

Received August 25, 2020, accepted September 3, 2020, date of publication September 8, 2020, date of current version September 22, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3022642

# Ship Target Discrimination in SAR Images Based on BOW Model With Multiple Features and Spatial Pyramid Matching

SHIYUAN CHEN<sup>1</sup>, XIAOJIANG LI<sup>1</sup>, SHAOQUAN CHI<sup>2</sup>, ZHILIANG LI<sup>3</sup>, (Member, IEEE), AND MAO YUXING<sup>1</sup>, (Member, IEEE)

<sup>1</sup>Department of Aerospace Science and Technology, Space Engineering University, Beijing 101416, China

<sup>2</sup>Service Support Center, Shijiazhuang Division of PLAA Infantry College, Shijiazhuang 050000, China

<sup>3</sup>School of Space Information, Space Engineering University, Beijing 101416, China

Corresponding author: Zhiliang Li (lemonslee@163.com)

This work was supported by the Youth Innovation Fund of the China High-Resolution Earth Observation under Grant MOST GFZX04061502.

**ABSTRACT** To eliminate the false alarms in the ship target detection effectively for synthetic aperture radar (SAR) images in complex scenes, this article presents a novel ship target discrimination algorithm based on bag of words (BOW) model with multiple features and spatial pyramid matching (SPM), which is named MF-SPM-BOW. The proposed discrimination method mainly contains three stages. First, the SAR scale-invariant feature transform (SAR-SIFT) descriptors and gray-level co-occurrence matrix (GLCM) descriptors are extracted as local features to describe the gradient information and texture information of local regions of an image chip. Then, the SPM technique considering its spatial location information-keeping capability is employed to generate global features with excellent discrimination ability. Finally, the support vector machine (SVM) discriminator based on multiple kernel learning is applied to realize feature fusion in image layer and thus identify targets and clutter. Experimental results show that compared with the traditional discrimination methods and the BOW model discrimination methods, the proposed SAR ship target discrimination algorithm achieves better discrimination performance, which can eliminate most of the false alarms in candidate ship target chips effectively.

**INDEX TERMS** SAR images, ship targets discrimination, bag of words (BOW), local features, global features, spatial pyramid matching (SPM).

## I. INTRODUCTION

With the rapid development of Synthetic aperture radar (SAR) imaging technology, SAR images are widely used in military and civil fields [1]–[3]. One of the most important applications is SAR ship target detection and recognition, which has attracted much more attention during the past decades [4], [5], which has attracted more and more attention. The ship target detection selects the candidate ship target chips from the whole image. Due to the complex conditions of sea surface and serious interference of clutter, there exist many false alarm chips in the selected target chips, namely clutter false alarms, land and island false alarms. The false alarms interfere with the subsequent ship target recognition and

reduce its efficiency. As a result, recently many algorithms and techniques have been utilized to high-resolution SAR ship target discrimination to eliminate the false alarms.

Traditional SAR ship target discrimination method consists of two stages: 1) discrimination features are extracted to describe the candidate chips and 2) discriminator is designed to make the decision. Many discrimination features have been developed over the past two decades for they have played a significant role in target discrimination. The classical discrimination features, e.g., the old Lincoln features [6], the new Lincoln features [7] and GAO's features [8], which mainly describe the differences between target and clutter in texture, shape, size and contrast. However, the above features only achieve impressive performance in some simple scenes for they just describe the candidate chips roughly. With the increase in spatial resolution, SAR images can

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

provide detailed information about texture of regions. Therefore extracting discrimination features which can represent the texture details of the image is the key to discriminate the target and clutter in complex scenes. As the bag of words (BOW) model can obtain image local features and reduce the gap between low-level features and high-level features [9], [10], it has been applied successfully to image classification and target detection [11]–[15]. Recently, some researchers have developed many methods to show the potential of its use for SAR image discrimination. A new discrimination method based on the BOW model with sample-reweighed category-specific and shared dictionary learning algorithm was first proposed and achieved good discrimination performance in complex scenes [16]. Aiming at improving the descriptive abilities of the low-level feature in the BOW model, an algorithm based on multiple low-level feature fusion was put forward and obtained better performance [17]. A multilevel and multidomain feature descriptor based on BOW model was designed to discriminate target and clutter in multiple targets environment at super-pixel level [18]. To cope with difficulty of lack of labeled training chips, the BOW model was applied to extract the mid-level features and experiments demonstrated the effectiveness in target discrimination [19], [20]. As scale invariant feature transform (SIFT) descriptors has outstanding ability in describing the gradient amplitude and direction of image local regions, it is always extracted as low-level features in the BOW model [21]. The above discrimination methods based on the BOW model extracted SAR-SIFT descriptors for its robustness to the speckle noise [22]. Although they have achieved some good discrimination performance, they still have the following issues. First, they employed SAR-SIFT features as the low-level features which only reflect the gradient information of the image local regions. Thus it may lead to discrimination accuracy loss as the texture information of the image local regions is also essential for target discrimination. Second, they neglected the image's spatial location information when extracting the global features, which may lead to divisibility loss of the discrimination feature.

Recently, the deep learning techniques are employed to ship target discrimination in SAR images. In [23], a neural network is used to re-examine the detection results at the discrimination stage so as to better implement CFAR detection. A computational framework based on deep neural networks is proposed for iceberg and ship discrimination [24]. And a very deep high network configuration is presented as a SAR ship discrimination stage [25]. All the above algorithms achieved good discrimination performance at the expense of much more training time and also require a lot of training data to train the network. So the use of those methods based on deep learning is limited. As the labeled training data are not easy to obtain sometimes, the methods based on deep learning are not researched in this article.

Based on the existing studies, high resolution SAR images can provide more texture information about its local region.

Meantime, target and clutter always have distinct texture features. Therefore, for target discrimination using BOW model, extracting features which can effectively describe the image texture characteristics is very important. As gray-level co-occurrence matrix (GLCM) feature can capture image texture information effectively [26], it is introduced to extracted low-level feature. According to the above analysis, in this article, we propose a new discrimination method based on the BOW model to deal with SAR ship target discrimination in complex scenes. In the local feature extraction stage, SAR-SIFT descriptors and GLCM descriptors are extracted to describe the difference of target and clutter. The spatial pyramid matching (SPM) technique [27] is also employed to generate the global feature, considering its excellent capability of capture the spatial location information of the image. What's more, the support vector machine (SVM) [28] discriminator based on multiple kernel learning is designed to discriminate target and clutter.

In summary, the main contributions of this article are twofold. 1) We employ the GLCM to extract the local features. The GLCM can be used to extract some statistical features related to textural information of the image [29]. Recently, in the remote sensing image classification field, studies [11], [26] used the GLCM to extract features, since it can provide textural information of the image. Nevertheless, the GLCM has not been employed for SAR ship target discrimination. In this article, we take the GLCM to extract the local features for SAR target discrimination. Moreover, by combining with the SAR-SIFT together, the local features can capture the gradient and texture information of the image chips, which give a comprehensive description of the differences of targets and clutter. 2) We apply the SPM model to generate the global features on the feature pooling stage. The SPM algorithm divides the image into different number of image blocks at different scales and obtains features in each image block. By exploiting the technique, the final discrimination features can obtain the spatial location information of the image chips, leading to an improvement of discrimination performance.

