Adaptive Hyperspectral Image Classification Based on the Fusion of Manifolds Filter and Spatial Correlation Features

JIANGSHANG LIAO AND LIGUO WANG, (Member, IEEE)

1. College of Information and Communications Engineering, Harbin Engineering University, Harbin 150001, China
2. College of Rail Transit, Guangdong Communication Polytechnic, Guangzhou 510650, China
3. College of Information and Communications Engineering, Dalian Nationalities University, Dalian 116600, China

Corresponding author: Liguo Wang (wangliguo@hrbeu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61675051, in part by the Natural Science Foundation of Guangdong Province under Grant 2018A030313195, and in part by the Science and Technology Project of Guangzhou under Grant 201804010262.

ABSTRACT In recent decades, the studies that obtain abundant spatial texture features, using a wide variety of filters for improving the performance of hyperspectral image (HSI) classification, have become a hotspot. However, the classification methods based on various filters are easy to fall into local feature extraction and neglect informative spatial correlation features. This paper presents an adaptive HSI classification method based on the fusion of adaptive manifold filter and spatial correlation feature (AMSCF). In which we use an adaptive manifold filter to extract spatial texture features, and use the domain transform normalized convolution filter and interpolated convolution filter to obtain the spatial correlation features. Besides, the spatial texture features and two correlation features are separately fused and classified by Large Margin Distribution Machine (LDM) to obtain the best classification. The experimental results demonstrate that the proposed AMSCF method is better than the other classification methods.

INDEX TERMS Hyperspectral image, large margin distribution machine, adaptive manifold filter, domain transform convolution filter, classification.

I. INTRODUCTION

Hyperspectral image (HSI) can carry comprehensive spectral characteristics that have been successfully applied to many remote-sensing applications fields, such as agriculture [1], [2], land area [3], [4], rock mines [5], [6], and waters [7], [8], playing an important role in the world.

Many scholars have applied a variety of machine learning methods to HSI and proposed various classifiers to improve classification accuracy, such as support vector machine (SVM) [9], [10]. Considering that SVM is a single interval and cannot represent the interval distribution of all training data, Zhang et al. propose a Large margin distribution machine (LDM) [11] that can maximize the marginal mean and minimize the margin variance to construct a new hyperplane for classification. In addition, many investigators attempt to progress the classification using LDM. For instance, Zhan et al. apply LDM to the hyperspectral image application and give an LDM-FL method with the spatial information extracted by the recursive filter (RF) to enhance the classification [12]. Moreover, Liao et al. perfect HSI classification using LDM that combines spatial information obtained by the bilateral filter (BF) and linear spatial correlation information [13]. Similarly, this paper also adopts LDM to enhance HSI classification.

In recent years, many investigators have proposed a wide variety of classification methods based on spatial information extraction, such as image segmentation [14], morphological feature extraction [15], and Markov random fields [16]. For example, Mehta et al. propose an approach for assimilating segmentation, clustering, and local band selection within its framework [17]. Li et al. merge the regions into a single classification framework by combining semi-supervised learning and conditional random fields [18]. To extract spatial features more deeply, many investigators use the superpixel-based segmentation method to obtain spatial information for classification. Jiang et al. propose a SuperPCA method that incorporates the spatial context information into the
dimensionality reduction and obtains the low-dimensional features by performing PCA on each homogeneous region partitioned by superpixel segmentation [19]. Chen et al. propose a SuperBF method that divides an HSI into some homogeneous regions obtained by superpixel segmentation and obtains spatial texture information on each homogeneous region by BF [20]. Xu et al. present a hypergraph-based low-rank subspace clustering to obtain a more complex manifold structure that is meaningful both spatially and spectrally for HSI classification [21]. Besides, morphological feature extraction can be applied to HSI classification. For instance, Zhou et al. propose a method based on morphological component analysis for decomposing original hyperspectral data into its morphological components. In each feature domain, the combination of active learning and semi-supervised learning magnifies the superpixel-based training data set [22]. Liao et al. utilize morphological profile filter and domain transform normalized convolution filter (DTNCF) [23] to extract spatial information and inputs the extracted information into a SVM, and thus implements a two-step optimization in the classification process [24]. Imani et al. propose a morphology-based feature extraction and classification framework, which includes the local neighborhood information in a spatial window for extension of the training set and tries to preserve the data structure in spectral-spatial feature space [25]. Besides, some researchers apply Markov random field (MRF) to HSI classification, e.g., Cao et al. present a supervised classification method that integrates spectral and spatial information into a unified Bayesian framework. Combined with MRF, a convolutional neural network is applied to obtain the spatial feature for enhancing the HSI classification [26]. Ghasrodashti et al. present a spectral-spatial classification method for the classification. The extended hidden MRF adopts Poisson distribution and weighted mean regularization to overcome Poisson noise, increasing the local neighborhood energy consistency of the classification with iterative maximum likelihood [27]. Although the extraction of spatial features for classification obtained good results in the past, most of them are a single spatial feature. Therefore, the spatial features of HSI have not been fully extracted and cannot be suitable for various HSI data sets.

To obtain better spatial information, a multitude of texture filters are employed to extract spatial information for enhancing HSI classification [28], such as Gabor filter (GF) [29]. Particularly, the two-dimensional GF is used to extract spatial features from HSI for the nearest regularized subspace classifier [30]. Moreover, Li et al. project Gabor feature of HSI onto the kernel induced space through the weighted summation composite kernel technique for SVM classifier [31]. Jia et al. provide a multi-feature learning algorithm that can learn proper dictionaries for Gabor feature and three other features until a stable code can be obtained in SVM [32]. Kang et al. use the first three components of principal component analysis (PCA) [33] to extract the spatial feature, forming a fusion feature for deep network on hyperspectral images by GF; and Ref. [34] presents a GF approach for HSI classification. Jia et al. propose a cascade superpixel regularized Gabor feature fusion method, which obtains spatial features by Gabor filters from the original HSI and exploits the decision information with SVM, then applies the quadrant bit coding and Hamming distance metric to encode the Gabor phase features, and finally obtain classification with a series of superpixel graphs with over-segmentation to under-segmentation [35]. Besides, edge-preserving filters can be applied to extract spatial information. For example, Wang et al. apply joint BF to smooth images on a probability map obtained by the SVM classifier. The retaining object boundary can be used to smooth the salt-and-pepper classification noise in the homogeneous region, which can contribute to the classification result by combining the segmentation map obtained by the minimum cut algorithm [36]. Sahadevan et al. integrate spatial contextual information obtained by BF into the spectral domain to improve SVM performance [37]. Qiao et al. extract spatial features by joint bilateral filtering with the first principal component as the guidance image, and the filtered image is classified with a sparse representation method [38]. Based on the work of Guo et al., guided filter (GDF) is used to extracting spatial information for optimizing the classification [39]. Liao et al. attempt to improve HSI classification according to the combination of spatial information with GDF and domain transform interpolated convolution filter (DTICF) [40]. To achieve a better classification, Kang et al. propose a classification method using edge-preserving filtering (EPF) and SVM [41], which can enhance the method with the proposed PCA-based EPFs (PCA-EPFs) classification [42]. Besides, based on image fusion with multiple subsets of adjacent bands and BF [43], a spatial texture feature is obtained by Kang et al. to improve the classification of HSI [44]. Moreover, adaptive manifold filter (AMF) [45] is used to realize high-dimensional filtering in real-time, and good spatial features can be extracted for HSI images. For example, Xie et al. try to enhance the extreme learning machine (ELM) classification of HSI images based on spatial information obtained by AMF [46]. It turns out that spatial features extracted by texture filters can improve the performance of HSI classification. However, the filters extracting spatial texture features tend to fall into local feature extraction by window functions, and it is difficult to attain the better spatial texture features to assist the classifier.

