

Received November 4, 2019, accepted November 27, 2019, date of publication December 4, 2019,
date of current version December 17, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2957547

A SAR Image Classification Algorithm Based on Multi-Feature Polarimetric Parameters Using FOA and LS-SVM

SHIYU LUO^{1,2}, (Member, IEEE), KAMAL SARABANDI², (Fellow, IEEE),
LING TONG¹, (Member, IEEE), AND LELAND PIERCE², (Senior Member, IEEE)

¹School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

²Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109, USA

Corresponding author: Shiyu Luo (shiyu_luo@126.com)

This work was supported by the Natural Science Foundation of China under Grant 41571333 and Grant 41901277.

ABSTRACT This paper presents a Synthetic Aperture Radar (SAR) image classification algorithm based on multi-feature using Fruit Fly Optimization Algorithm (FOA) and Least Square Support Vector Machine (LS-SVM). First, pixel-based information derived from three elements of coherency matrix, six parameters obtained by $H/\alpha/A$ decomposition and Freeman decomposition techniques, and three polarimetric parameters including the total receive power (SPAN), pedestal height, and Radar Vegetation Index (RVI), as well as region-based information derived from eight texture parameters obtained by Grey Level Co-occurrence Matrix (GLCM) are combined to use as the features of land cover. Second, Kernel Principal Component Analysis (KPCA) is used to reduce the dimensionality of the multi-feature data derived from the integration of the pixel-based and region-based information. Third, LS-SVM is used as the classifier in this study due to its fast solving speed and desirable classification capability. Since the input parameters of LS-SVM significantly affect the classification performance, we employ FOA to obtain the optimized input parameters. Finally, the experiments on two fully polarimetric SAR images of various crops with a limited number of samples are implemented by the proposed method and other commonly used methods, respectively. The results show that the proposed method can attain better classification performances compared with other methods.

INDEX TERMS Polarimetric SAR image, classification, multi-feature, FOA.

I. INTRODUCTION

Land cover information is important for land development and management. Land cover classification is also the first step in remote sensing of vital global parameters such as soil moisture. Remotely sensing data obtained from various sensors provides an economical way to characterize land cover information. Optical remote sensing is an effective approach but by itself is limited by weather conditions. Synthetic Aperture Radar (SAR), which can obtain information under different weather conditions, is therefore used for acquiring land cover information in various regions.

Significant research aiming at land cover classification has been reported by many researchers. In the early years, most studies were developed based on single-polarization

data [1], [2]. Since single-polarization data does not contain all the polarization information of ground objects, such methods were most likely to create confusion among similar ground objects and thereby were only suited for coarse classification [3]–[5]. With the rapid development of the SAR techniques, many methods utilizing multi-polarization or full-polarization data were explored for attaining a better classification [6]–[8]. The critical procedure for these methods is polarimetric decomposition, which provides a way to obtain the physical features of natural media. Many polarimetric decomposition methods have been explored by many researchers [9]–[14]. However, Shimoni *et al.* indicated that the sensitivities of different polarimetric decomposition methods to land cover types are various [15], also different features have different metric contributions [16]. They recommended to use various decomposition methods for

The associate editor coordinating the review of this manuscript and approving it for publication was Gerardo Di Martino¹.

land cover classification to achieve a better classification performance.

Thus far, however, the polarimetric parameters obtained by decomposition methods are pixel-based features. Some studies have demonstrated that substantial improvements in classification can be attained by integrating polarimetric and regional information [17], [18]. E.g. [19] utilizes Markov random field to obtain regional information for a better SAR image classification performance. In general, textural feature based on grey level is commonly used to characterize regional information. Grey Level Co-occurrence Matrix (GLCM) is the most used tool to enable the acquisition of textural and spatial features [20], which effectively compensates for the pixel-based information from polarimetric information. Haralick proposed fourteen statistical features obtained from the GLCM to describe the texture information [21]. Usually, eight statistical features of them are enough to characterize the texture information [22]. For a full-polarization SAR image, 24 statistical features need to be calculated with respect to hh, vv, and hv polarizations.

In general, the main steps of a supervised classification method are feature extraction and model training based on machine learning. The use of as many features as possible theoretically helps to improve the classification accuracy. However, the increase in feature dimension leads to two significant issues, curse of dimensionality and feature redundancy [23]. For the first issue, when the dimension of data is high, much more training data is required for estimating the distribution of the data. Since the estimation of the data distribution from finite samples is essential in machine learning, the ability of machine learning is reduced with the increase in feature dimension. Besides, the algorithm complexity rises sharply with the increase in feature dimension. As regards to the second issue, feature redundancy can add noise without providing any additional information. In machine learning, the features are given equal significances, which brings about the exaggeration of the influence of the redundant information and thus it may lead to wrong results especially if the data is accompanied by noise or error. Therefore, optimal dimensionality reduction is necessary. The common descending dimension algorithms include Principal Component Analysis (PCA) [24], Independent Component Analysis (ICA) [25], and Linear Discriminant Analysis (LDA) [26]. However, the above algorithms are based on linear characteristics. They cannot be applied to the non-Gaussian data set. In order to solve this problem, a kernel function is adopted to map such datasets nonlinearly into high dimensional space, which changes a linearly inseparable problem into a separable one. The corresponding improved methods are Kernel Principal Component Analysis (KPCA) [27], Kernel Independent Component Analysis (KICA) [28], and Kernel Linear Discriminant Analysis (KLDA) [29]. Li concluded that the performance of the three methods is nearly the same, whereas KPCA is less time-consuming [30]. Therefore, we employ KPCA to reduce the dimensionality of the data in this study.

The classifiers utilized in SAR image classification can be divided into two categories. One is statistical distribution-based models, and the other is non-parametric-based models. The common method for the former case is under the assumption that the coherency matrix of each pixel follows Wishart distribution [6], [31]–[33]. However, since the coherency matrices of pixels for each class are not completely modelled by this distribution [34], [35], the performance suffers. The most effective classifiers for the latter case are neural networks and Support Vector Machines (SVM) [2], [36], [37]. Since neural networks tend to get trapped in local minima, SVM is much more used for SAR image classification in recent studies [38]. However, since the complexity of SVM based on the quadratic programming problem is strongly related to the number of training data, the approach is too slow. Therefore, Suykens *et al.* proposed Least Squares Support Vector Machine (LS-SVM) based on linear equation [39], which significantly improves the speed and reduces the required computational resources. In SVM and LS-SVM, the input parameters strongly affect the classification result [40], whereas such parameters are set by past experience for most common cases, resulting in the requirement of large training dataset for ensuring the classification accuracy. However, as noted by Chi *et al.*, definition and acquisition of a sufficiently large amount of training samples from a SAR image is often a critical problem [41]. Therefore, parametric optimization is another approach to improve the classification performance if the training dataset is limited. The typical optimized algorithms are Ant Colony Algorithm (ACA) [42], Fish Swarm Algorithm (FSA) [43], Genetic Algorithm (GA) [44], and Particle Swarm Optimization (PSO) [45]. These algorithms are of either high computational complexity or poor global search capability [46]. Recently, Pan proposed a novel optimized algorithm called Fruit Fly Optimization Algorithm (FOA) [47], which is inspired by a fruit fly looking for food. Compared with other swarm intelligence algorithms, FOA is of low computational complexity, good global search capability, and high precision. The employment of FOA contributes to an improved classification accuracy while reducing the demand for training data.

In this study, a method for SAR image classification based on multi-feature using FOA and LS-SVM is proposed. The contributions of this paper are summarized as follows. First, we employ various polarimetric decomposition techniques and GLCM to obtain polarimetric and texture information, which are then integrated as the features of land cover. Second, we utilize KPCA to reduce the dimensionality of the multi-feature dataset obtained by the integration of polarimetric and texture information. Third, to attain a better classification result under the condition of a limited set of training data, we introduce FOA to LS-SVM. By combining FOA with LS-SVM, the optimized parameters of LS-SVM classifier can be obtained. In the experiments, the classification of two SAR images is implemented by the proposed method and other three commonly used methods, respectively.

II. METHODOLOGY

The main steps of the proposed method include extraction of the polarimetric and textual features using polarimetric decomposition methods and GLCM, respectively, dimensionality reduction of multi-feature dataset using KPCA, and input parameters optimization with respect to LS-SVM using FOA. Before applying the method to an image, the refined Lee speckle filter with a 5×5 window is used to reduce the noise in the image [48].

A. CHARACTERISTIC PARAMETERS OF LAND COVER

1) POLARIMETRIC INFORMATION

The polarimetric information of each pixel in a radar image is completely represented by a backscattering matrix with four complex elements S_{hh} , S_{hv} , S_{vh} and S_{vv} , where h and v denote horizontal and vertical polarization respectively. Generally, it is assumed that the backscattering matrix is symmetric under the assumption that the transmit and receive antennas coincide, thereby leading to $S_{hv} = S_{vh}$. For the natural terrain surfaces, since there is more than one scattering center in a resolution cell of a SAR image, many decomposition methods for the incoherent cases have been developed, which are based on the analysis of the coherency matrix or the covariance matrix. Given a vector form of the elements of the backscattering matrix $\mathbf{k} = 1/\sqrt{2} [S_{hh} + S_{vv}, S_{hh} - S_{vv}, 2S_{hv}]^T$ or $\mathbf{t} = [S_{hh}, \sqrt{2} S_{hv}, S_{vv}]^T$, the coherency matrix \mathbf{T}_3 and the covariance matrix \mathbf{C}_3 are obtained respectively by $\mathbf{T}_3 = \mathbf{k}\mathbf{k}^{*T}$ and $\mathbf{C}_3 = \mathbf{t}\mathbf{t}^{*T}$, where T denotes transposition, $*$ denotes conjugation.

The most used incoherent decomposition methods are Freeman decomposition and $H/\alpha/A$ decomposition [9], [49]. Freeman method decomposes the covariance matrix into three components related to different scattering mechanisms as follows,

$$\langle \mathbf{C} \rangle = f_s \langle \mathbf{C}_s \rangle + f_d \langle \mathbf{C}_d \rangle + f_v \langle \mathbf{C}_v \rangle \quad (1)$$

where $\langle \rangle$ denotes multi-look processing, f_s, f_d, f_v are the weights related to surface, double-bounce, and volume scattering, respectively. The corresponding scattering powers P_s, P_d, P_v are then obtained based on f_s, f_d, f_v .

$H/\alpha/A$ method decomposes the coherency matrix of eigenvectors and eigenvalues $\lambda_1, \lambda_2, \lambda_3$ ($\lambda_1 \geq \lambda_2 \geq \lambda_3$),

$$\langle \mathbf{T} \rangle = \mathbf{U} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \mathbf{U}^{*T} \quad (2)$$

where \mathbf{U} is the unitary matrix. The column vectors of \mathbf{U} are the eigenvectors.

The scattering entropy H , the average scattering angle $\bar{\alpha}$, and the anisotropy degree A can be obtained in terms of the elements derived from (2). We refer readers to literature [49] for the detailed procedure.

Each decomposition method has disadvantages. For example, Freeman decomposition cannot obtain the random

fluctuations of the media, and the parameters obtained by $H/\alpha/A$ decomposition are not of any profound physical meanings [50]. Furthermore, different polarimetric decomposition techniques aim at different ground objects [15]. The integration of various decomposition methods therefore attributes to improve land cover classification.

