Received January 28, 2020, accepted February 7, 2020, date of publication February 11, 2020, date of current version February 19, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2973246

A Novel Multi-Feature Joint Learning Method for Fast Polarimetric SAR Terrain Classification

JUNFEI SHI, HAIYAN JIN, AND XIAOHUA LI
School of Computer Science and Technology, Xi'an University of Technology, Xi'an 710048, China

Corresponding author: Junfei Shi (shijunfei1@163.com)

This work was supported in part by the National Basic Research Program (913 Program) of China under Grant 2013CB329402, in part by the National Natural Science Foundation of China under Grant 61113090, Grant 61472319, Grant 61703333, and Grant 61502382, in part by the Natural Science Foundation of Shaanxi Province under Grant 2018JQ6055 and Grant 2019JQ-740, in part by the Research Project of Shaanxi Provincial Education Department under Grant 19JK0566, and in part by the Technology Innovation Project of Xi'an University of Technology under Grant 112/256081620.

ABSTRACT Polarimetric synthetic aperture radar (PolSAR) image classification is one of the most important study areas for PolSAR image processing. Many kinds of PolSAR features can be extracted for PolSAR image classification, such as the scattering, polarimetric or image features. However, it is difficult to improve the classification accuracy of PolSAR images by using all these low-level features directly, since they may conflict with each other for classification. Hence, how to joint learn these low-level features to obtain high-level discriminating features is a challenging task. To solve this problem, a novel fast multi-feature joint learning method(fMF-JLC) is proposed for PolSAR image classification. The proposed method extract three kinds of low-level features of PolSAR data at first. Then, a multi-feature joint sparse representation model(MF-JSR) is proposed by designing joint sparse constraints on the extracted features above. Moreover, the joint sparse features are further compressed to overcome the dimension curse and acquire semantic features by the topic model. By this way, the low-level features are fused and discriminating high-level features are acquired. However, the pixel-wise feature learning method is time consuming. To speed the proposed method, a superpixel-based fast learning method is designed by involving the contextual relationship. Experiments are taken on three sets of real PolSAR data with different sensors and bands, and several compared methods are used to verify the effectiveness of the proposed method. The experimental results illustrate that the proposed method can obtain better performance than the state-of-art methods, especially for the heterogenous areas.

INDEX TERMS Polarimetric SAR classification, joint multi-feature sparse representation, joint multi-feature learning, fast classification method.

I. INTRODUCTION

Polarimetric synthetic aperture radar (PolSAR) terrain classification is the key of PolSAR image interpretation. PolSAR terrain classification has attracted many researchers' attention since it can provide more scattering characteristics and polarimetric information than the SAR data. Abundant PolSAR classification methods have been proposed and achieved superior performance. However, there are still some challenges for PolSAR terrain classification. Firstly, speckle is one of the main factors that causes misclassification. In addition, the heterogenous areas, such as the urban area, are difficult to be classified into semantic homogenous regions, since pixels within the heterogenous terrain object have obvious intensity variations, and the corresponding low-level features are great different. It is the gap between the low-level feature and the high-level semantic.

To suppress the speckle noises, many PolSAR classification approaches have been proposed, which are mainly categorized into three types: 1) statistical distribution-based methods [1], [2], which assume the PolSAR speckle data as Gaussian or non-Gaussian distribution. Specifically, under the homogeneity assumption, the scattering matrix obeys Gaussian distribution, such as the Wishart-based classification method [3] and the improved Cloude-Pottier decomposition method [4]. However, the homogeneity assumption is not satisfied for high-resolution and heterogeneous PolSAR images any more. So, some non-Gaussian modeling methods [5], [6] were proposed, such as K-Wishart, G0 and KummerU distributions. 2)statistical and scattering information based

The associate editor coordinating the review of this manuscript and approving it for publication was Wenming Cao.

methods, which added the scattering information to improve the classification accuracy, such as Cloude decomposition [7], Freeman decomposition [8], Huyen decomposition [9] and so forth. These approaches can obtain a fine classification result by considering polarimetric information, but produce an salt-and-pepper classification map without the structural and spatial information. References [10], [11] 3)image feature-based PolSAR classification methods [12]–[14], which include approaches based on textural modeling [15], such as gray level co-occurrences matrices (GLCM) and Gabor or wavelet features [16], and approaches with regularization criterion, such as Markov Random Field(MRF) [17], [18] and contour criterion [19]. Compared with classification methods without spatial information, these approaches can achieve more homogenous classification results by considering the local spatial dependency in an image.

These methods above are mainly based on a certain kind of features for PolSAR image classification. However, the PolSAR image include various terrain types, and many kinds of features can be extracted from different perspectives. It is difficult to classify all the terrain types well with a single feature. Multiple kinds of features [20] can provide more discriminating information for PolSAR classification, since they can describe the PolSAR image from different viewpoints. However, how to combine these features to improve classification accuracy has been an important issue for PolSAR image classification.

Sparse representation is a promising method to fuse multiple features, which has attracted many researchers. Sparse representation model [21] has been widely used in face recognition [22], hyperspectral image classification [23] and object detection [24]. Sparse learning [25] assumes the input signal can be approximately represented by a linear combination of atoms in a dictionary. The sparse constraints promise that many of coefficients are zero since the dictionary is over-complete. Later, sparse representation methods were extended to dealing with multi-feature representation based classification, and some multi-feature joint sparse representation methods [26]–[28] were proposed. In 2015, Zhang et al. [27] proposed a fast multi-feature joint sparse representation method for hyperspectral image classification. In 2017, Yang *et al.* [28] proposed a joint multi-feature dictionary learning method for face recognition. These methods can fuse multiple features effectively and improve classification accuracy. However, all these methods are used for optical images, which are totally different from PolSAR images. Considering polarimetric characteristics [29], in 2015, an improved PolSAR classification method [30] is proposed by using multi-feature combination. In 2017, a nearest-regularized subspace classification method(NRS-MRF) [31] was proposed for PolSAR images by using polarimetric feature vector and spatial information. In 2018, a spatial multi-attribute graph was constructed for PolSAR classification [32]. Recently, a multi-level feature extraction method [33] was proposed for PolSAR image classification. All these methods can combine multiple kinds of PolSAR features well. However, the joint sparse coefficients are obtained by stacking different features, which will cause dimension curse. Moreover, without considering semantic information, they are difficult to classify the heterogenous areas into the same class, such as the urban area. It is because that pixels within them have obvious bright-dark intensity changes. These intensity changes are mainly caused by their imaging characteristics. For low-resolution images, the intensity change is caused by the scattering wave of the object and the ground nearby. For high-resolution images, it is produced by scattering waves of ground objects and their shadows. This variation makes two pixels nearby produce different sparse feature representation. And different features will produce different classes within the heterogenous area. Therefore, higher-level features [34], [35] should be exploited to obtain a semantically consistent region within the heterogenous area.

Topic model [36] is an efficient tool to establish a bridge between the low-level features and high-level semantic ones. High-level semantic is learned by a multi-level generative model in a latent space. In 1990, a latent semantic analysis (LSA) [37] model is proposed. Later, Hofmann extended it into probabilistic LSA (pLSA) [38], which assumes that a document and a word are conditionally independent from the topic given. However, it always causes overfitting and is difficult to assign probability to a previously unseen document. In 2003, the latent Dirichlet allocation (LDA) model [39] is proposed, which induces hidden random variable to reduce overfitting. The LDA is formulated by a three-level Bayesian model and variation inference is used to estimate parameters.

Inspired by the sparse representation and topic model, in this paper, a novel multi-feature joint learning method is proposed for the PolSAR image classification. Compared with the conventional classification methods, the proposed method has three characteristics as follows: 1) A joint multi-feature sparse learning method is proposed to combine three types of features, which are the polarimetric data, scattering characteristics and image contextual features respectively. 2)According to the learned joint sparse features, higher-level features are acquired by the topic model to suppress the dimension curse. 3) A superpixel-based fast learning method is exploited to involve spatial information and reduce computing time. Experimental results illustrate the effectiveness of the proposed method.