The remainder of this article is organized as follows. Section II reviews the traditional BOW model. Section III introduces the research framework and technical details of the ship target discrimination based on the BOW model using multi-feature and SPM technique (MF-SPM-BOW). Section IV outlines the experimental results and analysis. Finally, Section V concludes this article.

## II. PRELIMINARIES ON BOW MODEL

In this section, we will review the traditional BOW model proposed in the literature.

The BOW model, first developed as a text classification tool, mainly includes local feature extraction stage, codebook generation stage, local feature coding stage, local feature pooling stage and classifier learning stage [10]. And those five stages are explained in detail in the following.

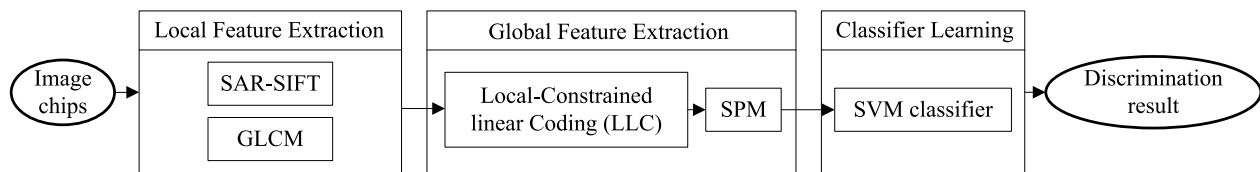


FIGURE 1. Flowchart of the ship target discrimination method based on MF-SPM-BOW model.

### A. LOCAL FEATURE EXTRACTION

Normally, there are two ways of obtaining the image patches, one is obtained by keypoints detection, and another is acquired by segmenting the image in regular grid method. The former can get sparse image patches and is suitable for target matching; the latter can get dense patches, which can be employed to image discrimination. Local features commonly used in SAR image target identification include SAR-SIFT descriptors, histogram of oriented gradient (HOG) and local binary pattern (LBP).

### B. VISUAL VOCABULARY CONSTRUCTION

Visual vocabulary offers a way to construct a new feature vector. And it is composed of visual words, which are the centers of clusters generated by local features clustering. The commonly used methods to generate visual vocabulary include K-Means and mean shift, etc.

### C. LOCAL FEATURE CODING

Local feature coding is to calculate the coding coefficient of each local feature according to the generated visual vocabulary. Gaussian mixture model (GMM) and Local Coordinate Coding (LLC) are the two common coding methods.

### D. LOCAL FEATURE POOLING

Local feature pooling is used to produce the final global features according to the coding coefficients. And the ways to fulfill feature pooling include max pooling and average pooling.

### E. CLASSIFIER LEARNING

With the global features of all the training chips obtained, the classifier is trained and then the decision value of each test chips is compared to a threshold to get the final discrimination result.

## III. METHODOLOGY

The framework and technical details of the ship target discrimination based on MF-SPM-BOW model is introduced in this section. Fig. 1 shows the flowchart of the proposed method. Similarity in traditional BOW model, the MF-SPM-BOW model contains three main steps: local feature extraction, global feature extraction and classifier learning. In the following, we describe each step of the proposed method in detail.

### A. LOCAL FEATURE EXTRACTION

Local feature extraction is the first and important step of the MF-SPM-BOW model. Superior features and their capability of describing target and clutter can improve the overall quality of the final discrimination. In this article, we first extract SAR-SIFT descriptors and then extract GLCM descriptors.

#### 1) SAR-SIFT DESCRIPTORS

The algorithm of SAR-SIFT descriptors extraction is introduced according to [22]. To deal with the speckle noise in the SAR image, gradient by ratio (GR) of the pixel is computed by using edge detection of ratio of exponential weighted averages (ROEWA), which can convert the multiplicative noise to additive noise to reduce the impact on edge detection.

The horizontal gradient  $G_{x,\alpha}$  of pixel  $(a, b)$  is defined as

$$M_{H1} = \int_{x=R} \int_{y=R^+} I(a+x, b+y) \times e^{-\frac{|x|+|y|}{\alpha}} \quad (1)$$

$$M_{H2} = \int_{x=R} \int_{y=R^-} I(a+x, b+y) \times e^{-\frac{|x|+|y|}{\alpha}} \quad (2)$$

$$R_H = \frac{M_{H1}}{M_{H2}} \quad (3)$$

$$G_{x,\alpha} = \log(R_H) \quad (4)$$

where  $\alpha$  is the exponential weight parameter,  $R$ ,  $R^+$ ,  $R^-$  represent the range of integration for exponential weight function in horizontal direction separately.

Similarly, the vertical gradient  $G_{y,\alpha}$  of pixel  $(a, b)$  is defined as

$$M_{V1} = \int_{y=R} \int_{x=R^+} I(a+x, b+y) \times e^{-\frac{|x|+|y|}{\alpha}} \quad (5)$$

$$M_{V2} = \int_{y=R} \int_{x=R^-} I(a+x, b+y) \times e^{-\frac{|x|+|y|}{\alpha}} \quad (6)$$

$$R_V = \frac{M_{V1}}{M_{V2}} \quad (7)$$

$$G_{y,\alpha} = \log(R_V) \quad (8)$$

So the gradient magnitude  $G_{n,\alpha}$  and orientation  $G_{o,\alpha}$  of the pixel  $(a, b)$  are regarded as

$$G_{n,\alpha} = \sqrt{(G_{x,\alpha})^2 + (G_{y,\alpha})^2} \quad (9)$$

$$G_{o,\alpha} = \arctan\left(\frac{G_{y,\alpha}}{G_{x,\alpha}}\right) \quad (10)$$

As introduced in Section II, dense extraction method can provide more abundant image information, and the SAR-SIFT descriptors using the dense points could more rapidly produce than those using sparse points. Therefore, we use dense SAR-SIFT for this study. As the image patch size can both influence the discrimination accuracy and computation complexity, we set the size of image patches to  $16 \times 16$  pixels to strike a balance between accuracy and complexity and it will be discussed in the experimental section. The overlapping method is applied to get sampling points and the sampling interval is set to 8 pixels. The SAR-SIFT descriptor is extracted as Algorithm 1.

---

**Algorithm 1** SAR-SIFT Descriptors Extraction
 

---

- Step 1:** Compute the gradient magnitude and orientation of each pixel in the image chip according to formula (1)-(10).
- Step 2:** Divide the image chip into  $N_1$  overlapping image patches according to the step length of 16 pixels and the interval of 8 pixels, and defined the center pixel of each patch as the key point.
- Step 3:** Mesh the image patch into 16 components which contain  $4 \times 4$  grids and calculate gradient histograms of each components in 8 orientations.
- Step 4:** Group the gradient histograms of 16 image components into a column vector with  $8 \times 16 = 128$  dimensions, which is regarded as local descriptors of the image patch.
- Step 5:** Form the column vectors of  $N_1$  image patches into a feature matrix, which is defined as SAR-SIFT descriptors.
- 

## 2) GLCM DESCRIPTORS

GLCM is defined as the probability of the occurrence of the pixel pairs  $(i, j)$  in a certain direction at a certain distance [26]. And it reflects the correlation of adjacent pixels and texture characteristic of the image. GLCM is represented by  $P_{\theta,s}$ , where  $\theta$  represents the direction of pixel  $j$  relative to  $i$ ,  $s$  represents the pixel distance of between  $i$  and  $j$ . Commonly used GLCM statistics are as follows.