In addition to the polarimetric decomposition methods, there are also some parameters characterizing polarimetric features. One is (SPAN) indicating the total received power from the four polarimetric channels. It is obtained by,

$$SPAN = |S_{hh}|^2 + 2|S_{hv}|^2 + |S_{vv}|^2 \quad (3)$$

The other two common parameters are polarimetric pedestal height P_H and Radar Vegetation Index (RVI). The latter one is for volume scattering media, such as vegetation canopies. They are given by,

$$P_H = \frac{\min(\lambda_1, \lambda_2, \lambda_3)}{\max(\lambda_1, \lambda_2, \lambda_3)} \quad (4)$$

$$RVI = \frac{4 \min(\lambda_1, \lambda_2, \lambda_3)}{\lambda_1 + \lambda_2 + \lambda_3} \quad (5)$$

where λ_i ($i = 1, 2, 3$) are the eigenvalues obtained from \mathbf{T}_3 .

In this study, we used three elements T_{11}, T_{22}, T_{33} from the coherency matrix, six parameters obtained by Freeman decomposition and $H/\alpha/A$ decomposition methods, and three polarimetric parameters including (SPAN), P_H , and (RVI) as input features.

2) REGIONAL INFORMATION

The grey level variability within the neighborhood of a pixel is an important image feature as well. The texture information is used to describe such spatial-based feature. Therefore, GLCM in this study is employed to obtain textural and spatial features.

GLCM utilizes the probability of replication with respect to a certain grayscale to describe texture information. It describes the integrated information concerning the direction, the interval of adjacency, and the variation range of the image value. The steps can be implemented as follows,

a. Given an image $I \in M \times M$, the elements of a co-occurrence matrix \mathbf{P} are defined as,

$$P(i, j) = \sum_{x=1}^M \sum_{y=1}^M \begin{cases} 1, & I(x, y) = i, I(x + \Delta x, y + \Delta y) = j \\ 0, & \text{other} \end{cases} \quad (6)$$

where Δ is the offset distance denoting the distance between the pixel and its neighbors. $P(i, j)$ counts the number of times that (i, j) occurs.

b. The co-occurrence matrix can be obtained in term of different offset directions. Usually four directions including horizontal direction, upper-right direction, vertical direction, and upper-left direction represented by $[0, \Delta]$, $[-\Delta, \Delta]$, $[-\Delta, 0]$, and $[-\Delta, -\Delta]$, respectively, are used to generate four co-occurrence matrices.

c. In order to improve the algorithm efficiency, the four matrices are processed by the normalization method by converting grey values to a certain grayscale G . In this case, probability $p(i, j)$ is used instead of number of times $P(i, j)$ in (6). Finally, the GLCM \mathbf{P} is obtained by averaging the four matrices.

In [21], fourteen statistical features obtained from the GLCM is used to describe texture information. Assuming that the normalized grayscale is G , the used features in this study are Contrast (CON), Angular Second Moment (ASM), Correlation (COR), Entropy (ENT), Inverse Difference Moment (IDM), Maximum probability (MAX), Dissimilarity (DIS), and Inversion (INV). Please see literature [21] for the detailed information.

For this algorithm, three parameters including window size, offset distance, and grayscale affect the result. A sliding window is used to obtain various textural parameters. A large window size contributes to reduce the influence of image noise while losing target characteristics. A small window size can preserve subtle characteristics whereas the results are strongly affected by noise. Offset distance should be small to guarantee the accuracy of co-occurrence matrices. The G grayscale generates a $G \times G$ co-occurrence matrix. A large G requires high computational resources. According to the conclusions in the literature [51], [52], we set window size to 5, offset distance Δ to 1, and grayscale G to 8 in this study.

B. KPCA FOR DIMENSIONALITY REDUCTION

Theoretically, if the number of samples is infinite, the curse of dimensionality will not occur. In practice, the required number of samples increases exponentially with the increase of dimensionality. In other words, if the number of samples is limited, it is better to use a small number of features [53]. In addition to 9 pixel-based features, 24 textural parameters for a full-polarization SAR image are included in the input features.

33-dimensional features are probably relatively large in terms of the used limited number of the samples in a SAR image. Furthermore, it is noted that the used features are correlated in some level, e.g. the three elements from the coherency matrix. The use of KPCA can significantly reduce the dimension of the data set while preserving as much variance as possible. The small number of input features contributes to improving the efficiency of the classifier. Consequently, KPCA is employed to reduce the dimensionality of this multi-feature dataset. The Cover theorem indicates that for n random data in Euclidean space, the linearly separable problem can be achieved by nonlinear mapping such data in the low dimensional space into a high dimensional space. Kernel function is one of the best mapping functions that meets the necessary requirements. Consequently, the key step of KPCA is to map the samples in the original input space nonlinearly into the characteristic space Ω and then utilize PCA in such space [27].

Considering an n -sample set $\mathbf{X} = \{x_1, x_2, \dots, x_n\} \in \mathbb{R}^d$ in a d dimensional space, we can map this set to a much higher dimensional space referred to the characteristic space by using a mapping function (kernel function) $\phi(\cdot)$, which is denoted by $\Phi(\mathbf{X}) = \{\phi(x_1), \phi(x_2), \dots, \phi(x_n)\}$. Because PCA requires the data set with zero-mean [24], in order to apply PCA in this characteristic space, $\Phi(\mathbf{X})$ is centralized using the equation,

$$\begin{aligned} \psi(\mathbf{X}) &= \{\psi(x_1), \psi(x_2), \dots, \psi(x_n)\} \\ \psi(x_j) &= \phi(x_j) - \frac{1}{n} \sum_{i=1}^n \phi(x_i), \quad j = 1, 2, \dots, n \end{aligned} \quad (7)$$

thus $\sum_{i=1}^n \psi(x_i) = 0$.

To implement PCA in this characteristic space [24], assuming that the eigenvectors of $\psi(\mathbf{X}) \psi(\mathbf{X})^T$ in the characteristic space are denoted by vectors ω_i ($i = 1, 2, \dots, d$), λ_i ($i = 1, 2, \dots, d$) are the corresponding eigenvalues. PCA is implemented by the formula,

$$\psi(\mathbf{X}) \psi(\mathbf{X})^T \omega_i = \lambda_i \omega_i \quad (8)$$

Since any vectors in a space can be linearly represented by all samples in such space, ω_i can be represented by the sample set $\psi(\mathbf{X})$,

$$\omega_i = \sum_{i=1}^n \alpha_i \psi(x_i) = \psi(\mathbf{X}) \alpha \quad (9)$$

Combining (8) with (9), we can obtain the formula after some algebraic manipulations:

$$\bar{\mathbf{K}} \alpha = \lambda_i \alpha \quad (10)$$

where $\bar{\mathbf{K}} = \psi(\mathbf{X})^T \psi(\mathbf{X})$ is defined as the $n \times n$ centralized kernel matrix. $\bar{\mathbf{K}}$ can be rewritten with respect to the mapping function $\phi(\cdot)$ by applying (7) as follows,

$$\bar{\mathbf{K}} = \mathbf{K} - \mathbf{I}_n \mathbf{K} - \mathbf{K} \mathbf{I}_n + \mathbf{I}_n \mathbf{K} \mathbf{I}_n \quad (11)$$

where $\mathbf{K} = \Phi(\mathbf{X})^T \Phi(\mathbf{X})$, \mathbf{I}_n is a $n \times n$ matrix where each element is $1/n$.

The dimensionality of \mathbf{X} is thus reduced to r -dimensional by using $\mathbf{X}_r = [\alpha_1, \dots, \alpha_r]^T \mathbf{X}$, ($r < d$), where α_i is obtained by (10) corresponded to the first r greatest λ_i ($i = 1, \dots, r$). r is determined when the ratio of $\sum_{i=1}^r \lambda_i$ ($r = 1, \dots, d$) to $\sum_{i=1}^d \lambda_i$ is greater than a threshold value ts . For most of cases, $ts = 90\%$ can attain a satisfactory result [54]. In our experiment, we set $ts = 90\%$. It is noted that the computational steps concerning kernel functions can be calculated by inner products. In this study, we choose polynomial Mercer kernel as the kernel function based on the facts that [55]–[57]: **a.** its performance is not significantly dependent on kernel parameters and **b.** it is efficient compared with other common used kernel (not considering linear kernel) when the dimensionality of data is very large. Polynomial Mercer kernel's inner product is $k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle = (x_i^T x_j + p)^q$. In the following experiment, we set $p = 1$, $q = 3$ based on the experimental experiences. It is noteworthy

that what measure of model performance that can be used in KPCA is still discussed by many researchers [55], [58]. This leads to that the optimal kernel including kernel types or kernel parameters is difficult to obtain. The selection of kernel and its parameters are not the main concern in this manuscript. We will consider this problem in our future work.

C. THE LS-SVM-BASED CLASSIFIER WITH OPTIMIZED PARAMETERS

1) THE LS-SVM CLASSIFIER

LS-SVM is an improved classifier based on SVM [39], it significantly improves the computational speed while reducing the required computational resources. LSSVM is robust to low-level feature noise due to its loss of sparsity while it is not robust to outliers. In the most of cases, the samples from a SAR image are obtained by the field investigation rather than visual examination. A point in a SAR image can be accurately associated with its corresponding land cover according to geographic coordinates. The case of wrong labelled points rarely occurs. Therefore, the effect of the robustness of LSSVM to outliers on the classification performance is not taken into account in this method. Given a n -training set $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \in \mathbb{R}^d \times \mathbb{R}$, where $x_i \in \mathbb{R}^d$ is the i -th input data with d dimensions and $y_i \in \mathbb{R}$ is the i -th output data, mapping this set into a high dimension space by utilizing kernel function $\phi(\cdot)$, the aim of LS-SVM is to find the optimal separating hyperplane in such space, which is expressed as,

$$y(x) = w \cdot \phi(x) + b \quad (12)$$

where w defines confidence interval, b is a real constant, $y(x)$ is the empirical risk.

Based on the structural risk minimization principle, this optimal separating hyperplane can be found by fixing empirical risk and minimizing confidence space. Since classification with zero error basically cannot be achieved, introducing non-negative slack variables ξ_i ($i = 1, \dots, n$) and the regularization factor $\gamma > 0$ to (12), consequently, w and b can be obtained by the equation with the constraint with respect to the training set as follows,

$$\begin{aligned} \min J(w, \xi) &= \frac{1}{2} \|w\|^2 + \gamma \sum_{i=1}^n \xi_i^2 \\ \text{s.t. } y_i &= w \cdot \phi(x_i) + b + \xi_i, \quad i = 1, \dots, n \end{aligned} \quad (13)$$

where (x_i, y_i) is the training set.

Equation (13) can be solved by Lagrangian dual method governed by,

$$L(w, b, \xi_i, \beta_i) = J(w, \xi) - \sum_{i=1}^n \beta_i (w \cdot \phi(x_i) + b + \xi_i - y_i) \quad (14)$$

where $\beta_i \in \mathbb{R}^{n \times 1}$ are Lagrange multipliers.

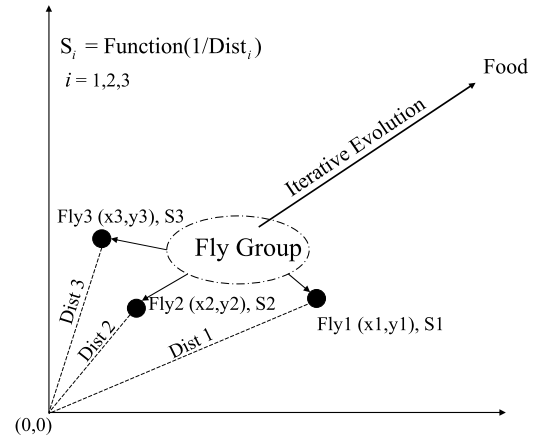


FIGURE 1. The sketch map of FOA.

