

Unsupervised PolSAR Image Classification Using Discriminative Clustering

Haixia Bi, Jian Sun, *Member, IEEE*, and Zongben Xu

Abstract—This paper presents a novel unsupervised image classification method for polarimetric synthetic aperture radar (PolSAR) data. The proposed method is based on a discriminative clustering framework that explicitly relies on a discriminative supervised classification technique to perform unsupervised clustering. To implement this idea, we design an energy function for unsupervised PolSAR image classification by combining a supervised softmax regression model with a Markov random field smoothness constraint. In this model, both the pixelwise class labels and classifiers are taken as unknown variables to be optimized. Starting from the initialized class labels generated by Cloude–Pottier decomposition and K -Wishart distribution hypothesis, we iteratively optimize the classifiers and class labels by alternately minimizing the energy function with respect to them. Finally, the optimized class labels are taken as the classification result, and the classifiers for different classes are also derived as a side effect. We apply this approach to real PolSAR benchmark data. Extensive experiments justify that our approach can effectively classify the PolSAR image in an unsupervised way and produce higher accuracies than the compared state-of-the-art methods.

Index Terms—Discriminative clustering, Markov random field (MRF), polarimetric synthetic aperture radar (PolSAR) image classification, softmax regression (SR) model.

I. INTRODUCTION

POLARIMETRIC synthetic aperture radar (PolSAR) is an active microwave remote sensing radar system that can obtain scattering characteristics of diverse terrains by transmitting and receiving radar signals in different polarization ways. It has been one of the most advanced and important environmental monitoring techniques due to its all-time and all-weather observation character and abundant high-resolution target information. PolSAR technique has been extensively applied in civil and military fields in recent years.

The successful applications of PolSAR rely heavily on PolSAR image interpretation techniques. The target information in images should be revealed and extracted by an effective interpretation and analysis approach on PolSAR data. PolSAR image classification is one of the most important steps in image

interpretation process. The classification result can be further employed in edge extraction, target detection, or recognition for subsequent decision making or directly taken as output to users as the final result. Due to its importance, PolSAR image classification problem has always been a hot research topic in PolSAR information processing in recent years. However, owing to the complexity and diversity of natural terrains and the inherent characteristics of PolSAR technique such as speckle noise, PolSAR image classification is still a challenging task.

PolSAR image classification methods can be categorized from three perspectives. Considering whether a training set is utilized, they can be categorized into supervised classification methods [1]–[12] and unsupervised classification methods [13]–[16], [18]–[23]. Considering the utilized scattering characteristics, they can be categorized into methods based on PolSAR statistical distribution [1], [15], [27]–[30], [34], polarimetric scattering mechanisms [12], [13], [26], [31], and both of them [14], [19], [22]. Considering the employed classification algorithms, they fall into methods based on neural network [2]–[5], support vector machine (SVM) [6]–[8], [12], Markov random field (MRF) [9], [12], [21], spectral clustering [16], [18], and fuzzy logic [4], [32], [33].

A. Related Works on PolSAR Image Classification

Machine learning approach has been a dominant approach in PolSAR image classification task in recent years. In the following, we will review the machine learning-based PolSAR classification approach from the perspective of supervised and unsupervised methods.

1) *Supervised Methods*: In supervised classification methods, a training set composed of labeled image regions for different terrain classes should be set in advance and classifiers are then trained by virtue of different learning algorithms. After that, the learned classifiers are applied to the test set. The research on supervised methods mainly focus on the classification and feature extraction algorithms.

Kong *et al.* [1] first integrated a Bayesian classifier that is based on polarimetric statistical assumption in PolSAR image classification. Neural network technique was first utilized in this problem by Pottier and Saillard [2]. After that, different novel neural network techniques [3]–[5] were employed in PolSAR image classification. SVM is the most extensively utilized supervised learning algorithm in PolSAR image classification and achieved favorable results due to its elaborate optimization architecture and well generalization ability.

Manuscript received September 2, 2016; revised December 3, 2016 and January 19, 2017; accepted February 13, 2017. Date of publication March 17, 2017; date of current version May 19, 2017. This work was supported in part by the National Basic Research Program (973 Program) of China under Grant 2013CB329404 and in part by the National Natural Science Foundation of China under Grant 61472313, Grant 11622106, and Grant 91330204.

The authors are with the Institute for Information and System Sciences, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: bhwzq@163.com; jiansun@mail.xjtu.edu.cn; zbxu@mail.xjtu.edu.cn).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2017.2675906

Fukuda and Hirosawa [6] first introduced SVM into PolSAR image classification. Then Lardeux *et al.* [7] and He *et al.* [8] extended the utilization of SVM in this area. The success of SVM-based classification method relies on effective features extracted from PolSAR data. The method in [8] extracted high-dimensional features using wavelet, textons, and sparse coding techniques, and then SVM classifier was utilized upon them.

Tao *et al.* [10] combined 48 PolSAR data features and applied Tucker tensor decomposition and independent component analysis techniques to get an intrinsic feature set, and then SVM and KNN classifiers were employed to classify the PolSAR image. Fisher linear discriminant technique was employed by Chen *et al.* [11] to adjust the weight of different parameters in PolSAR image classification, which significantly improved the classification results. To make use of the anisotropic information embedded among different looks, Xu *et al.* [24] decomposed both polarimetric and anisotropic features simultaneously. Experimental results demonstrated the effectiveness of the proposed anisotropic entropies.

MRF technique was employed by Wu *et al.* [9] in PolSAR image classification, making use of the spatial relationship between adjacent pixels in classification process. It resulted in smooth classification label map by enforcing the connectivity of class labels within each class. Masjedi *et al.* [12] incorporated texture features and contextual information realized by MRF technique in PolSAR image classification, which significantly increased the accuracy.

With the booming development of deep learning in recent years, many advanced deep learning architectures have been proposed [35], [36], which provide a new way for PolSAR image classification. Deep belief network was applied to PolSAR image classification by Liu *et al.* [37]. Zhang *et al.* [38] employed stacked sparse autoencoder to learn deep spatial sparse features of PolSAR. Jiao and Liu [39] designed a PolSAR image classification method based on deep stacking network, which established a novel connection between PolSAR image interpretation and deep learning.

For supervised algorithms, if the chosen training pixels are enough to cover all terrain types and well express the terrain characteristics, they can achieve more precise and reliable classification results than unsupervised algorithms. This means that the performance of supervised algorithms largely relies on the manual selection of a good training set with ground-truth class labels. However, ground-truth class labels are not always available for PolSAR images, and the homogeneity and integrality of the selected training sets cannot be guaranteed through manual operation in many cases. All of these factors make supervised methods nonautomatic and uncontrollable.

2) *Unsupervised Methods*: Different from supervised classification methods, unsupervised methods aim at labeling all pixels without the aid of any manually labeled image pixels with ground-truth class labels. The key point of unsupervised methods is to find physically explicable and effective polarimetric characteristics that can discriminate different terrain classes. Unsupervised methods mostly start from an initialization, which is generated by executing an initial segmentation rule, and then perform classification according to different

strategies, which are derived from diverse PolSAR data statistical laws.

Most unsupervised PolSAR image classification methods are based on target decomposition and polarimetric data distribution theories. The most classical unsupervised method is the one proposed by Cloude and Pottier [13], which is based on Cloude–Pottier decomposition theory. In this method, Cloude–Pottier target decomposition is first performed on the complex coherency matrix T , and then the 2-D feature plane constructed by scattering entropy H and scattering angle $\bar{\alpha}$ is divided into eight subspaces, and hence all scattering mechanisms can be classified into eight basic zones. Initialized with the classified basic zones, an iterative maximum *a posteriori* classification based on Wishart distance is employed on the PolSAR data [14].

Along with the research on polarimetric scattering theory, it reveals that Wishart distribution, which is based on Gaussian model, is suitable only for describing evenly distributed regions. However, K -Wishart distribution based on non-Gaussian model is the generalization form of Wishart distribution. It well fits both Gaussian and non-Gaussian real PolSAR cases [15], [25], [34]. According to this result, many unsupervised classification algorithms based on non-Gaussian K -Wishart distribution were presented [15], [28].

There are extensive unsupervised image classification algorithms based either on different target decomposition theories, such as deorientation theory [23], Freeman decomposition [26], four-component decomposition [31], or on different PolSAR data statistical theories, such as \mathcal{G} distribution [27]. Many machine-learning-based methods have been utilized in unsupervised PolSAR classification in recent years. Spectral clustering technique was first applied to segmentation of SAR images [17], and then extended to PolSAR image classification by Ersahin *et al.* [16] and Lin *et al.* [18]. Wang *et al.* [19] proposed a novel unsupervised PolSAR image classification algorithm based on scattering power, entropy, and copolarized ratio. Liu *et al.* [20] designed an unsupervised classification method based on Wishart triplet Markov field (TMF). MRF was also applied in unsupervised PolSAR image classification by Doulgeris [21].

Unsupervised methods are fast and totally automatic. These methods do not require the ground-truth class labels in training, and the classification results are physically explicable. However, the classification accuracy (CA) relies heavily on the design of the unsupervised algorithm. It is still a challenging task to design an effective unsupervised algorithm for PolSAR image classification.

B. Related Works on Discriminative Clustering

Discriminative clustering is a framework with ingenious design, which relies explicitly on supervised classification techniques to perform unsupervised clustering. It was first introduced by Xu *et al.* [40], where an SVM model is employed to find maximum margin hyperplanes through data. To solve this combinatorial optimization problem over labels, a convex relaxation in terms of a semidefinite program is considered. Bach and Harchaoui [41] presented a discriminative clustering framework using linear discriminative cost

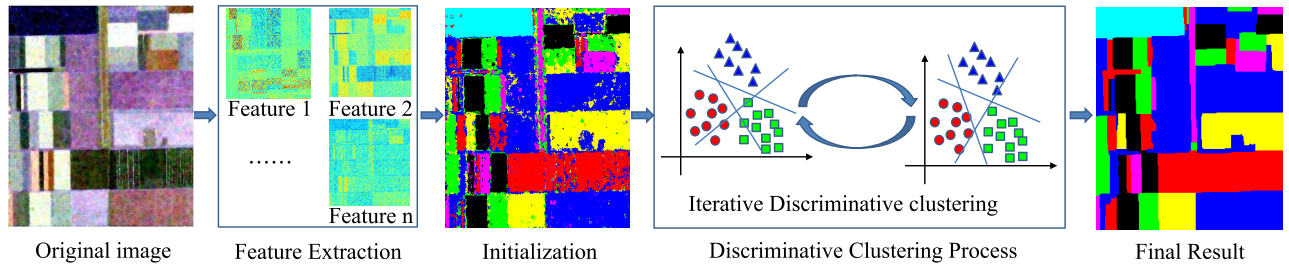


Fig. 1. Overview of our PolSAR image classification method based on discriminative clustering.

function based on a convex relaxation of a combinatorial problem. Joulin *et al.* [42], [43] and Sun and Ponce [44] introduced discriminative clustering to image cosegmentation problem that segments the common objects from a collection of images. The discriminative clustering method was also applied to investigate discriminative image part templates for different image classes, which are then taken as features for image classification [45].

