Union of Class-Dependent Collaborative Representation Based on Maximum Margin Projection for Hyperspectral Imagery Classification

Haoyang Yu, *Member, IEEE*, Xiaodi Shang, *Student Member, IEEE*, Meiping Song, *Member, IEEE*, Jiaochan Hu, Tong Jiao, Qiandong Guo, and Bing Zhang, *Fellow, IEEE*

***Abstract*—This article proposed a novel spectral-spatial classification framework for hyperspectral image (HSI) through combining collaborative representation (CR) and maximum margin projection (MMP). First, class-dependent CR classifier (CDCRC) is used on HSI classification to fully make use of self-information contained in each class. Second, the MMP is included into the framework to discover local manifold structure. Combined with CDCRC, it formed the classifier named CDCRC based on MMP (CMCRC), which aims to reduce band redundancy. Finally, a comprehensive spectral-spatial classifier, called union of CMCRC, is proposed to optimize the classification map through integrating cumulative probability of residuals instead of applying strong constraints to maintain the spatial consistency. Experimental results on three real hyperspectral datasets demonstrate the effectiveness and practicality of the proposed methods over other related models for HSI classification tasks.**

***Index Terms*—Collaborative representation (CR), dimensionality reduction, hyperspectral remote sensing, image classification, maximum margin projection (MMP).**

NOMENCLATURE

List of Acronyms

HSI	Hyperspectral image.
HSIC	Hyperspectral imagery classification.
SVM	Support vector machine.
SR	Sparse representation.
SRC	SR-based classifier.
CR	Collaborative representation.
CRC	CR-based classifier.
DR	Dimensionality reduction.
MMP	Maximum margin projection.
JSRC	Joint SRC.
JCRC	Joint CRC.
CDCRC	Class-dependent collaborative representation classifier.
CMCRC	CDCRC based on MMP.
UCMCRC	Union of CMCRC.
SCP	Spatial cumulative probability.
OA	Overall accuracy.
AA	Average accuracy.

Manuscript received August 20, 2020; revised October 19, 2020; accepted November 7, 2020. Date of publication November 17, 2020; date of current version January 6, 2021. This work was supported in part by the National Science Foundation of China under Grant 61971082 and Grant 61890964, and in part by the Fundamental Research Funds for Central Universities under Grant 3132020218 and Grant 3132019341. *(Corresponding author: Xiaodi Shang.)*

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

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

Tong Jiao is with the Graduate School of Geography, Clark University, Worcester, MA 01610 USA (e-mail: tjiao@clarku.edu).

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

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

Digital Object Identifier 10.1109/JSTARS.2020.3038456

I. INTRODUCTION

HYPERSEPCTRAL remote sensing is a multidimensional information acquisition technology that combines imaging technology and spectral technology [1]–[3]. Compared to multispectral images, hyperspectral images (HSIs) have higher spectral resolution, which enhances their capability to describe and distinguish the ground objects. Nowadays, HSIs have been widely used in various fields such as target recognition, disaster monitoring, resource exploration, land-use, and land-cover mapping [4]–[6].

Hyperspectral imagery classification (HSIC), which assigns the pixels to the available classes [7], is a widely used technique in hyperspectral imagery processing. Various pixel-wise spectral classifiers have been developed [8]–[15], such as support vector machine (SVM) [8], [9], neural networks [10], [11], and sparse representation (SR) [12]–[15]. Recently, deep learning [16] has drawn significant attention and proved to be effective in HSIC [17]. It can be considered as an extension of the hand-crafted neural network. In contrast to traditional classification methods, deep learning can automatically learn complex features of HSIs with a large number of hierarchical layers [18]. However, training a deep network is quite expensive and requires a large number of training samples [19] and [20]. Additionally, SR-based classifier (SRC), which represents each tested pixel sparsely by a few labeled atoms via l_0 or l_1-normed regularization takes the low-rank characteristic of HSI [21] into consideration, and has been shown to improve HSIC [22]. However, its competitive nature imposed by sparseness constraint [23] is not effective as

expected in improving the accuracy of classification. In contrast, collaborative representation (CR) [24], which incorporates all the atoms for representation and makes each atom be selected with equal chance, achieves a better performance than SR. Despite of advances in spectral classifiers as mentioned above, Hughes phenomenon produced by the unbalance between limited training samples and high-dimensionality continues to limit the accuracy of classification [25].

Dimensionality reduction (DR) has been used to solve this issue. In general, there are two approaches to reduce the dimensionality of HSI, namely feature extraction and band selection. Specifically, band selection preserves the most discriminative bands while discarding some redundant and noisy bands to reduce the band number [26]–[32]. The classical methods such as maximum variance principal component analysis (MVPCA) [26], constrained band selection (CBS) [27], entropy-based [28]–[30], and evolution-based band selection [31] have been shown to be effective. Recently, a Boltzmann entropy-based band selection was developed, which can characterize both statistical information of HSI and spatial distribution of pixels [29]. Moreover, a band selection method based on a variant of entropy has also proved efficient [30]. As a limitation, the band subsets selected by most of these traditional band selection techniques are not enough to describe the discriminants for different types of classes to some extent [32].

Feature extraction, as a significant and effective approach for DR [33], is mainly composed of linear or nonlinear equations which aims to transform original data from a high-dimensional space to an optimal low-dimensional space [34]. Some classical algorithms have shown their capability to optimize spectral features of HSI and demonstrated their effectiveness in HSIC, including principal components analysis (PCA) [35], maximum noise fraction (MNF) [36], and independent component analysis (ICA) [37]. Manifold learning, as a basic approach for pattern recognition, has also been employed for spectral feature extraction and HSIC [38], which assumes that high-dimensional data can be approximately represented in manifold low-dimensional subspaces [39]. Based on manifold learning, He *et al.* [40] proposed a linear and semi-supervised algorithm called maximum margin projection (MMP) to extract features for image retrieval tasks. Unlike most DR algorithms, which rely on global Euclidean structure but ignore the difference in data distribution, MMP is designed to discover the local structure through characterizing both geometrical and discriminant features of data in manifold spaces. Its advantage of seeking class-specific subspace allows it to be successfully applied to DR of HSI. However, spectral features of similar objects could differ as a result of noise and spectral variability [41], [42], which limits the accuracy of HSIC based on spectral information alone and thus makes it necessary to incorporate other information that helps refine classification.

Spatial information is an example of such information. It has been widely used in various classification methods [43]–[50], the basic idea of which is to assume that pixels within an adjacent region share similar spectral characteristics and, thus, are more likely to belong to same category [43]. To integrate spectral and spatial information, Tarabalka [44] proposed a classification model, which combines SVM with MRF (SVM-MRF). The SVM-MRF model can effectively modify initial results in classification through incorporating the spatial correlation between pixels into the posterior probability of spectral features. In addition, spatial correlation has also been introduced into the SR-based and CR-based models of second generation [48]–[50], usually known as joint SRC (JSRC) [49] and joint CRC (JCRC) [50], which represents an object with a local area around a tested pixel rather than a single pixel. Although with a better performance in HSIC, the joint representation-based methods tend to be time-consuming as a result of finding solutions for strong constraint vectors.

In this article, we proposed a framework that incorporated recent advances in HSIC including CR, manifold learning, and spatial correlation. First, we applied the class-dependent collaborative representation classifier (CDCRC) to explore the impact of intraclass training samples on optimizing the sparse coefficient. Second, MMP is introduced to extract the class-specific features for HSIC by obtaining the projection matrix according to the dictionary of each class. The new classifier named CDCRC based on MMP (CMCRC), which represents the MMP-reduced data sparsely by CDCRC, aims to reduce band redundancy and improve classification performance. Finally, a spatial information integration method with consideration of spatial coherence was incorporated into the framework of classification, called the union of CMCRC (UCMCRC). In summary, UCMCRC optimizes the classification map and improves the classification performance by combining residual accumulation probability image with a softer spatial information postprocessing method. Major contributions of this article are summarized as follows.

1) Suitable mechanism and simple calculation based on class-dependent CR

CDCRC simplifies the SR process from the perspective of self-information of each class. It represents the tested pixel by aggregating all samples of a single class in the dictionary, which not only simplifies the calculation but also achieves better classification performance than CRC.

2) Effective feature extraction and efficient hyperspectral classification via MMP

CMCRC has fully considered the advantage of CDCRC representing the target vector by class, and finally unifies the samples required for MMP with the labeled samples in the dictionary used by the CDCRC to perform feature optimization and SR to achieve final classification. In a word, CMCRC can reduce time consumption and further improve the classification.

3) Relaxed spatial integration and precise classification based on the cumulative probability of residuals

UCMCRC calculates the cumulative probability of the central pixel for each class according to the reciprocal image of residuals, and then integrates the spatial information to achieve HSIC, which is effective and practical in classification for real hyperspectral scenes.

The remainder of this article is organized as follows. Section II introduces SRC, CRC as well as MMP algorithms. Section III describes the proposed CDCRC, CMCRC, and UCMCRC in detail. Section IV validates the proposed framework on three real hyperspectral datasets by drawing comparisons with other

state-of-the-art HSIC approaches. Section V concludes this article with some remarks.

II. Related Model Description

Before the description of the related models, we first list some of the notations that will be used throughout this article.