The rest of this paper is organized as follows. In Section II, the PolSAR data is introduced. In Section III, the proposed method is described in detail. The experimental results are shown and discussed in Section IV. Finally, the conclusion is drawn in Section V.

II. POLAREIMTERIC SAR DATA

A. POLSAR DATA REPRESENTATION

PolSAR data is the scattering echo waves of radar system. The electromagnetic waves are transmitted and received by two ways: horizonal and vertical. Under the reciprocal

TABLE 1. The target information for nine elements of Huyen decomposition.

Elements	Target Information
A_0	the symmetry of targets
$B_0 - B$	the non-symmetry of targets
$B_0 + B$	Non-regularity of target
C	linear characteristic of target
D	local curvature difference
E	surface torsion
F	helicity of target
G	the glue between symmetry and non-symmetry
H	orientation of target

backscattering condition, PolSAR data can be represented by the scattering vector k [40] in the linear basis as:

$$k = [\, S_{hh} \quad \sqrt{2}S_{hv} \quad S_{vv} \,]^T \tag{1}$$

where S_{hh} is the scattering element by horizontal transmitting and horizontal receiving polarization, and the other three elements are similarly defined. $\sqrt{2}$ is used to keep the computing consistency of the total power.

According to the scattering vector k, polarimetric SAR data can also be expressed as a covariance matrix C:

$$C = kk^{*T} \tag{2}$$

where the superscript $*$ is the complex conjugate.

In addition, polarimetric information can be represented by a coherency matrix T which can describe the scattering procedure of terrain objects well. It is obtained by the linear transformation of the covariance matrix C. Generally, the diagonal elements of T are used as RGB channels respectively to illustrate the PolSAR PauliRGB image.

B. TARGET DECOMPOSITION OF POLSAR DATA

Multiple of target decomposition methods have been proposed for decades. We will introduce the Cloude and Pottier, Freeman and Durden and Huyen decomposition respectively in this section.

1) CLOUDE AND POTTIER DECOMPOSITION

According to the Cloude and Pottier decomposition, the scattering entropy H,the anti-entropy A and the average of scattering angles α are obtained by the eigenvalue decomposition of coherency matrix T. The details can be found in reference [7].

2) FREEMAN AND DURDEN DECOMPOSITION

Freeman and Durden decomposition can divide the covariance matrix into surface, double-bounce and volume scattering. So, the powers of three scattering categories can be calculated respectively. The details can be found in reference [8].

3) HUYEN DECOMPOSITION

Huyen decomposition [9] can represent the coherency matrix as 9 independent elements $\{A_0, B_0, B, C, D, E, F, G, H\}$. Their target information is given in Table 1.

III. PROPOSED METHOD

For considering the spatial relationship and semantic information, a fast multi-feature joint sparse learning method is proposed for the PolSAR image classification. The main procedure is illustrated in Fig. 1. Firstly, three types of features are extracted, which includes features from the original data, target decomposition and the PolSAR image. Secondly, based on the three types of features, a joint multi-feature sparse representation method is proposed by constructing three sets of codewords. This method mainly includes that a dictionary is formulated by combining these codewords, and the joint sparse features are extracted from the dictionary. Thirdly, to reduce computing time, superpixels are extracted by the simple linear iterative cluster(SLIC) method, and sparse representation is obtained for each superpixel by adding the similar sparse constraint in the superpixel. Then, topic model is used to learn superpixels' high-level features. Finally, based on the learned high-level features, the SVM method is used to classify the PolSAR image.

A. FEATURE EXTRACTION

The PolSAR features can be extracted from various perspectives, such as polarimetric features, scattering information and image contextual features. Each feature can provide helpful information for PolSAR image classification. However, traditional classification methods don't make full use of all the features. It is because some features are contradictory to each other for classification, and it may reduce the classification performance by simply connecting them together. In this paper, we propose a joint multi-feature sparse learning method to utilize different features. Firstly, three types of features, which can provide complementary information for image classification, are extracted from PolSAR images. One type of feature is extracted from the original PolSAR data named *Feature 1* in Fig. 1. The second one is extracted from the target decomposition named *Feature 2*, and the last one is extracted from the PolSAR images named *Feature 3*, such as the textural and contour features, which can describe the heterogenous area effectively.

1) FEATURES FROM ORIGINAL DATA (16 FEATURES)

The first type of feature, named Feature 1, is extracted from the scattering matrix, the coherency matrix and the SPAN data respectively.

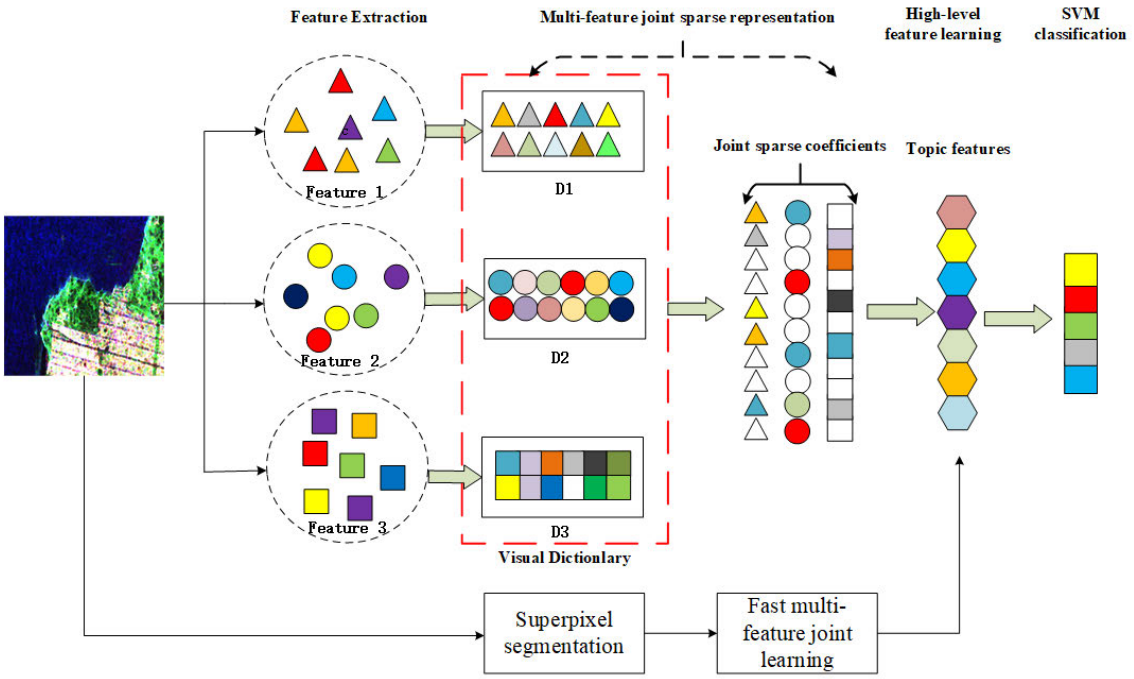

FIGURE 1. Procedure of the proposed fast multi-feature joint learning method for PolSAR terrain classification.

Scattering matrix elements(6 features):

$$\{real(S_{hh}), imag(S_{hh}), real(S_{hv}), imag(S_{hv}), real(S_{vv}), imag(S_{vv})\} \quad (3)$$

Coherency matrix elements(9 features):

$$\{T_{11}, T_{22}, T_{33}, real(T_{12}), imag(T_{12}), real(T_{13}), imag(T_{13}), real(T_{23}), imag(T_{23})\} \quad (4)$$

Features from SPAN data(1 feature):

$$span = |S_{hh}|^2 + 2|S_{hv}|^2 + |S_{vv}|^2 \quad (5)$$

2) FEATURES FROM TARGET DECOMPOSITION (17 FEATURES)

According to the target decomposition of PolSAR data, three types of decomposition parameters and two polarimetric parameters are extracted as Feature 2.

Cloude and Pottier decomposition(3 features): the entropy H, the anti-entropy A and the average of scattering angle α.

Freeman decomposition (3 features): the surface, double-bounce and volume scattering power.