The discriminative clustering approach has two advantages that motivate us to apply it to PolSAR image classification. First, discriminative clustering approach incorporates the discriminative power of supervised classification method into the automatic unsupervised classification problem. Reusing of the existing supervised learning tools has been proved to be advantageous in weakly supervised tasks [42]–[44] and unsupervised task [45]. Second, discriminative clustering is a general framework allowing for incorporating different discriminative loss functions or the other domain-specific constraints into a single loss function, which is suitable and flexible for different applications.

C. Our Contributions

In this paper, we introduce a novel discriminative clustering-based model to PolSAR image classification, which solves an unsupervised classification problem taking advantage of the discriminative learning power of supervised classification method. We design an energy function combining a cross-entropy loss term and a class label smoothness term. The first term is responsible for learning classifiers to discriminate different classes and the second term enforces label smoothness. The classifiers and class labels are both taken as variables to be optimized.

Given a rough initialization of labels based on the Cloude–Pottier decomposition theory and K -Wishart polarimetric statistical distribution, we propose to iteratively and alternately optimize the classifiers and class labels. The optimization of classifiers is implemented by solving a softmax regression (SR) problem, and the optimization of class labels boils down to a combinatorial optimization problem effectively solved by a belief propagation (BP) algorithm. By alternately optimizing the classifiers and class labels in an iterative manner, the class labels will be updated step by step until the termination criterion is met.

This paper first introduces the idea of discriminative clustering to PolSAR image classification. For this application, our designed energy model differs from the original discriminative

clustering model [41] in two points. First, instead of using a linear regression model as in [42] and [43], an SR model is employed to perform multiclass classification based on a cross-entropy loss. Second, to get contextually smooth classification result, spatial smoothness constraint is incorporated, which is realized by an MRF model in optimization. The experiments on benchmark data clearly show the advantages of our proposed method over other state-of-the-art unsupervised PolSAR classification methods.

II. MODEL AND ITS OPTIMIZATION

A. Overview of Our Method

We now present the basic pipeline of the proposed discriminative clustering-based PolSAR image classification method. Given a PolSAR image with N pixels, the input of each pixel is a coherency matrix T . A 58-D pixelwise feature vector x_i for each pixel i ($i \in \{1, \dots, N\}$) is extracted from coherency matrix T . The algorithm is designed to assign class label y_i ($y_i \in \{1, \dots, K\}$) to each pixel i , where there are K classes in total. We further denote $X = \{x_i\}$ and $Y = \{y_i\}$ in the following sections.

Taking the AIRSAR Felvoland PolSAR image for example, Fig. 1 illustrates an overview of the proposed PolSAR image classification approach. It comprises three steps.

- 1) *Feature Extraction*: Pixelwise feature vectors X are extracted from coherency matrix T through implementing mathematical transforms and target decompositions on it.
- 2) *Initialization*: An initialized class label map is produced as the input of the following step.
- 3) *Discriminative Clustering*: An iterative optimization algorithm is run to optimize the discriminative clustering-based model for refining the classification label.

The discriminative clustering steps and initialization will be presented from Sections II-B–II-D, and the feature extraction step will be discussed in Section III. We organize the discussions of feature extraction after discriminative clustering because the discriminative clustering method is the main contribution of this paper.

B. Discriminative Clustering-Based Model

The proposed approach aims at estimating pixelwise class labels using pixelwise features of the given PolSAR image. Based on the idea of discriminative clustering, we define

a novel loss function for the unsupervised PolSAR image classification, and the class labels Y and classifiers W can be derived by minimizing it. The loss function is defined as

$$E(Y, W|X) = E_c(Y, W|X) + E_s(Y, W|X) \quad (1)$$

where $E_c(Y, W|X)$ is named as discriminative clustering term and $E_s(Y, W|X)$ is the label smoothness term.

1) *Discriminative Clustering Term*: The discriminative clustering term $E_c(Y, W|X)$ is defined following the formula of SR model (i.e., multinomial logistic regression model). The SR model is the generalized form of logistic regression, which aims at resolving the multiclass classification problem. The hypothesis of SR model is defined as

$$h_W(x_i) = \begin{bmatrix} P(y_i = 1|x_i; W) \\ P(y_i = 2|x_i; W) \\ \vdots \\ P(y_i = K|x_i; W) \end{bmatrix} = \frac{1}{\sum_{j=1}^K e^{W_j^T x_i}} \begin{bmatrix} e^{W_1^T x_i} \\ e^{W_2^T x_i} \\ \vdots \\ e^{W_K^T x_i} \end{bmatrix} \quad (2)$$

where $W = [W_1, W_2, \dots, W_K]^T$. W_i is the classifier parameter vector of the i th category, which will be learned through the supervised training process. The output of SR model $h_W(x_i)$ is a normalized vector that exactly interprets the probability of pixel i belonging to different classes [46]. Due to this favorable character, its easy implementation, and fast computation speed, the SR model has been successfully applied in many fields [47]–[52].

The loss function E_c for a regularized SR model is defined as

$$E_c(Y, W|X) = L(Y, W|X) + R(W) \quad (3)$$

where $L(Y, W|X)$ is the softmax loss function and $R(W)$ is the regularization item. The softmax loss function is defined as the cross-entropy loss

$$L(Y, W|X) = -\sum_{i=1}^N \sum_{j=1}^K \frac{1}{N_j} 1\{y_i = j\} \log \frac{e^{W_j^T x_i}}{\sum_{l=1}^K e^{W_l^T x_i}} \quad (4)$$

where $1\{y_i = j\}$ is an indicator function that is equal to 1 when $\{y_i = j\}$ is true and is equal to 0 otherwise. Different from the classical softmax loss function, a parameter N_j is added, which is defined as the number of pixels whose labels are equal to j . The purpose of this parameter is to reduce the sample imbalance caused by discrepancy among the number of pixels belonging to different categories. The softmax loss function measures the agreement between the predictions of classifiers W and target class labels. Both classifiers and target class labels are taken as variables to be learned since we do not have ground-truth class labels for training. They will be alternately and iteratively optimized as shown in Section II-C.

To prevent possible overfitting caused by data noise and outliers, a regularization item $R(W)$ is incorporated after the

cross-entropy loss. L2-norm regularizer is employed in this paper, which is defined as

$$R(W) = \alpha_c \sum_{i=1}^K \sum_{j=1}^M W_{ij}^2 \quad (5)$$

where α_c is the regularization parameter and M is the number of vector dimension. With this convex regularization term (for any $\alpha_c > 0$), the loss function E_c is strictly convex and guaranteed to have a unique solution theoretically. In this paper, α_c is fixed as 5×10^{-5} .

2) *Label Smoothness Term*: The label smoothness term $E_s(Y, W|X)$ is defined to enforce that the neighboring pixels in an image should have similar class labels. We define it as

$$E_s(Y, W|X) = \alpha_s \sum_{i=1}^N \sum_{j \in \mathcal{N}(i)} S_{ij} \quad (6)$$

where α_s is the label smoothness factor and $\mathcal{N}(i)$ is the set of neighboring pixels of pixel i . S_{ij} is defined as

$$S_{ij} = |y_i - y_j| \exp\left(-\frac{\|v_i - v_j\|_2^2}{2\sigma}\right) \quad (7)$$

where v_i is a feature vector located at pixel i , which should be selected as the features significantly changing values across the edges in image. We take it as combination of Pauli matrix components in this paper. σ is the mean squared distance between features of two adjacent pixels i and j . Label smoothness term tends to make adjacent pixels have the same label and encourages the label boundaries to align with strong edges. For neighboring pixels i and j within flat regions, $\exp(-(\|v_i - v_j\|_2^2)/2\sigma)$ in (7) is large, then minimizing S_{ij} strongly enforces that labels y_i and y_j should take the same value. However, for neighboring pixels across strong edges, $\exp(-(\|v_i - v_j\|_2^2)/2\sigma)$ is smaller (or even near to zero) and thus the discrepancy between labels of neighboring pixels i and j is allowable in optimization.

3) *Discussion on the Final Loss Function*: Summarized from the above formulations, the final integrated loss function is defined as

$$E(Y, W|X) = -\sum_{i=1}^N \sum_{j=1}^K \frac{1}{N_j} 1\{y_i = j\} \log \frac{e^{W_j^T x_i}}{\sum_{l=1}^K e^{W_l^T x_i}} + \alpha_c \sum_{i=1}^K \sum_{j=1}^M W_{ij}^2 + \alpha_s \sum_{i=1}^N \sum_{j \in \mathcal{N}(i)} |y_i - y_j| \exp\left(-\frac{\|v_i - v_j\|_2^2}{2\sigma}\right) \quad (8)$$

where the classifier parameter matrix W and class labels Y are the variables to be optimized.

It should be noted that although the discriminative clustering loss function $E_c(Y, W|X)$ follows the formula of SR model, they are substantially different. In the softmax classification method, the labels in Y are given as constant, which are the ground-truth class labels in a training data set. However, in the discriminative clustering method, Y is one of the variable sets

to be optimized in an unsupervised setting. In this way, our proposed approach solves unsupervised classification problem by virtue of discriminative learning idea, which interprets the essence of the discriminative clustering method.

C. Optimization

The classifiers W for different classes and the class labels Y can be solved by minimizing the loss function defined in (8). This optimization problem can be decomposed into two subproblems, which could be solved alternately and iteratively.