$\mathbf{X}=\{\mathbf{x}\in\Re^L\}$	HSI with L spectral bands
$\mathbf{x}=(x^1,...,x^L)$	Tested pixel
$\mathbf{D}=(\mathbf{D}_1,...,\mathbf{D}_m,...,\mathbf{D}_M)$	Dictionary with M sub-dictionaries
$\mathbf{D}_m=\{\mathbf{x}_t^m \mid t\in[1,n^m]\}$	$\mathbf{D}_m$ with n^m labeled samples
$n=\sum_{m=1}^{M} n^m$	Total of labeled samples
$\boldsymbol{\alpha}=(\boldsymbol{\alpha}_1,...,\boldsymbol{\alpha}_m,...,\boldsymbol{\alpha}_M)$	Sparse coefficient of $\mathbf{x}$
$\boldsymbol{\alpha}_m=(\alpha_{m+1},...,\alpha_{m+n})$	Sparse coefficient of $\mathbf{x}$ with class m
$\mathbf{A}=(\mathbf{a}_1,\mathbf{a}_2,...,\mathbf{a}_d)\in\Re^{L\times d}$	Projection matrix
$\mathbf{Y}=\{\mathbf{y}\in\Re^d\}$	Dimension-reduced HSI with d spectral bands
$\mathbf{R}=\{\mathbf{r}\in\Re^M\}$	Residual image with M classes
$w\times w$	$w\times w$ pixel-sized neighborhood
$\mathbf{R}_{i,j}=(\mathbf{r}_{i-w,j-w},.,\mathbf{r}_{i,j},.,\mathbf{r}_{i+w,j+w})$	$w\times w$ pixel-sized neighborhood centered on $\mathbf{r}_{i,j}$
$\hat{\mathbf{R}}_{i,j}=(\hat{\mathbf{r}}_{i-w,j-w},.,\hat{\mathbf{r}}_{i,j},.,\hat{\mathbf{r}}_{i+w,j+w})$	Reciprocal of $\mathbf{R}_{i,j}$
$\mathbf{P}_{i,j}$	Normalized probability map of $\hat{\mathbf{R}}_{i,j}$

A. SR-Based Classifier

As a generation of the nearest neighbor classifier as well as the nearest subspace classifier [59], SRC is a simple but effective classifier, which codes a tested sample $\mathbf{x}$ over a dictionary $\mathbf{D}$ such that $\mathbf{x}$ can be nearly represented by $\mathbf{x}\approx\mathbf{D}\alpha$. SRC adopts the l_0-norm to measure the sparsity of α, which determines the number of nonzero values in α. Because the combinatorial l_0-norm minimization turns to be a nonconvex NP-hard problem, the l_1-norm minimization as the closest convex function is utilized to replace it in SR formulation. Then, the class label of the tested pixel (i.e., check class by class) is determined by representing it using all training samples in $\mathbf{D}$. Therefore, SRC can be formulated as a Lagrange formulation as follows:

$$\hat{\alpha}=\arg\min_{\boldsymbol{\alpha}}\{||\mathbf{x}-\mathbf{D}\alpha||_2^2+\lambda||\boldsymbol{\alpha}||_1\} \tag{1}$$

where $||\boldsymbol{\alpha}||_1=\sum_{m=1}^{n}|\alpha_m|$ denotes the l_1-norm, and λ is defined as a regularization parameter that measures the relative importance of minimizing the vector α. Then, the residual of $\mathbf{x}$ is calculated. Then, the SR of $\mathbf{x}$ using $\mathbf{D}$ and the optimal solution of coefficient $\hat{\alpha}$ is given by

$$\mathbf{r}_m(\mathbf{x})=||\mathbf{x}-\mathbf{D}_m\hat{\alpha}_m||_2^2 \tag{2}$$

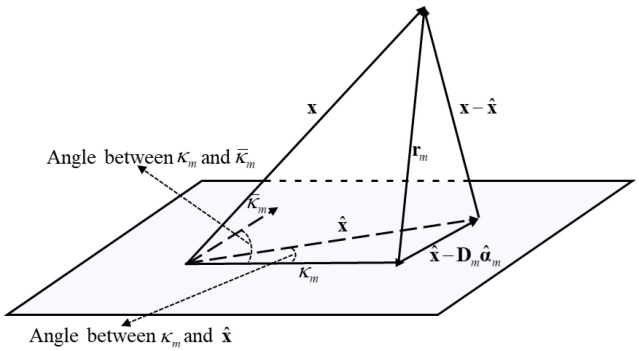


Fig. 1. Geometric illustration of the representation of $\mathbf{x}$ over $\mathbf{D}$.

where $\hat{\alpha}_m$ is defined as the corresponding coefficient vector related to the class m in $\hat{\alpha}$. Then, the final class label of $\mathbf{x}$, which is determined based on the minimum residual error between $\mathbf{x}$ and its approximation $\mathbf{D}_m\hat{\alpha}_m$ is assigned by

$$\text{class}(\mathbf{x})=\arg\min_m \mathbf{r}_m(\mathbf{x}). \tag{3}$$

In this case, if $\mathbf{x}$ belongs to class m, $\mathbf{D}_m\hat{\alpha}_m$ may represent $\mathbf{x}$ well while the most coefficients in $\hat{\alpha}$ with $m\neq n$ are zeros, and only $\hat{\alpha}_m$ has a value. That is, the nonzero value in $\hat{\alpha}$ will ultimately determine the class label of $\mathbf{x}$.

B. CR-Based Classifier (CRC)

Compared to the competitive mechanism imposed by sparseness constraints in SRC, the collaborative mechanism is more suitable for HSIC [23]. Therefore, the concept of CR and the CRC is introduced in this section. As shown in Fig.1, if considering the classification criterion of SRC from the perspective of geometric projection with letting $\hat{\mathbf{x}}=\mathbf{D}\hat{\alpha}$ as the perpendicular projection of $\mathbf{x}$ onto the space spanned by $\mathbf{D}$, the reconstruction residual $\mathbf{r}_m(\mathbf{x})$ of mth class in (2) in SRC can be further derived as

$$\mathbf{r}_m(\mathbf{x})=||\mathbf{x}-\hat{\mathbf{x}}||_2^2+||\hat{\mathbf{x}}-\mathbf{D}_m\hat{\alpha}_m||_2^2. \tag{4}$$

Since $||\mathbf{x}-\hat{\mathbf{x}}||_2^2$ is a constant, $\mathbf{r}_m^*(\mathbf{x})=||\hat{\mathbf{x}}-\mathbf{D}_m\hat{\alpha}_m||_2^2$ named as CR error plays a conclusive role in classification. For simplicity, via donating the parallel vectors $\kappa_m=\mathbf{D}_m\hat{\alpha}_m$ and $\bar{\kappa}_m=\sum_{n\neq m}\mathbf{D}_n\hat{\alpha}_n$, $\mathbf{r}_m^*(\mathbf{x})$ can be expressed as

$$\mathbf{r}_m^*(\mathbf{x})=\frac{\sin^2(\hat{\mathbf{x}},\kappa_m)||\hat{\mathbf{x}}||^2}{\sin^2(\kappa_m,\bar{\kappa}_m)} \tag{5}$$

where $(\hat{\mathbf{x}},\kappa_m)$ represents the angle between $\hat{\mathbf{x}}$ and κ_m. Similarly, $(\kappa_m,\bar{\kappa}_m)$ is the angle between κ_m and $\bar{\kappa}_m$. It can be concluded from (5) that, CR considers not only whether the angle between $\hat{\mathbf{x}}$ and κ_m is small, but also whether the angle between κ_m and $\bar{\kappa}_m$ is large. Due to eliminating the l_1-norm constraint, such a mechanism of "double checking" in CR can make HSIC more effective and robust.

Differing from SRC with l_1-norm that is time-consuming and cannot exploit the sparsity of α with insufficient data samples, CRC believes that a tested sample can be collaboratively represented by the l_2-norm via all the training samples in $\mathbf{D}$.

The regularized least square is utilized in CRC such that its closed-form solution can be easily achieved. Therefore, the CRC can be modeled as

$$\hat{\alpha} = \arg\min_{\boldsymbol{\alpha}} \{||\mathbf{x} - \mathbf{D}\alpha||_2^2 + \lambda||\boldsymbol{\alpha}||_2^2\} \tag{6}$$

where $||\boldsymbol{\alpha}||_2 = (\sum_{m=1}^{n} |a_m|^2)^{1/2}$ is defined as an l_2-norm constraint and λ is a regularization parameter, which makes the solution stable and controls the sparsity of $\hat{\alpha}$. Eventually, the sparsity by l_2-norm is weaker than that by l_1-norm. Then, the optimal solution of CRC can be expressed as

$$\hat{\alpha} = (\mathbf{D}^T\mathbf{D} + \lambda\mathbf{I})^{-1}\mathbf{D}^T\mathbf{x}. \tag{7}$$

Once the representation coefficient $\hat{\alpha}$ is obtained, the combination of the dictionary and the weight vector is further considered to represent the tested pixel $\mathbf{x}$. Furthermore, we can compute residuals and obtain the final class label of the tested pixel as the same way did by (2) and (3).

C. Maximum Margin Projection

MMP is a linear and semisupervised DR method proposed by He *et al.* [40] based on *a priori* consistency assumption. Different from conventional DR algorithms such as PCA and LSDA, which only focus on global Euclidean structure, MMP could discover the local structure and characterize both geometrical and discriminant structures of the data manifold. The algorithm constructs a within-class graph G_w and a between-class graph G_b by exploiting both intraclass and interclass information, and finds a linear transformation matrix $\mathbf{A}$ that maps image data $\mathbf{X}$ to a subspace via

$$\mathbf{Y} = \mathbf{A}^T\mathbf{X} \tag{8}$$

where $\mathbf{Y} \in \Re^{d\times N}$, $\mathbf{X} \in \Re^{L\times N}$, and $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_d) \in \Re^{L\times d}$ with $d \ll L$. Thus, if map all the points of graph $G_w$and $G_b$to a line so that the points in G_w stay as closer as possible while the points in G_b keep as far as possible, the margin between relevant and irrelevant data samples will be maximized at each local neighborhood. Therefore, the two objective functions of MMP can be described as follows:

$$\min \sum_{ij} (y_i - y_j)^2 W_{w,ij} \tag{9}$$

$$\max \sum_{ij} (y_i - y_j)^2 W_{b,ij} \tag{10}$$

where W_w and W_b are the weight matrices of G_w and G_b defined as

$$W_{w,ij} = \begin{cases} \gamma, & \text{if } x_i \text{ and } x_j \text{ share the same label} \\ 1, & \text{if } x_i \text{ and } x_j \text{ is unlabeled} \\ & \text{but } x_i \in N_w(x_j) \text{ or } x_j \in N_w(x_i) \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

$$W_{b,ij} = \begin{cases} 1, & \text{if } x_i \in N_b(x_j) \text{ or } x_j \in N_b(x_i) \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

where $N(x_i)$ is defined as the set of k nearest neighbors of x_i. Furthermore, $N_b(x_i)$ contains the neighbors with different labels, and $N_w(x_i)$ is consist of all the remainder of the neighbors, with $N_b(x_i) \cup N_w(x_i) = N(x_i)$ and $N_b(x_i) \cap N_w(x_i) = \emptyset$. γ is a weighted adjustment parameter.