Huynen decomposition(9 features):

$$\{A_0, B_0, B, C, D, E, F, G, H\} \quad (6)$$

Polarimetric parameters(2 features):

co-polarization ratio:

$$r_o = \frac{\langle S_{vv} S_{vv}^* \rangle}{\langle S_{hh} S_{hh}^* \rangle} \quad (7)$$

where $\langle \cdot \rangle$ is an operator of multi-look processing, and r_o measure the difference between S_{vv} and S_{vv}.

cross-polarization ratio:

$$r_x = \frac{\langle S_{hv} S_{hv}^* \rangle}{\langle S_{hh} S_{hh}^* \rangle} \quad (8)$$

where r_x is sensitive to the volume scattering.

3) FEATURES FROM POLSAR IMAGE(20 FEATURES)

According to the PolSAR image, the texture and contour features are extracted as Feature 3.

a: TEXTURE FEATURES (4 FEATURES)

Gray-level co-occurrence matrix (GLCM) [16] is an effective tool to describe the texture of the SAR images by calculating the local spatial relationship of an image. It can reflect the orientation, space and variation mode in a local area. Based on GLCM, four textural features are defined as:

Contrast: It measures the local variation of the image. The stronger the local variations, the larger the contrast value. A higher contrast reflects the texture is clearer. So, the contrast reflects the clarity of the texture [41].

$$con = \sum_i \sum_j (i-j)^2 P(i,j) \quad (9)$$

where $P(i, j)$ is the GLCM value between pixel i and j.

Energy: It measures the stability of image texture, and reflects the distribution of image intensity.

$$Asm = \sum_i \sum_j P(i,j)^2 \quad (10)$$

Entropy: It measures the information randomness. Large entropy means more complex intensity distribution.

$$Ent = -\sum_i \sum_j p(i,j) \log p(i,j) \tag{11}$$

Relativity: It measures the similarity of the image intensity.

$$Corr = \left[\sum_i \sum_j ((i,j)p(i,j)) - \mu_x \mu_y \right] \Big/ \sigma_x \sigma_y \tag{12}$$

b: CONTOUR FEATURES (16 FEATURES)

In this paper, three channels and SPAN images are used to compute the contour energy values respectively. To calculate contour accurately, anisotropic Gaussian kernel filters are used, which ensure pixels along the edge have more weights. Filters with three scales and 18 orientations are designed to obtain the contour features. In addition, a ratio operator is utilized to compute edge-line energy. Since the diagonal terms of the coherency matrix have been proved to be multiplicative noise model [40], the ratio operator can suppress the speckle noises of the three channels' and SPAN data. Hence, the edge and line energies [42] are defined as:

$$E_{edge} = \left(\frac{1}{n} \sum_{i=1}^{n} w_i x_i \right) \Big/ \left(\frac{1}{m} \sum_{j=1}^{m} w_j x_j \right) \tag{13}$$

$$E_{line} = \min\{E_{edge}^{ij}, E_{edge}^{jk}\} \tag{14}$$

where x_i and x_j are the intensity values in pixel i and j respectively. n and m are pixel numbers in two regions of a filter respectively, and w_i is the anisotropic Gaussian kernel. In Equation 14, i,j and k are the three adjacent regions in a line filter, and j is the middle region. E_{edge}^{ij} stands for the edge energy between regions i and j. This equation means that a line can be detected when two strong edges with very close distance appear simultaneously. Edge and line energies are computed with 3 scales and 18 orientations, and then maximum energy is selected as the the contour feature for each pixel.

B. MULTI-FEATURE JOINT SPARSE REPRESENTATION

Three types of features have been extracted from PolSAR images. However, how to fuse them effectively to improve the classification performance is an important issue. To solve this problem, a multi-feature joint sparse representation method(MF-JSR) is proposed for PolSAR image classification. In addition, a superpixel-based fast MF-JSR method(MF-SJSR) is proposed to reduce the computation time.

1) MULTI-FEATURE JOINT SPARSE REPRESENTATION

Sparse representation is based on the assumption that a signal can be linearly combined by a set of atoms in a given dictionary [43]. Due to its excellent performance, the sparse learning methods have attracted many researchers' attention and been widely applied to face recognition, image classification and SAR image processing. It is noted that the sparse representation model can be expressed as: $\min_{\alpha} \|y - D\alpha\|_2^2 + \lambda\|\alpha\|_0$. The first and second terms are the residual and the sparseness terms respectively. Here, y is the test data, D is the given dictionary, and α is the sparse coefficients. $\|\alpha\|_0$ is the 0-norm sparseness, which makes sure the obtained coefficients are sparse. In addition, λ is the scalar constant.

According to the sparse representation model, a multi-feature visual dictionary is needed to formulate. Firstly, the extracted feature descriptors for each category are formulated as a feature vector f^k. It is noted that the feature vectors should be normalized to make sure the same value range. Second, for each kind of feature, visual codewords are extracted by random sampling for each class of the image. Finally, a multi-feature dictionary is formulated by combining three set of visual dictionaries extracted from three types of features.

Based on the multi-feature visual dictionary, the joint sparse representation method is derived for each pixel. According to the multi-feature case, we suppose each pixel can extract K different kinds of features. For each pixel, the feature vector is defined as $y = [y^1, y^2, \ldots, y^K]$, and $y^k \in \Re^{f^k}$ is the kth kind of feature vector, and f^k is the dimension number of the kth kind of feature vector. The dictionary is composed as $D = [D^1, D^2, \ldots, D^K]$, and $D^k = [D_1^k, D_2^k, \ldots, D_C^k]$ is the subdictionary for the kth feature($D_c^k \in \Re^{f^s \times N_c}, c = 1, \ldots, C$). Here,$C$ is the number of classes in an image. D_c^k is the cth class atom set in the subdictionary D^k, and N_c is the number of atoms in D_c^k. Therefore, the multi-feature joint sparse representation model(MF-JSR) is defined as:

$$\min_{\alpha} \sum_{k=1}^{K} \left(\left\| y^k - D^k \alpha^k \right\|_2^2 + \lambda \|\alpha\|_0 \right) \tag{15}$$

This sparse representation model can obtain multi-feature joint sparse coefficients. However, it fails to consider the relationship of sparse coefficients for different kinds of features. Since each kind of feature is extracted to construct the subdictionary D^k, and different features should achieve the similar discriminate result for the same pixel. So, it is reasonable that the nonzero coefficients should share some similarity. Specifically, we assume that the positions of nonzero coefficients tend to be the same. While, due to great differences between different kinds of features, the sparse coefficients need not to be identical. To achieve this similarity, the $l_{row,0}$-norm penalty is utilized to enforce joint sparsity constraints across multiple features. This regularization encourages the coefficients to share a common sparsity pattern, which can preserve the cross-feature information. By this way, the joint sparse model is modified as a squared reconstruction error term and a sparse $l_{row,0}$-norm regularization term. Thus, different features can not only vote the same class, but also provide additional complementary information. Under this assumption, the multi-feature joint sparse representation

model (MF-JSR) can be modified as:

$$\min_{A} \sum_{k=1}^{K} \left(\left\| y^k - D^k \alpha^k \right\|_2^2 \right) + \lambda \|A\|_{row,0} \tag{16}$$

where $A = \left[\alpha^1, \alpha^2, \ldots, \alpha^K\right]$. The first term is trying to reduce the representation errors for multiple features, and the second one can not only make sure the sparseness for each feature, but also preserve the similar sparse pattern for different features. However, the $l_{row,0}$-norm optimization problem is a NP-hard problem. Here, we use $l_{2,1}$-norm instead of the $l_{row,0}$-norm for optimizing conveniently. We can use greedy algorithms, such as the orthogonal matching pursuit(OMP) method [44], to approximately work out this function. The OMP method has been widely used to solve the sparse optimization problem since it has strong convergence characteristics for optimization. In addition, the OMP method can reduce the computational complexity, since it can only select parts of dictionaries for computation to avoid the large matrix inverse operation. Therefore, we select the OMP algorithm to optimize the MF-JSR problem.