The iterative optimization algorithm is discussed as follows. Initial class labels Y are generated in advance. In each iteration, classifiers W are first solved by optimizing (8) with fixed class labels Y (*subproblem 1*). Then class labels Y are derived by optimizing (8) with fixed classifiers W (*subproblem 2*). *subproblem 1* learns classifiers based on the current class labels, and *subproblem 2* updates the class labels using the updated classifiers considering the label smoothness constraint.

1) Subproblem 1: Solve W with fixed Y .

The purpose of *subproblem 1* is to learn classifiers W from the currently estimated class labels Y , which act as fixed variables in this subproblem. Therefore, in an iteration t , the currently estimated class labels are taken as the initialized class labels for iteration $t = 1$ and the class labels estimated in iteration $t - 1$ for iteration $t > 1$.

In this subproblem, since class labels Y are taken as constant, the loss function in (8) can be rewritten as

$$E(W|X, Y) = -\sum_{i=1}^N \sum_{j=1}^K \frac{1}{N_j} 1\{y_i = j\} \log \frac{e^{W_j^T x_i}}{\sum_{l=1}^K e^{W_l^T x_i}} + \alpha_c \sum_{i=1}^K \sum_{j=1}^M W_{ij}^2. \quad (9)$$

We optimize W by minimizing the above SR problem. It cannot be solved with a closed-form solution, and thus an iterative optimization algorithm is employed. Taking derivatives of $E(W|X, Y)$ with respect to W_j , the gradient can be computed as

$$\nabla_{W_j} E(W|X, Y) = -\sum_{i=1}^N \left[x_i \left(\frac{1}{N_j} 1\{y_i = j\} - P(y_i = j|x_i; W) \right) \right] + 2\alpha_c W_j. \quad (10)$$

It should be noted that ∇_{W_j} is itself a vector. Its k th element $((\partial E(W|X, Y))/(\partial W_{jk}))$ is the partial derivative of $E(W|X, Y)$ with respect to the k th element of W_j . With this formula for the derivative, we then use the L-BFGS [55] optimization algorithm¹ to minimize $E(W|X, Y)$ in (9).

With the optimized classifiers W , the hypothesis function defined in (2) exactly gives the probabilities of a pixel belonging to different classes.

2) Subproblem 2: Solve Y with fixed W .

Based on the updated classifiers W in *Subproblem 1*, *Subproblem 2* aims to estimate pixelwise class labels Y . Label smoothness constraint is taken into account in this process.

Since classifiers W are taken as constant, the loss function in (8) can be rewritten as

$$E(Y|X, W) = -\sum_{i=1}^N \sum_{j=1}^K \frac{1}{N_j} 1\{y_i = j\} \log \frac{e^{W_j^T x_i}}{\sum_{l=1}^K e^{W_l^T x_i}} + \alpha_s \sum_{i=1}^N \sum_{j \in \mathcal{N}(i)} |y_i - y_j| \exp \left(-\frac{\|v_i - v_j\|_2^2}{2\sigma} \right). \quad (11)$$

Class labels Y can be solved by minimizing (11) with fixed W . This label assignment problem is a combinatorial optimization problem in essence. The model in (11) can be regarded as an MRF model. An undirected graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ can be established over the image, where image node set \mathcal{V} corresponds to pixels and undirected edge set \mathcal{E} represents the neighborhood relationship between the pixels [53]. We then define the label variable y_i as a random variable on each node $v_i \in \mathcal{V}$, and then the class labels Y over the graph constitute an MRF, and $E(Y|X, W)$ is the energy function on it. The first term of the energy function reflects the cost of a pixel to be assigned to different categories. The larger the probability of a pixel belonging to a certain category, the more probable that the pixel to be assigned with the corresponding label. The second term constrains that labels of neighboring pixels should be close.

Labeling problem in an MRF has been demonstrated to be a nondeterministic polynomial time hard problem, and a graph cut [53], or BP [54] algorithm can be used to approximately achieve the optimal solution. The BP algorithm is employed in this paper. The advantage of BP algorithm is that labeling information can be quickly propagated across the image. The scheme of “top-bottom-left-right” message passing schedule makes the BP algorithm converge fast.

3) *Updating N_j ($j = 1 \dots K$):* The pixel number N_j of each category changes after updating class labels in subproblem 2. The pixel numbers for different categories should be updated accordingly and act as inputs to the next iteration.

D. Initialization

Given the pixelwise coherency matrix T for a PolSAR image, the initialization step is to generate initial estimates of the class labels, acting as the input to the subsequent iterative optimization procedures. The initialization step is based on the Cloude–Pottier decomposition theory and K -Wishart polarimetric statistical distribution.

Cloude–Pottier decomposition proposed by Cloude and Pottier [13] is employed first to generate initialized class labels of image pixels. The process is detailed in the following section.

Coherency matrix T is decomposed into a weighted sum of the matrices of three targets through eigenvalue analysis as

$$T = \lambda_1 e_1 e_1^* + \lambda_2 e_2 e_2^* + \lambda_3 e_3 e_3^* \quad (12)$$

¹<https://www.cs.ubc.ca/~schmidtm/Software/minFunc.html>.

where λ_i and e_i are eigenvalues and eigenvectors of T , respectively. The superscript $*$ denotes the complex conjugate transpose operation. Each eigenvector can be parameterized as

$$e_i = e^{i\phi_i} [\cos \alpha_i \quad \sin \alpha_i \cos \beta_i e^{i\gamma_i} \quad \sin \alpha_i \sin \beta_i e^{i\gamma_i}]' \quad (13)$$

where angle α_i represents the scattering mechanism of the i th decomposed target and angle β_i is twice the polarization orientation angle of the i th decomposed target. The superscript $'$ denotes the matrix transpose operation. Average scattering angle $\bar{\alpha}$ and entropy H are further defined as

$$\bar{\alpha} = P_1 \alpha_1 + P_2 \alpha_2 + P_3 \alpha_3 \quad (14)$$

$$H = \sum_{i=1}^3 -P_i \log_3 P_i \quad \text{where} \quad P_i = \frac{\lambda_i}{\sum_j \lambda_j}. \quad (15)$$

It can be observed from the above definition that angle $\bar{\alpha}$ represents the average scattering mechanism of random targets and entropy H reveals the randomness of scattering targets. Then the $H - \bar{\alpha}$ space is divided into eight subspaces by predefined horizontal and vertical straight lines [13], where each subspace represents a terrain type ranging from high entropy multiple scattering to low entropy surface scattering.

The initial division (i.e., classification) gives a rough interpretation of terrain types, but the division is arbitrary. A maximum likelihood classifier based on polarimetric statistical distribution is then applied to refine the classification map. The most commonly used polarimetric statistical distributions include complex Wishart distribution [14] and K -Wishart distribution [34]. K -Wishart distribution is employed in this paper, under which the probability density function for the coherency matrix T can be written as

$$P(T) = \frac{2|T|^{L-d} (L\alpha_L)^{\frac{1}{2}(\alpha_L+Ld)}}{R(L,d)|V|^L \Gamma(\alpha_L)} \frac{K_{\alpha_L-Ld}(2\sqrt{L\alpha_L \text{Tr}(V^{-1}T)})}{\text{Tr}(V^{-1}T)^{-(\alpha_L-Ld)/2}} \quad (16)$$

where $\text{Tr}(\cdot)$ and $|\cdot|$ denote the trace operation and the determinant, respectively. $K_\alpha(\cdot)$ is the modified Bessel function of the second kind, $\Gamma(\cdot)$ is the standard Gamma function, V is the mean coherency matrix, L represents the number of looks, d is the vector dimension, and α_L is the shape parameter that represents the data distribution of land cover.

According to Bayesian classification method, if the following condition is met, T will be assigned to class m :

$$d_m(T) < d_j(T) \quad \forall j \neq m \quad (17)$$

where $d_m(T)$ is the K -Wishart distance defined as

$$d_m(T) = L \ln |V_m| + \ln(\Gamma(\alpha_L)) - \frac{\alpha_L + Ld}{2} \ln(L\alpha_L) - \frac{\alpha_L - Ld}{2} \ln(\text{Tr}(V_m^{-1}T)) - \ln K_{\alpha_L-Ld}(2\sqrt{L\alpha_L \text{Tr}(V_m^{-1}T)}) \quad (18)$$

where V_m is the cluster center of mean coherency matrix of class m . The above K -Wishart distance is defined to reveal the similarity of a pixel to a certain terrain cluster. The initial cluster center is set as the cluster mean of the initial

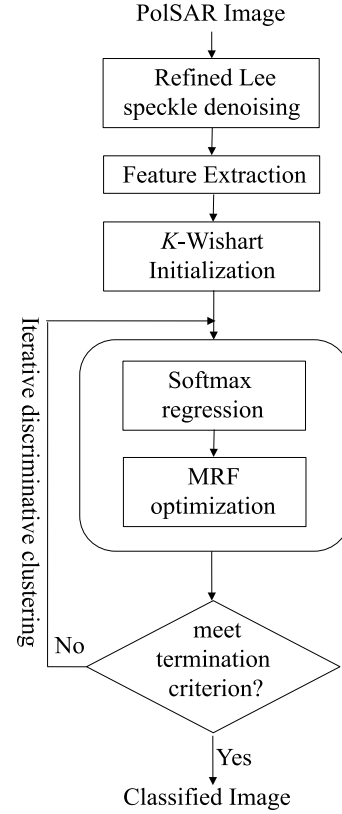


Fig. 2. Flowchart of the proposed algorithm.

division generated by Cloude–Pottier decomposition. Then the K -Wishart distances of each pixel to all cluster means are computed. A pixel will be assigned to a cluster with smallest K -Wishart distance. Then the new cluster centers are generated. We iterate the above process until convergence. If the desired class number is less than 8, a merging process will be implemented according to the distance between classes measured by the K -Wishart distances between cluster means. Two classes with smaller cluster distance will be merged with higher priority until the class number reaches the expected one. If the desired number of classes K is larger than 8, a finer initial $H - \bar{\alpha}$ subspace division should be carried out. Each of the original eight subspaces is divided to n^2 average parts by equivalently adding n horizontal and n vertical lines on it, where n is defined as the nearest ceiling integer of square root of $K/8$. Then K -Wishart clustering based on the finer division is iteratively executed until the desired class number is reached.