To make full use of spatial features for improving classification capabilities, we proposed a new method. Considering that AMF can obtain rich texture features in local and global optimization, we adopt AMF for spatial texture feature extraction in this work. However, AMF can only extract effective spatial texture features but may ignore informative spatial correlation feature. Consequently, some investigators use DTNCF and DTICF to obtain the optimal spatial correlation feature to compensate for the deficiencies of spatial texture features for enhancing HSI classification [24], [40]. Nevertheless, DTNCF and DTICF have different effects on
spatial texture features extraction for different HSI, resulting in different classification outcomes. To deal with these problems, this paper presents a new method based on the fusion of adaptive manifold filter and spatial correlation features (AMSCF) for adaptive hyperspectral image classification. Specifically, spatial texture features can be extracted by AMF, and spatial correlation features can be acquired by DTNCF and DTICF. The two kinds of spatial features, separately classified by LDM with the maximum marginal mean and the minimal margin variance, are fused, and output optimal classification after comparison. The contributions of this work are as follows:

(1) AMF is used to extract the spatial texture feature of the hyperspectral image, and the obtained spatial correlation feature of DTNCF is used for classification.

(2) To achieve the optimal classification, the proposed method uses DTNCF and DTICF to extract the most suitable spatial correlation feature and adapt to types of HSI for improving classification performance.

(3) To enhance the performance of the LDM classification, the two kinds of spatial information are combined, and the experimental results indicate that the proposed AMSCF method is better than other classification methods.

The rest of this paper is organized as follows. Section II describes the proposed method. Section III shows the hyperspectral image dataset to verify the effectiveness of the method and analyses the experimental results with AMSCF. Section IV states the conclusions.

II. METHODOLOGY
A. CLASSIFICATION METHODS
1) LDM-BASED CLASSIFICATION METHODS
The LDM proposed by Zhang et al. attempted to maximize the marginal means and minimize the marginal variances simultaneously for the construction of a new hyperplane to improve the SVM classification performance with maximizing the minimum margin. A training set can be defined as \( S = \{x_1, y_1, \ldots, x_n, y_n\} \), in which, \( x_1 \) is the training sample labeled by \( y_i \in \{+1, -1\} \), \( i = 1, 2, \ldots, n \), where \( n \) is the number of the training data. The goal of SVM is to obtain a predictive function with strong generalization performance and predict unlabeled data using the obtained function. To show the margin of SVM, \( w \) can be represented as the linear decision function and \( f \) is a linear model as follows

\[
f(x) = w^T \omega(x)
\]

where \( \omega(x) \) is a feature mapping of \( x \) by a kernel \( k \), i.e., \( k_{ij} = \omega(x_i)^T \omega(x_j) \).

With the maximization of the minimum margin, the margin of instance \( (x_i, y_i) \) for SVM in individual cases can be formulated as

\[
\gamma_i = y_i w^T \omega(x), \quad i = 1, 2, \ldots, n
\]

\[
\min \frac{1}{2} w^T w
\]

\[
\text{s.t.} \quad y_i w^T \omega(x_i) \geq 1
\]

\[
\xi > 0, \quad i = 1, 2, \ldots, n
\]

where \( \xi_i \) is a slack variable that measures the degree of misclassification of \( x_i \).

To improve the single interval of SVM, the marginal mean and the marginal variance were proposed to construct a new hyperplane for LDM and can be characterized as Eq. 3 and Eq. 4, respectively.

\[
\gamma = \sum_{i=1}^{n} y_i w^T \omega(x_i) = 1 - \frac{1}{n}(Xy)^T w, \quad i = 1, 2, \ldots, n
\]

\[
\hat{\gamma} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (y_i w^T \omega(x_i) - y_j w^T \omega(x_j))^2
\]

\[
= \frac{1}{n^2}(nw^T XX^T w - w^T y y^T X^T w)
\]

where \( X = [\omega(x_1) \cdots \omega(x_n)] \) and \( y = [y_1 \cdots y_n] \).

For the inseparable conditions, the soft-margin LDM can be expressed as

\[
\min_{w, \xi} \frac{1}{2} w^T w + \alpha_1 \hat{\gamma} - \alpha_2 \gamma + E \sum_{i=1}^{n} \xi_i
\]

\[
\text{s.t.} \quad y_i w^T \omega(x_i) \geq 1 - \xi_i,
\]

\[
\xi > 0, \quad i = 1, 2, \ldots, n
\]

where \( \alpha_1 \) and \( \alpha_2 \) are the parameters corresponding to the trading-off the marginal variance and the marginal mean, respectively; \( E \) is a trading-off parameter. The terms of \( \min_{w} \frac{1}{2} w^T w - \alpha_2 \gamma \) and \( \min_{w} \frac{1}{2} w^T w + \alpha_1 \hat{\gamma} \) in Eq. 5 maximizes the mean margin and minimizes the margin variance, respectively. Meanwhile, due to the maximization of the mean margin and minimization of the margin variance, LDM can more effectively achieve generalization performance for the classification.

The classification hyper-planes of SVM and LDM are shown in Fig.1, and two categories of HSI grounds are plotted in “” and “” respectively. The SVM margin hyper-plane \( H_{SVM} \) is drawn in a red line which is the vertical line of the dotted red line, and the LDM margin hyper-plane \( H_{LDM} \) is drawn in the blue line. Fig.1 shows that the hyper-plane of SVM maximizes the minimum margin for all samples and the hyper-plane of LDM maximizes the mean margin and minimizes the margin variance simultaneously for all samples. We can conclude that the hyper-plane of LDM is more effective than that of SVM for the classification.

To improve the effectiveness and solve the nonlinear problem of LDM, substituting Eq. 3 and Eq. 4 into Eq. 5 can yield

\[
\min_{w, \xi} \frac{1}{2} w^T w + \frac{\alpha_1}{n^2}(nw^T XX^T w - w^T y y^T X^T w)
\]

\[
- \frac{\alpha_2}{n} (Xy)^T w + E \sum_{i=1}^{n} \xi_i
\]

\[
\text{s.t.} \quad y_i w^T \omega(x_i) \geq 1 - \xi_i,
\]

\[
\xi > 0, \quad i = 1, 2, \ldots, n
\]
The linear filtering \([48]\) of the hyperspectral image can be expressed as

\[
O_i = R_i + E_i, \tag{11}
\]

where \(R_i\) is the \(i\)th band of the hyperspectral image, \(E_i\) is the filtering result of the image \(R_i\), and \(O_i\) is the result of a linear combination filter of image.

The AMF [45] proposed by Gastal et al. is a high-dimensional filtering method that consists of three steps (i.e., splatting, blurring, and slicing), and the filtering for the hyperspectral image can be shown in the following parts.

\(a: \text{SPLATTING}\)

Pixels of the image can be projected onto the current manifold, and a Gaussian distance weighting can be performed on each manifold, such as

\[
\phi_{\text{splat}}(\alpha_{ki}) = \xi_{S/2}(\alpha_{ki} - f_i) f_i, \tag{12}
\]

where \(f_i\) is the pixel of the hyperspectral image, \(\alpha_{ki}\) is the sampled pixel, and \(\xi_{S/2}(\alpha_{ki} - f_i)\) controlled by the covariance matrix \(S_{ij}\) calculates the Gaussian distance weight of \(\alpha_{ki}\) for a position in the band, which can completely utilize the spatial correlation of the pixel and have the global diffusion and spatial correlation retention characteristics.