Energy is the quadratic sum of all elements in the GLCM and represents the uniformity of the image grayscale change. Energy is defined as:

$$ASM = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p^2(i, j) \quad (11)$$

Entropy is the quantity of information of an image and describes the texture complexity of an image. Entropy is regarded as

$$ENT = - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p(i, j) \log p(i, j) \quad (12)$$

Homogeneity measures the local variation of image texture and reflects the regular degree of image texture. Homogeneity is defined by

$$HOM = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p(i, j) / [1 + (i - j)^2]^2 \quad (13)$$

Dissimilarity measures the similarity of image texture and is regarded as

$$DIS = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} |i - j| p(i, j) \quad (14)$$

Contrast reflects the clarity of image texture and is represented as

$$CON = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i - j)^2 p(i, j) \quad (15)$$

Correlation measures the similarity of elements in the row direction or the column direction, and is defined as

$$COR = \frac{1}{\sigma_1 \sigma_2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i - u_1)(j - u_2) p(i, j) \quad (16)$$

where  $p(i, j)$  is the element of matrix  $P_{\theta,s}$  in Formula (11)-(16),  $u_1, u_2, \sigma_1, \sigma_2$  are separately defined by:

$$u_1 = \sum_{i=0}^{N-1} i \sum_{j=0}^{N-1} p(i, j) \quad (17)$$

$$u_2 = \sum_{i=0}^{N-1} j \sum_{j=0}^{N-1} p(i, j) \quad (18)$$

$$\sigma_1^2 = \sum_{i=0}^{N-1} (i - u_1)^2 \sum_{j=0}^{N-1} p(i, j) \quad (19)$$

$$\sigma_2^2 = \sum_{j=0}^{N-1} (j - u_2)^2 \sum_{i=0}^{N-1} p(i, j) \quad (20)$$

Since there may exist single target or multiple targets in the candidate image chips, the image texture is complex and changes a lot for different regions. If GLCM feature is extracted directly from the whole image chip, the texture characteristic of the image chip may not be effectively acquired. To deal with the situation in this article, the image chip is first divided into many small superpixels by using superpixel segmentation algorithm and then GLCM feature is generated from the superpixels. GLCM descriptors are extracted as Algorithm 2.

## B. GLOBAL FEATURE EXTRACTION

Global feature, directly applied to discriminate target and clutter, is an important factor for the discrimination performance of the method. Specially, there are two ways to fuse SAR-SIFT descriptors and GLCM descriptors to obtain global features. One is to splice those two features directly

**Algorithm 2** GLCM Descriptors Extraction

- Step 1:** Divide the image chip into  $N_2$  superpixels by using simple linear iterative cluster (SLIC) method [30].
- Step 2:** Normalize the image amplitude to the interval  $[0, T]$ ,  $\theta$  is set to one of the four directions:  $0^\circ, 45^\circ, 90^\circ, 135^\circ$ , where  $T$  denotes the gray level of the quantized image.
- Step 3:** Calculate GLCM  $P_{\theta,s}$  ( $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$ ) from 4 directions in each superpixel.
- Step 4:** Compute GLCM features according to formula (11)-(16), namely *ASM*, *ENT*, *HOM*, *DIS*, *CON* and *COR*.
- Step 5:** Group all the GLCM features into a column vector, which is regarded as local GLCM descriptors of the superpixel.
- Step 6:** Form the column vectors of  $N_2$  superpixels into a feature matrix, which is defined as GLCM descriptors.

after extracting local features, and it is regarded as feature fusion in regional layer. Another is to generate mid-level features using those two local features separately according to BOW model and then splice mid-level features together to get the final global feature. It is called feature fusion in image layer. The former is easy to accomplish, but it treats all features equal and ignores different roles that different features may play. So it is suitable for simple features fusion. The latter can get individual mid-level features on the basis of low-level features and the weights of mid-level features can be calculated by training classifier. So we choose to fuse features in image layer. To alleviate the unrecoverable loss of discriminative information, we employ LLC model [31] to encode the SAR-SIFT descriptors and GLCM descriptors to get local codes. Then we use SPM model to process those codes to generate global feature. The flowchart of global feature extraction is shown in Fig. 2.

## 1) LLC MODEL

With strong reconstruction ability and high computational efficiency, LLC model is applied to encode all the extracted local features. LLC model has employed the local constraints to solve the objective function and the local correlation is better used than other codes. It is clear that encoding method using LLC model is the same, no matter what descriptors are. Thus to save pace we only introduce the encoding method

**Algorithm 3** Encoding Using LLC Model

- Step 1:** SAR-SIFT descriptors are clustered by K-Means algorithm [32] to generate a codebook  $\mathbf{CB} = [cb_1, cb_2, \dots, cb_M]$ ,  $M$  is size of  $\mathbf{CB}$ . Let  $F = [f_1, f_2, \dots, f_N]$  represents SAR-SIFT descriptors in an image chip and  $N$  denotes feature dimension.
- Step 2:** For the SAR-SIFT descriptor  $f_n$ ,  $n \in [1, N]$ , calculate the Euclidean distance from  $f_n$  to each element of  $\mathbf{CB}$ , and then find five nearest neighbors in  $\mathbf{CB}$  to form the local base  $\mathbf{LB}_n$ .
- Step 3:** The local code coefficient  $\tilde{c}_n$  of  $f_n$  can be acquired by solving the objective function:

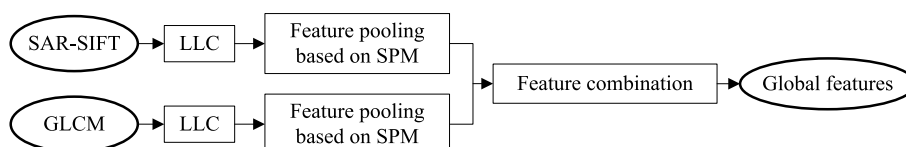
$$\min_C \sum_{n=1}^N \|f_n - \mathbf{LB}_n \tilde{c}_n\|^2 \quad s.t. \quad \sum_{n=1}^N \tilde{c}_n = 1 \quad (21)$$

- Step 4:** The coefficients of all elements in  $\mathbf{CB}$  are grouped into a column vector, which is denoted as code  $c_n$  of  $f_n$ , and the coefficients of all elements are zeros, except for those elements in  $\mathbf{LB}_n$ .
- Step 5:** All SAR-SIFT descriptors in the image chip are coded one by one, and finally code coefficient matrix  $\mathbf{c} = [c_1, c_2, \dots, c_N]$  can be obtained.

of SAR-SIFT descriptors in the following steps and omit the introduction of the GLCM descriptors encoding method.

## 2) FEATURE POOLING BASED ON SPM MODEL

The spatial location information of the image chips is very essential for the extraction of global features with good ability to distinguish targets and clutter. Thus in the feature pooling stage, SPM model is applied to generate the final global features. SPM algorithm considers the spatial location information of the image chip and divides it into different number of image blocks at different scales. Then in each image block, the code coefficient matrix is counted to get mid-level features and the final global features of the whole image chip are obtained by splicing the mid-level features in all image blocks. As we all known, the discrimination ability and dimensions of features varies with the number of image blocks. So the SPM three-tier model is used here, as shown in Fig. 3. Besides, considering the LLC method chosen previously, we use max pooling to extract global features in this article. In summary, the steps of feature pooling based on SPM model is as Algorithm 4.