E. Summary of the Algorithm

Fig. 2 gives the flowchart of the proposed algorithm, which will be summarized as follows. The refined Lee *et al.*'s filter [56] is first employed to reduce the speckle noise and then feature vectors are extracted from the PolSAR data. After that, an initialization map is generated based on Cloude–Pottier decomposition and K -Wishart distribution hypothesis as discussed in Section II-D. Next, the classification result is solved through a discriminative clustering framework,

TABLE I
FEATURES EMPLOYED IN A DISCRIMINATIVE CLUSTERING METHOD

Polarimetric Feature category	Designation	Physical Description
Polarimetric matrices and mathematical transforms	$T_{ij}(i, j = 1, 2, 3, i \leq j)$	Elements of coherency matrix under horizontal and vertical linear polarization way (modulus and argument)
	$Lin45T_{ij}(i, j = 1, 2, 3, i \leq j)$	Elements of coherency matrix under $+45^\circ/-45^\circ$ linear polarization way (modulus and argument)
	$CirT_{ij}(i, j = 1, 2, 3, i \leq j)$	Elements of coherency matrix under left and right circular polarization way (modulus and argument)
	$\frac{I_{hv}}{I_{hh}}, \frac{I_{hv}}{I_{vv}}, \frac{I_{hh}}{I_{vv}}, \frac{I_{rr}}{I_{lr}}, \frac{I_{ll}}{I_{lr}}, \frac{I_{ll}}{I_{rr}}, \frac{I_{mn}}{I_{mn}}, \frac{I_{mn}}{I_{nn}}, \frac{I_{mn}}{I_{nn}}$	Intensities ratios under horizontal and vertical linear, $+45^\circ/-45^\circ$ linear, and left and right circular polarization ways
	SPAN	Polarimetric total power
Target decomposition features	Pauli matrix components	Pauli decomposition parameters
	P_s, P_d, P_v, α_L	Freeman decomposition parameters
	$\bar{\alpha}, H, A, \beta, (1-H)(1-A), (1-H)A, H(1-A), HA$	Cloude-Pottier decomposition parameters

as discussed in Sections II-B and II-C. The features used in this paper will be discussed in Section III. The parameter selection and termination condition of the proposed method will be discussed in Section IV-C.

III. FEATURES

To carry out discriminative learning and prediction, an exquisitely defined polarimetric feature vector is essential. In general, polarimetric features fall into two categories: original polarimetric matrices and their direct mathematical transforms and features derived from classical target decomposition methods. The original polarimetric matrices generally include scattering matrix S , coherency matrix T , and covariance matrix C . The target decomposition methods employed here include Pauli decomposition, Cloude–Pottier decomposition, and Freeman decomposition. To clearly interpret the choice of the features, the derivations of the matrices S and T will be detailed as follows.

For coherent scattering targets, based on the horizontal and vertical polarimetric basis, scattering matrix S can be expressed as

$$S = \begin{bmatrix} S_{hh} & S_{hv} \\ S_{vh} & S_{vv} \end{bmatrix} \quad (19)$$

where S_{pq} denotes the scattering matrix element corresponding to the p - q polarization of a receiving transmitting wave (p and q refer to horizontal h or vertical v for linear polarization). Under reciprocity condition, $S_{hv} = S_{vh}$. However, for distributed targets, the backscatter waves are not fully polarized and matrix S cannot well represent the scattering characteristics of complex targets. One possible representation is the second-order polarimetric representation T , which is based on the Pauli basis of scattering matrix S [57]

$$K_p = \frac{1}{\sqrt{2}} \begin{bmatrix} S_{hh} + S_{vv} \\ S_{hh} - S_{vv} \\ 2S_{hv} \end{bmatrix} \quad T = \langle K_p K_p^* \rangle \quad (20)$$

where $\langle \cdot \rangle$ stands for the spatial averaging during the multilook processing.

In total, 58 features are extracted from PolSAR data in this paper. All of them are listed in Table I and discussed as follows.

- 1) Intensities and phases of coherence matrices under three different polarimetric ways: a) horizontal and vertical linear polarization; b) $+45^\circ/-45^\circ$ linear polarization; and c) left and right circular polarization.
- 2) The ratios between different intensity channels in horizontal and vertical linear, $+45^\circ/-45^\circ$ linear, and left and right circular polarization. I_{pq} stands for the intensity of p - q polarization, where p and q belong to the subscript set $\{h, v, l, r, m, n\}$ representing horizontal linear, vertical linear, left circular, right circular, $+45^\circ$ linear, and -45° linear polarization, respectively.
- 3) The SPAN is defined as $SPAN = I_{hh} + 2I_{hv} + I_{vv}$. SPAN represents the total polarimetric power, which is an important PolSAR discriminator.
- 4) Nine Pauli decompositions parameters based on Pauli basis of horizontal and vertical linear, $+45^\circ/-45^\circ$ linear, and left and right circular polarization ways. For each polarization way, there are three Pauli decomposition channels, which represent isotropic odd, even, and $\pi/4$ even scattering power, respectively.
- 5) Parameters P_s, P_d, P_v , and α_L are extracted from Freeman decomposition, where P_s, P_d , and P_v represent powers of surface scatter, double bounce scatter, and volume scatter, respectively, and α_L is the shape parameter. The smaller the value, the stronger the non-Gaussianity.
- 6) Eight parameters are extracted from Cloude–Pottier decomposition, where H represents the randomness of targets, $\bar{\alpha}$ reveals the average scattering mechanism of targets, A is the anisotropy, which reflects the ratio of the two weaker scattering mechanisms, and β shows the target orientation. Other four parameters are mathematical combinations of H and A .

In the above features, features in (1)–(3) belong to the first category, i.e., polarimetric matrices and their mathematical transforms, and features in (4)–(6) fall into the second category, i.e., target decomposition features.

IV. EXPERIMENTS

A. Experimental Data and Settings

Experiments are carried out on three real PolSAR images, which are shown in Fig. 3. The first data are for Flevoland

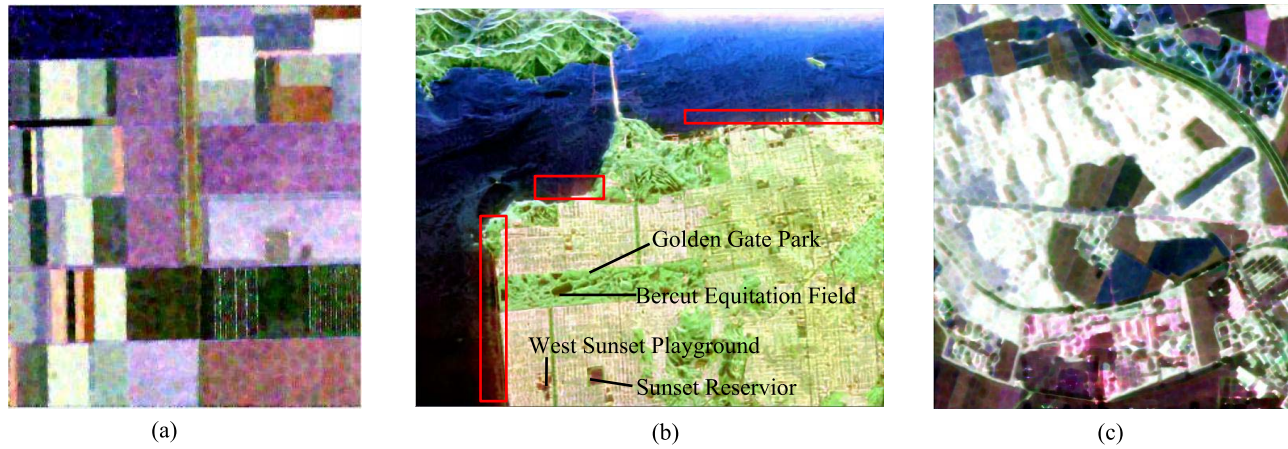


Fig. 3. Original images of (a) Flevoland, (b) San Francisco, and (c) Oberpfaffenhofen areas.

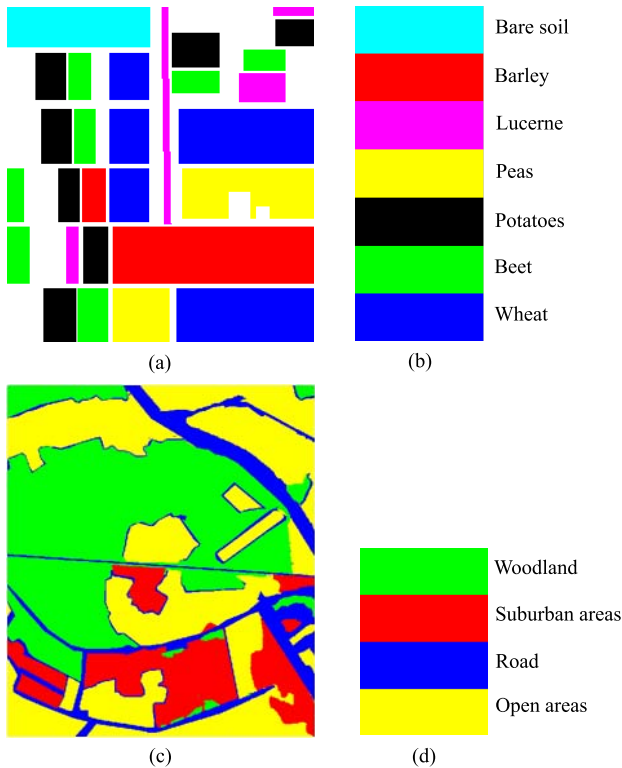


Fig. 4. Ground-truth class labels and color codes on Flevoland area data and Oberpfaffenhofen area data. (a) Ground-truth class labels of Flevoland area data. (b) Color code of Flevoland area data. (c) Ground-truth class labels of Oberpfaffenhofen area data. (d) Color code of Oberpfaffenhofen area data.

area in Netherlands, which was acquired by NASA/JPL AIRSAR on August 16, 1989. The size of the image we used is 300×270 , as shown in Fig. 3(a). The ground-truth class labels and the corresponding color codes are shown in Fig. 4(a) and (b), respectively. As observed from the ground-truth class labels of the Flevoland area data, there are seven classes in the PolSAR image including *Bare soil*, *Barley*, *Lucerne*, *Peas*, *Potatoes*, *Beet*, and *Wheat*. The second data are four-look NASA/JPL AIRSAR L-band data of the San Francisco area. The size of the image is 900×1024 , as shown in Fig. 3(b). In these data, the terrain types include vegetation,

sea, suburban areas, and bare land. The last one is an E-SAR data of Oberpfaffenhofen area in Germany, which is provided by the German Aerospace Center. The size of this image is 800×700 , as shown in Fig. 3(c). The ground-truth class labels and color codes are shown in Fig. 4(c) and (d). There are four classes in the image: woodland, suburban areas, road, and open areas.