When the high-dimensional feature space is projected to a low-dimensional subspace, minimizing (9) attempts to ensure the nearest neighbors within-class closer after dimensionality reduction. Besides, maximizing (10) is to make samples outside the k-nearest neighbors as dispersive as possible. Formulas (9) and (10) can be further expressed by (13) and (14) as follows:

$$\min \frac{1}{2}\sum_{ij} (y_i - y_j)^2 W_{w,ij} = \mathbf{a}^T\mathbf{X}\Lambda_w\mathbf{X}^T\mathbf{a} - \mathbf{a}^T\mathbf{X}\mathbf{W}_w\mathbf{X}^T\mathbf{a} \tag{13}$$

$$\max \frac{1}{2}\sum_{ij} (y_i - y_j)^2 W_{b,ij} = \mathbf{a}^T\mathbf{X}\Lambda_b\mathbf{X}^T\mathbf{a} - \mathbf{a}^T\mathbf{X}\mathbf{W}_b\mathbf{X}^T\mathbf{a} \tag{14}$$

where $\mathbf{a}$is a projection vector. Λ_w and Λ_b are within-class and between-class diagonal matrices, respectively. With a constraint $\mathbf{a}^T\mathbf{X}\Lambda_w\mathbf{X}^T\mathbf{a} = 1$, (13) can be simplified as $\min_{\mathbf{a}} 1 - \mathbf{a}^T\mathbf{X}\mathbf{W}_w\mathbf{X}^T\mathbf{a}$ equivalent to $\max_{\mathbf{a}} \mathbf{a}^T\mathbf{X}\mathbf{W}_w\mathbf{X}^T\mathbf{a}$. Thus, the objective function of MMP can be further reduced as

$$\arg\max_{\mathbf{a}} \mathbf{a}^T\mathbf{X}[\beta(\Lambda_b - \mathbf{W}_b) + (1-\beta)\mathbf{W}_w]\mathbf{X}^T\mathbf{a} \tag{15}$$

where β is a constant with $\beta \in [0, 1]$. The vector $\mathbf{a}$ given in (15) can be obtained by a maximum eigenvalue solution. Finally, the process of MMP for dimensionality reduction can be implemented by $\mathbf{y}_i = \mathbf{A}^T\mathbf{x}_i$.

Algorithm 1: CDCRC.

Input: X, D, **x**
Step 1: Compute $\hat{\alpha}_m$ according to (16) with $\mathbf{D}_m$.
Step 2: Obtain residual image cube **R** according to (2).
Step 3: Identify the class label of **x** according to (3).
Output: *class*(**x**)

III. Proposed Methods

The original CRC adopts all the labeled samples from different classes in **D** for solving the sparse coefficients and representation. Although CRC considers the information of all classes in the process of solving sparse coefficients, it does not fully consider the discrimination information contained in each class, which may influence the representation, the calculation of residuals, and lead to misclassification. Accordingly, an alternative way, such as the class-dependent SRC [60] and the nearest subspace classifier [59], which estimates the sparse coefficients through within-class samples to represent the tested samples and generate residuals, was developed. Therefore, drawing on the abovementioned methods, this section applies CDCRC, which focuses on the initial representation information for each class into HSIC, introduces a dimensionality reduction method using the spectral information of dictionary to extract the class-specific features effectively, and proposes an effective decision mechanism by integrating the spatial information of a residual image.

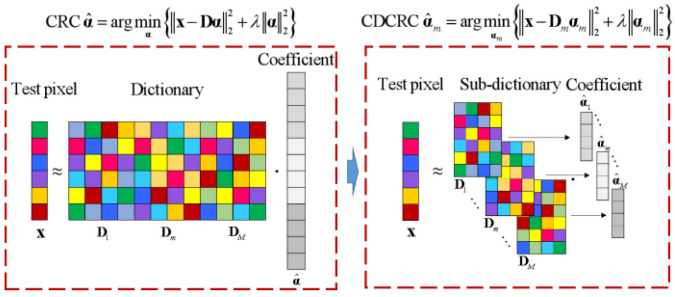


Fig. 2. Collaborative representation of CRC and CDCRC.

Algorithm 2: CMCRC.

Input: X, D, x
Step 1: Perform MMP on the dictionary **D** to obtain the projection matrix **A**.
Step 2: Reduce the dimensionality of the data **X** by (8) to obtain the dimension-reduced data **Y**.
Step 3: Implement CDCRC on **Y** for HSIC.
Output: *class*(**x**)

A. Class-Dependent CRC

Unlike CRC keeps **D** as a whole to obtain optimal vector $\hat{\alpha}$ for the tested pixel **x** in (6), CDCRC utilizes each subdictionary to obtain a representation coefficient $\hat{\alpha}_m$ of class m, which is given by

$$\hat{\alpha}_m = \arg\min_{\boldsymbol{\alpha}_m} \left\{ \|\mathbf{x} - \mathbf{D}_m\boldsymbol{\alpha}_m\|_2^2 + \lambda \|\boldsymbol{\alpha}_m\|_2^2 \right\}. \tag{16}$$

It should be noted that the main purpose of the CDCRC is to explore the interaction between the intrinsic information of single class and coefficient α. Thus, the key difference between original CRC and its improved method CDCRC illustrated in Fig. 2 is whether **D** is divided into M classes for the solution of $\hat{\alpha}$. As mention in (2), CDCRC utilizes each subdictionary $\mathbf{D}_m$ as well as corresponding weight vector $\hat{\alpha}_m$ to represent the tested pixel approximately. Also, the image classification can be finished by (3) that minimizes the residual error gained in (2). The process of CDCRC is summarized in Algorithm 1.

B. CDCRC Based on MMP

CDCRC defines a different way of obtaining the weight coefficient, which takes into account the global sample information and the local discrimination information of a single class. However, CDCRC still uses all bands of HSI for HSIC, and does not specifically select features that are helpful for the classification task. When the dimension of the HSI is large and does not match the number of training samples, the classification accuracy may be reduced, which is called the Hughes phenomenon. To address this issue, we choose a DR method originated from image retrieval, MMP, for effective feature extraction.

The characteristic of MMP is that it can minimize the compact graph within the class and maximize the separation graph between classes by differently constructing weights of different classes, so that the similar data points are closer and the heterogeneous data points are more distant after embedding the low-dimensional manifold of high-dimensional data. It cannot only use a small number of labeled samples to describe the manifold structure of the data, but also use a large number of unlabeled samples to avoid overfitting problems, thereby effectively solving the problem of high dimensionality and less labeled information in HSIC. More importantly, the projected samples used in MMP are very similar to the labeled training samples that form the dictionary in the CDCRC model, which is also the main reason why we combine MMP and CDCRC to design a new classification model called CMCRC.

Algorithm 3: UCMCRC.

Input: X, D, **x**, w, τ
Step 1: Obtain residual image cube **R** with regard to **Y** according to CMCRC. Let $\hat{\mathbf{R}}$ be the reciprocal image of **R**.
Step 2: Extract $w \times w$ pixel-sized window centered on (i, j) from $\hat{\mathbf{R}}$, $\hat{\mathbf{R}}_{i,j}$.
Step 3: Normalize the value of $\hat{\mathbf{R}}_{i,j}$ to form a probability map, $\mathbf{P}_{i,j}$.
Step 4: Calculate the SCP of the tested pixel for each class according to $\mathbf{P}_{i,j}$ by (14).
Step 5: Identify the class label of **x** with maximum SCP

$$class(\mathbf{x}) = \arg\max_m \mathrm{SCP}_m(\mathbf{x})$$

Output: *class*(**x**)

The proposed CMCRC algorithm first takes the labeled samples in the dictionary constructed by CDCRC and all the tested samples in the image as the labeled samples and unlabeled samples for MMP to map high-dimensional data into the low-dimensional data space. Then, the dimension-reduced image is classified by CDCRC to obtain the final classification map. Detailed steps of CMCRC is presented in Algorithm 2.

C. Union of CMCRC

CMCRC can effectively obtain spectral information through feature extraction, so that CDCRC can perform accurate CR in a typical feature domain and improve classification accuracy. However, both CDCRC and CMCRC belong to the classifiers in the spectral domain, only considering the spectral information of HSI. In fact, the tested pixel may produce spectral variation, that is to say, the same objects may have different spectral characteristics, which will negatively affect the classification performance of spectral classifiers, leading to in misclassification.