2) SUPERPIXEL-BASED MULTI-FEATURE JOINT SPARSE REPRESENTATION

The proposed MF-JSR method is an effective tool to fuse the multiple types of features. However, the sparse coefficients will be computed for each pixel, and this pixel-wised method is time-consuming. In addition, without considering the neighborhood relationship of pixels, the MF-JSR method is also sensitive to speckle noises and easily causes misclassification. Therefore, the use of contextual information are necessary. Traditional neighbor-based joint sparse representation methods [45] always use the fixed neighborhood window. However, pixels in the fixed neighborhood window not always share the same class label, especially for the edge area. In this paper, we extract the superpixel as the adaptive neighborhood to learn the joint sparse feature. So, to speed the feature learning procedure, a superpixel-based fast feature learning method(MF-SJSR) is proposed for PolSAR image classification. Considering spatial information, we assume all the pixels in a superpixel have the same class label, and their sparse coefficients share the same sparse pattern. By this way, we can learn a superpixel's sparse representation simultaneously for one time, and the computational complexity is reduced greatly.

Superpixels are regions obtained by the initial segmentation of PolSAR image. To obtain superpixels [46], an initial over segmentation is required to partition the SPAN image into a lot of homogenous regions. Many low-level segmentation methods such as watershed [47], mean-shift algorithm [48], SLIC method [49], level set method [50] can be used in this step. Here we choose the SLIC method for initial segmentation since it can produce superpixels with similar size and obtain better edge preservation. For detail information about SLIC, please refer to [49], [51], [52]. To suppress speckle noises and reduce computation time, a superpixel-based joint sparse feature representation is extracted as the initial features of the topic model. Each superpixel can extract a joint sparse representation by the MF-SJSR method.

C. TOPIC MODEL-BASED HIGH-LEVEL FEATURE LEARNING

The joint sparse representation can describe the terrain features well. However, adjacent superpixels may produce great different feature representations within the heterogenous area, such as the urban area, although they represent the same terrain type. Therefore, high-level semantic features should be learned for describing heterogenous areas. Latent Dirichlet allocation (LDA),also known as topic model, is a popular feature learning method. It is based on a generative probabilistic model, which is proposed for document classification at first. The basic idea is to learn a set of latent topics, and each document can be represented as the mixture of the topics. In addition, topics are learned from a distribution of words by a generative probabilistic model. Recently, LDA has been used for the scene classification of the natural images. In this paper, we propose a topic model-based multi-feature learning method for PolSAR terrain classification. It is proposed to learn high-level semantic features for complex terrain types, such as the urban area or the forest. The topic model-based multi-feature learning procedure is illustrated in Fig. 2. Firstly, joint sparse representation features are obtained for each superpixel by the proposed MF-SJSR method. Then, the LDA model is utilized to learn the high-level features for each superpixel. Furthermore, the topic model can reduce the sparse feature dimension effectively.

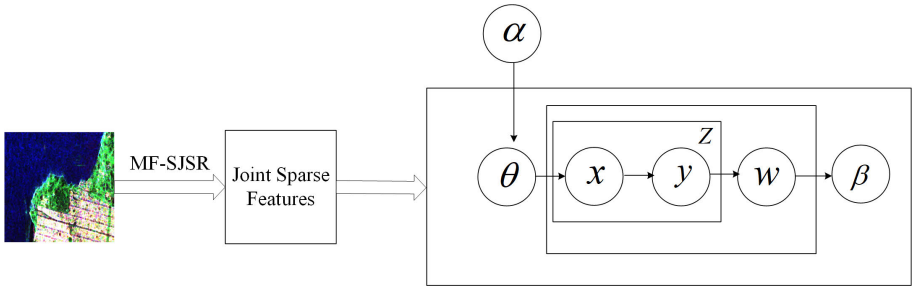

FIGURE 2. The topic model-based multi-feature learning procedure.

LDA assumes the following generative process for each image w in a set D:

1) Choose $N \sim Poisson(\xi)$

2) Choose $\theta \sim Dir(\alpha)$

3) for each of the N words w_n:

(a)Choose a topic $z_n \sim Multinomial(\theta)$

(b)Choose a word w_n from $p(w_n|z_n, \beta)$, a multinomial probability conditioned on the topic z_n.

According to this assumption, an LDA model is constructed. Based on the probability model, the marginal distribution of an image is defined as:

$$p(w|\alpha, \beta) = \int p(\theta|\alpha)\left(\prod_{n=1}^{N}\sum_{z_n} p(z_n|\theta)p(w_n|z_n, \beta)\right) d\theta \quad (17)$$

Given the probability model, the key problem is the inference and parameter estimation. The posterior distribution of the hidden variables is denoted as:

$$p(\theta, z|w, \alpha, \beta) = \frac{p(\theta, z, w|\alpha, \beta)}{p(w|\alpha, \beta)} \quad (18)$$

The parameter estimation is an optimization problem. An iterative method [53] is used to optimize parameters. After variation inference, parameters are updated by:

$$\phi_{ni} \propto \beta_{iw_n} \exp\left\{E_q\left[\log(\theta_i)|\gamma\right]\right\} \quad (19)$$

$$\gamma_i = \alpha_i + \sum_{n=1}^{N} \phi_{ni} \quad (20)$$

After the superpixel-based sparse representation, the number of samples is greatly reduced. We take 10% of superpixels as the training samples to train the topic model. The remaining superpixels are used as the test samples. The maximum likelihood methods are used to calculate the model's parameters. Thus, high-level features are learned by the latent topic model.

D. SVM CLASSIFICATION

The support vector machines(SVM) classifier finds a hyperplane which separates two-class data with maximal margin [54]. The margin is defined as the distance of the closest training point to the separating hyperplane. Given a training set of M data points $\{y_i, x_i\}_{i=1}^{M}$ where $x_i \in R^n$ is the ith input pattern, and $y_i \in R$ is the output pattern. The SVM method tries to construct a classifier defined as:

$$y(x) = sign\left[\sum_{i=1}^{M} \alpha_i y_i \psi(x, x_i) + b\right] \quad (21)$$

where α_i is positive real constant and b is a real constant. $\psi(x, x_i)$ has multiple of choices. If $\psi(x, x_i) = x_i^T x$, it is a linear SVM; If $\psi(x, x_i) = (x_i^T x + 1)^d$, it is a polynomial SVM of degree d; If $\psi(x, x_i) = \exp\left\{-\|x - x_k\|_2^2\right\}/\sigma^2$, it is a RBF SVM, where σ is a constant. Since most of data do not satisfy the linear distribution, we use the Gaussian kernel SVM to conduct nonlinear classification in this paper. Finally, the procedure of the proposed fMF-JLC method is illustrated in Table 2.

TABLE 2. The algorithm procedure of the proposed fMF-JLC.

Algorithm 1: the proposed fMF-JLC
step 1: Extract three types of features from PolSAR images.
step 2: Construct the three visual dictionaries.
step 3: Extract the superpixels of the PolSAR image by the SLIC method.
step 4: Learn the joint sparse representation coefficients for each superpixel by equation 16.
step 5: Learn high-level features by the topic model based on the joint sparse features.
step 6: Classify the high-level features of the PolSAR image by the SVM classifier.

IV. EXPERIMENTAL STUDY

A. EXPERIMENTAL SETTINGS

In this section, three real PolSAR data in different bands and sensors are used to test the effectiveness of the proposed method. The first one is RADARSAT-2 C band San Francisco area 4-look fully polarimetric data with the resolution of 5m. The second one is a PolSAR image from the Flevoland Area, which is NASA/JPL AIRSAR L-band four-look fully polarimetric SAR data with the resolution of 8m. The last one is the PolSAR image from the Oberpfaffenhofen area which is 8-look L-band ESAR data. In addition, the ground truth of each PolSAR image is given, and the classification accuracy and the confusion matrix are calculated to measure the performance of the proposed method.

