*Research Article*

# Polarimetric SAR Image Classification Using Multifeatures Combination and Extremely Randomized Clustering Forests

## Tongyuan Zou,[1] Wen Yang,[1, 2] Dengxin Dai,[1] and Hong Sun[1]

[1] *Signal Processing Lab, School of Electronic Information, Wuhan University, Wuhan 430079, China*
[2] *Laboratoire Jean Kuntzmann, CNRS-INRIA, Grenoble University, 51 rue des Mathématiques, 38041 Grenoble, France*

Correspondence should be addressed to Wen Yang, yangwen@whu.edu.cn

Terrain classification using polarimetric SAR imagery has been a very active research field over recent years. Although lots of features have been proposed and many classifiers have been employed, there are few works on comparing these features and their combination with different classifiers. In this paper, we firstly evaluate and compare different features for classifying polarimetric SAR imagery. Then, we propose two strategies for feature combination: manual selection according to heuristic rules and automatic combination based on a simple but efficient criterion. Finally, we introduce extremely randomized clustering forests (ERCFs) to polarimetric SAR image classification and compare it with other competitive classifiers. Experiments on ALOS PALSAR image validate the effectiveness of the feature combination strategies and also show that ERCFs achieves competitive performance with other widely used classifiers while costing much less training and testing time.

## 1. Introduction

Terrain classification is one of the most important applications of PolSAR remote sensing which can provide more information than conventional radar images and thus greatly improve the ability to discriminate different terrain types. During last two decades, many algorithms have been proposed for PolSAR image classification. The efforts mainly focus on the following two areas: one is mainly on developing new polarimetric descriptor based on statistical properties and scattering mechanisms; the other is to employ some advanced classifiers originated from machine learning and pattern recognition domain.

In the earlier years, most works were focused on the statistical properties of PolSAR data. Kong et al. [1] proposed a distance measure based on the complex Gaussian distribution for single-look polarimetric SAR data and used it in maximum likelihood (ML) classification framework. Lee et al. [2] derived a distance measure based on complex Wishart distribution for multilook polarimetric SAR data. With the progress of research on scattering mechanism, many unsupervised algorithms have been proposed. In [3], van Zyl proposed to classify terrain types as odd bounce, even bounce, and diffuse scattering. In [4], for a refined classification with more classes, Cloude and Pottier proposed an unsupervised classification algorithm based on their $H/\alpha$ target decomposition theory. Afterwards, Lee et al. [5] developed an unsupervised classification method based on Cloude decomposition and Wishart distribution. In [6], Pottier and Lee further improved this algorithm by including anisotropy to double the number of classes. In [7], Lee et al. proposed an unsupervised terrain and land-use classification algorithm based on Freeman and Durden decomposition [8]. Unlike other algorithms that classify pixels statistically and ignore their scattering characteristics, this algorithm not only uses a statistical classifier but also preserves the purity of dominant polarimetric scattering properties. Yamaguchi et al. [9] proposed a four-component scattering model based on Freeman's three-component model, and the helix scattering component was introduced as the fourth component, which often appears in complex urban areas whereas disappears in almost all natural distributed scenarios.

PolSAR image classification using advanced machine learning and pattern recognition methods has shown exceptional growth in recent years. In 1991, Pottier et al. [10] firstly introduced the Neural Networks (NNs) to PolSAR image

Table 1: Polarimetric parameters considered in this work.

| Feature[ref] | Expression |
|---|---|
| Amplitude of HH-VV correlation coeff. [22, 23] | $\left\| \dfrac{\langle S_{HH} S_{VV}^* \rangle}{\sqrt{\|S_{HH}\|^2 \|S_{VV}\|^2}} \right\|$ |
| Phase difference HH-VV [23, 24] | $\arg(\langle S_{HH} S_{VV}^* \rangle)$ |
| Copolarized ratio in dB [25] | $10 \cdot \log\left( \dfrac{\|S_{VV}\|^2}{\|S_{HH}\|^2} \right)$ |
| Cross-polarized ratio in dB [25] | $10 \cdot \log\left( \dfrac{\|S_{HV}\|^2}{\|S_{HH}\|^2} \right)$ |
| Ratio HV/VV in dB [22] | $10 \cdot \log\left( \dfrac{\|S_{HV}\|^2}{\|S_{VV}\|^2} \right)$ |
| Copolarization ratio [24] $\dfrac{\sigma_{VV}^0}{\sigma_{HH}^0}$ | $\dfrac{\langle S_{VV} S_{VV}^* \rangle}{\langle S_{HH} S_{HH}^* \rangle}$ |
| Depolarization ratio [23, 24] $\dfrac{\sigma_{HV}^0}{\sigma_{HH}^0 + \sigma_{VV}^0}$ | $\dfrac{\langle S_{HV} S_{HV}^* \rangle}{\langle S_{HH} S_{HH}^* \rangle + \langle S_{VV} S_{VV}^* \rangle}$ |

classification. In 1999, Hellmann [11] further introduced fuzzy logic with Neural Networks classifier; Fukuda et al. [12] introduced Support Vector Machine (SVM) to land cover classification with higher accuracy. In 2007, She et al. [13] introduced Adaboost for PolSAR image classification; compared with traditional classifiers such as complex Wishart distribution maximum likelihood classifier, these methods are more flexible and robust. In 2009, Shimoni et al. [14] investigated the Logistic regression (LR), NN, and SVM for land cover classification with various combinations of the PolSAR and PolInSAR feature sets.

The methods based on statistical properties and scattering mechanisms are generally pixel based with high computation complexity, and the employed polarimetric characteristics are also limited. The methods with advanced classifiers are usually implemented on patch level, and they can easily incorporate multiple polarimetric features. At present, with the development of polarimetric technologies, PolSAR can capture abundant structural and textural information. Therefore, classifiers arise from machine learning and pattern recognition domain such as SVM [15], Adaboost [16], and Random Forests [17] have attracted more attention. These methods usually can handle many sophistical image features and usually get remarkable performance.