Several approaches [3],[7], [9], [10], [13], [15], [36], [37], [53], [54], [46]–[50], [53]–[58] such as edge-preserving filtering [9], Boltzmann entropy [29], SVM-MRF [44], JSRC [49], and JCRC [50], focusing on the utilization of spatial information are proposed in HSIC. In information theory, the improved Shannon entropy [51]–[53] and Boltzmann entropy [54]–[56] capture the configuration and composition information of hyperspectral data

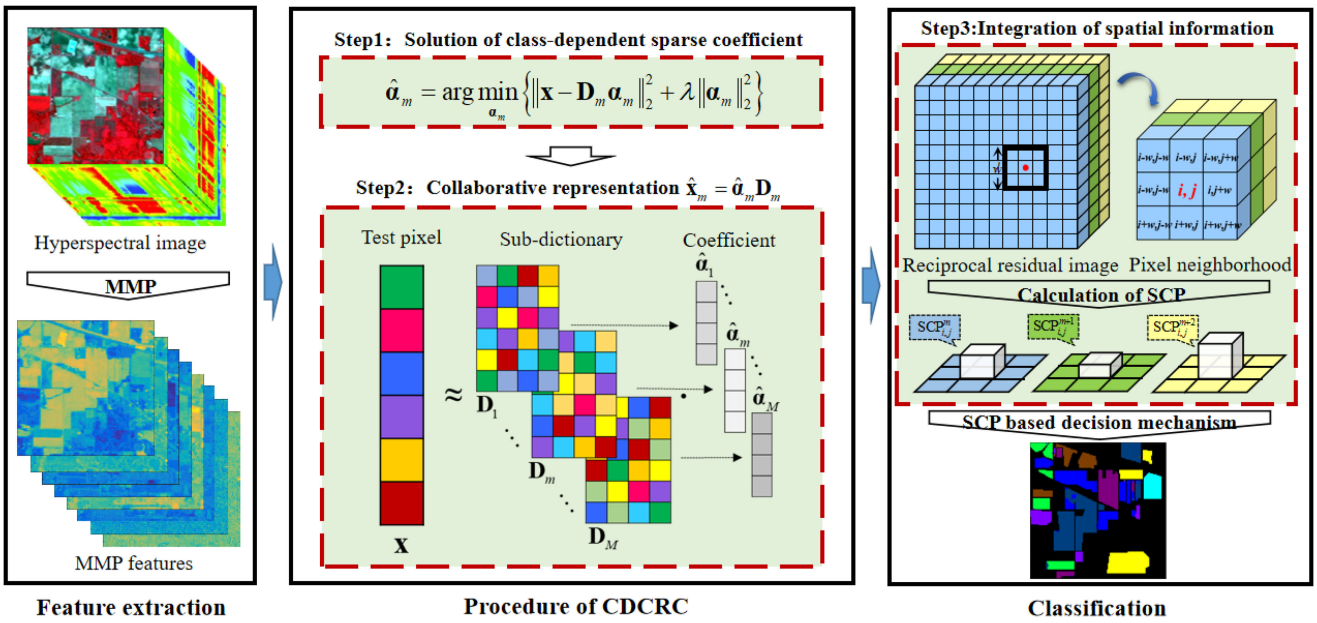


Fig. 3. Schematic diagram of the proposed UCMCRC model.

to obtain spatial information. In this article, the basic assumption is that the spatially adjacent pixels are more likely to belong to the same class. Therefore, the pixels with spectral variation will be corrected based on the spatial consistency assumption. It is noticeable that most CR-based spectral-spatial classifiers are too restrictive on spatial information. In practical applications, it is difficult and time-consuming to obtain the optimal sparsity coefficient under such strong constraints. Motivated by this point, a novel spectral-spatial framework with an extension of the aforementioned CMCRC model, called the UCMCRC, is further proposed.

The schematic diagram of UCMCRC is shown in Fig. 3. The implementation process can be divided into three steps. First, MMP makes use of the atoms in the dictionary of CDCRC to reduce the original HSI to a low-dimensional feature image $\mathbf{Y}$ by (8). Second, for the abovementioned feature image $\mathbf{Y}$, we can use CDCRC to solve the optimal coefficient value $\hat{\alpha}_m$ obtained from the subdictionary $\mathbf{D}_m$, and gain the residual value of the tested pixel according to (2). Finally, most of CR-based models determine the class label of the tested pixel according to the minimum residual error. Unfortunately, in some cases, the spectral variation may cause the issue that the class where the tested pixel gets the smallest residual is not correct. According to the assumption with spatial consistency, that is, the closer the pixels, the higher the probability of the same class, UCMCRC considers the impact of neighborhood pixels on central pixels and proposes a new concept of spatial cumulative probability (SCP). For a tested pixel $\mathbf{x}_{i,j}$, $\mathbf{R}_{i,j}$ represents a $w \times w$ pixel-sized window centered on it from residual image $\mathbf{R}$. $\hat{\mathbf{R}}_{i,j}$ is denoted as the reciprocal image of $\mathbf{R}_{i,j}$, and is normalized to generate a probability map, $\mathbf{P}_{i,j}$. In this case, we can use the SCP of the tested pixel to update and correct the residual value in spatial domain before the final decision. The SCP of $\mathbf{x}_{i,j}$ for class m is calculated as follows:

$$\mathrm{SCP}_m(\mathbf{x}_{i,j}) = \mathbf{p}_m + \tau \sum_{v=1}^{\widehat{w}^2} \mathbf{p}_m{}^{(v)} \tag{17}$$

where τ is a penalty parameter, which controls the significance of neighborhood. $\mathbf{p}_m$ represents the value of central pixel for class m in $\mathbf{P}_{i,j}$ and $\mathbf{p}_m{}^{(v)}$ is the neighborhood value of $\mathbf{p}_m$ where v denotes the index of vector in the neighborhood. If the label of $\mathbf{x}_{i,j}$ is determined only by residual $\mathbf{r}(\mathbf{x}_{i,j})$ in CMCRC with $\mathrm{r}_{i,j}^{m+1} < \mathrm{r}_{i,j}^{m+2} < \mathrm{r}_{i,j}^{m}$, the $\mathbf{x}_{i,j}$ belongs to class $(m+1)$. Inversely, if we consider the neighborhood pixels via SCP in UCMCRC, it is obvious that the class label of the tested pixel $\mathbf{x}_{i,j}$ should be $(m+2)$ with $\mathrm{SCP}_{i,j}^{m+2} > \mathrm{SCP}_{i,j}^{m} > \mathrm{SCP}_{i,j}^{m+1}$ as shown in Fig. 3. The process of UCMCRC is given in Algorithm 3.

IV. Experimental Results and Analysis

In this section, CDCRC, CMCRC, and UCMCRC are evaluated via three real hyperspectral datasets provided in Section IV-A. The settings of the related parameters including the number of training samples n, the dimension d, the window size of neighborhood w and the penalty parameter τ in UCMCRC are investigated in Section IV-B. For comparison, SVM and the SR-based models including SRC and CRC, are first utilized in the spectral domain. Furthermore, in the spatial domain, a postprocessing model for spatial information called SVM-based Markov random field (SVM-MRF), and two joint representation-based models, JSRC and JCRC, are analyzed. Finally, the classification results of various tested methods are discussed from the perspective of d, n, overall accuracy (OA), and average accuracy (AA) in Section IV-C. In the meantime, the computational complexity of CR-based models comprising CRC, CDCRC, CMCRC, and UCMCRC is also calculated.

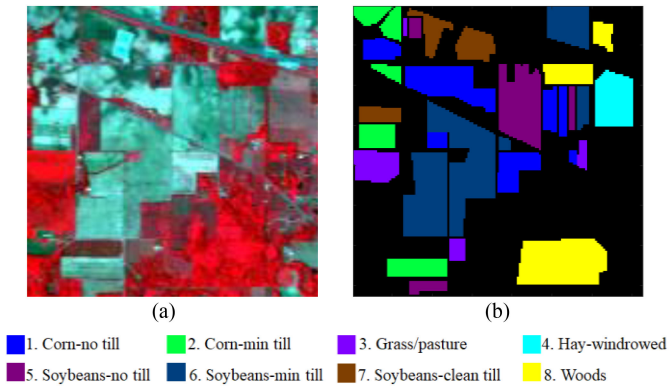


Fig. 4. Indian Pines dataset. (a) False-color composite image. (b) Reference map.

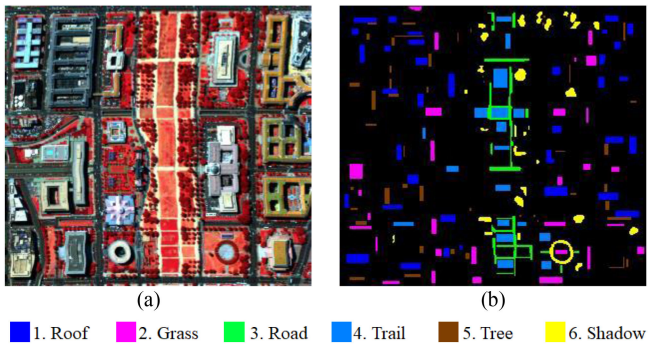


Fig. 5. Washington DC Mall dataset. (a) False-color composite image. (b) Reference map.

A. Datasets

1) Indian Pines: The Indian Pines scene was collected by the Airborne Visible/Infrared Imaging Spectrometer sensor with the size of $145 \times 145 \times 220$, the spatial resolution of 20 m, and the spectral range from 0.4 to 2.5 μm. Following [24], [61], [62], there are eight mutually exclusive classes used in this study so as to satisfy the requirement of sparsity and sample selection, which are Corn-no till, Corn-min till, Grass/Pasture, Hay-windrowed, Soybean-no till, Soybean-min till, Soybean-clean till, and Woods. Fig. 4(a) and (b) shows the false-color composite image and the reference map, respectively.

2) Washington DC Mall: The Washington DC Mall scene is a hyperspectral airborne data with 210 bands and 2 m spatial resolution. After removing noisy bands due to water absorption, the size of this scene is $280 \times 307 \times 191$ with spectra ranging from 0.4 to 2.4 μm. In order to satisfy the requirement of sparsity and sample selection, six classes of interest are considered for HSIC. The false-color image is shown in Fig. 5(a) and its reference map is shown in Fig. 5(b).

3) University of Pavia: The University of Pavia scene acquired by the reflective optics system imaging spectrometer sensor in Italy. Its size is $610 \times 340 \times 103$ with 12 bands removed due to the high noise and water absorption. Its spatial resolution is 1.3 m and the spectral range is from 0.43 to 0.86 μm. Its false-color image is shown in Fig. 6(a). There are nine classes provided in the reference data shown in Fig. 6(b).