During the experiment, 53 features are extracted from there kinds of features in the PolSAR data, and a dictionary with 300 atoms is formulated for each kind of features. So, a multi-feature dictionary with 900 atoms are formulated, and each superpixel is represented by the sparse features with 900 dimension. In addition, the number of topic is selected as 20. Therefore, the sparse features are further reduced to high-level features with 20 dimension. Furthermore, we select 10% of the superpixel number as the training data, and only one pixel is selected as the training sample in each superpixel. Therefore, only several training pixels are needed. A computer with Intel Core i3 CPU and 4G RAM is used, and all the experiments are conducted on Window7 system with Matlab 2016a.

To evaluate the effectiveness of the proposed method, five methods are used to compare their performance. The first one is a multi-feature SVM method, shorted for 3F + SVM, which is conducted by extracting the proposed three types of features, and then followed by the SVM classification. The second one is the learned joint sparse features-based classification method, noted by 3F + JSR. This is a pixel-wised classification method. The third one is the superpixel-based joint sparse feature learning method, shorted by 3F + SJSR. This method is designed to verify the "fast" characteristic of the proposed method by involving superpixel. The time reduction can be seen by comparing running time between 3F + JSR and 3F + SJSR, which are pixel- and superpixel-wised classification methods respectively. The three methods above extract low- and mid-level features of the proposed

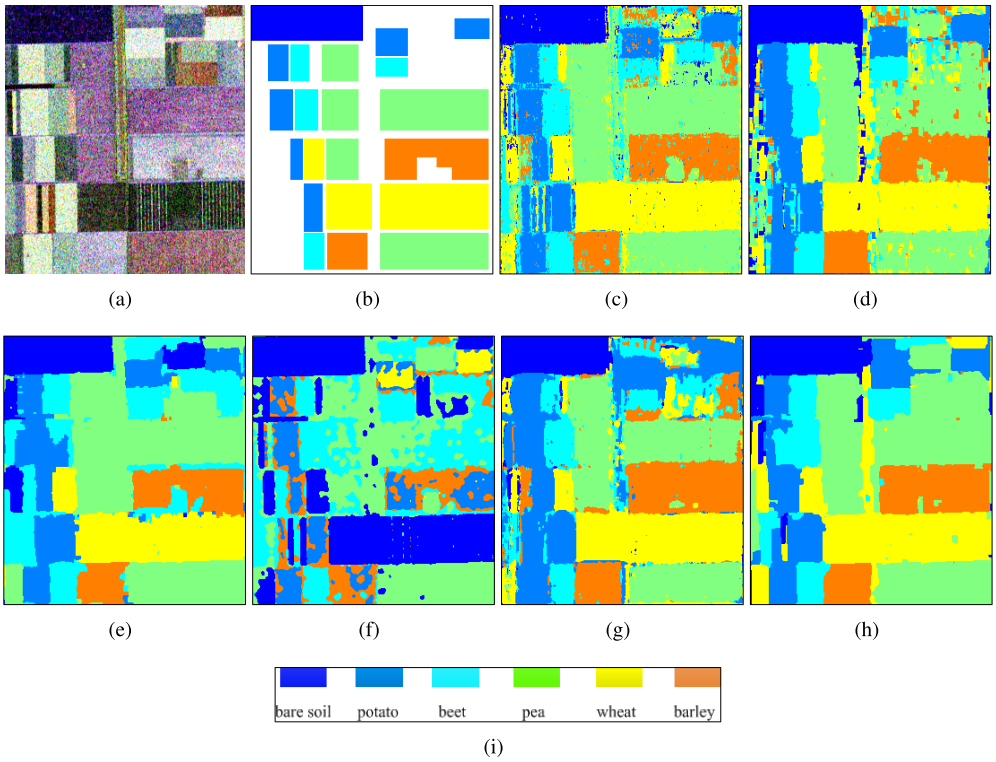

FIGURE 3. Classification results on Flevoland area. (a) Pauli image of Flevoland area. (b) Ground truth of (a). (c) Classification result by the 3F + SVM method. (d) Classification result by 3F + JSR method. (e) Classification result by 3F + SJSR method. (f) Classification result by the Wishart MRF method. (g) Classification result by the NRS-MRF method. (h) Classification result by the proposed method. (i) Corresponding terrain objects for different labels.

method for classification, and aim to verify the effect of the low-, mid- and high-level features in the proposed method by ablation experiments. The fourth method is the Wishart MRF method [3], which uses the MRF to involve the spatial information to reduce the speckle noises. The last one is a nearest-regularized subspace based MRF method, noted by NRS-MRF [31], which applies the nearest-regularized subspace method to extract the sparse representation of multiple features, and obtain initial classification result. Then, the MRF is used to optimize the classification result. To make it fair, all the methods use the proposed three types of features as the initial multi-features.

B. EXPERIMENTAL RESULTS OF FLEVOLAND DATA SET

The PauliRGB image of Flevoland area is shown in Fig. 3(a) with the image size of 300 × 270. The PauliRGB image is obtained by considering $|S_{hh} - S_{vv}|$,$|S_{hh} + S_{vv}|$ and $\sqrt{2}\,|S_{hv}|$ as RGB channels. The ground truth is given in Fig. 3(b). The ground truth is labeled by referring the ground truth map from Ref. [55]. There are 6 categories of crops labeled with different colors in the ground truth map. They are *peas*, *potatoes, wheat, barley, beet* and *bare soil* respectively as shown in Fig. 3(i). During calculating the classification accuracy, the white area is not taken into account due to the lack of true labels in these areas.

The experimental results by five compared and proposed methods are illustrated in Figs. 3(c)-(h) respectively. The 3F + SJSR method in (e) can also obtain better classification result than 3F + SVM and 3F + JSR methods. It shows the effectiveness of the superpixel-based joint sparse feature learning. In addition, the Wishart MRF method confuse the *potatoes* and the *barley* classes. The NRS-MRF method also produces a speckle-noise result. The proposed method can obtain more homogenous regions due to the feature learning and superpixel scheme.

To evaluate the proposed method quantitatively, the classification accuracy is calculated to measure the performance of these methods in Table 3. It demonstrates the proposed method can obtain better performance than other compared methods.

Moreover, the computing time of different methods are illustrated in Table 5. It can be shown that the 3F + SJSR method takes less time than 3F + JSR method, and the proposed method costs less time than the NRS-MRF method. Furthermore, to verify the "fast" characteristic of the proposed method, three sets of PolSAR data are conducted to test the time reduction between 3F + JSR and 3F + SJSR methods in Table 5. We can see 3F + SJSR cost less time than 3F + JSR in all the data sets. Also, the time increasing rate is dramatically reduced with the increasing of the image size.

TABLE 3. Classification accuracy of different methods on Flevoland Data Set(%).

	3F+SVM	3F+JSR	3F+SJSR	Wishart MRF	NRS-MRF	Proposed method
bare soil	95.34	97.86	**99.98**	99.70	97.38	99.92
potato	93.93	94.27	99.49	56.36	**99.98**	86.09
beet	95.55	83.17	99.68	82.91	92.28	**99.68**
pea	97.87	**99.89**	94.28	68.28	96.04	99.62
wheat	97.51	98.19	98.51	0.00	98.54	**98.60**
barley	95.43	96.12	99.58	51.70	**99.78**	99.30
Overall accuracy	96.50	96.74	97.61	59.83	97.42	**97.68**
Kappa	95.57	95.86	96.98	58.92	96.74	**97.06**

TABLE 4. Computing time of different methods on Flevoland Data Set(s).