Using Karush-Kuhn-Tucker (KKT) condition, four linear equations with respect to (14) can be obtained by $\partial L / \partial w = 0$, $\partial L / \partial b = 0$, $\partial L / \partial \xi_i = 0$, and $\partial L / \partial \beta_i = 0$, respectively. The solutions of β_i and b are therefore obtained by,

$$\begin{bmatrix} 0 & -\mathbf{y}^T \\ \mathbf{y} & \Phi \Phi^T + \gamma^{-1} \mathbf{I} \end{bmatrix} \begin{bmatrix} b \\ \boldsymbol{\beta} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{1}_n \end{bmatrix} \quad (15)$$

where $\Phi = [\phi(x_1), \dots, \phi(x_n)]^T$, $\mathbf{y} = [y_1, \dots, y_n]^T$, $\boldsymbol{\beta} = [\beta_1, \dots, \beta_n]^T$, \mathbf{I} is the unit matrix, and $\mathbf{1}_n = [1, \dots, 1]^T$. The solutions of β_i and b are only related to the training set and the regularization factor γ .

Consequently, the LS-SVM classifier is obtained by,

$$f(x) = \sum_{i=1}^n \beta_i k(x_i, x) + b \quad (16)$$

where β_i and b are obtained by (15), $k(x_i, x)$ is the inner product with respect to $\mathbf{K} = \Phi(X)^T \Phi(X)$. It is noted that in addition to parameters of kernel functions, (16) is related to the regularization factor γ , which is optimized in the following Sections.

2) FRUIT FLY OPTIMIZATION ALGORITHM

FOA is a swarm intelligence technique for global optimization [47]. Compared with other swarm intelligence algorithms, FOA is of low computational complexity, good global search capability, and high precision. The basic theory is as follows. First, smelling is used to fly towards an approximate location of the food. Second, when the fly approaches the food within a certain distance, vision is used to move towards the specific food position of the food. The sketch map is shown in Fig. 1. Based on this figure, the iterative procedures are described as follows,

a. Determining the necessary parameters, including the group size (g_size), the maximum iterations, the distance (dis), and the initialized location (X_in, Y_in). The group size is defined as the number of fruit flies in each iteration,

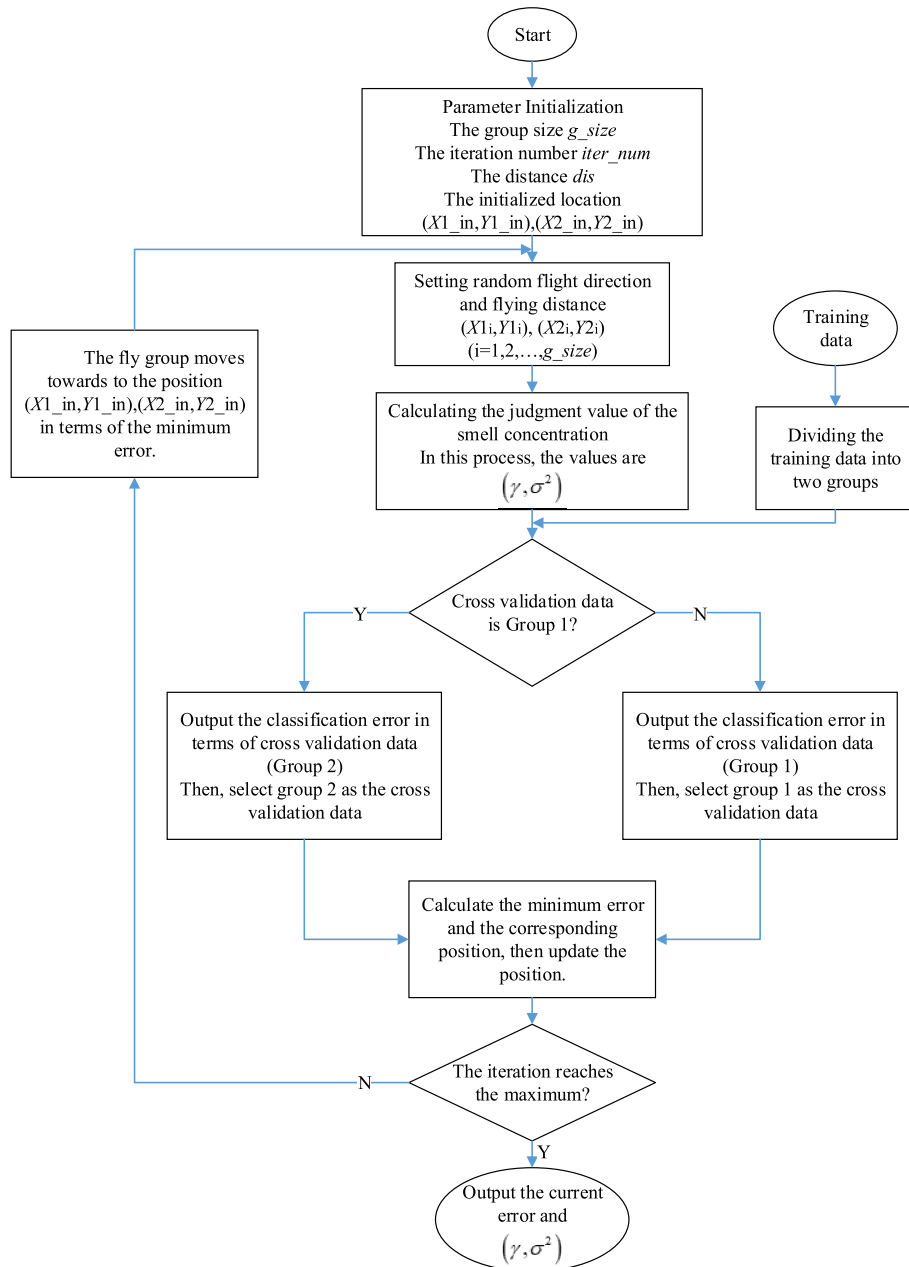


FIGURE 2. The flow chart of LS-SLM with FOA.

the larger group size speeds up the convergence, however, the algorithm requires more time. The distance denotes the maximum distance moved for each fruit fly in every iteration. The distance and the initialized location both affect the algorithm efficiency.

b. Setting flight direction and flying distance. Defining a random number $rand \in [0, 1]$, the current position is expressed by,

$$\begin{cases} X_i = X_in + 2 \times dis \times rand - dis \\ Y_i = Y_in + 2 \times dis \times rand - dis \end{cases}, \quad i = 1, \dots, g_size \quad (17)$$

c. Calculating the distance to the origin.

$$Dist_i = \sqrt{X_i^2 + Y_i^2} \quad (18)$$

Thus, the judgment value of the smell concentration is defined as the reciprocal of such distance $1/dist$.

d. Calculating the smell concentration S_i for each fruit fly. The smell concentration is obtained by solving the fitness function with the input variable (the judgment value of the smell concentration), where the fitness function is decided in terms of the objective problem:

$$S_i = fitness(1/dist_i) \quad (19)$$

e. Finding out the fruit fly with the optimized solution among the fly group and the corresponding position for the next iteration. Finally, the rest of fruit flies move towards to such position.

$$\begin{aligned} [BestS, position] &= optimization(S) \\ SBest &= BestS \\ X_in &= X(position), \quad Y_in = Y(position) \end{aligned} \quad (20)$$

f. Repeating b-e until the iterative number is reached.

3) INTRODUCTION OF FOA TO LS-SVM

The classification performance is affected by the input variables of the kernel function and LS-SVM [40]. In some cases, the performance of the classification with the default input variables cannot attain a good result. Therefore, FOA is adopted to obtain the optimized parameters for the classifier. In this process, we use Radial Basis Function (RBF) as the kernel function. Because RBF has good ability to cope with nonlinear dataset when its parameter is appropriate. Also, it has only one parameter, which requires less computational resources compared with other kernel functions. RBF is defined as,

$$k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (21)$$

where σ^2 is the bandwidth in RBF.

In addition to σ^2 , the regularization parameter γ in LS-SVM also needs to be optimized, which provides a trade-off between the fitting error minimization and smoothness. Consequently, we need to optimize two parameters, σ^2 and γ . σ^2 and γ are assumed to be the judgment value of the small concentration. Furthermore, the range of γ basically is ten times larger than that of σ^2 . The current position and the judgment value of the small concentration in step b and c thus are defined respectively as follows,

$$\begin{cases} X1_i = X1_in + 2 \times dis \times rand - dis \\ Y1_i = Y1_in + 2 \times dis \times rand - dis \\ X2_i = X2_in + 2 \times dis \times rand - dis \\ Y2_i = Y2_in + 2 \times dis \times rand - dis \end{cases} \quad i = 1, \dots, g_size \quad (22)$$

$$\gamma = 10 \times \frac{1}{\sqrt{X1_i^2 + Y1_i^2}}, \quad \sigma^2 = \frac{1}{\sqrt{X2_i^2 + Y2_i^2}} \quad (23)$$

In order to avoid the over-learning problem, we divide the training data into two groups for cross validation. Moreover, the fitness function is defined as the classification error. The optimized result has minimum error. The fitness function is defined as,

$$S_i = \frac{error(\gamma, \sigma^2)}{num} \quad (24)$$

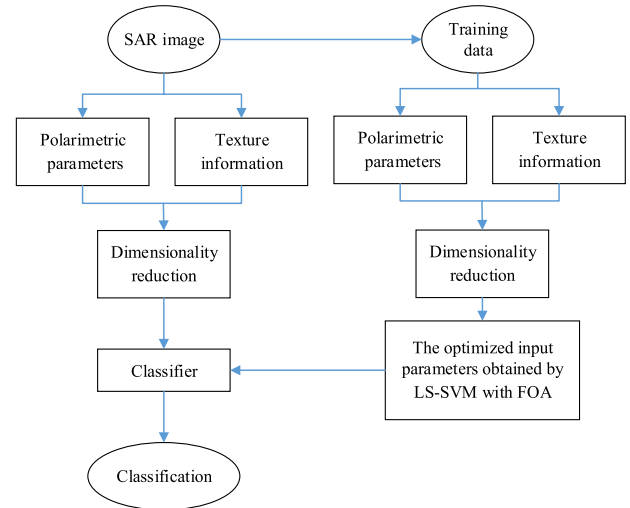











FIGURE 3. The flow chart of the proposed method.

TABLE 1. The number of the pixels for the training data and the validation data.

	Class	Training	Ratio (%)	Validation
	Barley (BA)	2500	1.91	43821
	Maize (MA)	2500	1.91	42498
	Potato (PO)	2500	1.91	27238
	Rapeseed (RA)	2500	1.91	17503
	Lucerne(LU)	1168	0.89	3095
	Grass (GR)	2500	1.91	5229
	Peas (PE)	553	0.42	970
	Beans (BE)	629	0.48	492
	Fruit (FR)	2078	1.59	6460
	Total	16928	12.93	147306

where num is the number of the cross validation data, $error$ is the number of the misclassified points obtained by,

$$\begin{cases} [\alpha, b] = TrainLS_SVM(\gamma, \sigma^2, Data_group(x, y)_k) \\ y_t = SimLS_SVM(\alpha, b, Data_group(x)_t), \end{cases} \quad (25)$$

$$error = error + 1, \quad y_t \notin Data_group(y)_t \quad (26)$$

where $(k = 1, t = 2)$ or $(k = 2, t = 1)$. $TrainLS_SVM$ is the training model used for calculating α and b in terms of input variables, $SimLS_SVM$ is the simulated model used for obtaining classification result, and $Data_group()$ is either the one or the other data group from the training data.