We first justify the validity of key components in our proposed method in Section IV-B. Then parameter analysis will be presented in Section IV-C. Finally, we demonstrate the effectiveness of the proposed method by comparison on three real PolSAR images with the other state-of-the-art algorithms: Wishart MRF algorithm [9], $H/\bar{\alpha}$ -Wishart algorithm [14], power entropy and copolarized ratio algorithm [19], Wishart TMF method [20], and Freeman Wishart algorithm [22]. Quantitative comparisons of Flevoland area data and Oberpfaffenhofen area data will be conducted, wherein the classification accuracies will be analyzed.

B. Validity Analysis

SR and MRF optimization (MO) are two subproblems in the proposed method. Comparisons will be made between the initialized class labels (in Section II-D), class labels after one round of SR, and class labels after one round of SR + MO.

Fig. 5(a), (d), and (g) shows initialized class labels of Flevoland area data, Oberpfaffenhofen area data, and San Francisco area data, respectively. Fig. 5(b), (e), and (h) shows the estimated class labels after one round of SR on three PolSAR data. Fig. 5(c), (f), and (i) shows class labels after one round of SR + MO. In Fig. 5(c), (f), and (i), α_s defining the weight of label smoothness term takes value of 1.

Comparing with the initialized class labels shown in the first column of Fig. 5, it can be seen from the second column that after performing the SR learning and prediction, many misclassified pixels are assigned with correct class labels, which exactly illustrates the discriminative character of the proposed method. Compared with the class labels of the first two columns, the class labels in the third column after MO are continuous with clearer edges and less isolate pixels, which shows the smoothing effect of MO.

TABLE II
CA VALUES (%) OF FLEVOLAND AREA DATA FOR THE PROPOSED METHOD

Object/Class	Bare soil	Barley	Lucerne	Peas	Potatoes	Beet	Wheat	<i>overall</i> CA value
Initialization	98.88	84.87	86.78	95.46	97.18	89.06	97.37	92.54
SR	100	93.85	94.84	95.28	97.18	89.18	96.92	95.45
SR+MO	100	99.76	95.07	98.09	98.38	99.91	99.13	98.99

TABLE III
CA VALUES (%) OF OBERPFAFFENHOFEN AREA DATA FOR THE PROPOSED METHOD

Object/Class	Woodland	Road	Suburban areas	Open areas	<i>overall</i> CA value
Initialization	38.17	64.44	46.65	71.78	55.69
SR	77.65	57.56	50.35	72.98	69.85
SR+MO	91.32	65.95	69.52	77.15	80.01

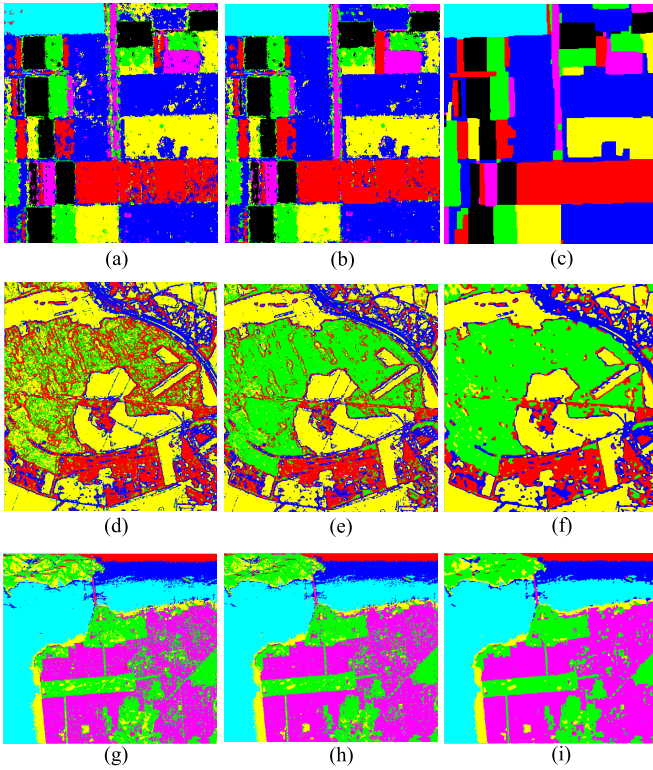


Fig. 5. Class labels on three real PolSAR images for the proposed method. (a) Initialized class labels of Flevoland area data. (b) Class labels of Flevoland area data after one round of SR. (c) Class labels of Flevoland area data after one round of SR + MO. (d) Initialized class labels of Oberpfaffenhofen area data. (e) Class labels of Oberpfaffenhofen area data after one round of SR. (f) Class labels of Oberpfaffenhofen area data after one round of SR + MO. (g) Initialized class labels of San Francisco area data. (h) Class labels of San Francisco area data after one round of SR. (i) Class labels of San Francisco area data after one round of SR + MO.

We define the *CA* of a class as the ratio of the number of pixels correctly classified for the class to the total number of pixels in this class. The *overall CA* is defined as the ratio of the number of correctly classified pixels in the whole image to the total number of pixels in the image. Tables II and III list the *CA* values of the initialized class labels, the class labels after one round of SR, and class labels after one round of SR + MO for Flevoland area data and Oberpfaffenhofen area data, respectively. For Flevoland area data, the *overall CA*

values of SR and SR + MO increase by 2.91% and 6.45% compared with the initialized class labels, while the increases of *overall CA* values for Oberpfaffenhofen area data are 14.16% and 24.32%, respectively, which demonstrates the validity of discriminative clustering using SR and label smoothing using MO.

C. Parameter Analysis

For the proposed algorithm, two parameters: α_s and the iteration number significantly influences the final result. α_s determines the strength of label smoothness constraint. A larger α_s implies that the label smoothness term plays more importance in the optimization and therefore tends to produce smoother class labels. The discriminative clustering algorithm can improve the *CA*; however, the number of iterations for the optimization influences the estimation accuracy. Also, parameters α_s and the iteration number are mutually influenced. Therefore, we next comprehensively analyze the effects of these parameters.

In the following, five typical values of α_s (0.1, 0.2, 0.5, 1, and 2) are selected for analysis. The classification results taking the five selected values after performing SR + MO for one to five iterations will be analyzed. Appropriate ranges of α_s and iteration number will be suggested after the analysis.

Fig. 6 shows the *overall CA* values under different α_s s and iteration numbers for Oberpfaffenhofen area data and Flevoland area data, respectively. The *overall CA* values of iteration number 0 correspond to the initialized label maps.

It can be obviously observed from the two plots in Fig. 6 that the *overall CA* values increase significantly in the first iteration. Stable and high *overall CA* values can be obtained when α_s takes larger values (e.g., 1) and the iteration number takes value of 2 or 3, which demonstrates the effectiveness of the proposed method. However, with the increase in iteration number, the *overall CA* values improve slightly or even marginally decreased. This is because for small α_s values (e.g., 0.1), the relatively weak label smoothness effect is unable to rectify wrong classified pixels due to speckle effect, while for big α_s values (e.g., 2), too many rounds of strong MO will make the classes with small and thin areas gradually disappear due to successive label smoothing. To demonstrate this point, we take Oberpfaffenhofen area data and San Francisco area

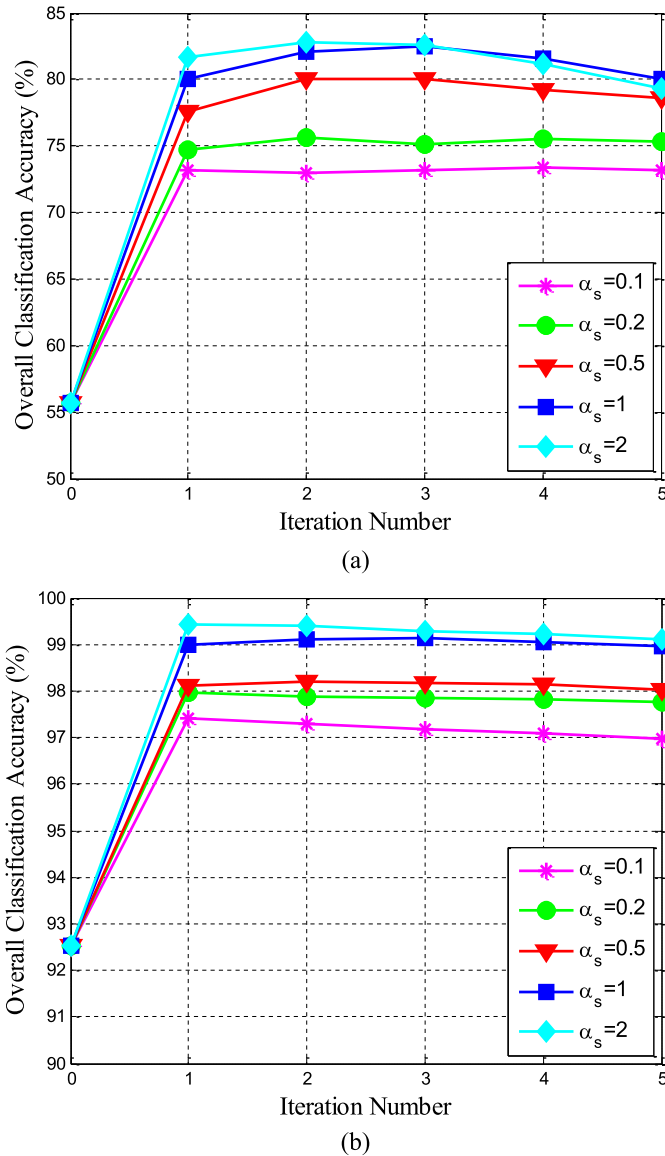


Fig. 6. Influence of α_s and iteration number on overall CA value. (a) Overall CA values of Oberpfaffenhofen area data under different α_s and iterations. (b) Overall CA values of Flevoland area data under different α_s and iterations.

data for illustration. Fig. 7 shows the class labels when α_s takes value of 2 under different iteration numbers. It can be seen from Fig. 7(a)–(c) that when α_s takes value of 2, with the increase in iteration number, the road class pixels are gradually eroded by the neighboring pixels and many narrow roads and edges (marked in black rectangles) even disappear in the fifth iteration. This also happens in San Francisco area data. With the increase of iteration number, many tiny details (e.g., marked in blue rectangle) are eroded by neighboring pixels. Table IV gives the CA values of Oberpfaffenhofen area data under different α_s s and iteration numbers. It can be observed from Table IV that when α_s takes value of 2, the CA value of road category evidently gets lower with the increase in iteration number. The CA value of road category of the fifth iteration decreases 39.76% compared with the first iteration.