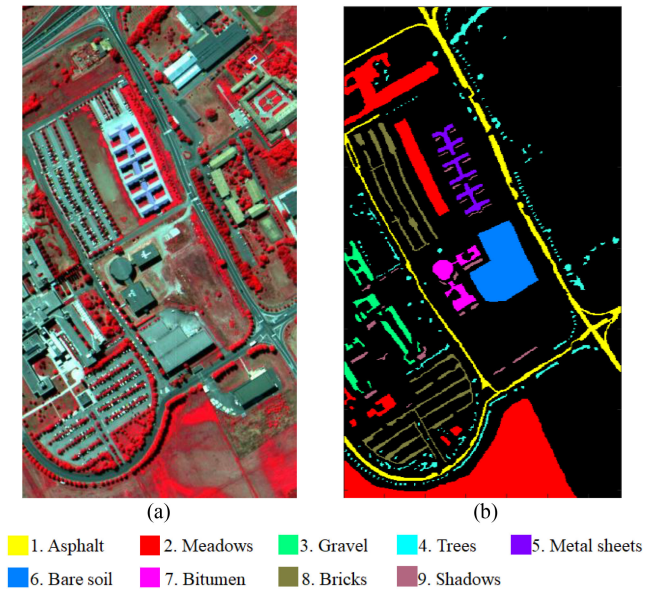


Fig. 6. University of Pavia dataset. (a) False-color composite image. (b) Reference map.

B. Parameter Setting

The relationship of four key parameters, n, d, w, and τ is investigated in detail for UCMCRC in our experiment. First, as can be seen from Fig. 7, the OA does not increase with an increasing n, especially for the Washington DC Mall and University of Pavia scenes.

Second, to a certain extent, the increase of data dimension d will improve OA, but too high d will lead to the deterioration of classification results, which also verifies the adverse impact of the mismatch between n and d on HSIC. With consideration of the balance between the extracted feature and the limited training samples, it is recommended that the maximum dimension should not be higher than $\lceil L/2 \rceil$, where L is the dimension of original HSI. Therefore, the maximum dimension of three datasets is 110, 100, and 60, respectively. When n, d, and OA are considered at the same time, it can be found that OA is very low and classification performance is poor with small n and d. Especially, the regularization parameters λ in the methods based on the representation models for three different datasets are set to the values ranging from 10^{-6} to 1 [63].

Third, when n and d are fixed, the different neighborhood window size w will also affect OA. In the experiments, w is assigned a total of seven different scales from 3 to 15 with a step size of 2. As a matter of fact, the optimal w will change according to datasets, which is related to the distribution characteristics of ground objects. As shown in Fig. 7, the OA mostly reaches a peak at first, and then decreases with the increase of w. For these three datasets, the optimal w is 5, 3, and 11, respectively. It indicates that different datasets generally have different optimal w according to the spatial coherence. Moreover, once w is too large, there will be overcorrection, resulting in misclassification.

Finally, given n, d, and w of the three datasets (the settings of n, d, and w for the Indian Pines, the Washington DC Mall, and the University of Pavia scenes are 100, 90, and 5, 50, 80, and 3,

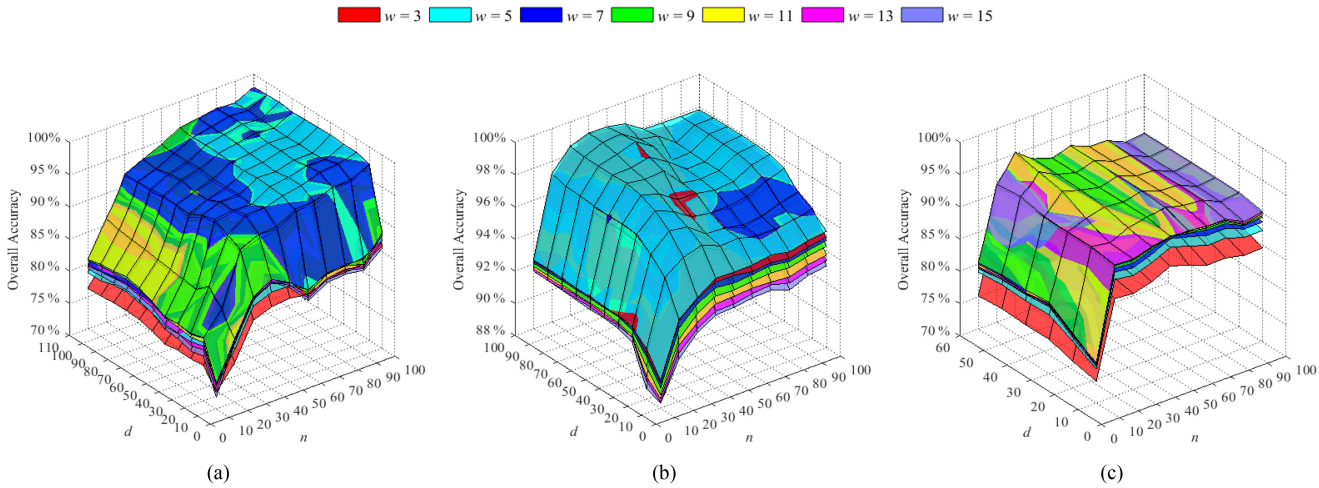


Fig. 7. Relationship between the number of training samples n, the dimension d, the size of neighborhood w, and the OA for the proposed UCMCRC using (a) Indian Pines dataset, (b) Washington DC Mall dataset, and (c) University of Pavia dataset.

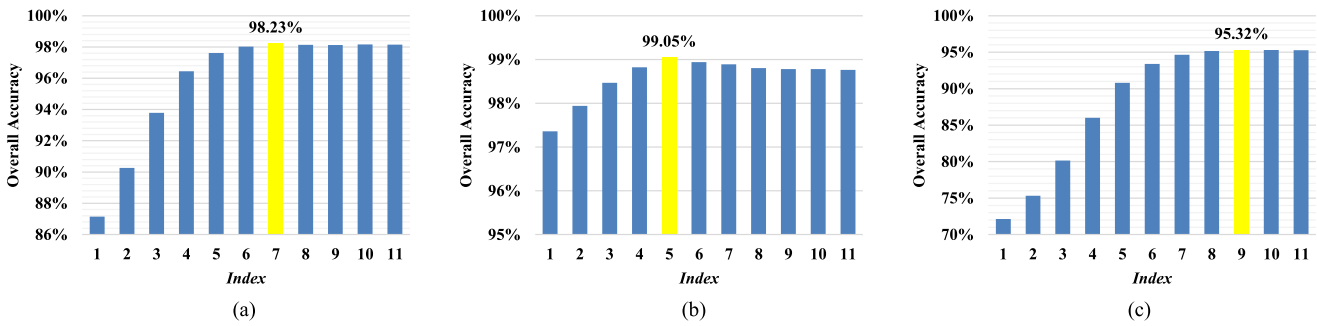


Fig. 8. Relationship between the penalty parameter τ and the OA with fixed n, d, and w for the proposed UCMCRC using (a) Indian Pines dataset, (b) Washington DC Mall dataset, and (c) University of Pavia dataset.

30, 60, and 11, respectively), Fig. 8 shows the influence of the penalty parameter τ in SCP on the classification performance of UCMCRC. For the three datasets, τ is consistent containing 11 different values with $\tau = \{2^{\Delta^{\text{index}}}/(w^2-1),\ \Delta^{\text{index}} = -2, -1, 0, 1, 2, 3, 4, 5, 6, 7, 8\}$ where index is range from 1 to 11. Fig. 8(a)–(c) shows that UCMCRC achieves the best OA when τ is equal to 2/3, 1/2, and 8/15 where the *index* is 7, 5, and 9.

C. Results Analysis

In experiments, the classifiers are evaluated by different numbers of training samples, of which are randomly selected (from 10 to 100 samples per class for the Indian Pines, the Washington DC Mall, and the University of Pavia scenes) for dimensionality reduction and dictionary construction. In addition, all the remaining samples as the test samples are used for validation. Tables I –III show the OA obtained by the nine tested classifiers. The best OA of each classifier is highlighted and the corresponding classification maps is shown in Figs. 9 –11 for three real datasets. It should be mentioned that under the different number of training samples, the corresponding dimensions of CMCRC and UCMCRC are different when the classification accuracy is optimal. Finally, the comparisons of computing time for four CRCs including CRC, CDCRC, CMCRC, and UCMCRC are shown in Fig. 12. As the results mentioned above, three conclusions can be drawn.

1) Comparison of the Spectral Classifiers: In the Indian Pines scene, the OA increases as n increases, and there is a positive relationship between them. This trend is obvious when n is small. As can be seen from Table I, when n is 100, both SRC and CRCs perform better than SVM whose performance is worst, which verifies the effectiveness of the representation-based framework in HSIC. It is worth noting that CRC and CDCRC do not have an advantage over SRC. Nevertheless, the mechanism of obtaining sparse coefficients by class is still effective, which can be demonstrated by CDCRC being superior to CRC.

After the data space characterized by MMP, the classification performance of CMCRC is greatly improved, with its OA being 3.71%, 14.15%, and 5.27% higher than that of SRC, CRC, and CDCRC, respectively. In addition, CMCRC achieved the highest AA among the spectral classifiers, which is 1.93%, 9.25%, and 4.54% higher than SRC, CRC, and CDCRC, respectively.

TABLE I
OVERALL ACCURACIES (IN PERCENT) OBTAINED BY THE DIFFERENT TESTED METHODS FOR THE INDIAN PINE SCENE

n	SVM	SRC	CRC	CDCRC	CMCRC	SVM-MRF	JSRC	JCRC	UCMCRC
10	42.34%	67.98%	59.28%	66.78%	71.76%	50.34%	71.39%	69.19%	81.86%
20	56.89%	69.27%	61.96%	68.45%	74.27%	74.14%	80.48%	78.76%	86.07%
30	60.68%	74.33%	66.03%	73.86%	80.67%	76.29%	82.69%	81.00%	93.63%
40	64.17%	75.82%	68.63%	74.81%	81.79%	79.01%	86.49%	84.86%	94.61%
50	64.77%	78.18%	71.89%	76.11%	83.38%	78.90%	89.37%	87.24%	94.57%
60	67.75%	80.16%	71.64%	78.63%	83.58%	84.39%	90.20%	88.99%	95.99%
70	68.97%	80.60%	71.78%	79.57%	84.32%	84.15%	90.90%	89.07%	96.41%
80	70.36%	81.11%	72.22%	80.67%	85.24%	85.99%	91.41%	89.74%	97.00%
90	70.52%	81.53%	72.21%	81.04%	86.48%	86.34%	92.12%	90.75%	97.23%
100	**71.76%**	**83.64%**	**73.20%**	**82.05%**	**87.35%**	**87.19%**	**93.19%**	**92.85%**	**98.23%**

In all cases, various from 10 to 100 labeled samples per class were used. The best results are highlighted in bold typeface.