**FIGURE 2.** Flowchart of global feature extraction.

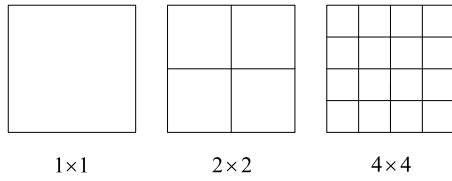


FIGURE 3. Typical SPM three-tier model.

**Algorithm 4** Feature Pooling Based on SPM Model

- Step 1:** Divide the input image chip into  $1 \times 1$ ,  $2 \times 2$  and  $4 \times 4$  image blocks separately according to Fig. 3 to get 21 image blocks together.
- Step 2:** In each block take the maximum values of code coefficient matrix  $\mathbf{c}$  obtained in feature code stage by columns and combine the maximum values of each column into a row vector as the mid-level feature of image block.
- Step 3:** The mid-level feature of image block is normalized by  $L2$  norm.
- Step 4:** The normalized features of all image blocks are spliced into a vector which is regarded as the final global features of the image chip.

**C. CLASSIFIER LEARNING**

The discrimination methods based on BOW model usually employ support vector machine (SVM) to distinguish target and clutter by learning the global features. SVM obtains the best classification interface by minimizing the structural risk. For sample  $z$  to be classified, the classification interface is defined as:

$$\mathbf{w}^T \cdot K(z) + b = 0 \tag{22}$$

where  $\mathbf{w}$  is weight coefficient vector, which describes the relevant parameters of classification interface,  $K(\cdot)$  is kernel function and  $b$  is the bias.

Normally, the local features extracted belong to the same type, the BOW models used a single kernel SVM with histogram cross kernel, and its function is denoted as:

$$K(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^N \min(x_i, y_i) \tag{23}$$

where  $\mathbf{x}$  and  $\mathbf{y}$  represent the  $N$  dimensional global features separately, and  $\mathbf{x} = [x_1, x_2, \dots, x_N]$ ,  $\mathbf{y} = [y_1, y_2, \dots, y_N]$ .

As SAR-SIFT descriptors capture the gradient information of image and GLCM descriptors reflect the texture information of image, they describe the differences of ship target and clutter in a different way and thus play a different role in discriminating target and clutter. Based on the above analysis, to accomplish the feature fusion of SAR-SIFT descriptors and GLCM descriptors in image layer, a SVM based on multiple kernel learning is applied as the discriminator. By classifier learning, the weights of the global features generated separately by SAR-SIFT descriptors and GLCM descriptors can be calculated.

Let  $\mathbf{S} = [s_1, s_2, \dots, s_{N_1}]$  denotes the global feature generated by SAR-SIFT descriptors,  $N_1$  is feature dimension,  $\omega_1$  represents weight. Similarly, Let  $\mathbf{G} = [g_1, g_2, \dots, g_{N_2}]$  denotes the global feature generated by GLCM descriptors,  $N_2$  is feature dimension,  $\omega_2$  represents weight. So the final global feature of the image chip is defined as  $\mathbf{F} = [\omega_1\mathbf{S}, \omega_2\mathbf{G}]$ . The following theorem can be deduced.

*Theorem 1:* The kernel function corresponding to image chip a and image chip b is denoted as:

$$K(\mathbf{F}_a, \mathbf{F}_b) = \omega_1 \sum_{i=1}^{N_1} \min(s_{ai}, s_{bi}) + \omega_2 \sum_{i=1}^{N_2} \min(g_{ai}, g_{bi}) \tag{24}$$

where  $\mathbf{F}_a$  and  $\mathbf{F}_b$  represent the global features of image chip a and image chip b separately,  $s_{ai}$  and  $s_{bi}$  represent the  $i$  dimensional feature of global feature  $\mathbf{S}$  separately, and  $g_{ai}$  and  $g_{bi}$  represent the  $i$  dimensional feature of global feature  $\mathbf{G}$  separately.

*Proof of Theorem 1:* Substitute the above definitions into Formula (23) and then expand it, the following can be deduced.

$$\begin{aligned} K(\mathbf{F}_a, \mathbf{F}_b) &= \min(\omega_1 s_{a1}, \omega_1 s_{b1}) + \min(\omega_1 s_{a2}, \omega_1 s_{b2}) \\ &\quad + \dots + \min(\omega_1 s_{aN_1}, \omega_1 s_{bN_1}) \\ &\quad + \min(\omega_2 g_{a1}, \omega_2 g_{b1}) + \min(\omega_2 g_{a2}, \omega_2 g_{b2}) \\ &\quad + \dots + \min(\omega_2 g_{aN_2}, \omega_2 g_{bN_2}) \\ &= \sum_{i=1}^{N_1} \min(\omega_1 s_{ai}, \omega_1 s_{bi}) + \sum_{i=1}^{N_2} \min(\omega_2 g_{ai}, \omega_2 g_{bi}) \\ &= \omega_1 \sum_{i=1}^{N_1} \min(s_{ai}, s_{bi}) + \omega_2 \sum_{i=1}^{N_2} \min(g_{ai}, g_{bi}) \end{aligned} \tag{25}$$

**Theorem 1** provides a conclusion that the feature fusion in image layer is equivalent to the weighted linear combination of kernel function corresponding to a single global feature. Therefore, we can complete feature fusion by multiple kernel learning. The method based on multiple kernel learning [33] has demonstrated its good discrimination performance, and it is used in this article to accomplish discrimination of ship target and clutter.

**IV. EXPERIMENTAL RESULTS AND ANALYSIS**

To better investigate the discrimination performance of the proposed method based on MF-SPM-BOW model and better understand the effects of the multiple features and the SPM technique, we conducted a series of experiments with different types of SAR images.

**A. EXPERIMENTAL DATA DESCRIPTION**

Two types of SAR images are used in this study. The first three images containing the ship targets and sea clutter with  $5m \times 5m$  resolution, from OpenSARShip dataset [34] are used in the experiments. All images from OpenSARShip dataset are the Sentinel-1 satellites images. As the images have larger amplitude and most of the images are the ocean, those images containing ship targets are cut out to form

new images with size of  $1500 \times 1500$  pixels for our experiments. For convenience, we refer to the three SAR images as Images I, Images II and Image III, respectively. The latter two images are collected from the AIR-SARShip-1.0 dataset [35] which comprises 31 images from Gaofen-3 satellite SAR images. The two images used here are  $1\text{m} \times 1\text{m}$  resolution and the size of images is  $3000 \times 3000$  pixels. And they are regarded as Images IV and Image V.

To validate the discrimination performance of the proposed method, we should extract the candidate image chips first. Normally, the size of candidate image chips is determined by SAR image resolution and target size. Considering that the length of ship targets is about 100 to 300 meters in Sentinel-1 images, in order to guarantee that there are both ship target and sea clutter in the candidate image chip, the size of candidate target chip is set to  $(2 \times 50 + 1) \times (2 \times 50 + 1) = 101 \times 101$  pixels for Images I-III. As AIR-SARShip-1.0 dataset is used for small ship target detection [32], the size of candidate target chip is set to  $(2 \times 70 + 1) \times (2 \times 70 + 1) = 141 \times 141$  pixels for Image IV-V. The candidate image chips are extracted as algorithm 5.