Summarized from the above analysis, taking both CA values and stability between iterations into account, α_s ranging

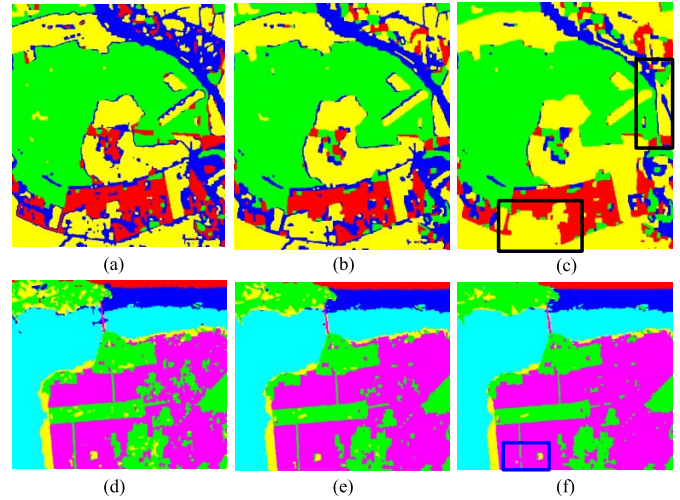


Fig. 7. Class labels under different α_s and iterations for Oberpfaffenhofen area data and San Francisco area data. (a)–(c) Class labels of Oberpfaffenhofen area data. $\alpha_s = 2$ and the numbers of iterations are 1, 3, and 5, respectively. (d)–(f) Class labels of San Francisco area data. $\alpha_s = 2$ and the numbers of iterations are 1, 3, and 5, respectively.

between 1 and 2 and the iteration number with a value of 2 or 3 are favorable parameter sets. In this paper, we set α_s to 1 and the iteration number to 3 for three real PolSAR images. However, in applications, α_s can be set based on personal preference. If results with stronger spatial connectivity are preferred, α_s should be adjusted to a larger value. However, if finer terrain details are required, α_s can be set to a smaller value.

D. Comparison With Other State-of-the-Art Algorithms

1) *Experimental Results on Flevoland Area Data:* Class labels of the Flevoland area data using the proposed method and five other algorithms are shown in Fig. 8. Table V gives the CA values of these six methods.

It can be observed from Fig. 8(a) that for the $H/\bar{\alpha}$ -Wishart classifier, many pixels are misclassified. For example, many pixels of the *Wheat* class are assigned with the *Lucerne* label [see the region marked by the white rectangle in Fig. 8(a)] and many pixels of *Potatoes* are assigned as *Barley* and *Beet* [see the region marked by the black rectangle in Fig. 8(a)]. Also, the boundaries between different classes are indistinct and unclear. The Wishart MRF algorithm employs MO in the classification process, which takes spatial relation into account. Compared with Fig. 8(a), Fig. 8(b) shows less misclassification pixels and a better pixel connectivity. The edges are clearer than the result in Fig. 8(a). Classification confusion is observed for both Freeman Wishart classifier and power entropy and copolarized ratio classifier. It can be seen from Fig. 8(c) and (d) that the bottom-right regions of the two images are severely contaminated. The pixels that should be classified as *Beet* are misclassified as *Lucerne*. Also, misclassification speckles are scattered in the two images. The Wishart TMF classification map shown in Fig. 8(e) has a better visual effect than Fig. 8(a)–(d). However, isolated pixel blocks still can be observed, especially in the bottom-right part of the image.

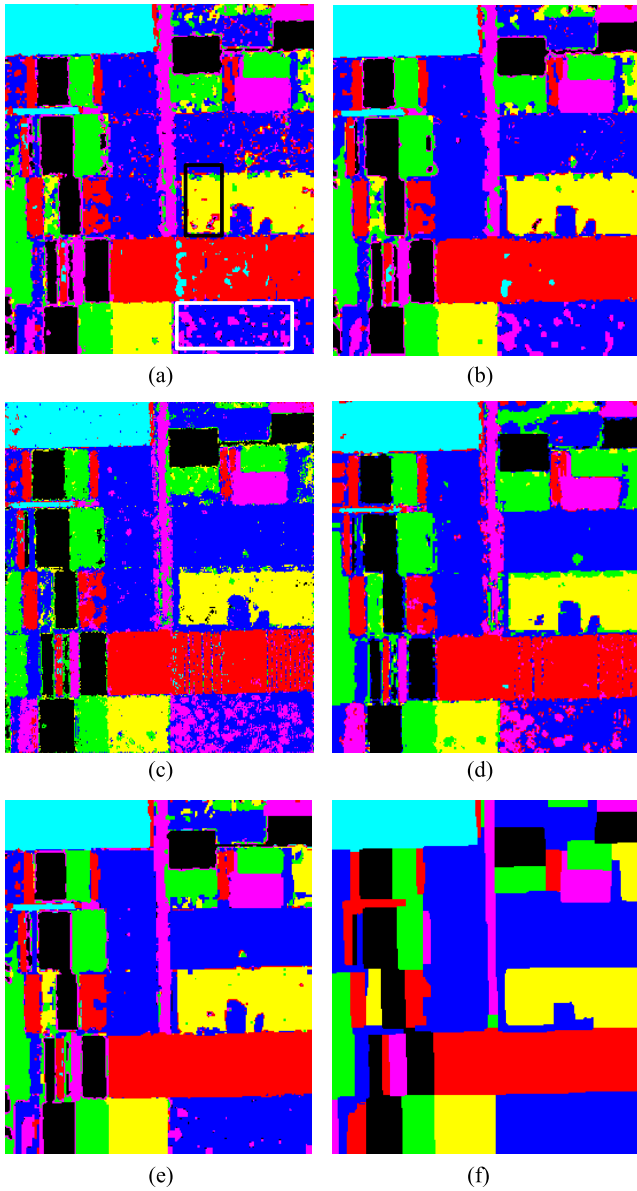


Fig. 8. Class labels of Flevland area data with different methods. (a) $H/\bar{\alpha}$ -Wishart classifier [14]. (b) Wishart MRF classifier [9]. (c) Freeman Wishart classifier [22]. (d) Power entropy and copolarized ratio classifier [19]. (e) Wishart TMF classifier [20]. (f) Proposed method.

The proposed method uses supervised SR and numerous features to carry out unsupervised classification, and many misclassification artifacts in the initialized label map are rectified during the iterative optimization. The introduced smoothness constraint using the MRF model leads to a smoother classification map. For the above reasons, compared with the other five methods, the proposed method shows a better performance with a lower misclassification ratio and a better visual effect. The *overall CA* values of the proposed method are 13.46%, 7.23%, 8.89%, 10.75%, and 3.19% higher than those of the compared methods, respectively.

2) *Experimental Results on Oberpfaffenhofen Area Data:* Class labels on Oberpfaffenhofen area data employing the proposed method and three other methods are shown in Fig. 9.

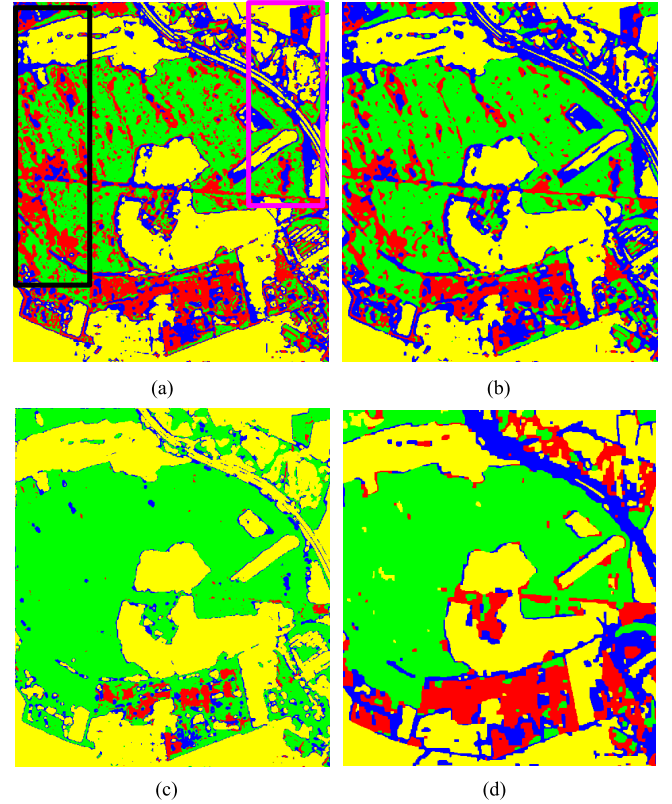


Fig. 9. Class labels of Oberpfaffenhofen area data with different methods. (a) $H/\bar{\alpha}$ -Wishart classifier [14]. (b) Wishart MRF classifier [9]. (c) Freeman Wishart classifier [22]. (d) Proposed method.

The *CA* values of these methods are listed in Table VI. The power entropy and copolarized ratio classifier and the Wishart TMF classifier are not compared because [19] and [20] did not apply them on these data in the original papers.