	3F+SVM	3F+JSR	3F+SJSR	Wishart MRF	NRS-MRF	Proposed method
time	**10.97**	112.59	54.68	402.97	213.29	63.75

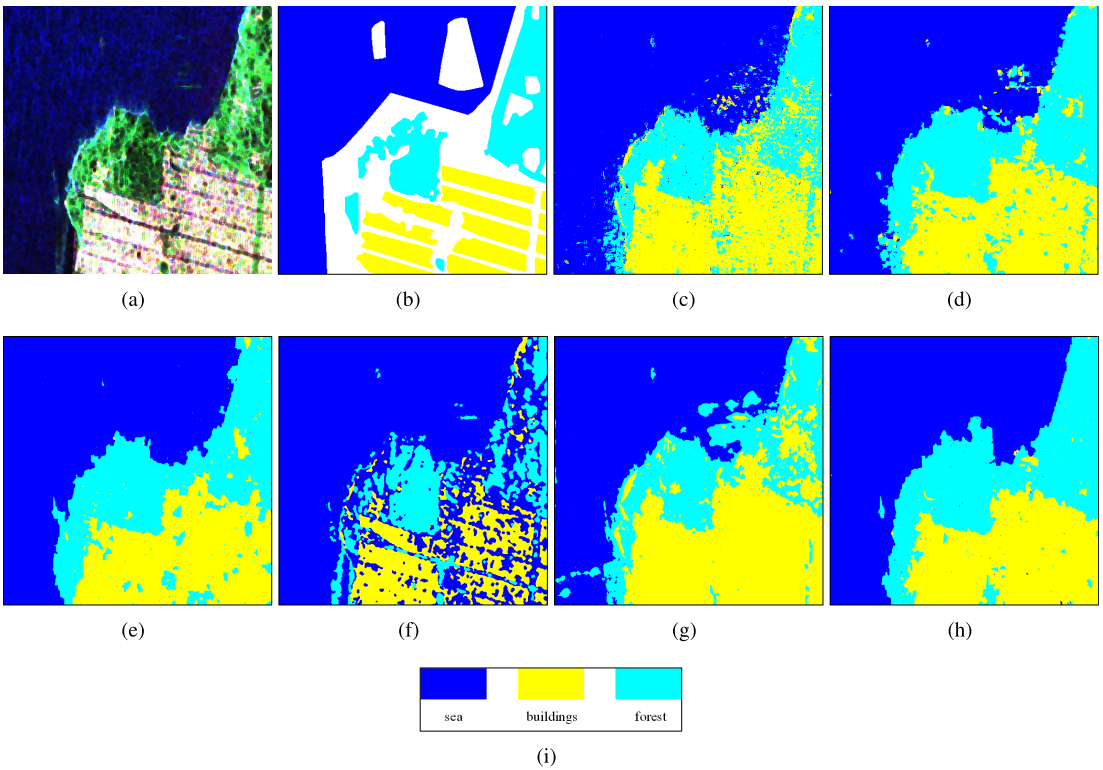

FIGURE 4. Classification results on San Francisco area. (a) Pauli image of San Francisco area. (b) Ground truth of (a). (c) Classification result by the 3F + SVM method. (d) Classification result by 3F + JSR method. (e) Classification result by 3F + SJSR method. (f) Classification result by the Wishart MRF method. (g) Classification result by the NRS-MRF method. (h) Classification result by the proposed method. (i) Corresponding terrain objects for different labels.

TABLE 5. Time reduction experiments of superpixel-based method on different Data Set(s).

	3F+JSR	3F+SJSR
Flevoland	112.59	54.68
San Francisco	234.18	68.32
Oberpfaffenhofen	8617.64	444.8

It mainly because the number of superpixel is not equal proportion increasing.

C. EXPERIMENTAL RESULTS OF SAN FRANCISCO DATA SET

The PolSAR image of San Francisco area is used to test the proposed method, and a subimage of San Francisco area is shown in Fig. 4(a) with the image size of 512 × 512. The corresponding ground truth is given in Fig. 4(b). The ground truth is labeled by referring the ground truth map from Ref. [15]. It can be seen that the PolSAR image is labeled as 3 categories as shown in Fig. 4(b). The area in blue is the *sea*, the cyan color represents the *forest* and the yellow color is the *buildings*. In addition, the other areas lacking true labels are colored in white. To calculate the classification accuracy, we only take the 3 categories of ground objects into account by labeling them with different colors in the experimental result.

The classification results by five compared methods are shown in Figs. 4(c)-(g) respectively. Fig. 4(h) is the

TABLE 6. Classification accuracy of different methods on San Francisco Data Set(%).

	3F+SVM	3F+JSR	3F+SJSR	Wishart MRF	NRS-MRF	Proposed method
sea	98.65	97.75	99.65	**100**	94.48	99.06
forest	95.34	95.33	97.04	67.01	82.44	**99.63**
buildings	95.04	96.18	96.81	64.36	**99.86**	97.86
Overall accuracy	96.76	96.75	97.91	77.12	93.17	**98.91**
Kappa	95.83	95.79	97.11	76.24	88.41	**98.13**

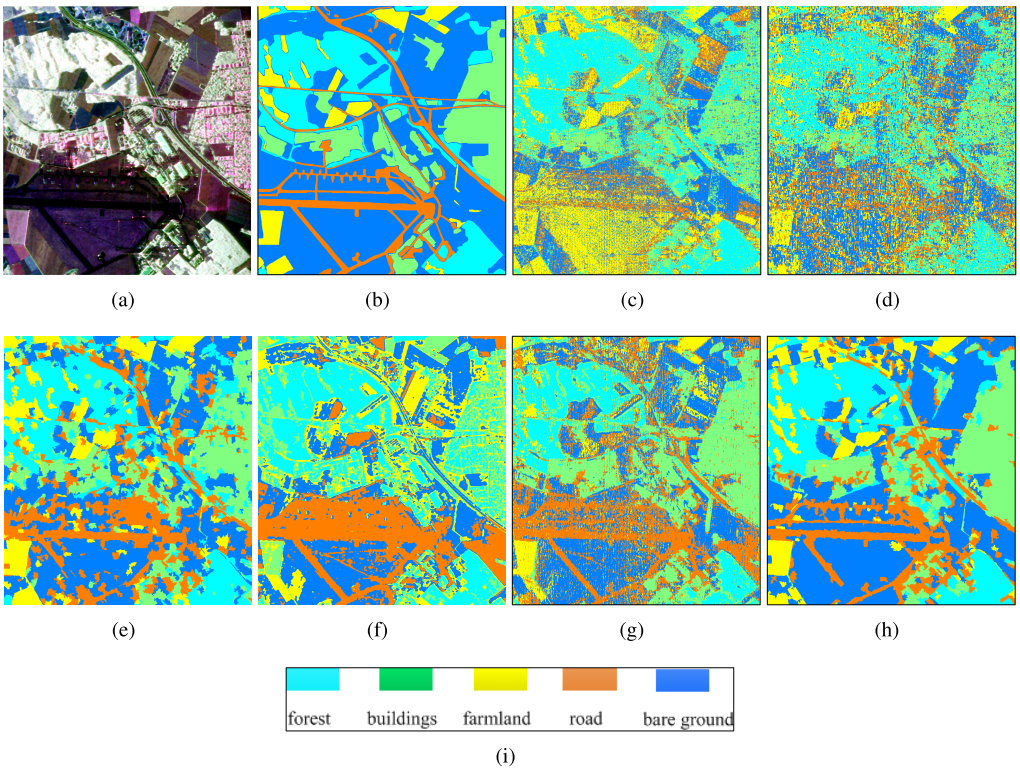

FIGURE 5. Classification results on Oberpfaffenhofen area. (a) Pauli image of San Francisco area. (b) Ground truth of (a). (c) Classification result by the 3F + SVM method. (d) Classification result by 3F + JSR method. (e) Classification result by 3F + SJSR method. (f) Classification result by the Wishart MRF method. (g) Classification result by the NRS-MRF method. (h) Classification result by the proposed method. (i) Corresponding terrain objects for different labels.

classification result by the proposed method. It can be seen that all the methods can classify the *sea* well. The *forest* and the *buildings* are heterogenous areas, and low-level feature based methods are difficult to classify them into semantic homogenous areas. So, the Wishart MRF method in Fig. 4(f) has many misclassifications in the *forest* and the *buildings*. The 3F + SVM and NRS-MRF methods also causes misclassifications without high-level features. Compared with other methods, the proposed method can obtain better performance in the *forest* and the *buildings* since high-level features can be learned by the topic model.