\(b: \text{BLURRING}\)

Smoothing the splatted values \(\phi_{\text{splat}}\), RF is adopted for filtering in this paper.

\[
\phi_{\text{blur}}(\beta_{ki}) = \text{RF}(\phi_{\text{splat}}) \tag{13}
\]

\(c: \text{SLICING}\)

The filtering response \(E_i\) of each pixel is obtained from \(\phi_{\text{blur}}\) by interpolation weights,

\[
E_i = \frac{\sum_k w_{ki} \phi_{\text{blur}}(\alpha_{ki}) \sum_{k=1}^K G_{w_{ki}} \phi_{\text{blur}}(\alpha_{ki})}{\sum_{k=1}^K G_{w_{ki}} \phi_{\text{blur}}(\alpha_{ki})} \tag{14}
\]

where \(K\) is the total number of adaptive manifolds, \(w_{ki}\) is the smoothing processing of \(\phi_{\text{splat}}(\alpha_{ki})\), which can be expressed as

\[
\phi_{\text{splat}}(\alpha_{ki}) = \xi_{S/2}(\alpha_{ki} - f_i) \tag{15}
\]

Eq. 14 is substituted into Eq. 11 to obtain a linear filtering result, which is manifold filtering of a hyperspectral image.

\[
O_i = R_i + \frac{\sum_{k=1}^K w_{ki} \phi_{\text{blur}}(\alpha_{ki}) \sum_{k=1}^K w_{ki} \phi_{\text{blur}}(\alpha_{ki})}{\sum_{k=1}^K w_{ki} \phi_{\text{blur}}(\alpha_{ki})} \tag{16}
\]

The adaptive manifold filtering algorithm for the hyperspectral image is the recursive process of splatting, blurring, and slicing; and the recursive process of manifold filtering is related to the height of the manifold. The height of the manifold tree determines the recursive hierarchy and the number of manifold tree nodes, and also greatly affects the filtering effect. The manifold tree height and the manifold node number are calculated as follows:

1) The tree’s height \(H\) is written as

\[
H = n + \max(2, [D_jD_r]), \tag{17}
\]
where $\lfloor \cdot \rfloor$ is the round-up operator, $n \in 1, 2, \cdots, N$,
\[
D_s = \lfloor \log_2 \sigma_s \rfloor - 1
\]  
(18)
where $\lceil \cdot \rceil$ is the round-down operator
\[
D_r = 1 - \sigma_r
\]  
(19)
where $D_s$ is a height computed from the spatial standard deviation $\sigma_s$ of the filter, and $D_r$ is a linear correlation computed from the range standard deviation $\sigma_r$.

Thus, the manifold number $K$ can be obtained by
\[
K = 2^H - 1.
\]  
(20)

2) Generate the first manifold, the first manifold $\mu_1(p_i)$ can be obtained by a low-pass filter,
\[
\mu_1(p_i) = (h_{\Sigma_S} \ast f)(p_i)
\]  
(21)
where $p_i \in S$ $h_{\Sigma_S}$ is a low-pass filter in $S$ with covariance matrix, $\Sigma_S$ which is a diagonal matrix that controls the decay of the Gaussian kernel.

3) Calculate eigenvector $v_1$. Solve the eigenvector corresponding to the largest eigenvalue of $(f_1 - \mu_1)(f_1 - \mu_1)^T$.

4) Segment the pixels into two clusters. The dot product $d$ is first calculated by Eq. 22. Then divide the pixels of HSI into two clusters according to Eq. 23.
\[
d = v_1^T(f_i - \mu_1)
\]  
(22)
\[
\begin{cases} p_i \in D_- & \text{if } d \leq 0 \\ p_i \in D_+ & \text{otherwise} \end{cases}
\]  
(23)

5) Calculate the new manifold. The new manifold $\mu_-(p_i)$ is calculated with a low-pass filter, similarly, $\mu_+(p_i)$ can be calculated, and $\eta_+(p_i), \mu_+(p_i)$ and $\mu_-(p_i)$ recursively calculated.

6) The next manifold and new manifold high-dimensional filtering can be recursively calculated and performed until it reaches the manifold height.

The manifolds generated by the adaptive manifold filter are always compatible with the input reference signal in the high-dimensional space, which makes the manifolds can well represent the input reference signal, with the adaptive manifold as the standard, so it can maintain clear edge features. In addition, the filtering process of the adaptive manifold filter will build a manifold tree while performing filtering. More detailedly, after calculating a manifold, then filter the input signal with the manifold, and then continue filtering recursively until all the manifolds are completed. It can be concluded that the result has more dimensional filtering directions.

Therefore, the manifold filtering has good global diffusivity and edge retention characteristics. The higher the height of the manifold tree, the more the manifold can adapt to the local population, and it has better local optimization. The manifold filtering can extract better spatial texture information for the local and global optimization, with obtaining excellent global spatial features. It can be concluded that the higher the manifold tree height, the more manifolds, and the local and global optimization capabilities will be stronger.

The filtering flow of the AMF for HSI is shown in Fig. 2.

3) DOMAIN TRANSFORM FILTER

In the past, the supervised HSI classification methods only obtained some useful features from spectral information, while the spatial-spectral classification methods focused on the extraction of spatial texture features. However, in the process of extracting texture features, spatial correlation features are often ignored.

DTNCF and DTICF [23] proposed by Gastal can be used for image filtering, and the two-dimensional image filtering can be transformed into one-dimensional image filtering. For a uniform discretization $S(\Omega)$ of the original domain $\Omega$, the energy function of DTNCF for the hyperspectral image $R$ at the $i$th band can be expressed as
\[
F_i(a) = \frac{1}{Z_a} \sum_{a,b \in S(\Omega)} R_i(b)G(\psi(a), \psi(b))
\]  
(24)
where $Z_a$ is the normalized factor of $a$ as shown in Eq. 25, and $G(\cdot)$ denotes the kernel filter. Furthermore, the implication of Eq. 26 is that the neighborhood pixels are the same ground, and $\delta(\cdot)$ is the Boolean function as shown in Eq. 27.
\[
Z_a = \sum_{b \in S(\Omega)} G(\psi(a), \psi(b))
\]  
(25)
\[
G(\psi(a), \psi(b)) = \delta(\|\psi(a) - \psi(b)\| \leq r)
\]  
(26)
\[
\delta(x) = \begin{cases} 1 & x \text{ is true} \\ 0 & \text{otherwise.} \end{cases}
\]  
(27)
where $\varphi(\cdot)$ is given in Eq. 28, which cumulatively integrals the partial differential of the image and then transforms it to
an increasing function. Therefore, the function \( \varphi(u) \) is used to reshape an image into a one-dimension vector. \( r \) is the radius of the filter as shown in Eq. \( 29 \), \( \sigma_s \) and \( \sigma_r \) are the space standard and range standard deviations, respectively. \( N \) is the total iteration number. \( \sigma_{J_i} \) is the standard deviation for the kernel used in the \( q^{th} \) iteration expressed in Eq. \( 30 \).

\[
\varphi(u) = \int_0^u 1 + \frac{\sigma_s}{\sigma_r} \sum_{i=1}^c |R_i'(x)| \, dx \tag{28}
\]

\[
r = \sqrt{3}\sigma_r \tag{29}
\]

\[
\sigma_{J_i} = \sigma_s\sqrt{\frac{2^n-q}{4^n-1}} \tag{30}
\]

At the \( i^{th} \) band, the energy function of DTICF for hyperspectral image \( R \) can be represented as

\[
F_i(u) = \int_{\Omega_w} D_w(x) Q(f(u), x) \, dx \quad i = 1, 2, \cdots, k; \ u \in \Omega_w \tag{31}
\]

Equation \( 31 \) shows a reconstructed signal \( D_w \) obtained by the linear interpolation (in \( \Omega_w \)) of the samples. Filtering \( D_w \) is performed by the continuous convolution, where \( Q \) is a normalized box kernel represented by the bool function and \( r \) is the filter radius as shown in Eq. \( 33 \). The significance is that the pixel neighborhood can be considered to be the same ground, so the spatial correlation of the hyperspectral image is preserved during the filtering process.

\[
Q(f(u), x) = \frac{1}{2r} \delta(|\varphi(u) - x| \leq r) \tag{32}
\]

Substituting Eq. \( 32 \) into Eq. \( 31 \), we can obtain

\[
F_i(u) = \frac{1}{2r} \int_{\varphi(u)-r}^{\varphi(u)+r} D_w(x) \, dx \tag{33}
\]

where \( \varphi(u) \) and \( r \) are given in Eq. \( 28 \) and Eq. \( 29 \), respectively.