In this paper, we focus on investigating multifeatures combination and employing a robust classifier named Extremely Randomized Clustering Forests (ERCFs) [18, 19] for terrain classification using PolSAR imagery. We first investigate the widely used polarimetric SAR features and further propose two feature combination strategies. Then in the classification stage we introduce the ERCFs classifier which has fewer parameters to tune and low computational

complexity in both training and testing, and it also can handle large variety of data without overfitting.

The organization of this paper is as follows. In Section 2, the common polarimetric features are investigated, and the two feature combination strategies are given. In Section 3, the recently proposed ERCFs algorithm is analyzed. The experimental results and performance evaluation are described in Section 4 and we conclude the paper in Section 5.

## 2. Polarimetric Feature Extraction and Combination

*2.1. Polarimetric Feature Descriptors.* PolSAR is sensitive to the orientation and characters of target and thus yields many new polarimetric signatures which produce a more informative description of the scattering behavior of the imaging area. We can simply divide the polarimetric features into two categories: one is the features based on the original data and its simple transform, and the other is based on target decomposition theorems.

The first category features in this work mainly include the Sinclair scattering matrix, the covariance matrix, the coherence matrix, and several polarimetric parameters. The classical $2 \times 2$ Sinclair scattering matrix $S$ can be achieved through the construction of system vectors [20]:

$$S = \begin{pmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{pmatrix}. \tag{1}$$

In the monostatic backscattering case, for a reciprocal target matrix, the reciprocity constrains the Sinclair scattering matrix to be symmetrical, that is, $S_{HV} = S_{VH}$. Thus, the two target vectors $k_p$ and $\Omega_l$ can be constructed based on the Pauli and lexicographic basis sets, respectively. With the two vectorizations we can then generate a coherency matrix $T$ and a covariance matrix $C$ as follows:

$$k_p = \frac{1}{\sqrt{2}} \begin{bmatrix} S_{HH} + S_{VV} \\ S_{HH} - S_{VV} \\ 2S_{HV} \end{bmatrix}, \qquad [T] = \left\langle k_p \cdot k_p^{*T} \right\rangle,$$

$$\Omega_l = \begin{bmatrix} S_{HH} \\ \sqrt{2} S_{HV} \\ S_V V \end{bmatrix}, \qquad [C] = \left\langle \Omega_l \cdot \Omega_l^{*T} \right\rangle, \tag{2}$$

where $*$ and $T$ represent the complex conjugate and the matrix transpose operations, respectively.

When analyzing polarimetric SAR data, there are also a number of parameters that have useful physical interpretation. Table 1 lists the considered parameters in this study: amplitude of HH-VV correlation coefficient, HH-VV phase difference, copolarized ratio in dB, cross-polarized ratio in dB, ratio HV/VV in dB, copolarization ratio, and depolarization ratio [21].

Polarimetric target decomposition theorems can be used for target classification or recognition. The first target decomposition theorem was formalized by Huynen based on the work of Chandrasekhar on light scattering with small anisotropic particles [26]. Since then, there have been many other proposed decomposition methods. In 1996, Cloude and Pottier [27] gave a complete summary of these different target decomposition methods. Recently, there are several new target decomposition methods that have been proposed [9, 28, 29]. In the next, we shall focus on the following five target decomposition theorems.

(1) *Pauli Decomposition*. The Pauli decomposition is a rather simple decomposition and yet it contains a lot of information about the data. It expresses the measured scattering matrix [S] in the so-called Pauli basis:

$$[S] = \alpha \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \beta \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + \gamma \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad (3)$$

where $\alpha = (S_{HH} + S_{VV})/\sqrt{2}$, $\beta = (S_{HH} - S_{VV})/\sqrt{2}$ and $\gamma = \sqrt{2}S_{HV}$.

(2) *Krogager Decomposition*. The Krogager decomposition [30] is an alternative to factorize the scattering matrix as the combination of the responses of a sphere, a diplane, and a helix; it presents the following formulation in the circular polarization basis $(r, l)$:

$$[S_{(r,l)}] = e^{j\varphi} \left\{ e^{j\varphi_s} k_s [S]_s + k_d [S]_d + k_h [S]_h \right\}, \quad (4)$$

where $k_s = |S_{rl}|$, if $|S_{rr}| > |S_{ll}|, k_d^+ = |S_{ll}|, k_h^+ = |S_{rr}| - |S_{ll}|$, and the helix component presents a left sense. On the contrary, when it is $|S_{ll}| > |S_{rr}|, k_d^- = |S_{rr}|, k_h^- = |S_{ll}| - |S_{rr}|$, and the helix has a right sense. The three parameters $k_s$, $k_d$, and $k_h$ correspond to the weights of the sphere, the diplane, and the helix components.