Consequently, the flow chart of LS-SLM with FOA is illuminated in Fig. 2. Fig. 3 shows the flow chart of the proposed method.

III. EXPERIMENTAL RESULTS AND DISCUSSION

The two tested images are the well-known fully polarimetric SAR images with 1279×1024 pixels at C band and L band respectively acquired by AIRSAR from Flevoland. It is noted that the sensitivities of the identical object to different frequencies are quite different, leading to that the polarimetric and textural features of the identical object are different at various bands. Therefore, the two tested images can be treated as two different images although they were acquired from

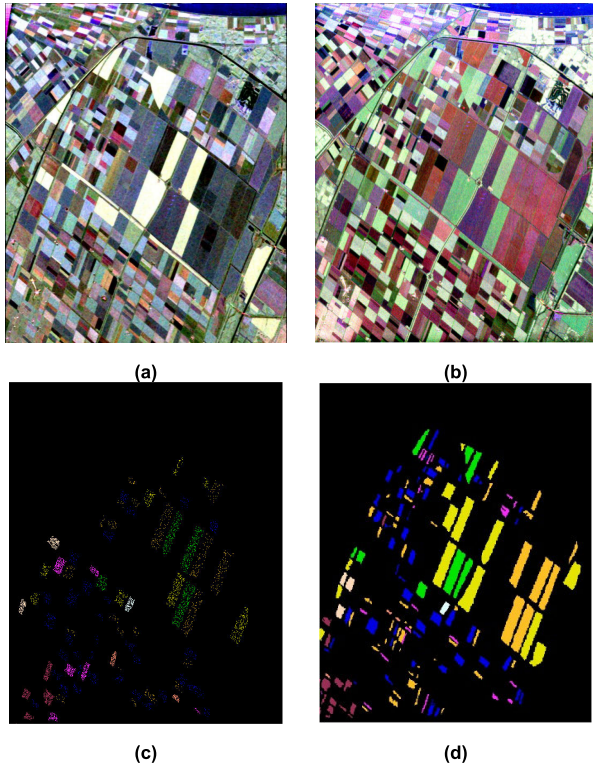


FIGURE 4. (a) RGB composite image obtained at C band by the Pauli decomposition method. (b) RGB composite image obtained at L band by the Pauli decomposition method. (c) The used training data. (d) The validation data.

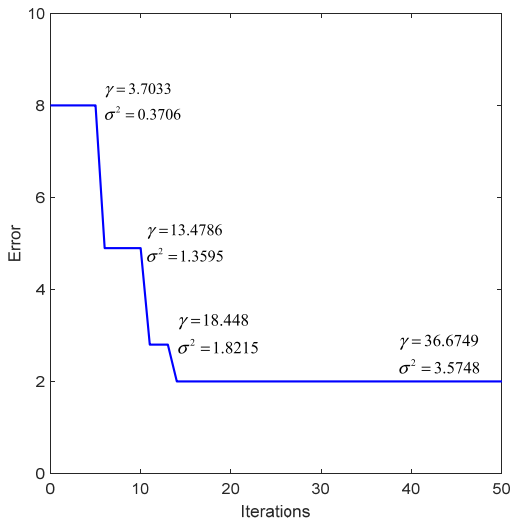


FIGURE 5. The optimization process using FOA with respect to C-image.

the same region. The observed region consists of a variety of crops in farmland. The crops are easily distinguished by the naked eye whereas good classification result is difficult to attain due to the similar characteristics of different crops. There are nine classes in the study area. Each class (crop) is represented by one color in the ground truth map. The ground truth data is divided into training data and testing data using the following method. In order to test the performance

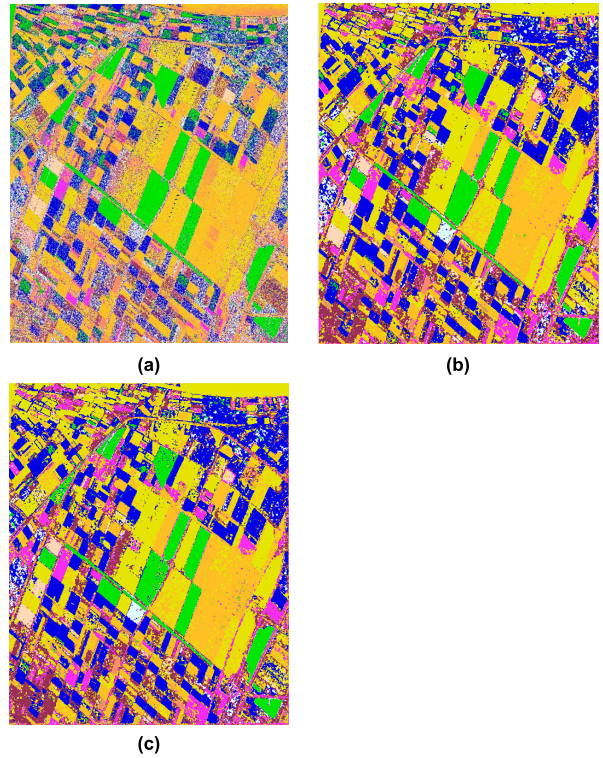


FIGURE 6. Classification accuracies of the proposed method with respect to C-image (a) using the default parameters, (b) using the empirical parameters, and (c) using the optimized parameters.

of the proposed method on the condition of a small number of samples, the used training data is selected randomly from the training dataset and the number of pixels in each class is limited to less than 2500. The number of the training and validation points in each class, the ratio of the training data over all the pixels, and the corresponding crops are described in Table 1. Fig. 4(a)(b) show the RGB composite images of the test images obtained by the Pauli decomposition method at C band and L band, respectively, Fig. 4(c) shows the used training data, and Fig. 4(d) shows the validation data. It can be seen that Figure 4(a) is different from Fig. 4(b) due to the feature differences. For convenience, the tested image acquired at C band will be referred to as C-image, and the tested image acquired at L band will be referred to as L-image. In the following experiments, the confusion matrix is used to evaluate the classification performance. The evaluation parameters include User Accuracy (UA), Producer Accuracy (PA), Overall Accuracy (OA), and the kappa coefficient. It is noted that kappa coefficient is an unbiased evaluation parameter, which is specifically used to evaluate the classification performance when the samples and validations from each class are unbalanced.

A. THE EFFECT OF INPUT PARAMETERS ON CLASSIFICATION PERFORMANCE

The result of the classification is significantly affected by the input variables of LS-SVM. The optimized parameters are obtained by FOA in this study. In FOA, some parameters need

TABLE 2. Classification accuracy of the proposed method using the default parameters with respect to C-image.

CLASSIFIED DATA	REFERENCE DATA										TOTAL	UA (%)
	BA	MA	PO	RA	LU	GR	PE	BE	FR			
BA	26019	5611	1097	3	153	429	12	1	647	33972	76.59	
MA	11227	32444	370	1	340	1004	0	26	169	45581	71.18	
PO	547	18	15546	84	24	75	132	0	247	16673	93.24	
RA	27	0	2207	17340	0	202	8	0	90	19874	87.25	
LU	1213	1230	401	3	2167	385	21	0	966	6386	33.93	
GR	3572	634	400	6	247	2252	25	0	256	7392	30.47	
PE	383	52	5558	62	54	437	645	0	1162	8353	7.72	
BE	15	2252	69	0	0	236	0	465	209	3246	14.33	
FR	818	257	1590	4	110	209	127	0	2714	5829	46.57	
Total	43821	42498	27238	17503	3095	5229	970	492	6460	147306		
PA (%)	59.38	76.34	57.08	99.07	70.02	43.07	66.50	94.45	42.01			
OA (%)	67.61											
Kappa	0.60											

TABLE 3. Classification accuracy of the proposed method using the empirical parameters with respect to C-image.

Classified data	Reference data										Total	UA (%)
	BA	MA	PO	RA	LU	GR	PE	BE	FR			
BA	37569	4711	935	513	465	423	0	146	467	45229	83.06	
MA	5916	36566	153	0	0	408	0	93	91	43227	84.59	
PO	284	116	24694	530	241	83	24	15	314	26301	93.89	
RA	0	189	365	16438	0	70	0	72	1	17135	95.93	
LU	11	20	43	0	2298	106	8	0	161	2647	86.82	
GR	13	299	55	0	42	3361	3	6	494	4273	78.66	
PE	0	1	137	11	1	62	913	0	8	1133	80.58	
BE	0	472	12	1	0	206	0	160	21	872	18.35	
FR	28	124	844	10	48	510	22	0	4903	6489	75.56	
Total	43821	42498	27238	17503	3095	5229	970	492	6460	147306		
PA (%)	85.73	86.05	90.66	93.92	74.25	64.28	94.12	32.52	75.90			
OA (%)	86.20											
Kappa	0.82											

TABLE 4. Classification accuracy of the proposed method using the optimized parameters with respect to C-image.

Classified data	Reference data										Total	UA (%)
	BA	MA	PO	RA	LU	GR	PE	BE	FR			
BA	38657	4384	1180	0	139	218	2	0	364	44944	86.01	
MA	5099	37468	164	1	0	528	0	30	140	43431	86.27	
PO	30	17	24506	3	231	1	42	0	250	25080	97.71	
RA	0	131	136	17433	0	73	0	2	0	17775	98.08	
LU	8	16	44	0	2396	79	6	0	148	2697	88.83	
GR	3	299	54	0	45	3455	0	24	413	4293	80.48	
PE	0	0	337	17	184	10	888	19	9	1464	60.66	
BE	0	105	15	40	46	228	0	417	17	868	48.04	
FR	24	78	802	9	54	637	32	0	5119	6755	75.78	
Total	43821	42498	27238	17503	3095	5229	970	492	6460	147306		
PA (%)	88.22	88.16	90.00	99.60	77.74	66.07	91.55	84.76	79.24			
OA (%)	88.57											
Kappa	0.85											

to be set initially. According to the aforementioned definitions in FOA and our experience with respect to LS-SVM, we set the group size to 10, maximum iteration is 50, the distance is 3, and the initialized location is $[X_1, Y_1] = [X_2, Y_2] = [0.5, 0.5]$. Fig. 5 shows the optimization process for the LS-SVM classifier using FOA with the input training data with respect to C-image. The output is the error in terms of the different parameters obtained by FOA in each iteration.