#### Algorithm 5 Image Chips Extraction

- Step 1:** Images I-V are detected by employing the classical two-parameter CFAR [36] to get candidate target pixels.
- Step 2:** Images I-V are divided into several superpixels by using SLIC method [30].
- Step 3:** For each superpixel if there is any candidate target pixel in the superpixel, then the superpixel is regarded as candidate target superpixel.
- Step 4:** Set the center of candidate target superpixel as the center of candidate image chip, and extract the corresponding candidate image chip with a certain size.
- Step 5:** Label the candidate image chips according to manual marking to obtain the target chips and clutter chips.

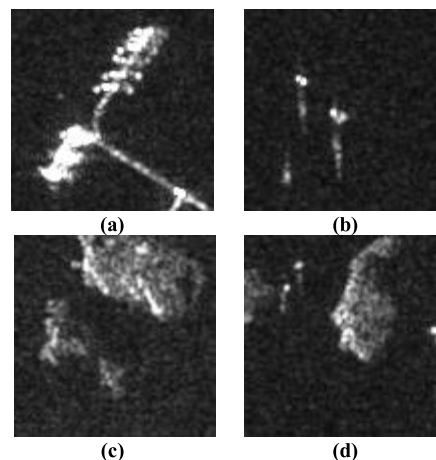
The numbers of target chips and clutter chips extracted are listed in Table 1.

**TABLE 1.** The numbers of target chips and clutter chips extracted from the five SAR images respectively.

	Image I	Image II	Image III	Image IV	Image V
Number of target chips	123	91	142	104	77
Number of clutter chips	398	258	307	349	206

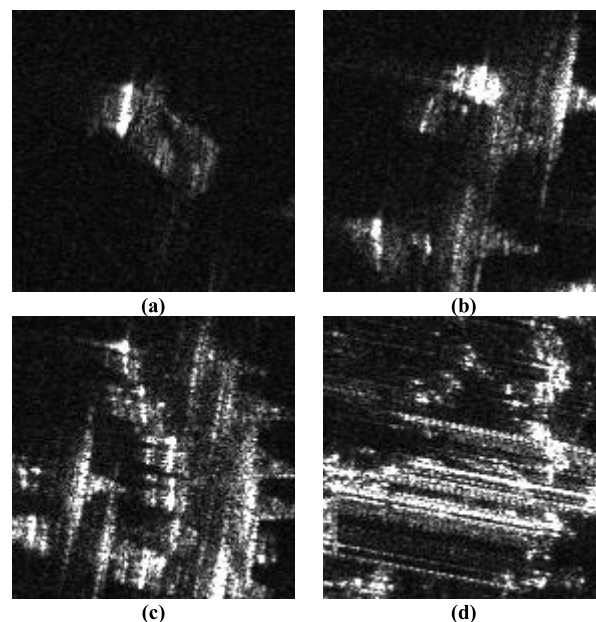
Before conducting experiments, we construct two datasets for different types of images: namely dataset for Sentinel-1 and dataset for Gaofen-3. For Images I-III from OpenSARShip dataset, dataset for Sentinel-1 is constructed. 200 chips are selected from target chips and clutter chips separately to construct training sample set. Then 150 chips are selected

from the remaining target chips and clutter chips separately to construct testing sample set. Some example image chips of dataset for Sentinel-1 used in the experiment are shown in Fig. 4.



**FIGURE 4.** Some example image chips of dataset for Sentinel-1 used in the experiments. (a) ship target chip; (b) ship target chip; (c) clutter chip; (d) clutter chip.

For Images IV-V from AIR-SARShip-1.0 dataset, dataset for Gaofen-3 is constructed. 120 chips are selected from target chips and clutter chips separately to construct training sample set. Then 60 chips are selected from the remaining target chips and clutter chips separately to construct testing sample set. Some example image chips of dataset for Gaofen-3 used in the experiment are shown in Fig. 5.



**FIGURE 5.** Some example image chips of dataset for Gaofen-3 used in the experiments. (a) ship target chip; (b) ship target chip; (c) clutter chip; (d) clutter chip.

It is clear that there exist single target and multiple targets in a true target chip, and the scenes in clutter chips are very complex. Therefore our two experimental datasets are valid

datasets for demonstrating the discrimination performance of the proposed methods.

## B. COMPARISON METHODS

To validate the discrimination performance of the proposed method, we compare it with five discrimination methods which include three classical methods and two methods based on BOW model. Each comparison method is described as follows.

- 1) Old Lincoln method: we use Old Lincoln features [6] as discrimination features and employ Gaussian kernel SVM as the discriminator.
- 2) New Lincoln method: we use New Lincoln features [7] as discrimination features and employ Gaussian kernel SVM as the discriminator.
- 3) GAO's method: we use GAO's features [8] as discrimination features and employ Gaussian kernel SVM as the discriminator.
- 4) the method based on SIFT-BOW: we use the classical SAR-SIFT descriptor [22] as local feature and generate the final features by BOW model as discrimination features. And histogram cross kernel SVM is employed as the discriminator.
- 5) the method based on MultiF-BOW [17]. We use MultiF-BOW as discrimination features and apply histogram cross kernel SVM based on multiple kernel learning as the discriminator.

The proposed method based on MF-SPM-BOW in the paper uses histogram cross kernel SVM based on multiple kernel learning as the discriminator.

## C. EVALUATION CRITERIA

The feature separability, as an essential factor for discrimination, is quantitatively measured by ratio of between-class distance to within-class distance (RBTW) [37]. RBTW is defined as

$$RBTW = \frac{S_B}{S_W} \quad (26)$$

between-class distance  $S_B$  is denoted as

$$S_B = \frac{1}{N_1 + N_2} \left( \sum_{i=1}^{N_1} dis(X_i - \mu_0) + \sum_{i=1}^{N_2} dis(Y_i - \mu_0) \right) \quad (27)$$

within-class distance  $S_W$  is defined by

$$S_W = \frac{1}{N_1 + N_2} \left( \sum_{i=1}^{N_1} dis(X_i - \mu_1) + \sum_{i=1}^{N_2} dis(Y_i - \mu_2) \right) \quad (28)$$

$$dis(X_i - \mu_k) = \frac{1}{\sqrt{(X_{i1} - \mu_{k1})^2 + (X_{i2} - \mu_{k2})^2 + \dots + (X_{iD} - \mu_{kD})^2}} \quad (29)$$

where  $N_1$  denotes the number of target chips,  $N_2$  denotes the number of clutter chips,  $\mu_k$  ( $k = 0, 1, 2$ ) represent the eigenvector mean of all image chips, the eigenvector mean of target chips, and the eigenvector mean of clutter chips,  $D$  is feature dimension in formula (26)-(29). It is drawn

from formula (26) that larger  $S_B$  and smaller  $S_W$  can lead to larger RBTW. And larger RBTW reflects good separability of features.