(3) *Freeman-Durden Decomposition*. The Freeman-Durden decomposition models [8] the covariance matrix as the contribution of three different scattering mechanisms: surface or single-bounce scattering, Double-bounce scattering, and volume scattering:

$$[C] = \begin{bmatrix} f_s|\beta|^2 + f_d|\alpha|^2 + \dfrac{3f_v}{8} & 0 & f_s\beta + f_d\alpha + \dfrac{f_v}{8} \\ 0 & \dfrac{2f_v}{8} & 0 \\ f_s\beta^* + f_d\alpha^* + \dfrac{f_v}{8} & 0 & f_s + f_d + \dfrac{3f_v}{8} \end{bmatrix}. \quad (5)$$

We can estimate the contribution on the dominance in scattering powers of $P_s$, $P_d$, and $P_v$, corresponding to surface, double bounce, and volume scattering, respectively:

$$P_s = f_s\left(1 + |\beta|^2\right), \quad P_d = f_d\left(1 + |\alpha|^2\right), \quad P_v = \frac{8}{3}f_v. \quad (6)$$

(4) *Cloude-Pottier Decomposition*. Cloude and Pottier [4] proposed a method for extracting average parameters from the coherency matrix $T$ based on eigenvector-eigenvalue Decomposition, and the derived entropy $H$, the anisotropy $A$, and the mean alpha angel $\underline{\alpha}$ are defined as

$$H = -\sum_{i=1}^{3} p_i \log_3(p_i), \quad p_i = \frac{\lambda_i}{\sum_{k=1}^{3} \lambda_k},$$

$$A = \frac{\lambda_2 - \lambda_3}{\lambda_2 + \lambda_3}, \quad (7)$$

$$\underline{\alpha} = \sum_{i=1}^{3} p_i \alpha_i.$$

(5) *Huynen Decomposition*. The Huynen decomposition [26] is the first attempt to use decomposition theorems for analyzing distributed scatters. In the case of coherence matrix, this parametrization is

$$[T] = \begin{bmatrix} 2A_0 & C - jD & H + jG \\ C + jD & B_0 + B & E + jF \\ H - jG & E - jF & B_0 - B \end{bmatrix}. \quad (8)$$

The set of nine independent parameters of this particular parametrization allows a physical interpretation of the target.

On the whole, the investigated typical polarimetric features include

(i) $F_1$: amplitude of upper triangle matrix elements of $S$;

(ii) $F_2$: amplitude of upper triangle matrix elements of $C$;

(iii) $F_3$: amplitude of upper triangle matrix elements of $T$;

(iv) $F_4$: the polarization parameters in Table 1;

(v) $F_5$: the three parameters $|\alpha|^2, |\beta|^2, |\gamma|^2$ of the Pauli decomposition;

(vi) $F_6$: the three parameters $k_s, k_d, k_h$ of the Krogager decomposition;

(vii) $F_7$: the three scattering power components $P_s, P_d, P_v$ of the Freeman-Durden decomposition;

(viii) $F_8$: the three parameters $H$-$\underline{\alpha}$-$A$ of the Cloude-pottier decomposition;

(ix) $F_9$: the nine parameters of the Huynen decomposition.

*2.2. Multifeatures Combination.* Recently researches [14, 31, 32] concluded that employing multiple features and different combinations can be very useful for PolSAR image classification. Usually, there is no unique set of features for PolSAR image classification. Fortunately, there are several common strategies for feature selection [33]. Some of them give only a ranking of features; some are able to directly select proper features for classification. One typical choice is the Fisher-score which is simple and generally quite effective. However,

it does not reveal mutual information among features [34]. In this study we present two simple strategies to implement the combination of different polarimetric features: one is by manual selection following certain heuristic rules, and the other is automatic combination with a newly proposed measure.

(1) *Heuristic Feature Combination.*

The heuristic feature combination strategy uses the following rules.

(i) Feature types are separately selected in the two category features.

(ii) In each category, the selected feature types should have better classification performance for some specific terrains.

(iii) Each feature should be little correlated with another feature within the selected feature sets.

(2) *Automatic Feature Combination.*

Automatic selection and combining different feature types are always necessary when facing a large number of feature types. Since there may exist many relevant and redundant information between different feature types, we need to not only consider the classification accuracies of different feature types but also keep track of their correlations. In this section, we propose a metric-based feature combination to balance the feature dependence and classification accuracy.

Given a feature type pool $F_i (i = 1, 2, \ldots, N)$, the feature dependence of the $i$th feature type is proposed to be defined as

$$ Dep_i = \frac{N-1}{\sum_{j=1, j \neq i}^{N} corrcoef\left(\overrightarrow{P_i}, \overrightarrow{P_j}\right)}. \tag{9} $$

$\overrightarrow{P_i}$ is the terrain classification accuracy of the $i$th feature type in feature type pool. $corrcoef(\cdot)$ is the correlation coefficient.

The $Dep_i$ is actually the reciprocal of average cross-correlation coefficient of the $i$th feature type, and it can represent the average coupling of the $i$th feature type and the other feature types. We assume that these two metrics are independent as done in feature combination, and then the selection metric of the $i$th feature type can be defined as

$$ R_i = Dep_i \cdot A_i, \tag{10} $$

where $A_i$ is the average accuracy of the $i$th feature type.

If the selection metric $R_i$ is low, the corresponding feature type will be selected with low probability. While the selection metric $R_i$ is high, it is more likely to be selected. After obtaining classification accuracy of each feature type, we propose to make feature combination by completely automatic combining method as Algorithm 1. The features with higher selection metric have higher priority to be selected, and the feature is finally selected only if it can improve the classification accuracy based on the selected features with a predefined threshold.

---

```
feature_combination(F, P)
Input: feature type pool F = {f₁, f₂, ..., f_N}
    classification accuracy Pᵢ with single feature type fᵢ
Output: a certain combination S = {f₁, f₂, ..., f_M}
    -Compute the selection metric R = {r₁, r₂, ..., r_N},
    rᵢ is the metric of the ith feature type;
    -S = empty_set
do
    -Find the correspond index i of the maximum of R
    if  add_to_pool(fᵢ, S) return true
        -select fᵢ for combining, S = {S, fᵢ};
        -remove fᵢ and Rᵢ from F and R;
    else
        return S;
while(true)
add_to_pool(fᵢ, S)
Input: a certain feature type fᵢ, a combination S
Output: a boolean
    -compute the classification accuracy Pₛ of S;
    -compute the classification accuracy P_c of {S, fᵢ};
if (P_c − Pₛ) > T
    return true;
else
    return false;
```

ALGORITHM 1: The pseudocode of automatic feature combining.