To measure the classification performance, the classification accuracies for each category and the average classification accuracy are given in Table 6. The percent value is given in this table, and the highest values are shown in bold. It can be seen that the Wishart MRF method has 100 percent accuracy in the sea area, while it has poor classification performance in both the *forest* and the *buildings*. It shows this method can classify the homogeneous regions well, but fails to classify the heterogeneous areas due to the lack of high-level features. In contrast, the proposed method can show better performance in both the *forest* and the *buildings* with multi-feature learning. Furthermore, the confusion matrix of the proposed method is illustrated in Table 7. It can be seen that the main misclassification is the *buildings* are confused as other classes in the proposed method.

TABLE 7. Confusion matrix of the proposed method on San Francisco Data Set.

	sea	forest	buildings
sea	104107	993	0
forest	60	37420	78
buildings	22	827	38792

TABLE 8. Classification accuracy of different methods on Oberpfaffenhofen Data Set(%).

	3F+SVM	3F+JSR	3F+SJSR	Wishart MRF	NRS-MRF	Proposed method
bare ground	36.94	45.16	57.39	46.33	50.04	**67.97**
forest	77.31	58.90	79.19	72.40	57.58	**83.10**
buildings	51.11	49.00	75.84	45.49	**84.67**	81.11
farmland	58.42	57.66	**89.08**	14.47	48.59	76.53
road	25.19	37.05	**67.32**	60.82	65.71	66.51
Overall accuracy	46.18	48.01	67.25	50.32	58.52	**73.04**
Kappa	31.70	32.05	56.44	35.87	45.75	**63.38**

D. EXPERIMENTAL RESULTS OF OBERPFAFFENHOFEN DATA SET

The PauliRGB image of Oberpfaffenhofen area is used to test the effectiveness of the proposed method in Fig. 5(a) with the image size of 1300 × 1200. The corresponding ground truth image is illustrated in Fig. 5(b). The ground truth is labeled by referring the satellite image from Google Earth and the ground truth map from Ref. [16]. There are 5 categories of ground objects as shown in Fig. 5(i). They are the *forest*, *buildings*, *farmland*, *road* and *bare ground* respectively.

Fig. 5(c)-(h) are the classification results by five compared and proposed methods respectively. It can be seen that the *buildings* and *forest* are difficult to be classified well. In addition, the road is difficult to be detected for all the methods. Specifically, the 3F + SVM and 3F + JSR are pixel-wised classification methods, which produce pepper-and-salt noise classification results in Figs. 5(c) and (d). The 3F + SJSR method can obtain more homogenous classification result by involving superpixels, while it causes many misclassifications without considering the contextual information of superpixels. The NRS-MRF method in (g) can improve classification performance, but causes the noise classification result. The proposed method can obtain semantically homogenous classification result, especially in *forest* and *buildings*. In addition, the proposed method can obtain better detection result in *road* than other methods, although it is easily confused by the *bare ground* class. In the future, linear features will be involved to improve the classification accuracy of the *road*. Moreover, the classification accuracy and kappa coefficients are calculated to measure the performance of these methods in Table 8. It can be seen that the proposed method can obtain better performance than other compared methods, especially in the *forest* and the *road* which are difficult to be distinguished in this PolSAR image.

V. CONCLUSION

In this paper, a fast multi-feature joint learning method was presented. Traditional PolSAR image classification methods are hardly to classify the heterogenous area into a semantic homogenous region. In this paper, high-level features were learned to resolve this problem from multiple low-level features. Firstly, to better represent the terrain objects, three types of features were extracted and further formulated a set of visual dictionary. Secondly, a multi-feature joint sparse representation model was proposed to fuse these features. Finally, according to the sparse representation, higher-level semantic features were learned by the topic model. In addition, superpixels were utilized to reduce the speckle noises and improved the computation speed. Several experiments were conducted on real PolSAR data, and the quantitative measurement was given to show the effectiveness of the proposed method.

In addition, the proposed method could combine multiple of PolSAR features effectively, and addressed the difficulty of feature selection. Moreover, high-level features were learned, and the learned features could represent the heterogenous terrain types well. By this way, the heterogenous terrain type was classified into a semantic homogenous area. However, the numbers of visual words and topics were given by the user. How to adaptively select the numbers of visual words and topics will be exploited in the further work.

REFERENCES

[1] J.-M. Beaulieu and R. Touzi, "Segmentation of textured polarimetric SAR scenes by likelihood approximation," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2063–2072, Oct. 2004.

[2] L. Bombrun and J.-M. Beaulieu, "Fisher distribution for texture modeling of polarimetric SAR data," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 512–516, Jul. 2008.

[3] J.-S. Lee, M. R. Grunes, T. L. Ainsworth, L.-J. Du, D. L. Schuler, and S. R. Cloude, "Unsupervised classification using polarimetric decomposition and the complex Wishart classifier," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 5, pp. 2249–2258, 1999.

[4] C. Fang, H. Wen, and Y. Wu, "An improved cloude-pottier decomposition using H/α/SPAN and complex Wishart classifier for polarimetric SAR classification," in *Proc. Int. Conf. Radar*, 2007, pp. 1–4.

[5] G. Vasile, J.-P. Ovarlez, F. Pascal, and C. Tison, "Coherency matrix estimation of heterogeneous clutter in high-resolution polarimetric SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 4, pp. 1809–1826, Apr. 2010.

[6] P. Formont, F. Pascal, G. Vasile, J.-P. Ovarlez, and L. Ferro-Famil, "Statistical classification for heterogeneous polarimetric SAR images," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 567–576, Jun. 2011.

[7] S. R. Cloude and E. Pottier, "An entropy based classification scheme for land applications of polarimetric SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 1, pp. 68–78, Jan. 1997.

[8] A. Freeman and S. L. Durden, "A three-component scattering model for polarimetric SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 3, pp. 963–973, May 1998.

[9] J. R. Huynen, "The stokes matrix parameters and their interpretation in terms of physical target properties," *Proc. SPIE*, vol. 1317, pp. 195–207, Oct. 1990.

[10] L. Zhang, W. Ma, and D. Zhang, "Stacked sparse autoencoder in PolSAR data classification using local spatial information," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 9, pp. 1359–1363, Jul. 2016.

[11] W. Xie, Z. Xie, F. Zhao and B. Ren, "POLSAR image classification via clustering-WAE classification model," *IEEE Access*, vol. 6, pp. 40041–40049, Aug. 2018.