It is observed in Fig. 5 that the different parameters lead to different errors. After 15 iterations, the error reaches the minimum and the current parameters $\gamma = 36.6749, \sigma^2 = 3.5748$ are supposed to be the optimized parameters for this classification model. Since the number of the training data for some classes is very small, the case with zero error is difficult to achieve. Similarly, the optimized parameters for L-image are $\gamma = 26.5363, \sigma^2 = 1.2655$. It indicates that

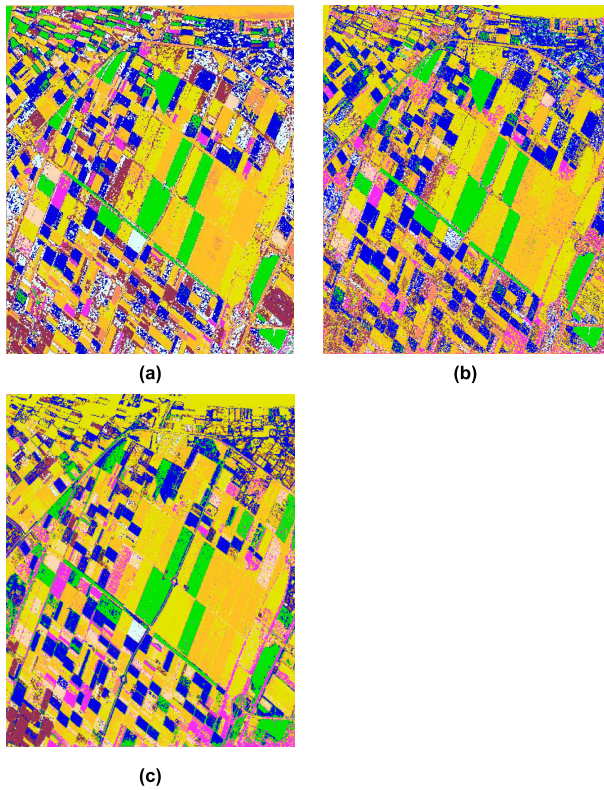


FIGURE 7. The classification results of C-image by (a) the supervised Wishart method, (b) the manifold-learning-based supervised graph embedding method, and (c) the sparse representation-based method.

the optimized input parameters of LS-SVM change with the kinds of the input land cover features rather than being fixed.

To assess the effect of input parameters on classification performance, we use the default input parameters ($\gamma = 1$, $\sigma^2 = 1$) and the empirical parameters ($\gamma = 10$, $\sigma^2 = 0.25$) [59] as the comparisons. The classification results of C-image with respect to various input parameters are shown in Fig. 6(a)(b)(c) and the corresponding confusion matrices are provided in Table 3, Table 3, and Table 4.

The results show that the proposed method with the default parameters shows the lowest overall accuracy of 67.61% and the kappa coefficient of 0.60. It demonstrates that the input parameters of LS-SVM has a significant influence on classification performance. The proposed method with the empirical parameters achieves a much better performance compared with that using default parameters, which attains the overall accuracy of 86.20% and the kappa coefficient of 0.82. The proposed method with the optimized parameters attains the best classification performance. Its overall accuracy and kappa coefficient increase 2.37% and 0.03 respectively compared with that using empirical parameters. Moreover, the proposed method with the optimized parameters achieves a higher user accuracy and producer accuracy for almost all classes, indicating that the optimized input parameters contribute to an improved classification accuracy.

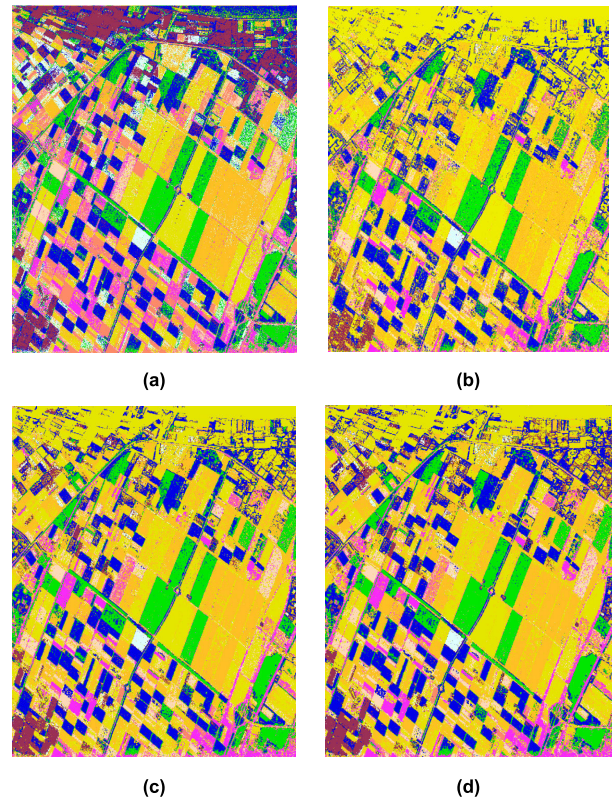


FIGURE 8. The classification results of L-image by (a) the supervised Wishart method, (b) the manifold-learning-based supervised graph embedding method, (c) the sparse representation-based method, and (d) the proposed method.

B. COMPARISONS TO THE THREE TYPICAL METHODS

In order to assess the performance of our method, we compare its performance with three other supervised approaches: a method based on the use of the complex Wishart distribution [6], a manifold-learning-based supervised graph embedding algorithm [60], and a method via sparse representation and polarimetric features [61]. The supervised Wishart classification is a pixel-based maximum likelihood classifier and is commonly used for polarimetric data classification [6], [31]. The manifold-learning-based supervised graph embedding algorithm uses a linear dimensionality reduction technology to obtain a low-dimensional subspace derived from polarimetric information, and classification is then implemented using such algorithm and neural network. The third comparison method is achieved on the basis of sparse representation-based techniques and simplified matching pursuit (SMP) algorithm.

The classification results of C-image obtained by the three comparison methods are shown in Fig. 7(a)(b)(c) and the corresponding confusion matrices are provided in Table 5, Table 6, and Table 7. The classification results of L-image obtained by the three comparison methods and the proposed method are shown in Fig. 8(a)(b)(c)(d) and the corresponding confusion matrices are provided in Table 8, Table 9, Table 10, and Table 11.

TABLE 5. Classification accuracy of the supervised Wishart method with respect to C-image.

Classified data	Reference data									Total	UA (%)
	BA	MA	PO	RA	LU	GR	PE	BE	FR		
BA	36394	6413	549	0	32	641	0	0	115	44144	82.44
MA	5599	34391	175	0	15	582	0	2	67	40831	84.23
PO	117	0	20437	22	0	8	1	0	37	20622	99.10
RA	107	0	160	17443	0	99	0	0	18	17827	97.85
LU	206	780	290	0	2707	728	0	0	370	5081	53.28
GR	1066	26	152	2	269	2327	0	0	1	3843	60.55
PE	1	2	3697	31	12	338	969	0	884	5934	16.33
BE	11	819	86	0	0	200	0	490	214	1820	26.92
FR	320	67	1692	5	60	306	0	0	4754	7204	65.99
Total	43821	42498	27238	17503	3095	5229	970	492	6460	147306	
PA (%)	83.05	80.92	75.03	99.66	87.46	44.50	99.90	99.59	73.59		
OA (%)	81.40										
Kappa	0.76										

TABLE 6. Classification accuracy of the manifold-learning-based supervised graph embedding method with respect to C-image.

Classified data	Reference data									Total	UA (%)
	BA	MA	PO	RA	LU	GR	PE	BE	FR		
BA	37671	4682	1215	161	254	246	0	38	546	44813	84.06
MA	6020	36549	152	93	0	578	0	122	113	43627	83.78
PO	8	16	23779	29	66	46	50	4	239	24237	98.11
RA	0	56	432	17201	0	67	0	17	0	17773	96.78
LU	2	19	86	1	2629	153	3	0	190	3083	85.27
GR	18	809	83	0	89	3341	0	51	408	4799	69.46
PE	0	0	499	12	1	126	899	0	46	1583	56.60
BE	3	274	34	0	7	154	0	260	19	751	34.62
FR	99	93	958	6	49	518	18	0	4899	6640	73.70
Total	43821	42498	27238	17503	3095	5229	970	492	6460	147306	
PA (%)	85.97	86.00	87.29	98.28	84.94	63.72	92.63	52.85	75.76		
OA (%)	86.36										
Kappa	0.82										

TABLE 7. Classification accuracy of the sparse representation-based method with respect to C-image.

Classified data	Reference data									Total	UA (%)
	BA	MA	PO	RA	LU	GR	PE	BE	FR		
BA	37979	4689	1304	2	178	260	1	25	455	44893	84.60
MA	5742	36831	149	14	2	666	0	71	143	43618	84.43
PO	7	10	23592	14	46	5	56	2	230	23962	98.45
RA	0	122	305	17439	0	83	0	4	0	17953	97.14
LU	3	13	61	0	2698	159	6	0	162	3102	86.97
GR	8	548	93	0	87	3190	0	86	348	4360	72.87
PE	0	0	779	21	19	75	887	10	34	1825	48.21
BE	3	180	83	8	21	226	0	294	17	802	36.66
FR	79	105	902	5	44	565	20	0	5071	6791	74.52
Total	43821	42498	27238	17503	3095	5229	970	492	6460	147306	
PA (%)	86.67	86.67	86.60	99.93	87.17	60.65	91.32	59.76	78.36		
OA (%)	86.87										
Kappa	0.83										

The results show that the proposed method performs better in the two tested examples than the other three methods. Comparing Table 4 with Table 5, Table 6, and Table 7, for C-image, the overall accuracy of the proposed method is 88.57%, whereas that of the supervised Wishart method, the manifold-learning-based supervised graph embedding algorithm, and the sparse representation-based method are 81.40%, 86.36% and 86.87%, respectively. Besides, the proposed method exhibits a kappa coefficient of 0.85, which is higher than that of other three methods, which are 0.76, 0.82 and 0.83,

respectively. For L-image, the proposed method also exhibits a higher overall accuracy and kappa coefficient compared with other three methods. From Table 8, Table 15, Table 10, and Table 11 we can see that the supervised Wishart method, the manifold-learning-based supervised graph embedding algorithm, the sparse representation-based method, and the proposed method achieve the overall accuracy of 73.14%, 76.96%, 78.90% and 81.13%, respectively, and achieve the kappa coefficient of 0.65, 0.70, 0.73, and 0.76, respectively. Comparing the classification of C-image with the

TABLE 8. Classification accuracy of the supervised Wishart method with respect to L-image.

Classified data	Reference data										Total	UA (%)
	BA	MA	PO	RA	LU	GR	PE	BE	FR			
BA	35148	9401	7468	505	1312	893	9	105	2100	56941	61.73	
MA	7999	32201	264	48	6	362	1	7	27	40915	78.70	
PO	21	100	17374	3627	122	113	19	138	163	21677	80.15	
RA	33	10	1379	13215	0	295	0	0	31	14963	88.32	
LU	535	286	279	16	1572	178	0	113	15	2994	52.51	
GR	61	69	137	75	78	3196	0	34	46	3696	86.47	
PE	1	165	6	0	0	1	931	4	57	1165	79.91	
BE	12	41	31	4	4	94	1	91	0	278	32.73	
FR	11	225	300	13	1	97	9	0	4021	4677	85.97	
Total	43821	42498	27238	17503	3095	5229	970	492	6460	147306		
PA (%)	80.21	75.77	63.79	75.50	50.79	61.12	95.98	18.50	62.24			
OA (%)	73.14											
Kappa	0.65											

TABLE 9. Classification accuracy of the manifold-learning-based supervised graph embedding method with respect to L-image.