There always exists a strong spatial correlation among pixels in HSI for the ground distribution with the appropriate uniformness. In addition, the spatially relevant feature is a related property of the reflection intensity between the adjacent pixels. However, spatial correlation feature is often ignored in the texture feature extraction with the edge-preserving filtering.

To analyze the spatial correlation characteristics of AMF, DTICF, and DTNCF, the spatial correlation of hyperspectral images can be measured by Moran’s \( I \) [50], [51] before and after the filtering, and the formula can be expressed as

\[
I = \frac{n \sum_{i=1}^n \sum_{j=1}^n \alpha_{ij}(X_i - \overline{X})(X_j - \overline{X})}{\sum_{i=1}^n \sum_{j=1}^n \alpha_{ij} \sum_{i=1}^n (X_i - \overline{X})^2} \tag{34}
\]

where \( X_i \) and \( X_j \) are the reflection intensities of two hyperspectral pixels, \( \overline{X} \) is the average of \( X \), \( n \) is the pixel number of one band, and \( \alpha_{ij} \) is the spatial weight.

The larger the \( I \), the stronger the spatial correlation, and vice versa. Subsection D of Section III provides an overview of the validation tests for DTNCF for space-related information extraction.

DTNCF and DTICF have excellent spatial correlation feature retention characteristics for HSI. However, these two filters have their advantages for different HSIs and can have different effects on the classification. Therefore, DTNCF and DTICF will be used in this work to adapt to different HSI datasets to obtain the best classification performance.

B. PROPOSED METHODS

To completely utilize two kinds of spatial features, a new classification approach (AMSCF) is proposed based on the combination of AMF, DTNCF, and DTICF. In which, spatial texture feature can be extracted with AMF, and spatial correlation features can be obtained with DTNCF and DTICF, separately. The fusion of two kinds of spatial features is classified by the LDM classifier, and the optimal classification can be exported.

The detailed flow chart of the AMSCF can be found in Fig. 3, which is roughly divided into 8 steps: 1) HSI is normalized; 2) The dimensionality of HSI is reduced by PCA; 3) Spatial texture feature can be extracted by AMF; and spatial correlation features can be obtained with DTNCF and DTICF, separately. The fusion of two kinds of spatial features is classified by the LDM classifier, and the optimal classification can be exported.

The systematic implementation process is given as follows.

III. EXPERIMENTS

A. HYPERSPECTRAL DATA DESCRIPTION

The effectiveness of AMSCF has been tested on three different hyperspectral image datasets.

1) Indian Pines [52] collected in 1992 from an Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over the Indian Pines region in Northwestern Indiana. This dataset included 220 spectral bands with a spatial size of \( 145 \times 145 \) pixels, with 200 bands remained and 20 spectral bands removed due to noise and water absorption. The image has 16 classes, each of which owns a specific type and number as shown in Table 2.

2) Salinas Valley [53] attained in 1990 by AVIRIS over Salinas Valley of Southern California and possesses a high spatial resolution of \( 3.7 \) m with an area of the spatial size of \( 512 \times 217 \) pixels and 206 spectral bands. Similarly, because of noise and water absorption phenomena, 200 bands are retained. This image also contains 16 classes, whose specific types and numbers are shown in Table 3.

3) Kennedy Space Center [54] from NASA AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) at the Kennedy Space Center in Florida on March 23, 1996. 224 bands are gotten with \( 10 \) nm width. The Kennedy Space Center dataset can be available at an

VOLUME 8, 2020
Algorithm 1 AMSCF

1. **Input**: Input the hyperspectral image $R$
2. **Normalization**: $W$ is obtained from the hyperspectral image $R$ that is Min-Max normalized
3. **Dimensionality reduction**: The normalized image $W$ is further lowered by PCA, while the principal component of the top 10% is chosen for AMF
4. for $i = 1: N$ % AMF filtering with $W$, and $N$ is the component number used to filter.
   (1) Compute the tree height $H$
   (2) Compute the manifold and recursive filtering
      for $h = 1: H$
      1) Compute the standard deviation
      2) Compute the Gaussian distance Weights $\beta_{ki}$
      3) Splatting by Eq. 12
      4) Blurring by Eq. 13
      5) Slicing by Eq. 14
      6) Compute the new manifolds by the low-pass filtering
   end for
   (3) Obtain the linear filtering result $K_{st}$ by Eq. 16.
   end for
5. for $j = 1: M$ % DTNCF filtering with $R$ and $M$ is the band number of $R$
   (1) Estimate the horizontal and vertical partial derivatives
   (2) Compute the distance of neighbor pixels
   (3) Compute the derivatives of horizontal and vertical domain
   (4) Perform DTNCF filtering
      for $q = 1: Q$
      1) Calculate the standard deviation $\sigma_s$
      2) Calculate the filter radius $r$
      3) Convolution filtering by Eq. 24
   end for
   Output filtering result $K_{nc}$.
6. for $j = 1: M$ % DTICF filtering with $R$ and $M$ is the band number of $R$
   (1) Estimate the horizontal and vertical partial derivatives
   (2) Compute the distance of neighbor pixels
   (3) Compute the derivatives of horizontal and vertical domain
   (4) Perform DTICF filtering
      for $q = 1: Q$
      1) Calculate the standard deviation $\sigma_s$
      2) Calculate the filter radius $r$
      3) Convolution filtering by Eq. 31
   end for
   Output filtering result $K_{ic}$.
7. **Fusion**: Linearly fuse $K_{nc}$, $K_{ic}$, and $K_{st}$ according to $H_1 = K_{nc} + K_{st}$ and $H_2 = K_{ic} + K_{st}$.

Algorithm (Continued.)

8. **Classification**: From the spatial information dataset $H_1$ and $H_2$, the two training sets are randomly selected in proportion for different kinds of HSI datasets, while the two remaining samples comprised the testing set, which is verified with the LDM classification, respectively.
9. **Output**: Obtain the classification results of $O_1$ and $O_2$, and output the optimal classification outcome after comparison.
10. **End**

altitude of approximately 20km with a spatial resolution of 18 m. After removal of water absorption and low SNR and, 176 bands are used for analysis. The image has 13 classes, and the specific types and numbers of each class are shown in Table 4.

B. COMPARED METHOD AND PARAMETERS SETTING

To verify the superiority of the proposed method, the following procedures are conducted to compare with the method of AMSCF.

(1) SVM [9]: According to the raw features of hyperspectral images, SVM can be applied with the Gaussian radial basis function kernel, and the main parameters include gamma $g = 0.18$ and penalty factor $C = 2500$.

(2) IFRF [44]: This method attains the classified results with SVM based on image fusion and RF. The spatial standard deviation $\sigma_s = 200$ and the range standard deviation $\sigma_r = 0.3$ are set for RF.

(3) LDM: Based on the raw features of hyperspectral images, a Gaussian radial basis function kernel is applied here. The margin variance and the margin mean are separately set as $\alpha_0 = 125000$ and $\alpha_1 = 50000$, while the trading-off parameter is set as $E = 100000$.

(4) LDM-FL [12]: This method obtains the classified results with LDM from RF. The parameter of LDM-FL is the same as the LDM method.

(5) PCA-EPFs [42]: The spatial information firstly constructed by applying edge-preserving filters is stacked to form the fused features, and the dimension is reduced by PCA for the classifier of SVM. The spatial standard deviation $\sigma_s = 3$ and the range standard deviation $\sigma_r = 0.1$ are set for PCA-EPFs.

(6) GFDN [34]: This method extracts the spatial features by GF on the first three principal components of the hyperspectral image to form the fused features, and then combines the original features used for deep network classification with the scale, orientation, row, and column number of to 5, 8, 55, and 55.

(7) SuperPCA [19]: In this method, the spatial features which acquired by performing PCA on some homogeneous region obtained with superpixel segmentation, were classified by SVM and got the final result by the majority voting.
(8) SuperBF [20]: In this method, spatial texture information obtained by BF on some homogenous region divided via superpixel segmentation, were classified by SVM.