To quantitatively measure the performance of different methods, the probability of detection  $p_d$ , the probability of false alarm  $p_f$ , and the probability of correct discrimination  $p_c$  are employed in this article. And they are calculated as follows.

$$p_d = \frac{N_{dt}}{N_t} \quad (30)$$

where  $N_{dt}$  denotes the number of target chips detected as target and  $N_t$  represents the total number of target chips.

$$p_f = \frac{N_{dc}}{N_c} \quad (31)$$

where  $N_{dc}$  denotes the number of clutter chips detected as target and  $N_c$  represents the total number of clutter chips.

$$p_c = \frac{p_d + (1 - p_f)}{2} \quad (32)$$

The larger value of  $p_d$ , the smaller value of  $p_f$  and the bigger value of  $p_c$  mean that the discrimination performance of the method is better.

## D. PARAMETERS SETTING

As the traditional BOW model always uses 128-dimension SAR-SIFT descriptors, the size of codebook  $\mathbf{CB}$  is set to 128 in this article. To avoid generating information loss and large computation cost, the dimension of GLCM descriptors should not too small or too big. So the number of superpixels when extracting GLCM descriptors is set to 110. For all methods use SVM as discriminator, the penalty factor  $C$  is set to 5.

In the proposed method, there exist several parameters in the local feature extraction stage to be discussed, which may affect the overall performance of the discrimination method. Thus, this section presents a discussion of the parameter selection by conducting experiments using the dataset for Gaofen-3.

### 1) SELECTION OF $k$

In the local feature extraction stage, the image patch size  $k$  when applying dense SAR-SIFT descriptors is an important factor for the discrimination accuracy. For Gaofen-3 images collected in this article, we adjust the value of  $k$  to calculate the probability of correct discrimination  $p_c$ . The relationship between  $k$  and  $p_c$  is shown in Fig. 6. This figure shows that the probability of correct discrimination increases with  $k$ , then it reaches its maximum when  $k$  is 16, and finally it gradually decreases with  $k$ . Therefore, we set  $k$  to 16 in our experiments.

### 2) SELECTION OF $T$ AND $s$

When extracting GLCM descriptors, the gray level  $T$  and pixel distance  $s$  both may affect the accuracy of the proposed approach. First, we fix  $T$  and conducting discrimination experiments by changing the value of pixel distance  $s$ .



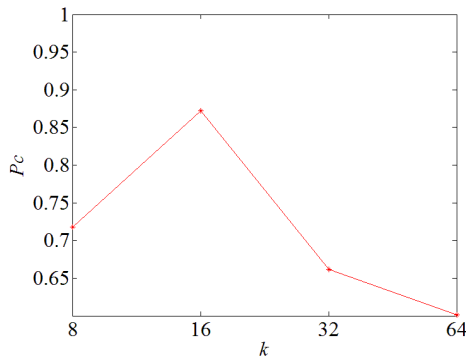


FIGURE 6. Relationship between  $k$  and  $p_c$ .

Then we change the value of  $T$ , and conducting the same experiments. The accuracy of the method with different  $T$  and  $s$  is shown in Fig. 7.

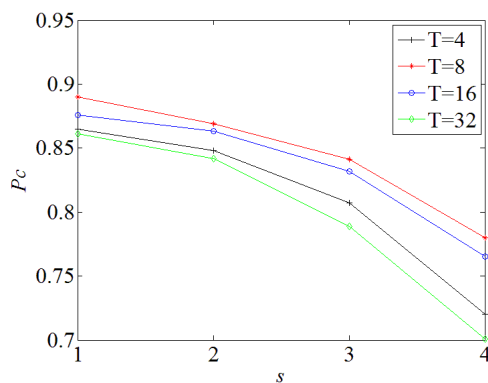


FIGURE 7. The accuracy of the method with different  $T$  and  $s$ .

From the Fig. 7, we can see that when  $T = 8$ , the proposed method achieves the best discrimination performance. And the accuracy of the method decreases with  $s$ , so  $s$  is set to 1. In summary, when  $T = 8$  and  $s = 1$ , the GLCM parameters in this article is the optimal parameters and can obtain the higher discrimination accuracy.

In summary, the parameters used in the paper are listed in Table 2.

TABLE 2. The values of parameters used in the experiments.

Parameters	Values
Size of codebook	128
Image patch size	16
Gray level	8
Pixel distance	1
Number of superpixels	110
Penalty factor	5

## E. COMPARISON OF DIFFERENT METHODS

### 1) ANALYSIS ON FEATURE SEPARABILITY

To measure the linear separability of the different features used in the above discrimination methods, RBTW is adopted

as the criterion. We randomly select 200 target chips and 200 clutter chips from dataset for Sentinel-1 to calculate the value of RBTW. In order to ensure the experiment accuracy, 5 experiments are conducted. Fig. 8 shows the RBTW values of different features and its average values. From Fig. 8, we can see that the RBTW value of the features based on BOW models are much larger than those of the classical discrimination features. Specifically, the RBTW value of features based on our proposed model is a little larger than those of the features based on SIFT-BOW and MultiF-BOW models. That is to say, our proposed features can linearly separate the target chips from the clutter chips more easily which is beneficial to the final discrimination.

### 2) ANALYSIS ON DISCRIMINATION RESULTS OF DIFFERENT METHODS

To quantitatively compare the discrimination performance of different methods, we conducted 100 discrimination experiments by using the datasets for Sentinel-1 and Gaofen-3, respectively. The quantitative comparison results are shown in Table 3 and Table 4.

TABLE 3. Performance comparison of different methods using the dataset for Sentinel-1.

Methods	$p_d$	$p_f$	$p_c$
Old Lincoln	0.755	0.309	0.723
New Lincoln	0.788	0.381	0.704
GAO	0.805	0.320	0.743
SIFT-BOW	0.823	0.206	0.809
MultiF-BOW	0.863	0.192	0.836
MF-SPM-BOW	<b>0.885</b>	<b>0.173</b>	<b>0.856</b>

TABLE 4. Performance comparison of different methods using the dataset for Gaofen-3.

Methods	$p_d$	$p_f$	$p_c$
Old Lincoln	0.696	0.342	0.677
New Lincoln	0.728	0.337	0.696
GAO	0.820	0.325	0.748
SIFT-BOW	0.837	0.195	0.821
MultiF-BOW	0.892	0.161	0.866
MF-SPM-BOW	<b>0.917</b>	<b>0.139</b>	<b>0.889</b>

Table 3 and Table 4 denote that the method based on SIFT-BOW model, the method based on MultiF-BOW model and our proposed MF-SPM-BOW method, which are all implemented by employing the BOW models, have much higher  $p_d$  and  $p_c$  than the Old Lincoln method, the New Lincoln method and the GAO's method, which are implemented by using the traditional target discrimination features. Higher  $p_d$  means the number of missed target chips is small and most of targets are detected. And higher  $p_c$  denotes smaller missed targets and false alarms. As shown in Table 3 and Table 4, the proposed method has larger  $p_c$  than both the method based on SIFT-BOW model and the method based on MultiF-BOW model. In addition, by comparing Table 3 with Table 4, we can see that the methods

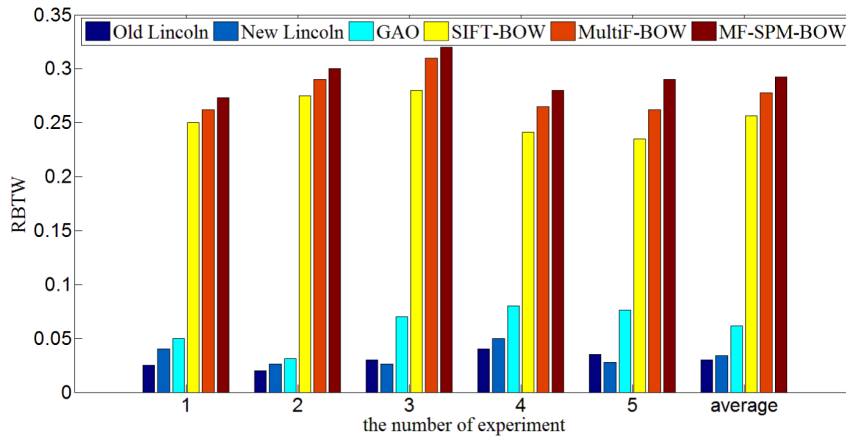


FIGURE 8. RBTW values of different discrimination features in the five experiments.

based on BOW models performs better when using Gaofen-3 image, which have a higher resolution. Thus it validates that the methods based on BOW models can obtain the detail information of high-resolution images when extracting discrimination features, while the traditional methods do not perform well when dealing with high-resolution images.