TABLE II
OVERALL ACCURACIES (IN PERCENT) OBTAINED BY THE DIFFERENT TESTED METHODS FOR THE WASHINGTON DC MALL SCENE

n	SVM	SRC	CRC	CDCRC	CMCRC	SVM-MRF	JSRC	JCRC	UCMCRC
10	88.43%	90.25%	86.15%	89.07%	90.32%	89.95%	84.63%	84.32%	92.31%
20	88.67%	93.63%	89.69%	92.51%	94.69%	94.18%	88.33%	87.40%	97.21%
30	90.41%	95.29%	91.67%	94.75%	96.15%	93.89%	91.71%	90.69%	98.49%
40	92.26%	96.12%	91.71%	94.62%	**97.16%**	96.01%	92.53%	92.55%	98.89%
50	92.56%	96.14%	93.00%	**94.85%**	96.42%	96.28%	93.90%	94.55%	**99.05%**
60	92.86%	96.40%	93.34%	94.17%	96.03%	95.90%	94.58%	94.72%	98.63%
70	93.83%	96.72%	93.76%	93.92%	95.71%	96.54%	95.33%	95.10%	98.00%
80	94.20%	96.95%	94.03%	94.14%	95.41%	96.90%	95.84%	95.97%	97.85%
90	95.26%	**97.59%**	94.76%	94.55%	96.31%	97.69%	96.33%	95.67%	97.83%
100	**95.36%**	97.49%	**94.98%**	94.49%	96.25%	**98.31%**	**96.38%**	**96.18%**	97.87%

In all cases, various from 10 to 100 labeled samples per class were used. The best results are highlighted in bold typeface.

TABLE III
OVERALL ACCURACIES (IN PERCENT) OBTAINED BY THE DIFFERENT TESTED METHODS FOR THE UNIVERSITY OF PAVIA SCENE

n	SVM	SRC	CRC	CDCRC	CMCRC	SVM-MRF	JSRC	JCRC	UCMCRC
10	51.36%	58.22%	52.45%	62.21%	66.24%	49.54%	51.57%	69.22%	81.02%
20	73.05%	71.74%	62.73%	70.85%	75.84%	81.66%	66.97%	75.17%	92.29%
30	79.75%	74.59%	66.84%	73.27%	76.20%	85.75%	68.09%	75.43%	**95.32%**
40	82.44%	75.50%	**68.66%**	75.90%	76.95%	88.88%	73.73%	77.87%	93.23%
50	85.25%	77.27%	68.47%	76.94%	79.31%	90.87%	78.90%	79.12%	92.36%
60	86.77%	78.06%	67.27%	77.27%	**79.69%**	91.92%	80.91%	80.84%	93.15%
70	88.15%	78.68%	67.50%	77.05%	77.76%	**93.10%**	83.77%	82.73%	91.99%
80	88.01%	79.51%	67.59%	76.53%	76.91%	92.61%	84.57%	84.18%	91.87%
90	**88.55%**	**80.00%**	68.14%	**77.53%**	76.46%	93.05%	85.74%	85.52%	90.96%
100	88.01%	79.77%	67.75%	77.25%	76.68%	92.56%	**86.49%**	**86.95%**	91.03%

In all cases, various from 10 to 100 labeled samples per class were used. The best results are highlighted in bold typeface.

This indicates that the combination of MMP and CDCRC can effectively solve the band redundancy under limited training samples and bring better characteristic representation results to improve the classification. For the other two datasets, among the five spectral classifiers, the OA value of CMCRC ranks second and third, respectively, as does AA. Although the classification results of CMCRC are not the best, it is still competitive with lower computational cost than other classifiers.

2) Comparison of the Spectral-Spatial Classifiers: Compared with SVM, SVM-MRF with spatial information can

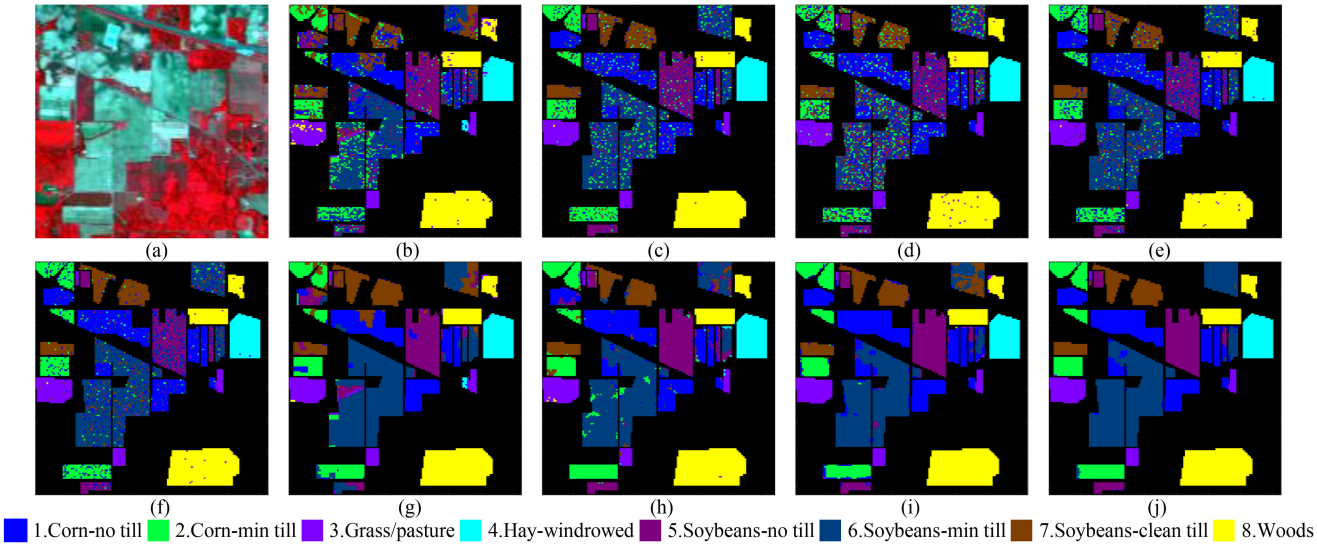


Fig. 9. Classification maps of best overall accuracy obtained by the different tested methods for the Indian Pines dataset. In all cases, various from 10 to 100 labeled samples per class were used. The overall/average accuracies are given in the parentheses. (a) False-color composite image. (b) SVM (71.76%/76.59%). (c) SRC (83.64%/87.51%). (d) CRC (73.20%/80.19%). (e) CDCRC (82.05%/84.90%). (f) CMCRC (87.35%/89.44%). (g) SVM-MRF (87.19%/89.36%). (h) JSRC (93.19%/94.69%). (i) JCRC (92.85%/94.92%). (j) UCMCRC (98.23%/98.48%).

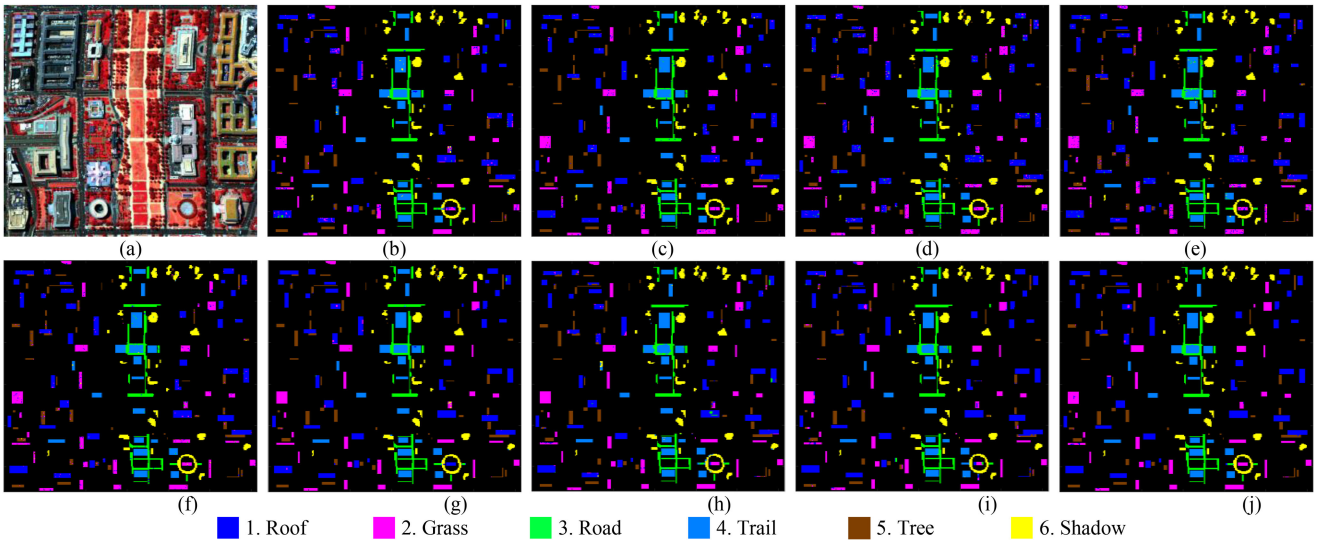


Fig. 10. Classification maps of best overall accuracy obtained by the different tested methods for the Washington DC Mall dataset. In all cases, various from 10 to 100 labeled samples per class were used. The overall/average accuracies are given in the parentheses. (a) False-color composite image. (b) SVM (95.36%/96.15%). (c) SRC (97.59%/97.73%). (d) CRC (94.98%/95.74%). (e) CDCRC (94.85%/94.85%). (f) CMCRC (97.16%/97.01%). (g) SVM-MRF (98.31%/98.35%). (h) JSRC (96.38%/96.49%). (i) JCRC (96.18%/95.44%). (j) UCMCRC (99.05%/99.10%).

remarkably improve the classification accuracy on the three data sets and obtain more homogeneous classification maps. It is proved that spatial information is helpful to HSIC. Similarly, this can be confirmed by the classification results of JSRC and JCRC whose OAs are superior to those of spectral classifiers SRC and CRC in the Indian Pines and the University of Pavia scenes. Regarding the result of the Washington DC Mall scene, the reason why the OA of JSRC is lower than that of SRC is the misclassification caused by the contradiction between strong sparsity constraint and fragment distribution.