(9) AM-SVM: The hyperspectral dimensionality is reduced with PCA, and the first 10% principal components are selected for SVM based on AMF, and the spatial standard deviation $\sigma_s$ and range standard deviation $\sigma_r$ are set to 5 and 1.0, respectively.

(10) AM-LDM: PCA lowers the dimensionality of the hyperspectral, and the first 10% principal components are diminished for LDM based on AMF, and the spatial standard deviation of $\sigma_s$ and the range standard deviation $\sigma_r$ are set to be the same as AM-SVM.

(11) DTNC-SVM: The hyperspectral dimensionality has been lessened with PCA, and the first 10% principal components are picked for SVM based on DTNCF, and the spatial standard deviation $\sigma_s$ and range standard deviation $\sigma_r$ are correspondingly set to the values of 200 and 0.3.

(12) DTNC-LDM: The hyperspectral dimensionality is diminished with PCA, and the first 10% principal components are taken for LDM based on DTNCF, and the parameters of $\sigma_s$ and $\sigma_r$ follows the values of DTNC-SVM.

(13) DTIC-SVM: The hyperspectral dimensionality is lessened with PCA, and the first 10% principal components are picked for SVM based on DTICF, and the values of spatial standard deviation $\sigma_s$ and range standard deviation $\sigma_r$ are 200 and 0.3, respectively.

(14) DTIC-LDM: PCA is also used to shorten the hyperspectral dimensionality, and the first 10% principal components are taken for LDM based on DTICF, and the parameters of $\sigma_s$ and $\sigma_r$ are equal to the values of DTIC-SVM.

(15) AMDTIC-SVM: The spatial texture feature is extracted from the top 10% principal components by AMF, and the spatial correlation feature can be obtained by DTICF. Last, the two kinds of spatial features are fused to be classified by SVM.

(16) AMDTNC-SVM: The classifying process is similar to the process of AMDTNC-SVM, other than that of the spatial correlation feature attained via DTICF.

(17) AMSCF: The spatial texture feature is extracted from the top 10% of principal components by AMF, and spatial correlation features can be obtained by DTNCF and DTICF. The two kinds of spatial features are fused and separately classified by LDM, and output optimal classification after comparison.

In this paper, Overall Accuracy (OA), Average Accuracy (AA), and Kappa statistic (Kappa) were adopted to measure classification accuracy. To avoid biased estimation, based on the configuration of i7-6700 CPU and 8GB RAM, 20 independent tests were carried out using a computer program of Matlab R2012b.

C. THE VALIDATION TEST OF AMF

To verify the effectiveness of AMF, the 15th, 90th, 125th, and 190th bands of Indian Pines were processed with AMF. As shown in Fig. 4, the four images of the first row are the original spectral image of Indian Pines, and another four images of the second row are the corresponding filtered images processed by AMF. It can be seen that good boundary features can be extracted by AMF from the hyperspectral images, with the great advantages of smoothing and removing hyperspectral image noise.

D. TEST OF SPATIAL CORRELATION INFORMATION

To compare the spatial correlation of AMF and DTICF, we calculated the mean of Moran’s $I$ for each band of Indian Pines and Salinas Valley datasets. The average Moran’s $I$ of the two filters is shown in Fig. 5. We can find...
that the average of Moran’s $I$ obtained from DTICF is higher than the average of AMF and raw spectral features. Besides, the average of Moran’s $I$ acquired by AMF is lower than that of the spectrum images, suggesting the weak spatial correlation information. Therefore, it can be illustrated that DTICF can extract good spatial correlation information and effectively compensate for the deficiency of AMF.
E. INVESTIGATION OF THE PROPOSED METHOD

1) OPTIMIZATION OF AMF

The parameters of spatial standard deviation $\sigma_s$ and range standard deviation $\sigma_r$ of AMF can influence the spatial information extraction, so the classification test of Indian Pines dataset was conducted to verify the effectiveness of parameter optimization. For Indian Pines dataset, 6% of the training samples were randomly selected, and the three optimal parameters were found using the exhaustive method to obtain the best LDM classification results. We first set the parameters to $\sigma_s \in 1, 2, \ldots, 20$ and the standard deviation $\sigma_r \in 0.1, 0.2, \ldots, 4.0$. Then, the experiments were sequentially performed for the classification. As shown in Fig. 5, when $\sigma_r \in 0.4, 0.5, \ldots, 1.5$ than other test values $\sigma_r$. Fig.6 shows the classification test, and $m$ is the number of iterations. Furthermore, the best classification can be achieved when $\sigma_s = 5$ and $\sigma_r = 1.0$. Therefore, to acquire a better classification, the parameters $\sigma_s = 5$ and $\sigma_r = 1.0$ were adopted in the following experiments.

To verify the importance of the manifold tree height for LDM classification, Table 1 shows the confirmation of the hyperspectral image classification, and $n$ is the height of the manifold tree. From the table, the optimal classification result of OA is 97.66%, whereas the height of the manifold tree is 6 and the number of manifolds is 255, respectively. On the contrary, the worst classification result corresponds to a manifold height of 2, which fully indicates that the number of manifolds is larger. Also, the spatial feature of extracting hyperspectral image is better and can significantly improve classification performance. However, the complexity of the algorithm increases due to the massive nodes in the manifold tree, and the appropriate height of the manifold tree is the key to extract good spatial texture features. Therefore, height of the manifold tree was selected as 6 in this paper for the following classification test.

2) EXPERIMENT OF INDIAN PINES

To assess the performance of AMSCF, sixteen methods were employed to classify and verify the data of Indian Pines, as shown below: The distribution of Indian Pines datasets are shown in Fig.7 (a). All 16 categories were selected, with 4% samples of the thirteen types of grounds number of Indian Pines as the training set and the remaining samples were employed as the test set, while 16% of the three types of grounds were insufficient for training and were selected as the training set. Table 2 shows the classification accuracy resulting from these thirteen classification methods, as shown in Fig. 7.

Table 2 shows the accuracy of OA, AA, and Kappa for each class using various methods, indicating that AMSCF achieves excellent accuracy, e.g., OA = 97.88%, AA = 96.28%, and Kappa = 97.58. In addition, the accuracy of AMSCF exceeds 99% in the 4 classes. This experiment shows that the classification performance of AMSCF is significantly enhanced compared with other approaches.

| $n$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|-----|---|---|---|---|---|---|---|---|---|
| Tree height | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Tree node | 3 | 7 | 15 | 31 | 63 | 127 | 255 | 511 | 1023 |
| OA (%) | 97.17 | 97.46 | 97.36 | 97.29 | 97.65 | 97.49 | 97.66 | 97.54 | 97.63 |
| AA(%) | 96.52 | 96.58 | 96.33 | 96.67 | 96.89 | 96.48 | 96.75 | 97.00 | 96.75 |
| Kappa | 96.78 | 97.11 | 96.99 | 96.91 | 97.32 | 97.13 | 97.33 | 97.20 | 97.29 |
FIGURE 7. Classification maps of different methods on the Indian Pines dataset (a) Ground (b) SVM (c) EPF (d) IFRF (e) PCA-EPFs (f) LDM (g) LDM-FL (h) GFDN (i) SuperPCA (j) SuperBF (k) AM-SVM (l) AM-LDM (m) DTIC-SVM (n) DTIC-LDM (o) DTNC-SVM (p) DTNC-LDM (q) AMDTIC-SVM (r) AMDTNC-SVM (s) AMSCF-IL (t) AMSCF-NL.
TABLE 2. Comparison of classification accuracies (in percent) provided by 9 methods for Indian Pines dataset (part A). Comparison of classification accuracies (in percent) provided by 7 methods for indian pines dataset (part B).