Classified data	Reference data										Total	UA (%)
	BA	MA	PO	RA	LU	GR	PE	BE	FR			
BA	34116	8628	2533	18	1190	609	8	112	384	47598	71.68	
MA	8951	33018	304	37	22	455	4	0	32	42823	77.10	
PO	4	69	20118	3217	12	47	2	95	153	23717	84.82	
RA	46	44	1620	14156	0	363	0	0	46	16275	86.70	
LU	407	356	342	2	1780	138	0	155	16	3196	55.70	
GR	109	133	247	66	91	3483	0	8	56	4193	83.05	
PE	0	167	1	0	0	2	949	50	88	1257	75.48	
BE	188	25	29	4	0	63	1	72	0	382	18.85	
FR	0	58	2044	3	0	69	6	0	5685	7865	72.76	
Total	43821	42498	27238	17503	3095	5229	970	492	6460	147306		
PA (%)	77.85	77.69	73.85	80.88	57.51	66.58	97.83	14.63	88.00			
OA (%)	76.96											
Kappa	0.70											

TABLE 10. Classification accuracy of the sparse representation-based method with respect to L-image.

Classified data	Reference data										Total	UA (%)
	BA	MA	PO	RA	LU	GR	PE	BE	FR			
BA	34534	8231	1536	3	913	645	8	39	182	46091	74.93	
MA	8785	33579	310	50	19	463	5	0	36	43247	77.64	
PO	1	66	21394	3109	7	49	8	147	152	24933	85.79	
RA	21	83	1503	14253	0	312	0	0	26	16198	87.99	
LU	351	239	365	4	2066	129	0	177	24	3355	61.58	
GR	100	96	231	60	90	3522	0	11	51	4161	84.63	
PE	0	141	8	9	0	0	941	56	106	1259	74.58	
BE	29	23	40	14	0	56	1	62	0	225	27.56	
FR	0	40	1851	1	0	53	7	0	5883	7835	75.05	
Total	43821	42498	27238	17503	3095	5229	970	492	6460	147306		
PA (%)	78.81	79.01	78.52	81.43	66.75	67.32	97.00	12.60	91.05			
OA (%)	78.90											
Kappa	0.73											

classification of L-image using each method, it is interesting to observe that both overall accuracy and kappa coefficient decrease even though the used classification method and the training data are exactly the same. This can be attributed to the fact that the polarimetric features of some crops obtained at C band are more distinguishable compared with that obtained at L band. This verifies the statement that a SAR image acquired by a single band is found to be insufficient to discriminate

some different ground objects which are of similar polarimetric characteristics at such band [62].

Each of the different features has a different influence on the classification result. Additional comparisons are made to explore the contributions of the elements of the coherency matrix, texture information, and polarimetric features. Furthermore, compared with the complex Wishart distribution-based classifier, the classification capability of

TABLE 11. Classification accuracy of the proposed method with respect to L-image.

Classified data	Reference data										Total	UA (%)
	BA	MA	PO	RA	LU	GR	PE	BE	FR			
BA	38887	5247	35	2	213	369	1	1	48	44803	86.80	
MA	2015	33007	120	13	29	579	14	0	26	35803	92.22	
PO	66	118	21533	2976	0	46	2	0	754	25495	84.45	
RA	9	174	1343	14263	1	777	5	0	79	16651	85.66	
LU	1772	791	32	0	2568	349	0	3	28	5543	46.33	
GR	254	392	482	182	214	2729	0	4	56	4313	63.27	
PE	735	2739	141	33	1	195	947	0	361	5152	18.38	
BE	1	17	997	0	69	156	0	484	16	1740	27.82	
FR	82	13	2555	34	0	29	1	0	5092	7806	65.23	
Total	43821	42498	27238	17503	3095	5229	970	492	6460	147306		
PA (%)	88.74	77.67	79.06	81.49	82.97	52.19	97.76	98.37	78.82			
OA (%)	81.13											
Kappa	0.76											

the LS-SVM-based classifier combining KPCA and FOA is tested as well.

C. CONTRIBUTIONS OF THE FEATURES TO CLASSIFICATION PERFORMANCES

1) CONTRIBUTION OF THE THREE ELEMENTS OF THE COHERENCY MATRIX

For some cases, the supervised Wishart method based on coherency matrix can attain a satisfactory result, which demonstrates that the elements of the coherency matrix have significant implications for SAR image classification. The proposed method without using the three elements T_{11} , T_{22} , T_{33} is used to investigate the contribution of such elements. The same method with the empirical input parameters ($\gamma = 10$, $\sigma^2 = 0.25$) is also employed in the comparison. Their classification results of C-image are shown in Fig. 9(a) and Fig. 9(b), and the corresponding confusion matrices are shown in Table 12 and Table 13.

Compared with Table 3 and Table 4, the overall accuracies for the two methods (from Table 12 and Table 13) decrease by 0.24% and 1.88%, respectively, and the kappa coefficients for the two methods decrease by 0 and 0.02, respectively. The performance of the proposed method without using the three elements is almost unchanged while that of the method with empirical parameters without using the elements is reduced. It indicates that with the decrease of types of input features, the classification capability of the proposed method without using FOA is reduced more than that using FOA. In other words, the classification accuracy is more influenced by the input parameters of the classifier when the types of input features decrease and thus the optimized input parameters are more critical for improving the accuracy under such case. This result agrees with the fact that when the input parameters of the classifier are inappropriate, more features or samples are required for improving classification accuracy.

It is observed from Table 5 and Table 12 that the proposed method still can achieve a better performance by using polarimetric characteristics and texture information, where the overall accuracy and kappa coefficient by the proposed

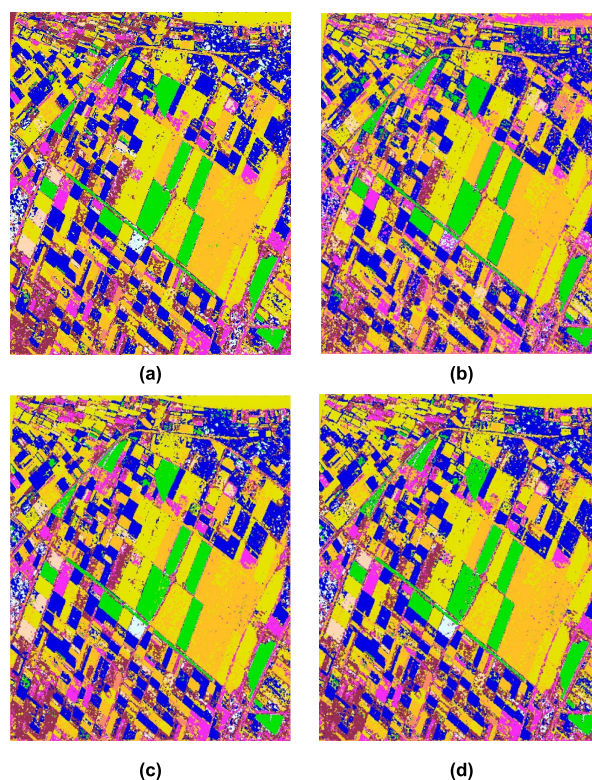


FIGURE 9. The classification results by (a) the proposed method without using the elements of the coherency matrix, (b) the proposed method with empirical parameters without using the elements of the coherency matrix, (c) the proposed method without using the texture information, and (d) the proposed method using only the elements of the coherency matrix.

method are greater than 6.93% and 0.09 respectively compared with that by the supervised Wishart method. However, although the performance of the supervised Wishart method is the worst among the three methods, the producer accuracies for some certain classes, such as beans, are the highest. The probable reason accounting for this fact is that the polarimetric characteristics and the texture features that we used in this study are extremely similar between beans and some other crops, resulting in misclassifying beans to other classes. Furthermore, it can be seen from the three confusion matrices

TABLE 12. Classification accuracy of C-image by the proposed method without using the elements of the coherency matrix.

Classified data	Reference data										Total	UA (%)
	BA	MA	PO	RA	LU	GR	PE	BE	FR			
BA	38451	4392	1398	1	112	231	0	0	340	44925	85.59	
MA	5148	37498	194	0	0	584	0	47	136	43604	85.99	
PO	33	20	24392	9	284	1	49	0	253	25041	97.41	
RA	0	133	90	17470	0	64	0	9	1	17767	98.33	
LU	7	16	50	0	2400	79	10	2	177	2741	87.56	
GR	4	254	45	0	40	3401	0	20	416	4150	81.95	
PE	0	0	300	14	159	2	874	27	7	1383	63.20	
BE	0	82	11	1	59	223	0	387	20	783	49.43	
FR	178	103	758	8	41	674	37	0	5100	6909	73.82	
Total	43821	42498	27238	17503	3095	5229	970	492	6460	147306		
PA (%)	87.75	88.24	89.55	99.81	77.54	65.04	90.10	78.66	78.95			
OA (%)	88.33											
Kappa	0.85											

TABLE 13. Classification accuracy of C-image by the method with empirical parameters without using the elements of the coherency matrix.

Classified data	Reference data										Total	UA (%)
	BA	MA	PO	RA	LU	GR	PE	BE	FR			
BA	37711	6561	933	535	411	349	1	163	504	49068	76.86	
MA	5853	33892	165	0	0	398	0	69	104	43481	84.85	
PO	207	120	24344	128	231	52	29	42	289	25443	95.68	
RA	1	149	281	16820	0	71	0	39	2	17363	96.87	
LU	7	23	42	1	1956	74	6	0	161	2270	86.17	
GR	17	303	449	0	450	3460	303	6	508	4696	73.68	
PE	0	1	159	10	3	25	604	1	7	810	74.57	
BE	0	340	10	0	0	206	0	164	20	740	22.16	
FR	25	109	855	9	44	594	27	7	4865	6535	74.45	
Total	43821	42498	27238	17503	3095	5229	970	492	6460	147306		
PA (%)	86.06	79.75	89.38	96.10	63.20	66.17	62.27	33.33	75.31			
OA (%)	84.32											
Kappa	0.80											

TABLE 14. Classification accuracy of C-image by the proposed method without using the texture information.

Classified data	Reference data										Total	UA (%)
	BA	MA	PO	RA	LU	GR	PE	BE	FR			
BA	37879	4467	954	74	303	272	2	123	370	44444	85.23	
MA	5731	37144	196	0	0	352	0	93	99	43615	85.16	
PO	155	109	24774	47	197	5	18	37	275	25617	96.71	
RA	1	56	145	17359	0	85	0	19	1	17666	98.26	
LU	8	21	46	1	2499	69	3	2	137	2786	89.69	
GR	18	384	53	0	39	3629	2	11	492	4628	78.41	
PE	0	0	192	13	7	30	918	4	16	1180	77.80	
BE	0	214	3	0	0	195	1	170	18	601	28.29	
FR	29	103	875	9	50	592	26	33	5052	6769	74.63	
Total	43821	42498	27238	17503	3095	5229	970	492	6460	147306		
PA (%)	86.44	87.40	90.95	99.18	80.73	69.40	94.64	34.55	78.20			
OA (%)	87.95											
Kappa	0.84											

(Table 5, Table 12, and Table 13) that less training data is more likely to lead to lower user accuracy or producer accuracy, because it is difficult to estimate accurate characteristics of ground objects from a limited number of samples.

2) CONTRIBUTION OF THE TEXTURE INFORMATION

The classification accuracy of C-image obtained by the proposed method without using the texture information is shown

in Table 14 and its classification result is shown in Fig. 9(c). Compared with the result in Table 4, the overall accuracy and the kappa coefficient increase by 0.62% and 0.01 when the texture information is used in the method, which demonstrates that texture information has a slight effect on enhancing the performance of the classification. The disadvantage of texture information is that the texture features of some pixels located around the boundary between two different

TABLE 15. Classification accuracy of C-image by the proposed method only using the elements of the coherency matrix.