## 3. Extremely Random Clustering Forests

The goal of this section is to describe a fast and effective classifier, Extremely Randomized Clustering Forests (ERCFs), which are ensembles of randomly created clustering trees. These ensemble methods can improve an existing learning algorithm by combining the predictions of several models. The ERCFs algorithm provides much faster training and testing and comparable accurate results with the state-of-the-art classifier.

The traditional Random Forests (RFs) algorithm was firstly introduced in machine learning community by Breiman [17] as an enhancement of Tree Bagging. It is a combination of tree classifiers in a way that each classifier depends on the value of a random vector sampled independently and having same distribution for all classifiers in the forests and each tree casts a unit vote for the most popular class at input. To build a tree it uses a bootstrap replica of the learning sample and the CART algorithm (without pruning) together with the modification used in the Random Subspace method. At each test node the optimal split is derived by searching a random subset of size $K$ of candidate attributes (selected without replacement from the candidate attributes). RF contains $N$ forests, which can be any value. To classify a new dataset, each tree gives a classification for that case; the RF chooses the classification that has the most out of $N$ votes. Breiman suggests that as the numbers of trees increase, the generalization error always converges and over fitting is not a problem because of the Strong Law of Large Numbers [17]. After the success of RF algorithm, several researchers have looked at specific randomization techniques

**Split_a_node**(S)
*Input*: labeled training set S
*Output*: a split [a < $a_c$] or nothing
**if** stop_split(S) **then** return nothing;
**else**
  *tries* = 0;
  **repeat**
    -*tries* = *tries* + 1;
    -selected an attribute number $i_t$ randomly
        and get the selected attribute $S_{i_t}$;
    -get a split $s_i$ = **Pick_a_random_split**($S_{i_t}$);
    -split S according $s_i$, and calculate the *score*;
  **until** (*score* ≥ $S_{min}$) or (*tries* ≥ $T_{max}$)
  **return** the split s* that achieved highest score;
**end if**
**Pick_a_random_split**($S_{i_t}$)
*Input*: an attribute $S_{i_t}$
*Output*: a split $s_i$
  -Let $s_{min}$ and $s_{max}$ denote the maximal
    and minimal value of $S_{i_t}$;
  -Get a random cut-point $s_i$ uniformly in [$s_{min}$ $s_{max}$];
  -return $s_i$;
**Stop_split**(S)
*Input*: a subset S
*output*: a boolean
  **if** |S| < $n_{min}$, **then** return true;
  **if** all attributes are constant is S, **then** return true;
  **if** all the training label is the same in S,
    **then** return true;
  **otherwise**, return false;

ALGORITHM 2: Tree growing algorithm of ERCFs.

for tree based on a direct randomization of tree growing method. However, most of these techniques just make litter perturbations in the search of the optimal split during tree growing, and they are still far from building totally random trees [18].

Compared with RF, the ERCFs [18] use consists in building many extremely randomized trees, which randomly pick attributes and cut thresholds at each node. The tree growing algorithm of ERCFs is shown as Algorithm 2. The main differences between ERCFs and RF are that it splits nodes by choosing cut-points fully at random and that it uses the whole learning sample (rather than a bootstrap replica) to grow the trees. At each node, the Extremely Clustering Trees splitting procedure is processing recursively until further subdivision is impossible, and the resulting node is scored over the surviving points by using the Shannon entropy as suggested in [18]. For a sample S and a split $s_i$, this measure is given by

$$Score\ (s_i, S) = \frac{2I_C^{s_i}(S)}{H_{s_i}(S) + H_C(S)}, \qquad (11)$$

where $H_C(S)$ is the (log) entropy of the classification in S, $H_{s_i}(S)$ is the split entropy, and $I_C^{s_i}(S)$ is the mutual information of the split outcome and the classification.

The parameters $S_{min}$, $T_{max}$, and $n_{min}$ have different effects: $S_{min}$ determines the balance of the grown tree; $T_{max}$ determines the strength of the attribute selection process, and it denotes the number of random splits screened at each node to develop. In the extreme, for $T_{max} = 1$, the splits (attributes and cut-points) are chosen in a totally independent way of the output variable. On the other extreme, when $T_{max} = N_s$, the attribute choice is not explicitly randomized anymore, and the randomization effect acts only through the choice of cut-points. $n_{min}$ is the strength of averaging output noise. Larger values of $n_{min}$ lead to smaller trees, higher bias, and smaller variance. In the following experiments, we set $n_{min} = 1$ in order to let the tree grow completely. Since the classification effect is not sensitive to the $S_{min}$ and $T_{max}$, we use $T_{max} = 50$ and $S_{min} = 0.2$.

Because of the extremely randomization, the ERCFs are usually much faster than other ensemble methods. In [18], the ERCFs are shown that they can perform remarkably on a variety of tasks and produce lower test errors than conventional machine learning algorithm. We adopt ERCFs mainly due to their three appealing features [19, 35]:

  (i) fewer parameters to adjust and do not worry about overfitting;

  (ii) higher computational efficiency in both training and testing;

  (iii) more robust to background clutter compared to state-of-the-art methods.

Since the polarimetric SAR images carry significantly more data capacity and can provide more features, the ERCFs are just put to good use.

## 4. Experimental Results

*4.1. Experimental Dataset.* The ALOS PALSAR polarimetric SAR data(JAXA) of Washington County, North Carolina, and the Land Use Land Cover (LULC) ground truth image (USGS) are used for feature analysis and comparison. The selected POLSAR image has 1236 × 1070 pixels with 8 looks and 30 m×30 m resolution. According to the LULC image data, the land cover mainly includes four classes: water, wetland, woodland, and farmland. Only the above four classes are considered in training and testing; the pixels of other classes are ignored. The classification accuracy on each terrain is used to evaluate the different feature types.