| Ground            | Sum | Training | SVM | IFRF | PCA-EPIs | LDM   | LDM-FL | GFDN  | SuperPCA | SuperBF | AM-SVM | AM-LDM |
|-------------------|-----|----------|-----|------|----------|-------|--------|-------|----------|---------|--------|--------|
| Alfalfa           | 54  | 9        | 55.47| 84.15| 71.27    | 82.76 | 93.45  | 100.00| 100.00   | 95.46   | 68.47  | 91.53  |
| Corn-no-till      | 1434| 57       | 67.82| 88.00| 85.24    | 74.20 | 92.87  | 90.63 | 91.29    | 92.98   | 93.02  | 92.41  |
| Corn-min-till     | 834 | 33       | 58.51| 70.81| 89.06    | 61.83 | 93.28  | 96.11 | 90.35    | 97.88   | 96.17  | 91.90  |
| Corn              | 234 | 9        | 17.65| 65.03| 69.99    | 41.27 | 84.78  | 94.12 | 76.55    | 92.84   | 60.40  | 91.90  |
| Grass-pasture     | 497 | 20       | 83.73| 87.98| 87.47    | 89.87 | 85.99  | 95.22 | 96.13    | 88.43   | 94.88  | 94.39  |
| Grass-trees       | 747 | 30       | 95.71| 97.66| 94.03    | 93.22 | 98.26  | 96.92 | 97.09    | 96.37   | 97.47  | 98.69  |
| Grass-pasture-ned | 26  | 4        | 22.21| 54.65| 27.37    | 78.39 | 73.81  | 95.45 | 95.83    | 53.39   | 53.92  | 91.30  |
| Hay-windrowed     | 489 | 20       | 95.49| 99.51| 98.98    | 96.58 | 100.00| 98.93 | 99.56    | 93.71   | 98.80  | 99.57  |
| Oats              | 20  | 3        | 20.59| 50.51| 45.50    | 85.82 | 86.76  | 93.75 | 100.00   | 68.64   | 54.14  | 92.11  |
| Soybeans-no-till  | 968 | 39       | 60.28| 83.74| 90.26    | 67.59 | 88.80  | 95.10 | 91.51    | 90.80   | 88.04  | 94.12  |
| Soybeans-min-till | 2468| 99       | 85.40| 92.52| 95.17    | 74.44 | 91.08  | 97.65 | 96.48    | 97.46   | 96.96  | 96.18  |
| Soybeans-clean-til | 614 | 25       | 68.01| 88.71| 88.71    | 68.20 | 87.38  | 94.22 | 90.48    | 92.64   | 96.83  | 97.42  |
| Wheat             | 212 | 8        | 85.05| 96.54| 73.58    | 99.38 | 95.06  | 95.57 | 99.49    | 83.21   | 98.22  | 99.52  |
| Woods             | 1294| 52       | 97.25| 97.58| 96.55    | 94.17 | 96.72  | 98.48 | 99.09    | 99.56   | 99.40  | 98.43  |
| Bldg-grass-tree   | 380 | 15       | 34.57| 71.97| 91.92    | 52.45 | 90.24  | 91.21 | 98.63    | 91.31   | 55.73  | 76.11  |
| Stone-stee-towers | 95  | 4        | 75.15| 82.48| 71.41    | 82.83 | 94.98  | 92.22 | 91.11    | 41.07   | 81.42  | 87.77  |

OA (%) - 76.11 88.61 90.67 77.05 92.36 95.70 94.58 94.21 93.05 95.22
AA (%) - 63.93 81.99 77.99 77.69 90.84 95.70 94.60 85.97 83.37 93.64
Kappa (%) - 72.42 86.97 89.34 73.79 91.32 95.10 93.82 93.38 92.05 94.56

3) EXPERIMENT OF SALINAS VALLEY

Likewise, the ground truth distribution of the Salinas Valley dataset is demonstrated in Fig. 8(a): all 16 classes were selected, with 0.7% of the samples as the training set, and the remaining 99.3% as the test set. Table 3 lists the classification accuracies of the Salinas Valley dataset under different methods, and the classification outputs are shown in Fig. 8.

In Fig. 8, the classification results of the Salinas Valley are shown, while Table 3 shows the OA, AA, Kappa, and accuracy of each method. The table also gives the best
TABLE 3. Comparison of classification accuracies (in percent) provided by 9 methods for Salinas Valley dataset (part A). Comparison of classification accuracies (in percent) provided by 7 methods for Salinas valley dataset (part B).

| Ground         | Sum | Training | SVM      | IFRF     | PCA-EPPs | LDM      | LDM-FL   | GFDN     | SuperPCA | SuperBF | AM-SVM | AM-LDM |
|----------------|-----|----------|----------|----------|----------|----------|----------|----------|----------|---------|--------|--------|--------|
| Alfalfa        | 54  | 14       | 97.59    | 99.84    | 99.58    | 99.15    | 99.55    | 93.73    | 100.00   | 96.91   | 99.30  | 99.94  |
| Corn-no-till   | 1434| 26       | 98.83    | 98.93    | 98.94    | 99.12    | 99.77    | 98.89    | 98.54    | 99.75   | 99.80  | 99.94  |
| Corn-min-till  | 824 | 14       | 82.51    | 100.00   | 99.62    | 95.95    | 100.00   | 94.60    | 100.00   | 96.18   | 89.83  | 100.00 |
| Grass-pasture  | 234 | 10       | 97.41    | 91.11    | 95.97    | 97.65    | 93.27    | 92.49    | 97.03    | 87.10   | 91.35  | 97.63  |
| Grass-trees    | 497 | 19       | 97.12    | 97.19    | 96.12    | 94.10    | 99.17    | 90.97    | 99.25    | 92.00   | 98.99  | 99.50  |
| Grass-pasture-mowed | 747 | 28     | 99.46    | 99.74    | 99.65    | 99.66    | 99.49    | 98.19    | 98.32    | 99.87   | 99.70  | 99.79  |

Hay-windrowed  26  25  99.22  98.96  98.54  99.56  98.04  99.61  96.99  99.43  99.35  99.63

Oats  20  43  98.55  99.99  100.00  98.79  99.94  100.00  99.03  99.18  99.25  99.81

Soybeans-no-till  968  23  97.78  98.81  97.60  96.67  99.68  95.39  94.76  99.06  88.63  95.68

Soybeans-min-till  2468  7  63.03  85.80  78.22  91.50  96.54  89.15  92.26  96.15  77.73  97.38

Soybeans-clean-till  614  13  93.80  98.39  98.79  99.59  97.38  95.30  100.00  99.55  97.71  100.00

Wheat  212  6  76.98  59.65  89.22  97.09  87.89  85.81  98.24  92.88  97.83  97.91

Woods  1294  7  86.47  96.34  77.56  93.24  90.96  96.80  97.35  79.09  83.81  95.93

Bldg-grass-tree  380  51  48.23  98.11  99.61  63.90  98.03  89.25  97.51  99.66  75.34  88.15

Stone-steel-towers  95  13  90.31  97.18  92.44  97.18  98.76  99.60  96.95  96.39  92.22  98.00

OA (%) - 86.61  97.15  96.45  88.39  98.49  95.58  97.92  98.01  92.26  95.91

AA (%) - 87.48  94.76  92.70  93.38  97.39  94.62  97.87  95.82  92.60  97.53

KAPPA (%) - 84.99  96.82  96.04  87.08  98.32  95.06  97.68  97.78  91.36  95.45

4) EXPERIMENT OF KENNEDY SPACE CENTER

Similarly, the distribution based on Kennedy Space Center dataset is shown in Figure 10 (a), and 5%(about 261) samples of all 16 classes are selected as the training set, and the remaining 95% is used as the test set. Table 4 lists the classification results of the Kennedy Space Center dataset for 16 methods. The effect of classification is shown in Fig. 9.

