td>C6</td>
<td>41</td>
<td>93.35</td>
<td>89.16</td>
<td>99.60</td>
<td>91.71</td>
<td>93.08</td>
<td>99.73</td>
<td>96.99</td>
<td>93.35</td>
<td>100</td>
<td>99.99</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

OA (%) | 90.8 | 95.76 | 98.70 | 96.54 | 94.02 | 98.91 | 96.79 | 97.21 | 99.54 | 99.02 | 99.30 | 99.64 | 99.72 |

AA (%) | 89.71 | 96.52 | 98.85 | 97.19 | 93.75 | 99.05 | 96.28 | 97.63 | 99.52 | 98.66 | 99.40 | 99.54 | 99.67 |

Kappa | 0.8883 | 0.9486 | 0.9842 | 0.9580 | 0.9273 | 0.9867 | 0.9609 | 0.9661 | 0.9943 | 0.9880 | 0.9914 | 0.9956 | 0.9966 |

Time(s) | 1.3 | 3.02 | 24.49 | 14.25 | 22.77 | 345.18 | 538.76 | 2.33 | 45.98 | 140.13 | 148.45 | 144.03 | 353.73 |

Figs. 14–16, with the increasing number of superpixels, the classification accuracy of the proposed methods also steadily improves. When $N_s$ increases from 50 to 200, the classification accuracy increases quickly. After that, the accuracy maintains steady growth and then reaches the best.

4) Time Analysis: The time for the proposed methods has been listed in Tables V–VII. It can be seen that R2MK is the least time-consuming. This is because the calculation speed can be quickly increased based on superpixel segmentation, while SRCR-BW requires more time updating the weights to find the optimal distribution. Nevertheless, the running time of the two proposed methods is different among different datasets. The complexity of SRCR-BW running on Pavia University data is less than that of CRC, even if these data have the largest sample size. Compared with
SRCR-BW, the complexity of MSRCR-BW has increased since the superposition of multiple features leads to a longer time of updating weights.

V. CONCLUSION

In this article, a new RCR algorithm with band weighting is proposed to measure the differences of features. The weight
and balance function that we used to explore the difference and similarity between bands simultaneously are determined according to the residual of each band. Two methods, single feature and multiple features, based on superpixel are proposed. The experimental results show that both can effectively improve classification accuracy. Moreover, the approach using multifeatures can further improve the accuracy although the degree of improvement is limited compared with the first method. The bands of the three types of features change differently, so the associated weights are also in great difference. Therefore, the multifeatures are mainly used to verify the impact of diverse weights.

Due to the limitation of the number of superpixels when constructing a dictionary, the classification accuracy may decrease slightly as the training samples increases, but the OA remains at a high level. We would like to further explore more effective methods to construct a discriminative dictionary based on spatial information.

Compared with a single feature, the method of multifeatures has indeed improved classification accuracy, but the computational cost is also increased. We would like to explore a more efficient approach to combine multiple features.

ACKNOWLEDGMENT

The authors would like to thank Prof. P. Gamba for providing the ROSIS University of Pavia and Prof. D. Landgrebe for providing the AVIRIS data. They also thank the reviewers for their valuable remarks and suggestions.

REFERENCES