*4.2. Evaluation of Single Polarimetric Descriptor.* We firstly represent PolSAR images as rectangular grids of patches at a single scale with the block size 12 × 12 and the overlap step 6. In the training stage, 500 patches of each class are selected as training data. Then, all the features are normalized to [0 1] by their corresponding maximum and minimum values across the image. We finally use the KNN and SVM classifier for evaluation of single polarimetric feature. KNN is a linear classifier. It selects the K nearest neighbours of the test patch within the training patches. Then it assigns to the new patch the label of the category which is most represented within

TABLE 2: Classification accuracies of single polarimetric descriptor using KNN and SVM classifier(%).

| Feature (dim) | Classifier | Water | Wetland | Woodland | Farmland | Ave.acc |
|---|---|---|---|---|---|---|
| $F_1(3)$ | KNN | 73.3 | 59.7 | 65.3 | 68.1 | 66.6 |
|  | SVM | 88.7 | 59.1 | 73.1 | 78.4 | 74.8 |
| $F_2(6)$ | KNN | 64 | 60.9 | 64.4 | 53.5 | 60.7 |
|  | SVM | 88.1 | 62 | 73.6 | 78.2 | 75.5 |
| $F_3(6)$ | KNN | 69.8 | 59.4 | 63.3 | 52 | 61.1 |
|  | SVM | 84.9 | 55.7 | 74.3 | 67.9 | 70.7 |
| $F_4(7)$ | KNN | 81.5 | 46.8 | 70.3 | 69.4 | 67 |
|  | SVM | 89.2 | 51.4 | 72.8 | 75.1 | 72.1 |
| $F_5(3)$ | KNN | 73.2 | 58.1 | 65 | 64.4 | 65.2 |
|  | SVM | 85.6 | 60.5 | 71.5 | 74.7 | 73.1 |
| $F_6(3)$ | KNN | 78.9 | 55.8 | 67.1 | 67.2 | 67.2 |
|  | SVM | 81.2 | 53.6 | 76.5 | 68 | 69.8 |
| $F_7(3)$ | KNN | 86.3 | 63 | 69 | 71.9 | 72.5 |
|  | SVM | 87.6 | 53.4 | 77.7 | 74.1 | 73.2 |
| $F_8(3)$ | KNN | 71.3 | 61.9 | 66.6 | 67.1 | 66.7 |
|  | SVM | 77.8 | 56.3 | 75.8 | 73.3 | 70.8 |
| $F_9(9)$ | KNN | 87.5 | 60.6 | 66.29 | 72.9 | 71.8 |
|  | SVM | 88.6 | 61.9 | 73.8 | 76.4 | 75.2 |

TABLE 3: Classification performances(%) of KNN and SVM with selected feature set and all features.

| Classifier | Features | Water | Wetland | Woodland | Farmland | Ave.acc |
|---|---|---|---|---|---|---|
| KNN | Selected features | 87.6 | 67.1 | 69.9 | 74.1 | 74.7 |
|  | All features | 86.1 | 67.7 | 68.1 | 74.8 | 74.2 |
| SVM | Selected features | 91.5 | 69.2 | 75.9 | 80.4 | 79.3 |
|  | All features | 91.3 | 70 | 75.2 | 79.7 | 79.1 |

the $K$ nearest neighbours. SVM constructs a hyperplane or set of hyperplanes in a high-dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data points of any class (so-called functional margin), since in general the larger the margin, the lower the generalization error of the classifier. In this experiment, for the KNN classifier, we use an implementation of fuzzy $k$-nearest neighbor algorithm [36] with $K = 10$ which is experimentally chosen. For the SVM, we use the LIBSVM library [37], in which the radial basis function (RBF) kernel is selected and optimal parameters are selected by grid search with 5-fold cross-validation. The classification accuracies of KNN and SVM using single polarimetric descriptor are shown in Table 2.

TABLE 4: The selection metric of the two categories features.

| Classifier | Features of category I | | | | Features of category II | | | | |
|---|---|---|---|---|---|---|---|---|---|
|  | $F_1$ | $F_2$ | $F_3$ | $F_4$ | $F_5$ | $F_6$ | $F_7$ | $F_8$ | $F_9$ |
| KNN | 1.39 | 1.6 | 1.03 | 1.27 | 0.67 | 0.69 | 0.75 | 0.69 | 0.75 |
| SVM | 0.76 | 0.77 | 0.73 | 0.73 | 0.76 | 0.72 | 0.75 | 0.74 | 0.78 |

TABLE 5: Classification performances(%) of SVM and ERCFs with Pset1, Pset2, and Pset3.

| Classifier | Features | Water | Wetland | Woodland | Farmland | Ave.acc |
|---|---|---|---|---|---|---|
| SVM | Pset1 | 88.3 | 63.0 | 73.9 | 78.3 | 75.9 |
|  | Pset2 | 89.4 | 67.5 | 72.2 | 81.2 | 77.6 |
|  | Pset3 | 91.5 | 69.2 | 75.9 | 80.4 | 79.3 |
| ERCFs | Pset1 | 89.3 | 64.2 | 74.5 | 78.7 | 76.7 |
|  | Pset2 | 89.8 | 69.1 | 72.3 | 80.9 | 78.0 |
|  | Pset3 | 91.5 | 69.6 | 76.4 | 80.9 | 79.6 |

TABLE 6: Time comsuming of SVM and ERCFs.

| Classifier | Training time (s) | Testing time (s) |
|---|---|---|
| SVM | 986.35 | 22.97 |
| ERCFs | 22.95 | 0.44 |

From Table 2, some conclusions can be drawn.