The classification results for Kennedy Space Center are shown in Fig. 9, while Table 4 shows the accuracies of OA, AA and Kappa for all methods, with the best accuracy of AMSCF as OA = 99.05%, AA = 98.37%, and Kappa = 98.94%. Moreover, AMSCF owned 100% accuracy in the seven classes. This experiment indicates that compared with
FIGURE 8. Classification maps of different methods on the Salinas Valley dataset (a)Ground (b)SVM (c) IFRF (d) PCA-EPFs (e) LDM (f) LDM-FL (g) GFDN (h) SuperPCA (i) SuperBF (j) AM-SVM (k) AM-LDM (l) DTIC-SVM (m) DTIC-LDM (n) DTNC-SVM (o) DTNC-LDM.
other classification methods, the classification performance is enhanced.

Also, the OA values of AMSCF are correspondingly larger than that of SVM, IFRF, PCA-EPFs, LDM, LDM-FL, GFDN, SuperPCA, SuperBF, AM-SVM, AM-LDM, DTIC-SVM, DTIC-LDM, DTNC-SVM, DTNC-LDM, AMDTIC-SVM and AMDTNC-SVM by 14.55%, 4.54%, 2.74%, 13.33%, 3.83%, 4.39%, 3.34%, 1.10%, 4.76%, 2.82%, 1.38%, 1.19%, 1.93%, 1.85%, 1.38% and 1.83%. The hyperspectral classification can completely verify the effectiveness of AMSCF.

5) ANALYSIS
First, the OA values of LDM for Indian Pines, Salinas Valley, and Kennedy Space Center are corresponding to 0.94%, 1.78%, and 1.22%, which are greater than SVM. In addition, the OA values of AM-LDM, DTIC-LDM, and DTNC-LDM for Indian Pines were 3.30%, 3.82%, and 2.90% larger than that of AM-SVM, DTIC-SVM, and DTNC-SVM. Besides, the OA values in Salinas Valley are 3.65%, 1.92%, and 1.02% superior to that of AM-SVM, DTIC-SVM, and DTNC-SVM, whereas the values in Kennedy Space Center are 1.94%, 0.19%, and 0.08% greater than that of AM-SVM, DTIC-SVM, and DTNC-SVM. It can be shown that LDM has the ability to maximize the margin mean and minimize the margin variance, which is better than SVM.

Second, the AM-SVM and AM-LDM OA values of Indian Pines are 16.93% and 18.17% higher than those of SVM and LDM, respectively. In addition, the OA values of AM-SVM and AM-LDM in Salinas Valley are 5.65% and 7.52% higher than those of SVM and LDM, separately. Moreover, the OA values of AM-SVM and AM-LDM for Kennedy Space Center are corresponding to 9.79% and 10.51% larger than those of SVM and LDM. This finding indicates that the spatial texture features extracted by AMF has good edge retention characteristics with the appropriate height of the manifold tree, and are effective for enhancing the classification performance of SVM and LDM.

Third, the OA values of DTNC-SVM were 16.16%, 11.21% and 12.63% that are higher than the OA values of SVM for Indian Pines, Salinas Valley and Kennedy Space Center, respectively. Similarly, the OA values of DTNC-LDM are 18.97%, 10.45% and 11.48% larger than that of LDM for the three datasets. In addition, the OA values of DTIC-SVM are 15.97%, 10.39% and 13.18% higher than that of SVM.
FIGURE 9. Classification maps of different methods on the Salinas Valley dataset (a) Ground (b) SVM (c) IFRF (d) PCA-EPFs (e) LDM (f) LDM-FL (g) GFDN (h) SuperPCA (i) SuperBF (j) AM-SVM (k) AM-LDM (l) DTIC-SVM (m) DTIC-LDM (n) DTNC-SVM (o) DTNC-LDM (p) AMDTIC-SVM (q) AMDTNC-SVM (r) AMSCF-IL (s) AMSCF-NL.

for Indian Pines, Salinas Valley and Kennedy Space Center, respectively. Approximatively, the OA values of DTIC-LDM were 18.85%, 10.53% and 12.14% huger than that of LDM for the three datasets. Thus, the spatial correlation feature extracted by DTNCF or DTICF in this work is efficient for enhancing the hyperspectral classification.
Fourth, the OA values of AMDTNC-SVM and AMSCF-NL in Indian Pines are 2.36% and 2.65% grander than those of AM-SVM and AM-LDM. Furthermore, the OA values of AMDTNC-SVM and AMSCF-NL in Salinas Valley are about 6.01% and 3.42% upper than those of AM-SVM and AM-LDM. In addition, the OA values of AMDTNC-SVM and AMSCF-NL in Kennedy Space Center are about 2.93% and 1.90% larger than those of AM-SVM and AM-LDM. Also, the OA values of AMDTIC-SVM and AMSCF-IL in Indian Pines are 1.87% and 2.11% grander than those of AM-SVM and AM-LDM. Therefore, it can be included that the spatial correlation feature extracted by DTNCF or DTICF in this work can compensate for the AMF deficiency in the hyperspectral classification.

Fifth, the OA values of AMSCF in Indian Pines, Salinas Valley, and Kennedy Space Center data sets are 7.20%, 2.89%, and 2.74% larger than that of PCA-EPFs, respectively. In addition, the OA values of AMSCF are 9.26%, 2.19%, and 4.54% greater than that of IFRF for Indian Pines, Salinas Valley, and Kennedy Space Center, respectively. It can be concluded that the fusion of multiple spatial features is very effective.

Sixth, the OA values of AMSCF for Indian Pines, Salinas Valley, and Kennedy Space Center data sets are 9.77%, 9.39%, and 9.95% larger than that of SuperPCA, respectively. In addition, the OA values of AMSCF are 10.33%, 9.04%, and 9.58% greater than that of SuperBF for Indian Pines, Salinas Valley, and Kennedy Space Center, respectively. It is shown that AMSCF has better performance than some classification methods with superpixel strategy.
Seventh, the OA values of AMDTNC-SVM and AMSCF-NL in Indian Pines are 0.49% and 0.54% larger than that of AMDTIC-SVM and AMDTIC-LDM. Similarly, the OA values of AMDTNC-SVM and AMSCF-NL in Salinas Valley are 0.66% and 0.15% greater than that of AMDTIC-SVM and AMDTIC-LDM. However, the OA values of AMDTNC-SVM and AMSCF-NL in Kennedy Space Center are 0.45% and 0.95% lower than that of AMDTIC-SVM and AMDTIC-LDM. It can be concluded that DTICF and DTNCF have different adaptability in HSI datasets, and the AMSCF algorithm can extract the optimal spatial correlation features and fuse them with AMF, demonstrating the adaptive classification performance with LDM in different HSI datasets.

Last but not least, the OA values of AMSCF in Indian Pines, Salinas Valley, and Kennedy Space Center are 5.51%, 0.84% and 3.83% larger than that of LDM-FL, respectively. Therefore, the spatial texture feature and spatial correlation feature extracted by AMF and DTNCF or DTICF in this paper are better for increasing the performance of LDM than that of RF. In addition, the OA values of AMSCF in Indian Pines, Salinas Valley, and Kennedy Space Center are 2.18%, 3.75% and 4.39% higher than that of GFDN, proving that AMSCF also has better performance than the classification methods with the certain deep learning.

To demonstrate the effect of training ratios on the classification, which had been used to test the different values in the three datasets, as shown in Fig.10. This figure indicates that the OA value of the proposed method can reach 96.38% if the training sample is 3% of the Indian Pines dataset. Also, the OA value can be enhanced to 98.31% when the ratio increases by 6%. In addition, if the training sample ratio of the Salinas Valley dataset varies from 0.1% to 0.2%, the OA value increases from 83.23% to 94.28%. It can achieve 99% when the ratio increases to 7%. Moreover, the OA value of the proposed method can be over 94% if the training sample is 2% of the Kennedy Space Center dataset. Also, the OA value can reach 99.05% when the ratio increases by 5%. Thus, the proposed method provided the best classification performance, with the stability of AMSCF of the different training ratios for HSI datasets.