To compare the performance of different discrimination methods more comprehensively, we give the receiver operating characteristic (ROC) curves in Fig. 9 and Fig. 10, where x-coordinate is the probability of detection  $p_d$  and y-coordinate is probability of false alarm  $p_f$ . As shown in Fig. 9 and Fig. 10, the ROC curve of our proposed method is the best among all discrimination methods. Based on these results, we can draw the conclusion that our proposed method based on MF-SPM-BOW model has superior discrimination performance than other methods.

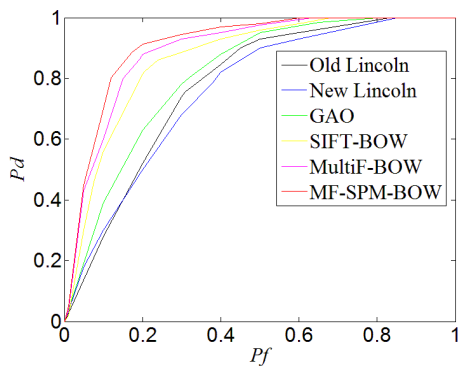


FIGURE 9. ROC curves of discrimination performance for different methods using the dataset for Sentinel-1.

The reasons for the superior performance of the proposed method are summarized as follows.

First, we use the BOW model to extract discrimination features. The traditional discrimination methods employ features which only give a rough description of target and clutter and neglect the difference in detail, thus it cannot achieve considerable performance as the methods based on

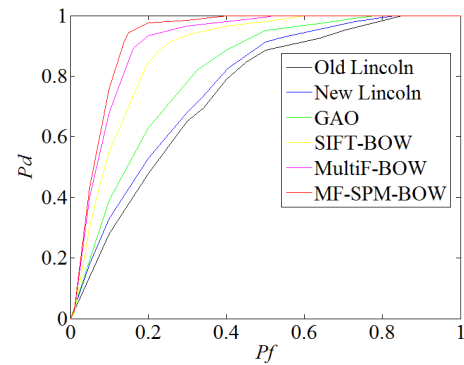


FIGURE 10. ROC curves of discrimination performance for different methods using the dataset for Gaofen-3.

BOW models, namely SIFT-BOW model, MultiF-BOW model and MF-SPM-BOW model.

Second, we combine the SAR-SIFT descriptors and GLCM descriptors together to extract local features. As SAR-SIFT descriptors mainly reflect the gradient information of image chips and GLCM descriptors represent the texture information of image chips, employing those features as local feature can give a more comprehensive description of the SAR image chips. Therefore our proposed method obtained better performance than the SIFT-BOW model.

Third, we employ SPM to generate global features. As the spatial location information of the image chips is very essential to the discrimination of target and clutter, applying SPM when generating global features could lead to the superior separability of the extracted discrimination features. Thus compared to MultiF-BOW model, our proposed MF-SPM-BOW model performs better in discrimination.

### 3) ANALYSIS ON COMPUTATIONAL LOADS OF DIFFERENT METHODS

We select one candidate image chip from the dataset for Sentinel-1 as test sample to compare the computational loads

of different methods. The testing time includes feature extraction time and classifier discrimination time. We run the codes on a computer with 32-GB RAM and an Inter Core i5-10210U CPU@ 3.7 GHz. The codes are written in MATLAB 2016b. Table 5 shows the testing time of individual sample for different methods.

As shown in Table 5, the methods based on BOW models cost much more time than the traditional discrimination methods. The reason is that feature extraction stage of those based on BOW models is much complex and cost more time. As GLCM descriptors extraction and feature pooling based on SPM model are added in the feature extraction stage, our proposed MF-SPM-BOW model takes a little more time than SIFT-BOW model does. Although the proposed method is time-consuming, it only takes approximately 0.33 times longer and it has little effect on the operation of the algorithm. When compared to MultiF-BOW model, the proposed MF-SPM-BOW model cost little time. MultiF-BOW model designs local feature descriptors based on traditional discrimination features and the feature extraction process is much more complex, so the testing time of it is the longest.

**TABLE 5. Testing time of individual sample for different methods.**

Methods	Running time(s)
Old Lincoln	0.38
New Lincoln	0.21
GAO	0.23
SIFT-BOW	0.51
MultiF-BOW	<b>0.71</b>
MF-SPM-BOW	0.67

## V. CONCLUSION

This article proposes a ship target discrimination method based on MF-SPM-BOW model in complex scenes. In the proposed method, on the basis of extracting SAR-SIFT descriptors, we employ GLCM descriptors as low-level features to give a comprehensive description of texture characteristic of image chips. Besides, we use SPM model to capture the spatial location information in the global feature generation stage. Finally we apply a SVM based on multiple kernel learning to complete feature fusion in image layer. Experimental results show the obvious discrimination performance improvement of the proposed method compared to the other methods. In the future work, we will try to employ the deep learning to discriminate target and clutter as the deep learning has great potential in image classification.