Features based on original data and its simple transform.

   (i) Sinclair scattering matrix has better performance in water and farmland classification.
   (ii) Covariance matrix has better performance in wetland classification.
   (iii) The polarization parameters in Table 1 have better performance in water, woodland, and farmland classification.

Features based on target decomposition theorems.

   (i) Freeman decomposition and Huynen decomposition have better performance in water and wetland classification.
   (ii) Freeman decomposition and Krogager decomposition have better performance in woodland classification.
   (iii) Huynen decomposition has better performance in farmland classification.

*4.3. Performance of Different Feature Combinations.* In this experiments, to obtain training samples, we first determine several "Training Area" polygons delineated with visual interpretation according to the ground truth data, and then we use a randomly subwindow sampling to build a certain number of training sets.

   Following the above mentioned three heuristic criterions and Table 2 we can obtain a combined feature set as
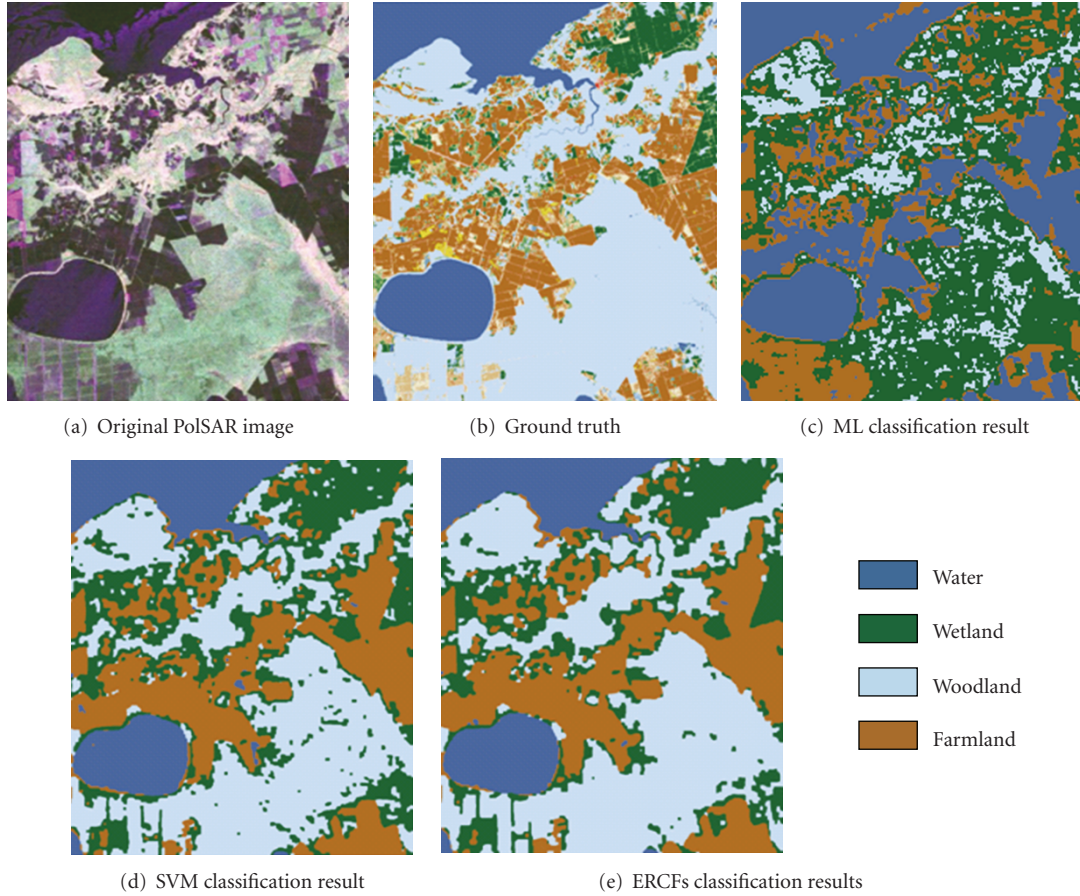
(a) Original PolSAR image    (b) Ground truth    (c) ML classification result



(d) SVM classification result    (e) ERCFs classification results

FIGURE 1: (a) ALOS PALSAR polarimetric SAR data of Washington County, North Carolina ($1236 \times 1070$ pixels, R: HH, G: HV, B: VV). (b) The corresponding Land use Land cover (LULC) ground truth. (c) Classification result using ML. (d) Classification result using SVM. (e) Classification result using ERCFs.
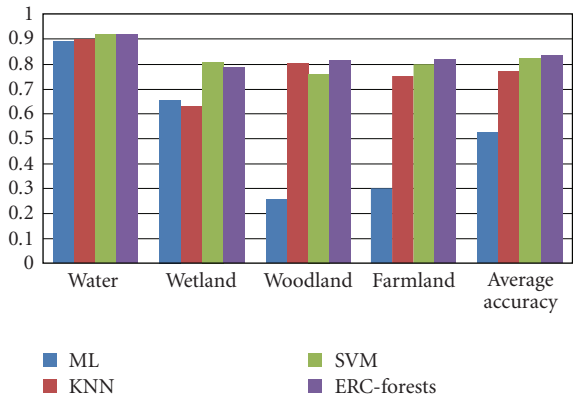


FIGURE 2: The quantitative comparison of different classifiers with features Pset3.

$\{F_1, F_2, F_4, F_7, F_9\}$, which is expected to get comparable performance than combination of all the features. Table 3 shows the performance comparison between the selected combining feature sets and the feature set by combination all of the feature type. It can be learned that the selected feature set gets a slightly higher average accuracy. Compared with single features performance in Table 2, we also find that the multifeatures combination can greatly improve the performance by $4 \sim 8\%$.