Compared with the established CMCRC model, the successively proposed UCMCRC has significant improvements in all datasets, which verifies that introducing a spatial item, SCP, into the final decision mechanism to adjust and correct the residual information of the tested pixel is effective for HSIC.

In general, compared with other classifiers used for testing, UCMCRC obtained the best classification results on the three datasets with the OA of 98.23%, 99.05%, and 95.32%, respectively, which is 5.04%, 0.74%, and 2.22% higher than the suboptimal classifiers (JSRC, SVM-MRF, and SVM-MRF),

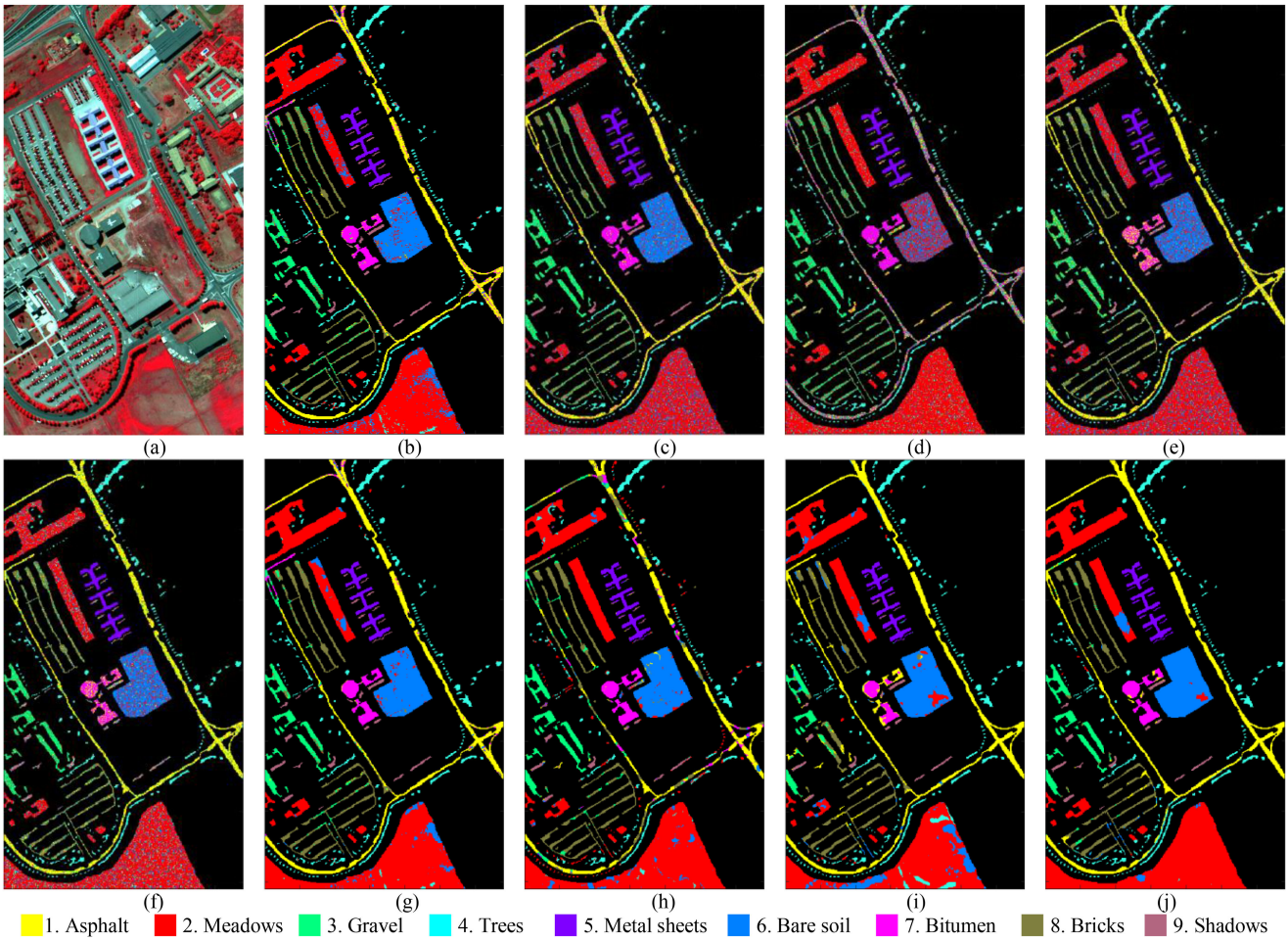
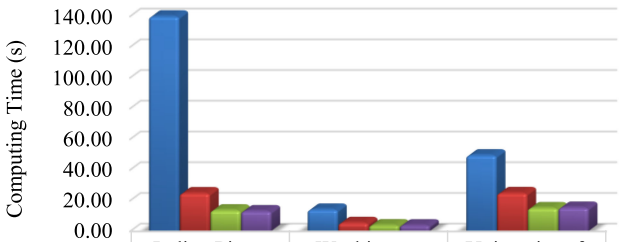


Fig. 11. Classification maps of best overall accuracy obtained by the different tested methods for the University of Pavia data set. In all cases, various from 10 to 100 labeled samples per class were used. The overall/average accuracies are given in the parentheses. (a) False-color composite image. (b) SVM (88.55%/91.13%). (c) SRC (80.00%/85.58%). (d) CRC (68.66%/67.48%). (e) CDCRC (77.53%/77.56%). (f) CMCRC (79.69%/80.01%). (g) SVM-MRF (93.10%/94.74%). (h) JSRC (86.49%/84.43%). (i) JCRC (86.95%/83.44%). (j) UCMCRC (95.32%/92.58%).



	Indian Pines	Washington DC Mall	University of Pavia
CRC	141.20	14.24	49.61
CDCRC	24.93	5.59	25.02
CMCRC	13.74	4.00	15.20
UCMCRC	13.21	3.92	15.50

Fig. 12. Comparison of computing time for four CR-based methods for three datasets. The numbers of the labeled samples correspond to the cases are 100, 50, and 30.

and 25.03%, 4.07%, and 26.66% higher than original CRC classifier. Furthermore, comparing the AA of all spectral-spatial classifiers, UCMCRC obtained the best value in the Indian Pines and Washington DC Mall datasets. For the University of Pavia dataset, UCMCRC also has a good classification result, and its AA is only lower than SVM-MRF. Therefore, the effectiveness and stability of the proposed spectral-spatial classification method for HSIs based on residual information are verified.

3) Computational Complexity of CRCs: Fig. 12 illustrates the computational time acquired by four CRCs including CRC, CDCRC, CMCRC, and UCMCRC, where the settings of n for three datasets are 100, 50, and 30, respectively. To be noticed that the time consumption of the original CRC model is largest compared with the other three proposed models. Most importantly, the time-consuming of the original CRC model is 5.6 times, 2.5 times, and 1.9 times than that of the established CDCRC model, which demonstrates that solving the coefficients class by class can effectively improve the speed of the classifier. For a clearer comparison, we have listed the specific values below the bar chart in Fig. 12. The time cost of CMCRC (13.74 s, 4 s, and 15.2 s) and UCMCRC (13.21 s, 3.92 s, and 15.5 s) is almost consistent both of which are lower than that of CDCRC (24.93 s, 5.59 s, and 25.02 s), which indicated that MMP could effectively reduce the complexity of data in the process of data

preprocessing and ultimately reduce the time consumption of the classifiers. In a word, CDCRC, CMCRC, and UCMCRC achieve relatively low computational cost, which confirms their practicability in HSIC.

V. Conclusion

This article proposes a novel spectral-spatial classification framework based on CDCRC and MMP. Through integrating spatial information of the pixel, the classification performance of CDCRC using MMP-reduced data improved dramatically. In general, the main advantages of UCMCRC can be summarized into the following three points.

1) The CDCRC can solve sparse coefficients more accurately and efficiently.
2) The organic combination of class-specific MMP and CDCRC in CMCRC reduces time consumption and obtains better classification performance.
3) The decision mechanism of SCP based on the residual accumulation probability image has faster processing speed and higher classification accuracy.

Experimental results on three real hyperspectral datasets demonstrate the superiority of the proposed method over other related methods.

References

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

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

[3] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, Mar. 2013.

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

[5] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 29–43, Jan. 2002.

[6] X. Jia and J. A. Richards, "Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 1, pp. 538–542, Jan. 1999.

[7] L. Fang, N. He, S. Li, A. J. Plaza, and J. Plaza, "A new spatial–spectral feature extraction method for hyperspectral images using local covariance matrix representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3534–3546, Jun. 2018.

[8] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

[9] X. Kang, S. Li, and J. A. Benediktsson, "Spectral–spatial hyperspectral image classification with edge-preserving filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2666–2677.

[10] C. Yu, R. Han, M. Song, C. Liu, and C. Chang, "A simplified 2D-3D CNN architecture for hyperspectral image classification based on spatial–spectral fusion," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2485–2501, Apr. 2020.

[11] Z. Li *et al.*, "Deep multilayer fusion dense network for hyperspectral image classification," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1258–1270, Mar. 2020.

[12] W. Li, J. Liu, and Q. Du, "Sparse and low rank graph-based discriminant analysis for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 4094–4105, Jul. 2016.

[13] C. Chen, N. Chen, and J. Peng, "Nearest regularized joint sparse representation for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 424–428, Mar. 2016.

[14] H. Yuan and Y. Tang, "Sparse representation based on set-to-set distance for hyperspectral image classification," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2464–2472, Jun. 2015.