Fig. 9(a) shows the class labels of the $H/\bar{\alpha}$ -Wishart algorithm. It can be observed from this image that a great deal of pixels are misclassified. For example, in the region marked by a black rectangle, many woodland pixels are misclassified as suburban areas and road. Inside the region that is marked by rose red rectangle, a lot of open area pixels are assigned with road labels. Also, the roads at the bottom of the image are invisible in Fig. 9(a). Besides, the whole image is full of speckle points, which leads to a poor visual effect. The Wishart MRF algorithm improves the speckle problem of Fig. 9(a) greatly, which could be seen clearly from Fig. 9(b), and the road at the top-right part of the image gets smoother and clearer. However, misclassification problem is still severe. The result of Freeman Wishart algorithm is shown in Fig. 9(c). Because the scattering mechanism of road and open areas are similar, these two classes are confused to a great extent, and the roads are barely visible in the whole image. For the suburban areas, only the areas that have the strongest double bounce scattering power are classified as suburban area category and the others are misclassified as woodland class.

The class labels of the proposed method are given in Fig. 9(d). Compared with the results of the other three methods, the proposed method generates less misclassification

TABLE IV
CA VALUES (%) OF OBERPFAFFENHOFEN AREA DATA UNDER DIFFERENT α_s s AND ITERATION NUMBERS

Iteration	α_s /Class	Woodland	Road	Suburban areas	Open areas	overall CA value
1	$\alpha_s=0.1$	79.76	60.28	61.62	74.98	73.14
	$\alpha_s=0.2$	82.34	61.66	63.52	75.43	75.11
	$\alpha_s=0.5$	87.38	63.88	66.82	76.28	77.60
	$\alpha_s=1.0$	91.32	65.95	69.52	77.15	80.48
	$\alpha_s=2.0$	94.02	63.40	72.71	78.75	81.70
2	$\alpha_s=0.1$	77.01	56.40	61.60	78.76	73.70
	$\alpha_s=0.2$	81.65	60.30	62.44	79.44	75.60
	$\alpha_s=0.5$	89.65	63.54	65.40	81.32	80.08
	$\alpha_s=1.0$	92.23	64.70	69.82	82.18	82.09
	$\alpha_s=2.0$	96.77	52.32	58.19	88.84	83.14
3	$\alpha_s=0.1$	76.40	56.75	60.38	80.01	73.13
	$\alpha_s=0.2$	79.70	59.53	60.90	80.87	75.11
	$\alpha_s=0.5$	90.06	63.33	58.36	83.32	80.02
	$\alpha_s=1.0$	93.65	62.18	60.78	84.41	82.25
	$\alpha_s=2.0$	96.17	46.42	63.07	88.36	82.64
4	$\alpha_s=0.1$	76.78	57.45	57.83	80.97	73.39
	$\alpha_s=0.2$	80.11	59.65	58.83	82.23	75.51
	$\alpha_s=0.5$	90.72	65.21	54.05	83.20	79.24
	$\alpha_s=1.0$	93.12	60.70	55.54	84.51	80.62
	$\alpha_s=2.0$	95.31	35.58	65.90	88.03	81.19
5	$\alpha_s=0.1$	76.25	57.32	56.68	82.82	73.21
	$\alpha_s=0.2$	80.19	60.76	56.47	82.21	75.36
	$\alpha_s=0.5$	91.68	66.64	52.73	82.38	78.64
	$\alpha_s=1.0$	92.23	58.25	53.75	85.01	79.52
	$\alpha_s=2.0$	91.98	23.64	71.52	88.29	79.27

TABLE V
CA VALUES (%) OF FLEVOLAND AREA DATA WITH DIFFERENT METHODS

Method/Class	Bare soil	Barley	Lucerne	Peas	Potatoes	Beet	Wheat	overall CA value
$H/\bar{\alpha}$ -Wishart classifier	99.94	86.60	84.63	88.50	81.44	84.00	82.48	85.59
Wishart MRF classifier	100.00	86.23	93.27	94.24	85.55	89.54	92.55	91.82
Freeman Wishart classifier	98.26	87.97	83.38	93.22	98.13	92.20	84.53	90.16
Power Entropy and Copolarized Ratio classifier	99.44	89.85	75.22	85.92	85.79	91.01	87.89	88.30
Wishart TMF classifier	99.81	97.60	86.10	98.27	90.00	95.86	97.33	95.86
The proposed method	100.00	100.00	95.07	98.21	98.57	98.63	99.93	99.05

pixels and a smoother label map, due to the discriminative clustering and MO. The *overall* CA values of the proposed method are 17.82%, 14.99%, and 8.90% higher than those of other methods, respectively.

3) *Experimental Results on San Francisco Area Data*: Class labels of the San Francisco area data using the proposed method and three other algorithms are shown in Fig. 10. Though ground-truth class labels are not available, we can still derive some conclusions by observing the classification results compared with the original image in Fig. 3(b). The areas colored with white, light blue, and dark blue are sea areas. The color variance of the sea area is due to the incident angle change of PolSAR during the observation. The green colored regions are covered with vegetation. The suburban areas are colored as yellow. The dark red areas are composed of beach (marked inside the red rectangles), and some grounds such as Bercut Equitation Field, West Sunset Playground, and Sunset Reservoir, which have been indicated in the image.

It can be observed from Fig. 10(a)–(c) that there are two big common problems for Wishart, Wishart MRF, and Freeman Wishart algorithms. The first one is the misclassification of suburban area as vegetation category, which is apparently visible in the bottom-right part of the image. The second one is the confusion of vegetation, suburban area, and ground for the mountains in the top-left corner of the image. These two problems also exist in the initialized label map of the

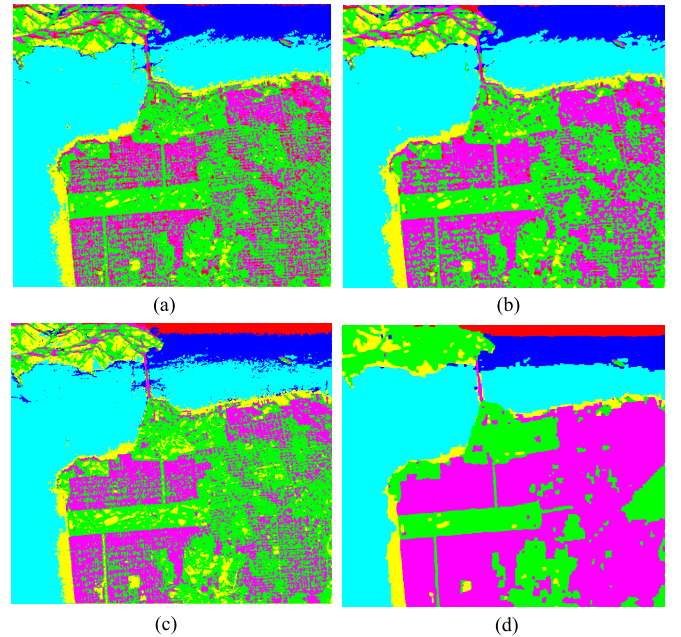


Fig. 10. Class labels of San Francisco area data with different methods. (a) $H/\bar{\alpha}$ -Wishart classifier [14]. (b) Wishart MRF classifier [9]. (c) Freeman Wishart classifier [22]. (d) Proposed method.

proposed method to some extent; however, after iterations of discriminative clustering and MO, the misclassification phenomenon is greatly reduced. Compared with the other

TABLE VI
CA VALUES (%) OF OBERPFAFFENHOFEN AREA DATA WITH DIFFERENT METHODS

Method/Class	Woodland	Road	Suburban areas	Open areas	overall CA value
$H/\bar{\alpha}$ -Wishart classifier	65.21	41.80	51.51	76.01	64.43
Wishar MRF classifier	74.43	55.52	42.28	72.93	67.26
Freeman Wishart classifier	97.71	18.13	16.80	87.85	73.35
The proposed method	93.65	64.70	60.78	84.41	82.25

three algorithms, the proposed method yields a better result, with better agreement with the original image, better pixel connectivity, and more distinct boundaries.

V. CONCLUSION

A novel unsupervised classification method based on discriminative clustering is presented in this paper. The contribution of our work is that we take advantage of supervised learning technique to perform unsupervised clustering, and contextual information is incorporated into the energy model. Initialized with a classification map based on Cloude–Pottier decomposition theory and K -Wishart distribution hypothesis, the classifiers and labels are updated by alternately solving an SR problem and an MRF combinatorial optimization problem. The optimization process is iteratively executed until the termination criterion is met. We apply the proposed method on three real PolSAR data. Comparing with several other state-of-the-art algorithms, higher accuracies and better connectivity are achieved by the proposed method.

Feature extraction and optimization is an important research topic in PolSAR image classification, because features are essential to the CA, computational complexity, and memory cost. We do not focus on feature design in this paper, but inspired by the excellent feature learning capacity of deep learning models, we plan to devise or learn more effective polarimetric features using deep learning techniques in discriminative clustering framework in the future work.

ACKNOWLEDGMENT

The authors would like to thank the National Aeronautics and Space Administration's Jet Propulsion Laboratory and the German Aerospace Center for providing the PolSAR data.