Based on the classification performance of single polarimetric feature in Table 2, the selection metric of each category features is given in Table 4. When selecting three feature types in the first category and two feature types in the second category using the KNN classifier, we can get the same combination result as Heuristic feature combination. When considering the SVM classifier, the result of selected combination is a slightly different with the former. The results say that the proposed selection parameter is a reasonable metric for feature combination.

After obtaining the classification performance of each feature type, we propose to make feature combination by completely automatic combining method as Algorithm 1. The features with higher selection metric have higher priority to be selected, and the feature is finally selected only if it can improve the classification accuracy based on the selected features with a predefined threshold. According to the selection metric in Table 2 and automatic feature combining as shown in Table 3, if threshold $T = 0.5$, automatic combination can get the same feature combination as the heuristic feature combination.

In the following experiment some intermediate feature combination states are selected to illustrate that the feature combination strategy can improve the classification performance step by step. The intermediate feature combination states include the following.

**Pset1**: select 1 feature type in the first category and 1 feature type in the second category; the combination features include $F_2$ and $F_9$.

**Pset2**: select 2 feature type in the first category and 1 feature type in the second category; the combination features include $F_1$, $F_2$ and $F_9$.

**Pset3**: the final selected feature set $\{F_1, F_2, F_4, F_7, F_9\}$.

Table 5 shows the classification performance of the three above intermediate feature combination states using SVM and ERCFs classifier, respectively. As expected, the averaged classification accuracy increases gradually with further multifeatures combination. The best single feature performance in Table 2 is 75.5%, while the classification accuracy using multifeatures combination is 79.3%, and both use SVM. ERCFs can provide a slightly higher accuracy with 79.6% based on the final combined feature set.

*4.4. Performance of Different Classifiers.* Now we further compare the performance of the ERCFs classifier with the widely used maximum likelihood (ML) classifier [2] and SVM classifier. The number of training and test patches is 2000 and 36 285, respectively.

The feature combination step can use heuristic selection to form a feature combination or use automatic combining to search an optimal feature combination. Here we recommend to use automatic combining since it is more flexible. When mapping the patch-level classification result to pixel-level, we take a smoothing postprocessing method based on the patch-level posteriors (the probability soft output of ERCFs or SVM classifier) [38]. We first assign each pixel posterior label probability by linearly interpolating of the four adjacent patch-level posteriors to produce smooth probability maps. Then we apply a Potts model Markov Random Field (MRF) smoothing process using graph cut optimization [39] on the final pixels labels to obtain final classification result. The classification results of ML classifier based on Wishart distribution, SVM, and ERCFs are shown in Figure 1.

Figure 2 is a quantitative comparison of the results based on the ground truth-LULC. It can be learned that ERCFs can get slightly better classification accuracy than SVM, and they both have much better performance than traditional ML classifier based on complex Wishart distribution.

In addition, ERCFs require less computational time compared to SVM classifier, which could be learned from the Table 6. SVM training time includes the time for searching the optimal parameters with a $10 \times 10$ grid search. ERCFs include 20 extremely clustering trees and we selected 50 attributes every time when making node splitting.

## 5. Conclusion

We addressed the problem of classifying PolSAR image with multifeatures combination and ERCFs classifier. The work started by testing the widely used polarimetric descriptors for classification, and then considering two strategies for feature combination. In the classification step, the ERCFs were introduced; incorporated with the selected multiple polarimetric descriptors, ERCFs have achieved satisfactory classification accuracies that as good as or slightly better than that using SVM at much lower computational cost, which shows that the ERCFs is a promising approach for PolSAR image classification and deserves particular attention.

## Acknowledgments

## 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," *Journal of Electromagnetic Waves and Applications*, vol. 2, no. 2, pp. 171–194, 1988.

[2] J. S. Lee, M. R. Grunes, and R. Kwok, "Classification of multilook polarimetric SAR imagery based on complex Wishart distribution," *International Journal of Remote Sensing*, vol. 15, no. 11, pp. 2299–2311, 1994.

[3] J. J. van Zyl, "Unsupervised classification of scattering mechanisms using radar polarimetry data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 27, pp. 36–45, 1989.

[4] S. R. Cloude and E. Pottier, "An entropy based classification scheme for land applications of polarimetric SAR," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 1, pp. 68–78, 1997.

[5] 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 Transactions on Geoscience and Remote Sensing*, vol. 37, no. 5, pp. 2249–2258, 1999.

[6] E. Pottier and J. S. Lee, "Unsupervised classification scheme of PolSAR images based on the complex Wishart distribution and the H/A/α. Polarimetric decomposition theorem," in *Proceedings of the 3rd European Conference on Synthetic Aperture Radar (EUSAR '00)*, Munich, Germany, May 2000.

[7] J. S. Lee, M. R. Grunes, E. Pottier, and L. Ferro-Famil, "Unsupervised terrain classification preserving polarimetric scattering characteristics," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 4, pp. 722–731, 2004.

[8] A. Freeman and S. Durden, "A three-component scattering model for polarimetric SAR data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 3, pp. 963–973, 1998.

[9] Y. Yamaguchi, T. Moriyama, M. Ishido, and H. Yamada, "Four-component scattering model for polarimetric SAR image decomposition," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 8, pp. 1699–1706, 2005.

[10] E. Pottier and J. Saillard, "On radar polarization target decomposition theorems with application to target classification by using network method," in *Proceedings of the International Conference on Antennas and Propagation (ICAP '91)*, pp. 265–268, York, UK, April 1991.

[11] M. Hellmann, G. Jaeger, E. Kraetzschmar, and M. Habermeyer, "Classification of full polarimetric SAR-data using artificial neural networks and fuzzy algorithms," in *Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS '99)*, vol. 4, pp. 1995–1997, Hamburg, Germany, July 1999.