[15] H. Yu *et al.*, "Global spatial and local spectral similarity-based manifold learning group sparse representation for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3043–3056.

[16] X. Chen and X. Lin, "Big data deep learning: Challenges and perspectives," *IEEE Access*, vol. 2, pp. 514–525.

[17] W. Zhao and S. Du, "Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544–4554, Aug. 2016.

[18] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, Sep. 2019.

[19] V. Singhal and A. Majumdar, "Row-sparse discriminative deep dictionary learning for hyperspectral image classification," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 11, no. 12, pp. 5019–5028, Dec. 2018.

[20] P. Liu, H. Zhang, and K. B. Eom, "Active deep learning for classification of hyperspectral images," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 10, no. 2, pp. 712–724, Feb. 2017.

[21] L. Gao *et al.*, "A new low-rank representation based hyperspectral image denoising method for mineral mapping," *Remote Sens*, vol. 9, no. 11, Nov. 2017, Art. no. 1145.

[22] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification via kernel sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 217–231, Jan. 2013.

[23] W. Li and Q. Du, "A survey on representation-based classification and detection in hyperspectral remote sensing imagery," *Pattern Recognit. Lett.*, vol. 83, no. 1, pp. 115–123, Nov. 2016.

[24] W. Li, Q. Du, F. Zhang, and W. Hu, "Hyperspectral image classification by fusing collaborative and sparse representations," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4178–4187, Sep. 2016.

[25] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 1, pp. 55–63, Jan. 1968.

[26] C.-I. Chang, Q. Du, T.-L. Sun, and M. L. G. Althouse, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 6, pp. 2631–2641, Nov. 1999.

[27] C.-I. Chang and S. Wang, "Constrained band selection for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1575–1585, Jun. 2006.

[28] M. Song, X. Shang, Y. Wang, C. Yu, and C. Chang, "Class information-based band selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8394–8416, Nov. 2019.

[29] P. Gao, P. Gao, J. Wang, H. Zhang, and Z. Li, "Boltzmann entropy-based unsupervised band selection for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 3, pp. 462–466, Mar. 2019.

[30] P. Gao, H. Zhang, D. Jia, C. Song, C. Cheng, and S. Shen, "Efficient approach for computing the discrimination ratio-based variant of information entropy for image processing," *IEEE Access*, vol. 8, pp. 92552–92564.

[31] H. Yang, Q. Du, and G. Chen, "Particle swarm optimization-based hyperspectral dimensionality reduction for urban land cover classification," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 544–554, Apr. 2012.

[32] J. Bai, S. Xiang, L. Shi, and C. Pan, "Semisupervised pair-wise band selection for hyperspectral images," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2798–2813, Jun. 2015.

[33] M. Zhang, M. Gong, Y. Mao, J. Li, and Y. Wu, "Unsupervised feature extraction in hyperspectral images based on wasserstein generative adversarial network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 5, pp. 2669–2688.

[34] J. Jiang, J. Ma, C. Chen, Z. Wang, Z. Cai, and L. Wang, "SuperPCA: A superpixelwise PCA approach for unsupervised feature extraction of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4581–4593, Aug. 2018.

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

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

[37] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Netw*, vol. 13, no. 4/5, pp. 411–430, Jun. 2000.

[38] L. Ma, M. M. Crawford, X. Yang, and Y. Guo, "Local-manifold-learning-based graph construction for semisupervised hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2832–2844.

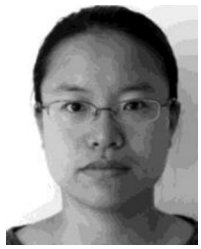
[39] D. Cai, X. He, and J. Han, "Document clustering using locality preserving indexing," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 12, pp. 1624–1637, Dec. 2005.
[40] X. He, D. Cai, and J. Han, "Learning a maximum margin subspace for image retrieval," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 2, pp. 189–201, Feb. 2008.
[41] J. M. Bioucas-Dias *et al.*, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.
[42] A. Zare and K. Ho, "Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 95–104, Jan. 2014.
[43] L. Shu, K. McIsaac, and G. R. Osinski, "Learning spatial–spectral features for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5138–5147, Sep. 2018.
[44] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson, "SVM- and MRF-based method for accurate classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 736–740, Oct. 2010.
[45] X. Shang, M. Song, and C.-I. Chang, "An iterative random training sample selection approach to constrained energy minimization for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, to be published. doi: 10.1109/LGRS.2020.3005078.
[46] H. Yu *et al.*, "Spectral-spatial hyperspectral image classification using subspace-based support vector machines and adaptive Markov random fields," *Remote Sens.*, vol. 8, no. 4, Apr. 2016, Art. no. 355.
[47] M. Song, X. Shang, and C.-I. Chang, "3-D receiver operating characteristic analysis for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 8093–8115, Nov. 2020.
[48] Q. Gao, S. Lim, and X. Jia, "Spectral–spatial hyperspectral image classification using a multiscale conservative smoothing scheme and adaptive sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7718–7730, Oct. 2019.
[49] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.
[50] W. Li and Q. Du, "Joint within-class collaborative representation for hyperspectral image classification," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2200–2208, Jun. 2014.
[51] Z. Li and P. Huang, "Quantitative measures for spatial information of maps," *Int. J. Geographic Inf.*, vol. 16, no. 7, pp. 699–709, Nov. 2002.
[52] T. Zhang, C. Cheng, and P. Gao, "Permutation entropy-based analysis of temperature complexity spatial-temporal variation and its driving factors in China," *Entropy*, vol. 21, no. 10, Oct. 2019, Art. no. 1001.
[53] P. Gao, Z. Li, and H. Zhang, "Thermodynamics-based evaluation of various improved Shannon entropies for configurational information of gray-level images," *Entropy*, vol. 20, no. 1, Jan. 2018, Art. no. 19.
[54] S. Cushman, "Calculating the configurational entropy of a landscape mosaic," *Landscape Ecol.*, vol. 31, no. 3, pp. 481–489, Nov. 2016.
[55] P. Gao and Z. Li, "Computation of the Boltzmann entropy of a landscape: A review and a generalization," *Landscape Ecol.*, vol. 34, no. 2, pp. 2183–2196.
[56] P. Gao and Z. Li, "Aggregation-based method for computing absolute Boltzmann entropy of landscape gradient with full thermodynamic consistency," *Landscape Ecol.*, vol. 34, no. 4, pp. 1837–1847, Jul. 2019.
[57] L. Sun, C. Ma, Y. Chen, H. J. Shim, Z. Wu, and B. Jeon, "Adjacent superpixel-based multiscale spatial-spectral kernel for hyperspectral classification," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 12, no. 6, pp. 1905–1919, Jun. 2019.
[58] L. Sun *et al.*, "Low rank component induced spatial-spectral kernel method for hyperspectral image classification," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 30, no. 10, pp. 3829–3842, Oct. 2020.
[59] L. Zhang, W. Zhou, and B. Liu, "Nonlinear nearest subspace classifier," in *Proc. 18th Int. Conf. Neural Inf. Process.*, Nov. 2011, pp. 638–645.
[60] M. Cui and S. Prasad, "Class-dependent sparse representation classifier for robust hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2683–2695.
[61] H. Yu, L. Gao, W. Liao, B. Zhang, A. Pižurica, and W. Philips, "Multiscale superpixel-level subspace-based support vector machines for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 2142–2146, Nov. 2017.
[62] W. Li, Y. Zhang, N. Liu, Q. Du, and R. Tao, "Structure-aware collaborative representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 7246–7261, Sep. 2019.
[63] H. Yu, X. Shang, X. Zhang, L. Gao, M. Song, and J. Hu, "Hyperspectral image classification based on adjacent constraint representation," *IEEE Geosci. Remote Sens. Lett.*, to be published. doi: 10.1109/LGRS.2020.2982706.

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

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

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

Her research includes hyperspectral image classification, band selection, and applications.



Meiping Song (Member, IEEE) received the Ph.D. degree in computer application technology from the College of Computer Science and Technology, Harbin Engineering University, Harbin, China, in 2006.

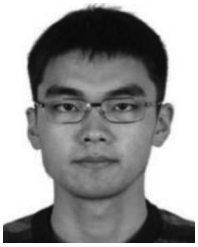
From 2013 to 2014, she was a Visiting Associate Research Scholar with the University of Maryland, Baltimore County, Baltimore, MD, USA. She is currently a Professor with the Information Science and Technology College, Dalian Maritime University, Dalian, China. Her research includes remote sensing and hyperspectral image processing.

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

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

Tong Jiao received the B.S. degree in geographic information system from Beijing Forestry University, Beijing, China, in 2011 and the M.S. degree in cartography and geographic information system from the Institute of Geographic Sciences and Natural Resources Research, Beijing, China, in 2014. She is currently working toward the Ph.D. degree with Clark University, Worcester MA, USA.

Her research includes hyperspectral image processing, classification and applications.



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

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

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

He is currently a Full Professor and the Deputy Director of the Aerospace Information Research Institute, CAS, where he has been leading key scientific projects in the area of hyperspectral remote sensing for more than 20 years. His research interests include the development of mathematical and physical models and image processing software for the analysis of hyperspectral remote sensing data in many different areas. He has developed five software systems in the image processing and applications. He has authored more than 300 publications, including more than 170 journal papers. He has edited six books/contributed book chapters on hyperspectral image processing and subsequent applications.

Dr. Zhang creative achievements were rewarded ten important prizes from Chinese government, and special government allowances of the Chinese State Council. He was the recipient of the National Science Foundation for Distinguished Young Scholars of China, in 2013, and the 2016 Outstanding Science and Technology Achievement Prize of the Chinese Academy of Sciences, the highest level of Awards for the CAS scholars. He is currently serving as the Associate Editor for IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING. He has been serving as Technical Committee Member of IEEE Workshop on Hyperspectral Image and Signal Processing, since 2011, and as the president of hyperspectral remote sensing committee of China National Committee of International Society for Digital Earth, since 2012. He is the Student Paper Competition Committee Member in IGARSS from 2015–2019.