**IV. CONCLUSION**

This paper combined two kinds of spatial features for LDM classification, and proposed a hyperspectral image classification method namely AMSCF. First, after splatting blurring and slicing, AMF with an appropriate height of manifold can extract better local and global spatial features and obtain better spatial texture features of HSI to improve the classification. Second, the spatial correlation features extracted by DTICF and DTNCF are effective to compensate for the deficiency of AMF in HSI classification. Last, the two spatial features were linearly fused for LDM classification, respectively. To verify the superior performance of AMSCF, three hyperspectral image datasets were tested. Although the training sample ratios are only 4%, 0.7%, and 5% of Indian Pines dataset, Salinas Valley dataset and Kennedy Space Center dataset, the OA values can reach 97.88%, 99.34% and 99.05%, respectively. It can be concluded that the proposed method can be greatly enhanced compared with other methods. The first advantage of this method is that the superior spatial texture feature obtained by AMF at the appropriate height of manifold, and local and global optimization capabilities can effectively assist the classification. The second benefit is that spatial correlation features extracted by DTICF and DTNCF can be adapted to different HSI datasets to attain better classification performance, with the stability of AMSCF of the different training ratios for HSI datasets.
the optimal spatial correlation feature and compensate the deficiency of AMF, thus effectively improving LDM classification for HSI. However, the disadvantage of the proposed method is that AMSCF needs too many parameters and cannot automatically adapt to all datasets. Therefore, our future work will focus on how to automatically determine the optimal parameters of AMSCF.

REFERENCES

[1] C. McCann, K. S. Repasky, R. Lawrence, and S. Powell, “Multi-temporal mesoscale hyperspectral data of mixed agricultural and grassland regions for anomaly detection,” ISPRS J. Photogramm. Remote Sens., vol. 131, pp. 121–133, Sep. 2017.

[2] C. M. Gevaert, J. Suomalainen, J. Tang, and L. Kooistra, “Generation of Spectral–Temporal response surfaces by combining multispectral satellite and hyperspectral UAV imagery for precision agriculture applications,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 8, no. 6, pp. 3140–3146, Jun. 2015.

[3] F. Chen, H. Jiang, T. Van De Voorde, S. Lu, W. Xu, and Y. Zhou, “Land cover mapping in urban environments using hyperspectral APEX data: A study case in baden, switzerland,” Int. J. Appl. Earth Observ. Geoinf., vol. 71, pp. 70–82, Sep. 2018.

[4] F. Chen, K. Wang, T. Van De Voorde, and T. F. Tang, “Mapping urban land cover from high spatial resolution hyperspectral data: An approach based on simultaneously unmixing similar pixels with jointly sparse spectral mixture analysis,” Remote Sens. Environ., vol. 196, pp. 324–342, Jul. 2017.

[5] F. J. A. van Ruitenbeek, H. M. A. van der Werff, W. H. Bakker, F. D. van der Meer, and K. A. A. Hein, “Measuring rock microstructure in hyperspectral mineral maps,” Remote Sens. Environ., vol. 220, pp. 94–109, Jan. 2019.

[6] Z. Asetl, J. V. Taranik, and D. N. Riley, “Mapping rock forming minerals at boundary canyon, death valley national park, california, using aerial SEBAS thermal infrared hyperspectral data,” Int. J. Appl. Earth Observ. Geoinf., vol. 64, pp. 326–339, Feb. 2018.

[7] L. Petrescu and E. Bilal, “Speciation of natural uranium and thorium in surface waters around a uranium mine (Bistrita Mts, Romania),” Geochimica et Cosmochimica Acta, vol. 70, no. 18, p. A486, Aug. 2006.

[8] X. Zha, A. K. Skidmore, R. Darvishzadeh, and T. Wang, “Estimation of forest leaf water content through inversion of a radiative transfer model from LiDAR and hyperspectral data,” Int. J. Appl. Earth Observ. Geoinf., vol. 27, pp. 120–129, Feb. 2015.

[9] F. Melgani, 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.

[10] M. Pal and G. M. Foody, “Feature selection for classification of hyperspectral data by SVM,” IEEE Trans. Geosci. Remote Sens., vol. 48, no. 5, pp. 2297–2307, May 2010.

[11] T. Zhang and Z.-H. Zhou, “Large margin distribution machine,” in Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD), Aug. 2014, pp. 313–322.

[12] K. Zhan, H. Wang, H. Huang, and Y. Xie, “Large margin distribution machine for hyperspectral image classification,” J. Electron. Imag., vol. 25, no. 6, Dec. 2016, Art. no. 063024.

[13] J. Liao, L. Wang, G. Zhao, and S. Hao, “Hyperspectral image classification based on bilateral filter with linear spatial correlation information,” Int. J. Remote Sens., vol. 37, no. 17, pp. 6861–6883, Sep. 2019.

[14] A. Meka and S. Chaudhuri, “A technique for simultaneous visualization and segmentation of hyperspectral data,” IEEE Trans. Geosci. Remote Sens., vol. 53, no. 4, pp. 1707–1717, Sep. 2014.

[15] M. D. Mura, A. Villa, J. A. Benediktsson, J. Chanussot, and L. Bruzzone, “Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis,” IEEE Geosci. Remote Sens. Lett., vol. 8, no. 3, pp. 542–546, Dec. 2010.

[16] P. Ghamisi, J. A. Benediktsson, and M. O. Ulfarsson, “Spectral-spatial classification of hyperspectral images based on hidden Markov random fields,” IEEE Trans. Geosci. Remote Sens., vol. 52, no. 5, pp. 2565–2574, Jun. 2013.

[17] A. Mehta and O. Dikshit, “Segmentation-based projected clustering of hyperspectral images using mutual nearest Neighbour,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 10, no. 12, pp. 5237–5244, Nov. 2017.
[38] T. Qiao, Z. Yang, J. Ren, P. Yuen, H. Zhao, G. Sun, S. Marshall, and J. A. Benediktsson, “Joint bilateral filtering and spectral similarity-based sparse representation: A generic framework for effective feature extraction and data classification in hyperspectral imaging,” *Pattern Recognit.*, vol. 77, pp. 316–328, May 2018.

[39] Y. Guo, H. Cao, S. Han, Y. Sun, and Y. Bai, “Spectral–Spatial Hyperspectral Image classification with K-Nearest neighbor and guided filter,” *IEEE Access*, vol. 6, pp. 18582–18591, Mar. 2018.

[40] J. Liao, L. Wang, S. Hao, and G. Zhao, “Hyperspectral image classification based on fusion of guided filter and domain transform interpolated convolution filter,” *Can. J. Remote Sens.*, vol. 44, no. 5, pp. 476–490, Sep. 2018.

[41] 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, Jul. 2013.

[42] X. Kang, X. Xiang, S. Li, and J. A. Benediktsson, “PCA-based edge-preserving features for hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7140–7151, Sep. 2017.

[43] L. Wang, S. Hao, and G. Zhao, “A deep network architecture for Super-Resolution-Aided hyperspectral image classification with classwise loss,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4650–4663, Jun. 2018.

[44] J. Liao and L. Wang, “Hyperspectral image classification based on fusion of curvature filter and domain transform recursive filter,” *Remote Sens.*, vol. 11, no. 7, p. 833, 2019.

JIANSHANG LIAO received the M.S. degree in communication and information systems from the Guangdong University of Technology, Guangzhou, China, in 2012. He is currently pursuing the Ph.D. degree in information and communication engineering with Harbin Engineering University, Harbin, China. He is also a Professor with the Guangdong Communication Polytechnic, Guangzhou. His current research interests include machine learning, pattern recognition, and hyperspectral image classification.

LIGUO WANG (Member, IEEE) received the M.S. and Ph.D. degrees in signal and information processing from the Harbin Institute of Technology, Harbin, China, in 2002 and 2005, respectively.

From 2006 to 2008, he was a Postdoctoral Researcher with the College of Information and Communications Engineering, Harbin Engineering University, Harbin, where he is currently a Professor. He is also a Distinguished Professor with Dalian Minzu University. His current research interests include remote sensing and machine learning. He has authored three books about hyperspectral image processing and more than 180 articles in journals and conference proceedings.