[12] S. Fukuda and H. Hirosawa, "Support vector machine classification of land cover: application to polarimetric SAR data," in *Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS '01)*, vol. 1, pp. 187–189, Sydney, Australia, July 2001.

[13] X. L. She, J. Yang, and W. J. Zhang, "The boosting algorithm with application to polarimetric SAR image classification," in *Proceedings of the 1st Asian and Pacific Conference on Synthetic Aperture Radar (APSAR '07)*, pp. 779–783, Huangshan, China, November 2007.

[14] M. Shimoni, D. Borghys, R. Heremans, C. Perneel, and M. Acheroy, "Fusion of PolSAR and PolInSAR data for land cover classification," *International Journal of Applied Earth Observation and Geoinformation*, vol. 11, no. 3, pp. 169–180, 2009.

[15] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, Berlin, Germany, 1995.

[16] Y. Freund and R. E. Schapire, "Game theory, on-line prediction and boosting," in *Proceedings of the 9th Annual Conference on Computational Learning Theory (COLT '96)*, pp. 325–332, Desenzano del Garda, Italy, July 1996.

[17] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[18] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine Learning*, vol. 63, no. 1, pp. 3–42, 2006.

[19] F. Moosmann, E. Nowak, and F. Jurie, "Randomized clustering forests for image classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 9, pp. 1632–1646, 2008.

[20] R. Touzi, S. Goze, T. Le Toan, A. Lopes, and E. Mougin, "Polarimetric discriminators for SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 30, no. 5, pp. 973–980, 1992.

[21] M. Molinier, J. Laaksonent, Y. Rauste, and T. Häme, "Detecting changes in polarimetric SAR data with content-based image retrieval," in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS '07)*, pp. 2390–2393, Barcelona, Spain, July 2007.

[22] S. Quegan, T. Le Toan, H. Skriver, J. Gomez-Dans, M. C. Gonzalez-Sampedro, and D. H. Hoekman, "Crop classification with multi temporal polarimetric SAR data," in *Proceedings of the 1st Workshop on Applications of SAR Polarimetry and Polarimetric Interferometry (POLinSAR '03)*, Frascati, Italy, January 2003, (ESA SP-529).

[23] H. Skriver, W. Dierking, P. Gudmandsen, et al., "Applications of synthetic aperture radar polarimetry," in *Proceedings of the 1st Workshop on Applications of SAR Polarimetry and Polarimetric Interferometry (POLinSAR '03)*, pp. 11–16, Frascati, Italy, January 2003, (ESA SP-529).

[24] W. Dierking, H. Skriver, and P. Gudmandsen, "SAR polarimetry for sea ice classification," in *Proceedings of the 1st Workshop on Applications of SAR Polarimetry and Polarimetric Interferometry (POLinSAR '03)*, pp. 109–118, Frascati, Italy, January 2003, (ESA SP-529).

[25] J. R. Buckley, "Environmental change detection in prairie landscapes with simulated RADARSAT 2 imagery," in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS '02)*, vol. 6, pp. 3255–3257, Toronto, Canada, June 2002.

[26] J. R. Huynen, "The Stokes matrix parameters and their interpretation in terms of physical target properties," in *Proceedings of the Journées Internationales de la Polarimétrie Radar (JIPR '90)*, IRESTE, Nantes, France, March 1990.

[27] S. R. Cloude and E. Pettier, "A review of target decomposition theorems in radar polarimetry," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, no. 2, pp. 498–518, 1996.

[28] R. Touzi, "Target scattering decomposition in terms of roll-invariant target parameters," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 1, pp. 73–84, 2007.

[29] A. Freeman, "Fitting a two-component scattering model to polarimetric SAR data from forests," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 8, pp. 2583–2592, 2007.

[30] E. Krogager, "New decomposition of the radar target scattering matrix," *Electronics Letters*, vol. 26, no. 18, pp. 1525–1527, 1990.

[31] C. Lardeux, P. L. Frison, J. P. Rudant, J. C. Souyris, C. Tison, and B. Stoll, "Use of the SVM classification with polarimetric SAR data for land use cartography," in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS '06)*, pp. 493–496, Denver, Colo, USA, August 2006.

[32] J. Chen, Y. Chen, and J. Yang, "A novel supervised classification scheme based on Adaboost for Polarimetric SAR Signal Processing," in *Proceedings of the 9th International Conference on Signal Processing (ICSP '08)*, pp. 2400–2403, Beijing, China, October 2008.

[33] A. L. Blum and P. Langley, "Selection of relevant features and examples in machine learning," *Artificial Intelligence*, vol. 97, no. 1-2, pp. C245–C271, 1997.

[34] Y. W. Chen and C. J. Lin, "Combining SVMs with various feature selection strategies," in *Feature Extraction, Foundations and Applications*, Springer, Berlin, Germany, 2006.

[35] F. Schroff, A. Criminisi, and A. Zisserman, "Object class segmentation using random forests," in *Proceedings of the 19th British Machine Vision Conference (BMVC '08)*, Leeds, UK, September 2008.

[36] J. M. Keller, M. R. Gray, and J. A. Givens Jr., "A fuzzy K-nearest neighbor algorithm," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 15, no. 4, pp. 580–585, 1985.

[37] C. C. Chang and C. J. Lin, "LIBSVM : a library for support vector machines," Software, 2001, http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[38] W. Yang, T. Y. Zou, D. X. Dai, and Y. M. Shuai, "Supervised land-cover classification of TerraSAR-X imagery over urban areas using extremely randomized forest," in *Proceedings of the Joint Urban Remote Sensing Event (JURSE '09)*, Shanghai, China, May 2009.

[39] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, 2001.