Classified data	Reference data										Total	UA (%)
	BA	MA	PO	RA	LU	GR	PE	BE	FR			
BA	37541	5661	1601	4	484	422	13	0	971	46697	80.39	
MA	5862	35872	153	2	13	843	0	159	144	43048	83.33	
PO	271	2	23056	132	1	30	153	0	298	23943	96.30	
RA	0	141	897	17332	0	69	0	47	0	18486	93.76	
LU	14	74	355	4	2192	808	54	0	1340	4841	45.28	
GR	70	429	211	15	45	2152	39	0	1056	4017	53.57	
PE	0	0	453	10	0	220	711	0	49	1443	49.27	
BE	2	316	49	0	0	195	0	286	25	853	33.53	
FR	61	3	483	4	360	490	0	0	2557	3958	64.60	
Total	43821	42498	27238	17503	3095	5229	970	492	6460	147306		
PA (%)	85.67	84.41	84.65	99.02	70.82	41.16	73.30	58.13	39.58			
OA (%)	82.74											
Kappa	0.78											

classes are considered as incorrect features, resulting in the inaccurate estimation of the texture features for the classes. The negative influence of this fact can be reduced or neglected by choosing a proper processing window size in terms of training data. It is also noted that the producer accuracy and the user accuracy of the beans class decreases by 50.21% and 19.75% respectively without using texture information, which indicates that texture information is helpful to distinguish some certain ground objects from other objects if these objects are of the similar polarimetric characteristics.

3) THE CLASSIFICATION CAPABILITY BETWEEN THE PROPOSED CLASSIFIER AND THE SUPERVISED WISHART CLASSIFIER, AND CONTRIBUTION OF THE POLARIMETRIC PARAMETERS

The LS-SVM classifier combining KPCA and FOA is considered as the proposed classifier in this study. To test the classification capability between the proposed classifier and the commonly used supervised Wishart classifier, the input feature types and the training data have to be the same. Since the supervised Wishart method is based on the elements of the coherency matrix, the proposed method only using the three elements of the coherency matrix (T_{11} , T_{22} , T_{33}) is implemented. The classification accuracy of C-image obtained by the proposed method only using the elements of the coherency matrix is shown in Table 15 and its classification result is shown in Fig. 9(d). Compared to the result from Table 5, the overall accuracy and the kappa coefficient are greater than 1.32% and 0.02 respectively by the proposed classifier. The result demonstrates that the classification capability of the proposed classifier is better than that of the supervised Wishart classifier. It is seen that the producer accuracies of the peas class and the beans class by the proposed classifier is much lower than that by the supervised Wishart classifier, which exhibit the difference values of 26.60% and 41.46% respectively. A possible reason for this is the fact that the estimation of the data distribution from the training data is essential in LS-SVM, whereas the number of the training data is very few for the peas class and the beans class, leading to

the difficulties in estimating the training data distribution of such classes. On the contrary, the supervised Wishart classifier is based on the assumption that the data follows complex Wishart distribution. Therefore, we can infer that the supervised Wishart method probably achieves a better classification if there is very little training data for each class.

Additionally, because it has been shown in a previous section that the texture information has a minor effect on the result for this study image, the contribution of the texture information can be ignored for the experimental image so that this method can roughly investigate the contribution of the polarimetric parameters as well. By comparing with Table 4, the overall accuracy and the kappa coefficient decreases by 5.83% and 0.07 respectively by the proposed method without using polarimetric parameters. Besides, the producer accuracies and user accuracies for all classes increase sharply when the polarimetric parameters are used in the classification. The comparison shows that the polarimetric parameters have a significant effect on the classification accuracy.

The experimental results show that the performance of the proposed method is better. The various comparisons prove that different features have a varying degree of influence on the classification result, and a small number of training data for each class probably brings about low classification accuracy by using this method. Therefore, some considerable ways to improve the classification accuracy include the employment of various features and the usage of as many samples as possible. In addition, the experimental result shows that a SAR image acquired by a single band is probably not able to discriminate some different ground objects, the use of multi-frequency sensors to acquire data is also likely to improve the accuracy.

IV. CONCLUSION

This paper proposes a classification method based on multi-feature using FOA and LS-SVM. The basic features for a SAR image are polarimetric parameters. As noted in [15], different polarimetric decomposition methods focus on different land cover type, therefore the use of as many various polarimetric parameters as possible is considered to enhance

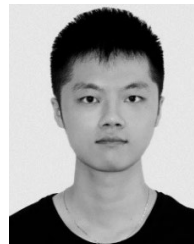
the classification performance. In this study, we use three elements from coherency matrix and six parameters obtained by $H/\alpha/A$ decomposition and Freeman decomposition, as well as the total receive power (SPAN), polarimetric pedestal height, and (RVI) as the features of land cover. In addition to pixel-based information from polarimetric parameters, region-based information is another significant feature for an image. GLCM is an efficient tool to obtain region-based information. Thus, eight representative textural parameters obtained by GLCM, which are CON, ASM, COR, ENT, IDM, MAX, DIS, and INV are introduced to obtain texture information. However, while the integration of multi-feature characterizes the ground object better, it results in the curse of dimensionality and feature redundancy. Zou et al. have demonstrated that the overall accuracy decreases if all features are used instead of a selected feature set [63]. Therefore, we adopt KPCA to reduce the dimensionality of data. KPCA is a nonlinear-based algorithm and thus is able to cope with non-Gaussian data set. The used classifier in this study is LS-SVM. Some studies have proven that the performance of SAR image classification by SVM is better than that by other classifiers [38]. Since the disadvantage of SVM is its long execution time, LS-SVM is employed in this method. Compared with SVM, LS-SVM significantly decreases the execution time and the required computational resources. However, the critical problem of LS-SVM is that the input parameters have noticeable influence on classification performance while such parameters are selected by the past experiences in most cases [64], [65]. FOA is thus utilized to obtain the optimized input parameters. Finally, an experiment on a full-polarization SAR image is carried out to test the performance of the proposed method.

Some experiments are carried out on the two fully polSAR images by various methods, as well as that the contributions of each of the different features are tested. The analysis on the experimental results offers some feasible approaches to enhance the classification performance. First, we recommend to using as many various polarimetric features as possible in the method if the selection of features for the classification of interest land covers is not known. Second, it is necessary to gain enough number of samples for each class. Third, the acquisition of data set by multi-frequency sensors contributes to improved classification accuracy.

REFERENCES

- [1] S. S. Saatchi, J. V. Soares, and D. S. Alves, "Mapping deforestation and land use in Amazon rainforest by using SIR-C imagery," *Remote Sens. Environ.*, vol. 59, no. 2, pp. 191–202, 1997.
- [2] L. E. Pierce, K. M. Bergen, M. C. Dobson, and F. T. Ulaby, "Multitemporal land-cover classification using SIR-C/X-SAR imagery," *Remote Sens. Environ.*, vol. 64, no. 1, pp. 20–33, 1998.
- [3] F. T. Ulaby, F. Kouyate, B. Brisco, and T. H. L. Williams, "Textural information in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. GRS-24, no. 2, pp. 235–245, Mar. 1986.
- [4] J.-S. Lee, M. R. Grunes, and E. Pottier, "Quantitative comparison of classification capability: Fully polarimetric versus dual and single-polarization SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 11, pp. 2343–2351, Nov. 2001.
- [5] S. Y. Luo, L. Tong, and Y. Chen, "A multi-region segmentation method for SAR images based on the multi-texture model with level sets," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2560–2574, May 2018.
- [6] J. S. Lee, M. R. Grunes, and R. Kwok, "Classification of multi-look polarimetric SAR imagery based on complex Wishart distribution," *Int. J. Remote Sens.*, vol. 15, no. 11, pp. 2299–2311, Jul. 1994.
- [7] L. E. Pierce, F. T. Ulaby, K. Sarabandi, and M. C. Dobson, "Knowledge-based classification of polarimetric SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 5, pp. 1081–1086, Sep. 1994.
- [8] L. Du and J. S. Lee, "Fuzzy classification of earth terrain covers using complex polarimetric SAR data," *Int. J. Remote Sens.*, vol. 17, no. 4, pp. 809–826, 1996.
- [9] 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.
- [10] 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.
- [11] E. Krogager and Z. H. Czyz, "Properties of the sphere, diplane, helix (target scattering matrix decomposition)," in *Proc. JIPR*, Mar. 1995, pp. 21–23.
- [12] 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.
- [13] J. J. van Zyl, "Application of Cloude's target decomposition theorem to polarimetric imaging radar data," *Proc. SPIE*, vol. 1748, pp. 184–191, Feb. 1993.
- [14] E. Krogager, "New decomposition of the radar target scattering matrix," *Electron. Lett.*, vol. 26, no. 18, pp. 1525–1527, Aug. 1990.
- [15] M. Shimoni, D. Borghys, R. Heremans, C. Perneel, and M. Acheroy, "Fusion of PolSAR and PolInSAR data for land cover classification," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 11, no. 3, pp. 169–180, Jun. 2009.
- [16] J. Ni, F. Zhang, Q. Yin, and H.-C. Li, "Robust weighting nearest regularized subspace classifier for PolSAR imagery," *IEEE Signal Process. Lett.*, vol. 26, no. 10, pp. 1496–1500, Oct. 2019.
- [17] C. F. Barnes and J. Burki, "Late-season rural land-cover estimation with polarimetric-SAR intensity pixel blocks and σ -tree-structured neighbor classifiers," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 9, pp. 2384–2392, Sep. 2006.
- [18] S. Luo, K. Sarabandi, L. Tong, and S. Guo, "Unsupervised multiregion partitioning of fully polarimetric SAR images with advanced fuzzy active contours," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2019.2947376.
- [19] 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, 2017.
- [20] L. S. Davis, S. A. Johns, and J. Aggarwal, "Texture analysis using generalized co-occurrence matrices," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 3, pp. 251–259, Mar. 1979.
- [21] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.
- [22] D.-C. He and L. Wang, "Texture unit, texture spectrum, and texture analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 28, no. 4, pp. 509–512, Jul. 1990.
- [23] R. E. Bellman, *Adaptive Control Processes: A Guided Tour*. Princeton, NJ, USA: Princeton Univ. Press, 1961.
- [24] I. Jolliffe, *Principal Component Analysis*. Hoboken, NJ, USA: Wiley, 2002.
- [25] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. Hoboken, NJ, USA: Wiley, 2004.
- [26] A. J. Izenman, "Linear discriminant analysis," in *Modern Multivariate Statistical Techniques*. Heidelberg, Germany: Springer, 2008, pp. 237–280.
- [27] B. Schölkopf, A. Smola, and K.-R. Müller, "Kernel principal component analysis," in *Artificial Neural Networks—ICANN*. Berlin, Germany: Springer, 1997, pp. 583–588.
- [28] F. R. Bach and M. I. Jordan, "Kernel independent component analysis," *J. Mach. Learn. Res.*, vol. 3, pp. 1–48, Jan. 2002.
- [29] V. Roth and V. Steinhage, "Nonlinear discriminant analysis using kernel functions," in *Proc. Adv. Neural Inf. Process. Syst.*, 1999, pp. 568–574.
- [30] J. Li, S. Zhou, and C. Shekhar, "A comparison of subspace analysis for face recognition," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 3, Apr. 2003, p. III-121.