REFERENCES

- [1] J. A. Kong, A. A. Swartz, H. A. Yueh, L. M. Novak, and R. T. Shin, "Identification of terrain cover using the optimum polarimetric classifier," *J. Electromagn. Waves Appl.*, vol. 2, no. 2, pp. 171–194, 1988.
- [2] E. Pottier and J. Saillard, "On radar polarization target decomposition theorems with application to target classification by using network method," in *Proc. ICAP*, vol. 1, Apr. 1991, pp. 265–268.
- [3] K. S. Chen, W. P. Huang, D. H. Tsay, and F. Amar, "Classification of multifrequency polarimetric SAR image using a dynamic learning neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 3, pp. 814–820, Mar. 1996.
- [4] Y. C. Tzeng and K. S. Chen, "A fuzzy neural network for SAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 1, pp. 301–307, Jan. 1998.
- [5] O. Antropov, R. Rauste, H. Astola, J. Praks, T. Hame, and M. T. Hallikainen, "Land cover and soil type mapping from spaceborne PolSAR data at L-Band with probabilistic neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5256–5270, Sep. 1998.
- [6] S. Fukuda and H. Hirokawa, "Support vector machine classification of land cover: Application to polarimetric SAR data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, vol. 1, Jul. 2001, pp. 187–189.
- [7] C. Lardeux et al., "Support vector machine for multifrequency SAR polarimetric data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 12, pp. 4143–4152, Dec. 2009.
- [8] C. He, S. Li, Z. X. Liao, and M. S. Liao, "Texture classification of PolSAR data based on sparse coding of wavelet polarization textures," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 8, pp. 4576–4590, Aug. 2013.
- [9] Y. Wu, K. Ji, W. Yu, and Y. Su, "Region-based classification of polarimetric SAR images using wishart MRF," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 668–672, Oct. 2008.
- [10] M. Tao, F. Zhou, Y. Liu, and Z. Zhang, "Tensorial independent component analysis-based feature extraction for polarimetric SAR data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2481–2495, May 2015.
- [11] B. Chen, S. Wang, L. Jiao, R. Stolkin, and H. Liu, "A three-component fisher-based feature weighting method for supervised PolSAR image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 4, pp. 731–735, Apr. 2015.
- [12] A. Masjedi, M. J. V. Zoej, and Y. Maghsoudi, "Classification of polarimetric SAR images based on modeling contextual information and using texture features," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 2, pp. 932–943, Feb. 2015.
- [13] S. R. Cloude and E. Pottier, "An entropy based classification scheme for land application of polarimetric SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 1, pp. 68–78, Jan. 1997.
- [14] 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, Sep. 1999.
- [15] A. P. Doulgeris, S. N. Anfinson, and T. Eltoft, "Classification with a non-Gaussian model for PolSAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 10, pp. 2999–3009, Oct. 2008.
- [16] 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.
- [17] X. Zhang, L. Jiao, F. Liu, L. Bo, and M. Gong, "Spectral clustering ensemble applied to SAR image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2126–2136, Jul. 2008.
- [18] L. Q. Lin, H. Song, P. P. Huang, W. Yang, and X. Xu, "Unsupervised classification of PolSAR data using large scale spectral clustering," in *Proc. IEEE IGARSS*, Jul. 2014, pp. 2814–2817.
- [19] S. Wang, K. Liu, J. Pei, M. Gong, and Y. Liu, "Unsupervised classification of fully polarimetric SAR images based on scattering power entropy and copolarized ratio," *IEEE Trans. Geosci. Remote Sens. Lett.*, vol. 10, no. 3, pp. 622–626, May 2013.
- [20] G. Liu, M. Li, Y. Wu, P. Zhang, L. Jia, and H. Liu, "PolSAR image classification based on Wishart TMF with specific auxiliary field," *IEEE Trans. Geosci. Remote Sens. Lett.*, vol. 11, no. 7, pp. 1230–1234, Jul. 2014.
- [21] A. P. Doulgeris, "An automatic u -distribution and Markov random field segmentation algorithm for PolSAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 1819–1827, Apr. 2015.
- [22] J.-S. Lee, M. R. Grunes, E. Pottier, and L. Ferro-Famil, "Unsupervised terrain classification preserving polarimetric scattering characteristics," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 4, pp. 722–731, Apr. 2004.
- [23] F. Xu and Y.-Q. Jin, "Deorientation theory of polarimetric scattering targets and application to terrain surface classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 10, pp. 2351–2364, Oct. 2005.
- [24] F. Xu, Y. Li, and Y.-Q. Jin, "Polarimetric-anisotropic decomposition and anisotropic entropies of high-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 9, pp. 5467–5482, Sep. 2016.

- [25] S. H. Yueh, J. A. Kong, J. K. Jao, R. T. Shin, and L. M. Novak, "K-distribution and polarimetric terrain radar clutter," *J. Electromagn. Waves Appl.*, vol. 3, no. 8, pp. 747–768, 1989.
- [26] 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.
- [27] C. C. Freitas, A. Frery, and A. H. Correia, "The polarimetric G distribution for SAR data analysis," *Environmetrics*, vol. 16, no. 1, pp. 13–31, 2005.
- [28] A. P. Doulgeris, S. N. Anfinson, and T. Eltoft, "Automated non-Gaussian clustering of polarimetric synthetic aperture radar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3665–3676, Oct. 2011.
- [29] A. C. Frery, A. H. Correia, and C. D. C. Freitas, "Classifying multifrequency fully polarimetric imagery with multiple sources of statistical evidence and contextual information," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3098–3109, Oct. 2007.
- [30] 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.
- [31] Y. Yamaguchi, T. Moriyama, M. Ishido, and H. Yamada, "Four-component scattering model for polarimetric SAR image decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 8, pp. 1699–1706, Aug. 2005.
- [32] P. R. Kersten, J.-S. Lee, and T. L. Ainsworth, "Unsupervised classification of polarimetric synthetic aperture Radar images using fuzzy clustering and EM clustering," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 519–527, Mar. 2005.
- [33] S. E. Park and W. M. Moon, "Unsupervised classification of scattering mechanisms in polarimetric SAR data using fuzzy logic in entropy and alpha plane," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 8, pp. 2652–2664, Aug. 2007.
- [34] J. S. Lee, D. L. Schuler, R. H. Langz, and K. J. Ransod, "K-distribution for multi-look processed polarimetric SAR imagery," in *Proc. IEEE IGARSS*, vol. 4, Aug. 1994, pp. 2179–2181.
- [35] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [36] H. Larochelle, Y. Bengio, J. Louradour, and P. Lamblin, "Exploring strategies for training deep neural networks," *J. Mach. Learn. Res.*, vol. 10, no. 1, pp. 1–40, Jan. 2009.
- [37] F. Liu, L. Jiao, B. Hou, and S. Yang, "POL-SAR image classification based on Wishart DBN and local spatial information," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3292–3308, Jun. 2016.
- [38] L. Zhang, W. Ma, and D. Zhang, "Stacked sparse autoencoder in PolSAR data classification using local spatial information," *IEEE Trans. Geosci. Remote Sens. Lett.*, vol. 13, no. 9, pp. 1359–1363, Sep. 2016.
- [39] L. Jiao and F. Liu, "Wishart deep stacking network for fast POLSAR image classification," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3273–3286, Jul. 2016.
- [40] L. Xu, J. Neufeld, B. Larson, and D. Schuurmans, "Maximum margin clustering," in *Proc. NIPS*, 2005, pp. 1537–1544.
- [41] F. Bach and Z. Harchaoui, "DIFFRAC: A discriminative and flexible framework for clustering," in *Proc. NIPS*, 2007, pp. 49–56.
- [42] A. Joulin, F. Bach, and J. Ponce, "Discriminative clustering for image co-segmentation," in *Proc. CVPR*, Jun. 2010, pp. 1943–1950.
- [43] A. Joulin, F. Bach, and J. Ponce, "Multi-class cosegmentation," in *Proc. CVPR*, 2012, pp. 542–549.
- [44] J. Sun and J. Ponce, "Learning discriminative part detectors for image classification and cosegmentation," in *Proc. ICCV*, Dec. 2013, pp. 3400–3407.
- [45] S. Singh, A. Gupta, and A. A. Efros, "Unsupervised discovery of mid-level discriminative patches," in *Proc. ECCV*, 2012, pp. 73–86.
- [46] D. Böning, "Multinomial logistic regression algorithm," *Ann. Inst. Stat. Math.*, vol. 44, no. 1, pp. 197–200, Mar. 1992.
- [47] F. P. S. Luus, B. P. Salmon, F. V. D. Bergh, and B. T. J. Maharaj, "Multiview deep learning for land-use classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 12, no. 4, pp. 2448–2452, Dec. 2015.
- [48] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image classification using soft sparse multinomial logistic regression," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 2, pp. 318–322, Mar. 2013.
- [49] S. Chen and H. Wang, "SAR target recognition based on deep learning," in *Proc. DSAA*, 2014, pp. 541–547.
- [50] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Trans. Ind. Electron.*, vol. 63, no. 5, pp. 3137–3147, May 2016.
- [51] N. Ahmed and M. Campbell, "Variational Bayesian learning of probabilistic discriminative models with latent softmax variables," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3143–3153, Jul. 2011.
- [52] G. Hinton *et al.*, "Deep neural networks for acoustic modeling in speech recognition," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012.
- [53] Y. Y. Boykov and M. P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images," in *Proc. ICCV*, Jul. 2001, pp. 105–112.
- [54] M. F. Tappen and W. T. Freeman, "Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters," in *Proc. ICCV*, 2003, pp. 900–906.
- [55] C. Y. Zhu, R. H. Byrd, P. H. Lu, and J. Nocedal, "L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization," *ACM Trans. Math. Softw.*, vol. 23, no. 4, pp. 550–560, Dec. 1997.
- [56] J.-S. Lee, M. R. Grunes, and G. D. Grandi, "Polarimetric SAR speckle filtering and its implication for classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 5, pp. 2363–2373, Sep. 1999.
- [57] S. R. Cloude and E. Pottier, "A review of target decomposition theorems in radar polarimetry," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 2, pp. 498–518, Mar. 1996.



Haixia Bi received the B.S. and M.S. degrees from the Ocean University of China, Qingdao, China, in 2003 and 2006, respectively. She is currently pursuing the Ph.D. degree with Xi'an Jiaotong University, Xi'an, China.

Her research interests include remote sensing image processing and machine learning algorithms.



Jian Sun (M'08) received the B.S. degree from the University of Electronic Science and Technology of China, Chengdu, China, in 2003, and the Ph.D. degree in applied mathematics from Xi'an Jiaotong University, Xi'an, China, in 2009.

He was a Visiting Student at Microsoft Research Asia, Beijing, China, from 2005 to 2008, a Post-Doctoral Researcher with the University of Central Florida, Orlando, FL, USA, from 2009 to 2010, and a Post-Doctoral Researcher with the Willow Project Team, Ecole Normale Supérieure de Paris, Paris, France, and with INRIA, Paris-Rocquencourt, France, from 2012 to 2014. He is currently an Associate Professor with the School of Mathematics and Statistics, Xi'an Jiaotong University. His research interests include the mathematical modeling and machine learning-based approaches for image processing, recognition, and medical image analysis.



Zongben Xu received the Ph.D. degree in mathematics from Xi'an Jiaotong University, Xi'an, China, in 1987.

He is currently the Chief Scientist of the National Basic Research Program of China (973 Project) and the Director with the Institute for Information and System Sciences, Xi'an Jiaotong University, where he is a member of the Chinese Academy of Sciences. His research interests include intelligent information processing and applied mathematics.

Prof. Xu was a recipient of the National Natural Science Award of China in 2007 and the CSIAM Su Buchin Applied Mathematics Prize in 2008. He delivered a 45 min talk on the International Congress of Mathematicians in 2010.