[12] H. Dong, W. Ma, Y. Wu, M. Gong, and L. Jiao, "Local descriptor learning for change detection in synthetic aperture radar images via convolutional neural networks," *IEEE Access*, vol. 7, pp. 15389–15403, 2019.
[13] S. Zhang, B. Hou, L. Jiao, Q. Wu, C. Sun, and W. Xie, "Context-based max-margin for PolSAR image classification," *IEEE Access*, vol. 5, pp. 24070–24077, 2017.
[14] K. Ersahin, I. G. Cumming, and R. K. Ward, "Segmentation and classification of polarimetric SAR data using spectral graph partitioning," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 1, pp. 164–174, Jan. 2010.
[15] S. Uhlmann and S. Kiranyaz, "Integrating color features in polarimetric SAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 4, pp. 2197–2216, Apr. 2014.
[16] C. He, S. Li, Z. Liao, and M. Liao, "Texture classification of PolSAR data based on sparse coding of wavelet polarization textons," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 8, pp. 4576–4590, Aug. 2013.
[17] P. Zhong, F. Liu, and R. Wang, "A new MRF framework with dual adaptive contexts for image segmentation," in *Proc. Int. Conf. Comput. Intell. Secur.*, Dec. 2007, pp. 351–355.
[18] J. Shi, L. Li, F. Liu, L. Jiao, H. Liu, S. Yang, L. Liu, and H. Hao, "Unsupervised polarimetric synthetic aperture radar image classification based on sketch map and adaptive Markov random field," *J. Appl. Remote Sens*, vol. 10, no. 2, May 2016, Art. no. 025008.
[19] J. M. Beaulieu, "Pseudo-convex contour criterion for hierarchical segmentation of SAR images," in *Proc. Can. Conf. Comput. Robot Vis.*, 2006, p. 29.
[20] Q. Chen, L. Li, Q. Xu, S. Yang, X. Shi, and X. Liu, "Multi-feature segmentation for high-resolution polarimetric SAR data based on fractal net evolution approach," *Remote Sens.*, vol. 9, no. 6, p. 570, Jun. 2017.
[21] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. Zhang, "A survey of sparse representation: Algorithms and applications," *IEEE Access*, vol. 3, pp. 490–530, 2015.
[22] C. Fan, L. Tian, Y. Ming, X. Hong, G. Zhao, and M. Pietikäinen, "Sparse projections matrix binary descriptors for face recognition," *Neurocomputing*, vol. 297, pp. 8–21, Jul. 2018.
[23] Y. Shao, N. Sang, C. Gao, and L. Ma, "Spatial and class structure regularized sparse representation graph for semi-supervised hyperspectral image classification," *Pattern Recognit.*, vol. 81, pp. 81–94, Sep. 2018.
[24] H. Yang, S. Qu, C. Chen, and B. Yang, "Multiple objects tracking with improved sparse representation and rank based dynamic estimation," *IEEE Access*, vol. 6, pp. 42264–42278, 2018.
[25] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Apr. 2009.
[26] X. Song, S. Jiang, and L. Herranz, "Joint multi-feature spatial context for scene recognition in the semantic manifold," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1312–1320.
[27] E. Zhang, X. Zhang, H. Liu, and L. Jiao, "Fast multifeature joint sparse representation for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 7, pp. 1397–1401, Jul. 2015.
[28] M. Yang, Q. Wang, W. Wen, and Z. Lai, "Multi-feature joint dictionary learning for face recognition," in *Proc. 4th IAPR Asian Conf. Pattern Recognit. (ACPR)*, Nov. 2017, pp. 629–633.
[29] Q. Song, F. Xu, and Y. Q. Jin, "Sar image colorization: Converting single-polarization to fully polarimetric using deep neural networks," *IEEE Access*, vol. 6, no. 99, 2018.
[30] L. Deng, Y.-N. Yan, and C. Wang, "Improved POLSAR image classification by the use of multi-feature combination," *Remote Sens.*, vol. 7, no. 4, pp. 4157–4177, Apr. 2015.
[31] F. Zhang, J. Ni, Q. Yin, W. Li, Z. Li, Y. Liu, and W. Hong, "Nearest-regularized subspace classification for PolSAR imagery using polarimetric feature vector and spatial information," *Remote Sens.*, vol. 9, no. 11, p. 1114, Nov. 2017.
[32] H. Liu, S. Yang, S. Gou, S. Liu, and L. Jiao, "Terrain classification based on spatial multi-attribute graph using Polarimetric SAR data," *Appl. Soft Comput.*, vol. 68, pp. 24–38, Jul. 2018.
[33] P. Han, B. Han, X. Lu, R. Cong, and D. Sun, "Unsupervised classification for PolSAR images based on multi-level feature extraction," *Int. J. Remote Sens.*, vol. 41, no. 2, pp. 534–548, Jan. 2020.
[34] X. Liu, L. Jiao, X. Tang, Q. Sun, and D. Zhang, "Polarimetric convolutional network for PolSAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 5, pp. 3040–3054, May 2019.
[35] J. F. Shi, F. Liu, Y. H. Lin, and L. Liu, "Polarimetric SAR image classification based on deep learning and hierarchical semantic model," *Acta Autom. Sinica*, vol. 43, no. 2, pp. 215–226, 2017.
[36] H. Xing, Y. Meng, D. Hou, J. Song, and H. Xu, "Employing crowdsourced geographic information to classify land cover with spatial clustering and topic model," *Remote Sens.*, vol. 9, no. 6, p. 602, Jun. 2017.
[37] S. Deerwester, "Indexing by latent semantic analysis," *J. Assoc. Inf. Sci. Technol.*, vol. 41, no. 6, pp. 391–407, 1990.
[38] T. Hofmann, "Probabilistic latent semantic indexing," in *Proc. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 1999, pp. 50–57.
[39] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.
[40] J. S. Lee and E. Pottier, *Polarimetric Radar Imaging*. Boca Raton, FL, USA: CRC Press, 2015.
[41] L. J. Zhao, Y. L. Qin, and G. Gao, "Detection of built-up areas from high resolution SAR images using the GLCM textural analysis," *J. Remote Sens.*, vol. 13, no. 3, pp. 135–142, 2008.
[42] J. Shi, H. Jin, and Z. Xiao, "A novel hybrid edge detection method for polarimetric SAR images," *IEEE Access*, vol. 8, pp. 8974–8991, 2020.
[43] 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.
[44] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, Mar. 2006.
[45] J. Li, H. Zhang, L. Zhang, X. Huang, and L. Zhang, "Joint collaborative representation with multitask learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5923–5936, Sep. 2014.
[46] Q. Xu, Q. Chen, S. Yang, and X. Liu, "Superpixel-based classification using K distribution and spatial context for polarimetric SAR images," *Remote Sens.*, vol. 8, no. 8, p. 619, Jul. 2016.
[47] L. Yan and P. Li, "Building footprints extraction methods based on marker-controlled watershed segmentation and local radon transformation," *Inf. Technol. J.*, vol. 13, no. 11, pp. 1903–1907, Nov. 2014.
[48] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 8, pp. 790–799, Aug. 1995.
[49] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.
[50] C. Liu, J. Yang, J. Yin, and W. An, "Coastline detection in SAR images using a hierarchical level set segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 11, pp. 4908–4920, Nov. 2016.
[51] Y. Liu, C. Yu, M. Yu, and Y. He, "Manifold SLIC: A fast method to compute content-sensitive superpixels," in *Proc. Comput. Vis. Pattern Recognit.*, 2016, pp. 651–659.
[52] H. Zou, X. Qin, S. Zhou, and K. Ji, "A likelihood-based SLIC superpixel algorithm for SAR images using generalized gamma distribution," *Sensors*, vol. 16, no. 7, p. 1107, Jul. 2016.
[53] C. He, T. Zhuo, D. Ou, M. Liu, and M. Liao, "Nonlinear compressed sensing-based LDA topic model for polarimetric SAR image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 3, pp. 972–982, Mar. 2014.
[54] J. Zhang, X. Lin, and X. Ning, "SVM-based classification of segmented airborne LiDAR point clouds in urban areas," *Remote Sens.*, vol. 5, no. 8, pp. 3749–3775, Jul. 2013.
[55] B. Liu, H. Hu, H. Wang, K. Wang, X. Liu, and W. Yu, "Superpixel-based classification with an adaptive number of classes for polarimetric SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 907–924, Feb. 2013.

JUNFEI SHI received the B.S. degree in computer science and technology from Henan Normal University, Henan, China, in 2009, and the Ph.D. degree in computer science and technology from Xidian University, Xi'an, China, in 2016. She is currently a Faculty with the School of Computer Science and Engineering, Xi'an University of Technology, Xi'an. Her current research interests include polarimetric SAR image classification, semantic modeling, and deep learning.

HAIYAN JIN received the B.S. degree from Xi'an Shiyou University, Xi'an, China, in 1999, and the M.S. and Ph.D. degrees from Xidian University, Xi'an, in 2005 and 2007, respectively. She is currently with the Xi'an University of Technology, Xi'an. She has published more than 30 articles in journals and international conferences. Her interests include intelligent information processing, image processing, and pattern recognition.

XIAOHUA LI received the B.S. degree in electronic and information engineering from Shanxi University, Taiyuan, China, in 2009, and the M.S. and Ph.D. degrees in underwater acoustic engineering from Northwestern Polytechnical University (NPU), Xi'an, China, in 2012 and 2016, respectively. She was a Visiting Student with the Department of Electrical and Computer Engineering, University of Connecticut (UCONN), Storrs, USA, from September 2012 to September 2014. She is currently a Faculty with the School of Computer Science and Engineering, Xi'an University of Technology, Xi'an. Her research interest is statistical signal processing, currently focusing on multitarget tracking and information fusion.