- [31] 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.
- [32] L. Ferro-Famil, E. Pottier, and J.-S. Lee, "Unsupervised classification of multifrequency and fully polarimetric SAR images based on the H/A/Alpha-wishart classifier," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 11, pp. 2332–2342, Nov. 2001.
- [33] K. Kimura, Y. Yamaguchi, and H. Yamada, "Unsupervised land cover classification using H/TP space applied to POLSAR image analysis," *IEICE Trans. Commun.*, vol. 87, no. 6, pp. 1639–1647, 2004.
- [34] C.-T. Chen, K.-S. Chen, and J.-S. Lee, "The use of fully polarimetric information for the fuzzy neural classification of SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 2089–2100, Sep. 2003.
- [35] S. Luo, K. Sarabandi, L. Tong, and S. Guo, "An improved fuzzy region competition-based framework for the multiphase segmentation of SAR images," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2019.2949742.
- [36] Y. Hara, R. G. Atkins, S. H. Yueh, R. T. Shin, and J. A. Kong, "Application of neural networks to radar image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 1, pp. 100–109, Jan. 1994.
- [37] M. A. Hearst, S. T. Dumais, E. Osman, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intell. Syst. Appl.*, vol. 13, no. 4, pp. 18–28, Jul./Aug. 2008.
- [38] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: A review," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 3, pp. 247–259, 2011.
- [39] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Process. Lett.*, vol. 9, no. 3, pp. 293–300, Jun. 1999.
- [40] V. Cherkassky and Y. Ma, "Practical selection of SVM parameters and noise estimation for SVM regression," *Neural Netw.*, vol. 17, no. 1, pp. 113–126, Jan. 2004.
- [41] M. Chi, R. Feng, and L. Bruzzone, "Classification of hyperspectral remote-sensing data with primal SVM for small-sized training dataset problem," *Adv. Space Res.*, vol. 41, no. 11, pp. 1793–1799, 2008.
- [42] M. Dorigo and L. M. Gambardella, "Ant colony system: A cooperative learning approach to the traveling salesman problem," *IEEE Trans. Evol. Comput.*, vol. 1, no. 1, pp. 53–66, Apr. 1997.
- [43] X.-L. Li, "An optimizing method based on autonomous animats: Fish-swarm algorithm," *Syst. Eng.-Theory Pract.*, vol. 22, no. 11, pp. 32–38, 2002.
- [44] Y.-C. Lin and K. Sarabandi, "A Monte Carlo coherent scattering model for forest canopies using fractal-generated trees," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 1, pp. 440–451, Jan. 1999.
- [45] J. Kennedy, "Particle swarm optimization," in *Encyclopedia of Machine Learning*. Heidelberg, Germany: Springer, 2010, pp. 760–766.
- [46] A. K. Qin, V. L. Huang, and P. N. Suganthan, "Differential evolution algorithm with strategy adaptation for global numerical optimization," *IEEE Trans. Evol. Comput.*, vol. 13, no. 2, pp. 398–417, Apr. 2009.
- [47] W. T. Pan, "Fruit fly optimization algorithm," in *The Latest Evolutionary Computing Technology*. Taipei, China: Tsang Hai Book Publishing Co., 2011, pp. 10–12.
- [48] J.-S. Lee, M. R. Grunes, and G. de Grandi, "Polarimetric SAR speckle filtering and its implication for classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 5, pp. 2363–2373, Sep. 1999.
- [49] 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.
- [50] G. Singh and G. Venkataraman, "Application of incoherent target decomposition theorems to classify snow cover over the Himalayan region," (in English), *Int. J. Remote Sens.*, vol. 33, no. 13, pp. 4161–4177, 2012.
- [51] A. H. S. Solberg and A. K. Jain, "Texture fusion and feature selection applied to SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 2, pp. 475–479, Mar. 1997.
- [52] U. Kandaswamy, D. A. Adjeroh, and M. C. Lee, "Efficient texture analysis of SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 9, pp. 2075–2083, Sep. 2005.
- [53] V. Spruyt, "The Curse of Dimensionality in classification," *Comput. Vis. Dummies*, vol. 21, no. 3, pp. 35–40, 2014.
- [54] J.-Q. Wan, Y. Wang, and Y. Liu, "Improvement of KPCA on feature extraction of classification data," *Comput. Eng. Des.*, vol. 31, no. 18, pp. 4085–4087, 2010.
- [55] M. A. Alam and K. Fukumizu, "Hyperparameter selection in kernel principal component analysis," *J. Comput. Sci.*, vol. 10, no. 7, pp. 1139–1150, 2014.
- [56] G. Valentini and T. G. Dietterich, "Bias-variance analysis of support vector machines for the development of SVM-based ensemble methods," *J. Mach. Learn. Res.*, vol. 5, no. 3, pp. 725–775, 2004.
- [57] A. Kocsor and L. Tóth, "Kernel-based feature extraction with a speech technology application," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2250–2263, Aug. 2004.
- [58] S. Mika, B. Schölkopf, A. J. Smola, K.-R. Müller, M. Scholz, and G. Rätsch, "Kernel PCA and de-noising in feature spaces," in *Proc. Adv. Neural Inf. Process. Syst.*, 1999, pp. 536–542.
- [59] J. Wang, Q. Chen, and Y. Chen, "RBF kernel based support vector machine with universal approximation and its application," in *Advances in Neural Networks—ISNN*. Berlin, Germany: Springer, 2004, pp. 512–517.
- [60] L. Shi, L. Zhang, J. Yang, L. Zhang, and P. Li, "Supervised graph embedding for polarimetric SAR image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 2, pp. 216–220, Mar. 2013.
- [61] L. Zhang, L. Sun, B. Zou, and W. M. Moon, "Fully polarimetric SAR image classification via sparse representation and polarimetric features," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 8, pp. 3923–3932, Aug. 2015.
- [62] V. Turkar, R. Deo, Y. S. Rao, S. Mohan, and A. Das, "Classification accuracy of multi-frequency and multi-polarization SAR images for various land covers," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 3, pp. 936–941, Jun. 2012.
- [63] T. Zou, W. Yang, D. Dai, and H. Sun, "Polarimetric SAR image classification using multifeatures combination and extremely randomized clustering forests," *EURASIP J. Adv. Signal Process.*, vol. 2010, Jan. 2010, Art. no. 4.
- [64] B. Waske and J. A. Benediktsson, "Fusion of support vector machines for classification of multisensor data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 3858–3866, Dec. 2007.
- [65] B. Waske, G. Menz, and J. A. Benediktsson, "Fusion of support vector machines for classifying SAR and multispectral imagery from agricultural areas," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2007, pp. 4842–4845.



SHIYU LUO (S'16–M'19) received the B.S. degree in technique and instrumentation of measurements and the Ph.D. degree in instrument science and technology from the School of Automation Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2010 and 2017, respectively.

He was a Visiting Scholar, from 2014 to 2016 and a Research Assistant, from 2016 to 2018 with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor. Since 2019, he has been a Postdoctoral Fellow with the School of Automation Engineering, UESTC. His research interests include synthetic aperture radar data processing, synthetic aperture radar image segmentation and classification, and information retrieval from SAR images.



KAMAL SARABANDI (S'87–M'90–SM'92–F'00) received the B.S. degree in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 1980, the M.S. degree in electrical engineering and the M.S. degree in mathematics from the University of Michigan, Ann Arbor, MI, USA, in 1986 and 1989, respectively, and the Ph.D. degree in electrical engineering from the University of Michigan, in 1989.

He was the Rufus S. Teesdale Endowed Professor of engineering with the Department of Electrical Engineering and Computer Science, University of Michigan, where he is currently the Director of the Radiation Laboratory. He led the Center for Microelectronics and Sensors, University of Michigan, supported by the Army Research Laboratory under the Micro-Autonomous Systems and Technology Collaborative Technology Alliance Program, from 2008 to 2018. He is also leading the Microwave Sensor Technology Center, University of Michigan. He has authored many book chapters and over 280 articles in refereed journals on miniaturized and on-chip antennas, meta materials, electromagnetic scattering, wireless channel modeling, random media modeling, microwave measurement techniques, radar calibration, inverse scattering problems, and microwave sensors. He has also had more than 650 articles and invited presentations in many national and international conferences and symposia on similar subjects. His research interests include microwave- and millimeter-wave radar remote sensing, meta-materials, electromagnetic wave propagation, and antenna miniaturization.

Dr. Sarabandi served as a member for the NASA Advisory Council appointed by the NASA Administrator for two consecutive terms, from 2006 to 2010. He served as the President for the IEEE Geoscience and Remote Sensing Society (GRSS), in 2015 and 2016. He was a member of the Editorial Board of the PROCEEDINGS OF THE IEEE. He is a member of Commissions F and B of URSI. He was a recipient of the Henry Russel Award from the Regent of The University of Michigan, in 1997; the GAAC Distinguished Lecturer Award from the German Federal Ministry for Education, Science, and Technology, in 1999; the 1996 EECS Department Teaching Excellence Award; the IEEE GRSS Distinguished Achievement Award and the University of Michigan Faculty Recognition Award, in 2005; the Best Paper Award at the 2006 Army Science Conference; the IEEE GRSS Symposium Best Paper Award, in 2008 and 2017; the Humboldt Research Award from The Alexander von Humboldt Foundation of Germany, in 2008; the 2010 Distinguished Faculty Achievement Award from the University of Michigan; the 2011 IEEE Judith A. Resnik Award by the IEEE Board of Directors; the IEEE GRSS with its 2013 Education Award; the College of Engineering Research Excellence Award, in 2004; the Stephen S. Attwood Award, in 2017; the Ted Kennedy Family Faculty Team Excellence Award, in 2018 from the College of Engineering, University of Michigan; and the NASA Group Achievement Award for his contributions to NASA SMAP Mission. In 2016, he was recognized as one of the top 50 Graduates of Sharif University of Technology. In the past several years, joint articles presented by his students at a number of international symposia (IEEE APS, in 1995, 1997, 2000, 2001, 2003, 2005, 2006, and 2007; 2016 IEEE IGARSS, in 1999, 2002, 2007, 2011, and 2014; IEEE IMS, in 2001; USNC URSI, in 2004, 2005, 2006, 2010, and 2011; AMTA, in 2006; URSI GA, in 2008 and 2014; and the Eastern Snow Conference, in 2016) have received best paper awards. He is serving as the Chair of the USNC URSI Commission F. He was an Associate Editor of the IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION and the IEEE SENSORS JOURNAL.



LING TONG (M'06) received the B.S. and M.S. degrees from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 1985 and 1988, respectively.

From 1988 to 1995, she was involved in researches with the National Institute of Measurement and Testing Technology, Chengdu. She is currently the Director of the Microwave Remote Sensing and Testing Laboratory, UESTC, where she is also a Professor (second grade) with the School of Automatic Engineering. She is also leading the Earth Observation Engineering Research Center of Sichuan Province, China, UESTC. Her research interests include mechanism of remote sensing, microwave remote sensing, data analysis and processing, and metrical information theory.



LELAND PIERCE (S'85–M'85–SM'01) received the B.S. degree in electrical and aerospace engineering and the M.S. and Ph.D. degrees in electrical engineering from the University of Michigan, Ann Arbor, MI, USA, in 1983, 1986, and 1991, respectively.

Since then, he has been the Head of the Microwave Image Processing Facility, Radiation Laboratory, Electrical Engineering and Computer Science Department, University of Michigan, where he is currently responsible for research into the uses of polarimetric SAR systems and other modalities for remote sensing applications, especially forest canopy parameter inversion.

• • •