## REFERENCES

- [1] X. Leng, K. Ji, S. Zhou, X. Xing, and H. Zou, "Discriminating ship from radio frequency interference based on noncircularity and non-gaussianity in Sentinel-1 SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 352–363, Jan. 2019.
- [2] L. Wan, T. Zhang, Y. Xiang, and H. You, "A robust fuzzy C-means algorithm based on Bayesian nonlocal spatial information for SAR image segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 896–906, Mar. 2018.
- [3] X. Wang and C. Chen, "Ship detection for complex background SAR images based on a multiscale variance weighted image entropy method," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 2, pp. 184–187, Feb. 2017.
- [4] T. Li and L. Du, "Target discrimination for SAR ATR based on scattering center feature and K-center one-class classification," *IEEE Sensors J.*, vol. 18, no. 6, pp. 2453–2461, Mar. 2018.
- [5] W. Ao, F. Xu, Y. Li, and H. Wang, "Detection and discrimination of ship targets in complex background from spaceborne ALOS-2 SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 2, pp. 536–550, Feb. 2018.
- [6] D. E. Kreithen, S. D. Halversen, and G. J. Owirka, "Discriminating targets from clutter," *Lincoln Lab. J.*, vol. 6, no. 1, pp. 25–51, Jan. 1993.
- [7] S. M. Verbout, A. L. Weaver, and L. M. Novak, "New image features for discriminating targets from clutter," in *Proc. Radar Sensor Technol. III*, Aug. 1998, pp. 120–137.
- [8] G. Gao, "An improved scheme for target discrimination in high-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 277–294, Jan. 2011.
- [9] Q. Zhu, Y. Zhong, B. Zhao, G.-S. Xia, and L. Zhang, "Bag-of-visual-words scene classifier with local and global features for high spatial resolution remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 6, pp. 747–751, Jun. 2016.
- [10] J. Feng, L. C. Jiao, X. Zhang, and D. Yang, "Bag-of-visual-words based on clonal selection algorithm for SAR image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 691–695, Jul. 2011.
- [11] S. Xu, T. Fang, D. Li, and S. Wang, "Object classification of aerial images with bag-of-visual words," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 2, pp. 366–370, Apr. 2010.
- [12] R. Bahmanyar, S. Cui, and M. Datcu, "A comparative study of bag-of-words and Bag-of-Topics models of EO image patches," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 6, pp. 1357–1361, Jun. 2015.
- [13] W. Yao, O. Loffeld, and M. Datcu, "Application and evaluation of a hierarchical patch clustering method for remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 6, pp. 2279–2289, Jun. 2016.
- [14] D. GLEICH and D. SIPOS, "Categorization based on sparse coding for SAR patch categorization," in *Proc. Eur. Conf. Synth. Aperture Radar*, Hamburg, Germany, 2016, pp. 1–4.
- [15] S. Cui, G. Schwarz, and M. Datcu, "Remote sensing image classification: No features, no clustering," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 11, pp. 5158–5170, Nov. 2015.
- [16] Y. Wang and H. Liu, "SAR target discrimination based on BOW model with sample-reweighted category-specific and shared dictionary learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 2097–2101, Nov. 2017.
- [17] W. Song, W. Yinghua, S. Lihui, L. Hongwei, and B. Zheng, "SAR target discrimination algorithm based on bag-of-words model with multi-feature fusion," *J. Electron. Inf. Technol.*, vol. 39, no. 11, pp. 2705–2715, Nov. 2017.
- [18] Z. Wang, L. Du, and H. Su, "Superpixel-level target discrimination for high-resolution SAR images in complex scenes," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 9, pp. 3127–3143, Sep. 2018.
- [19] L. Du, H. Dai, Y. Wang, W. Xie, and Z. Wang, "Target discrimination based on weakly supervised learning for high-resolution SAR images in complex scenes," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 461–472, Jan. 2020.
- [20] L. Du, Y. Wang, W. T. Xie, and Z. C. Wang, "A semisupervised infinite latent Dirichlet allocation model for target discrimination in SAR images with complex scenes," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 666–678, Jan. 2020.
- [21] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
- [22] F. Dellinger, J. Delon, Y. Gousseau, J. Michel, and F. Tupin, "SAR-SIFT: A SIFT-like algorithm for SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 453–466, Jan. 2015.
- [23] Q. An, Z. Pan, and H. You, "Ship detection in Gaofen-3 SAR images based on sea clutter distribution analysis and deep convolutional neural network," *Sensors*, vol. 18, no. 2, p. 334, Jan. 2018.
- [24] X. Yang and J. Ding, "A computational framework for iceberg and ship discrimination: Case study on Kaggle competition," *IEEE Access*, vol. 8, pp. 82320–82327, May 2020.

- [25] C. P. Schwegmann, W. Kleynhans, B. P. Salmon, L. W. Mdakane, and R. G. V. Meyer, "Very deep learning for ship discrimination in synthetic aperture radar imagery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2016, pp. 104–107.
- [26] X. Li, L. Zhang, L. Wang, and X. Wan, "Effects of BOW model with affinity propagation and spatial pyramid matching on polarimetric SAR image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 7, pp. 3314–3322, Jul. 2017.
- [27] F.-X. Lu and J. Huang, "Beyond bag of latent topics: Spatial pyramid matching for scene category recognition," *Frontiers Inf. Technol. Electron. Eng.*, vol. 16, no. 10, pp. 817–828, Oct. 2015.
- [28] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, Apr. 2011.
- [29] D. A. Clausi, "An analysis of co-occurrence texture statistics as a function of grey level quantization," *Can. J. Remote Sens.*, vol. 28, no. 1, pp. 45–62, Jan. 2002.
- [30] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.
- [31] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *Proc. 33rd IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 3360–3367.
- [32] M.-S. Yang and K. P. Sinaga, "A feature-reduction multi-view k-means clustering algorithm," *IEEE Access*, vol. 7, pp. 114472–114486, Aug. 2019.
- [33] S. S. Bucak, R. Jin, and A. K. Jain, "Multiple kernel learning for visual object recognition: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 7, pp. 1354–1369, Jul. 2014.
- [34] L. Huang, B. Liu, B. Li, W. Guo, W. Yu, Z. Zhang, and W. Yu, "Open-SARShip: A dataset dedicated to sentinel-1 ship interpretation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 1, pp. 195–208, Jan. 2018.
- [35] X. Sun, Z. Wang, and Y. Sun, "AIR-SARShip-1.0: high-resolution SAR ship detection dataset," *J. Radars*, vol. 8, no. 6, pp. 852–862, 2019, doi: [10.12000/JR19097](https://doi.org/10.12000/JR19097).
- [36] L. M. Novak, S. D. Halversen, G. J. Owirka, and M. Hiatt, "Effects of polarization and resolution on the performance of a SAR automatic target recognition system," *Lincoln Lab. J.*, vol. 8, no. 1, pp. 49–68, 1995.
- [37] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.



**SHIYUAN CHEN** received the B.S. degree in electronic science and technology and the M.S. degree in image interpretation from PLA Information Engineering University, Zhengzhou, China, in 2013 and 2016, respectively. She is currently pursuing the Ph.D. degree in armament science and technology with Space Engineering University, Beijing, China.

Her research interests include remote sensing image processing and SAR image interpretation, especially on ship target detection and discrimination.



**XIAOJIANG LI** received the B.S. degree in computer science and technology and the M.S. and Ph.D. degrees in computer software and theory from Northwestern Polytechnical University, Xi'an, China, in 1995, 1998, and 2002, respectively.

From 2003 to 2008, he was an Associate Professor with the Department of Space Equipment, Space Engineering University. Since 2008, he has been a Professor with the Department of Aerospace Science and Technology, Space Engineering University. His research interests include SAR image processing and pattern recognition.



**SHAOQUAN CHI** received the B.S. degree in electronic science and technology from PLA Information Engineering University, Zhengzhou, China, in 2013.

Since 2013, he has been a Teaching Assistant with the Service Support Center, Shijiazhuang Division of PLAA Infantry College, Shijiazhuang, China. His research interests include machine learning and data mining.



**ZHILIANG LI** (Member, IEEE) received the B.S. and M.S. degrees in image interpretation from PLA Air Force Aviation University, Changchun, China, in 2011 and 2013, respectively, and the Ph.D. degree in information and communication engineering from Space Engineering University, Beijing, China, in 2017.

Since 2017, he has been a Lecturer with Space Engineering University. His research interests include image interpretation, machine learning, and evolutionary computing.



**MAO YUXING** (Member, IEEE) received the B.S. degree from Army Logistics University, Wuhan, China, in 2014, and the M.S. degree from the PLA Naval University of Engineering, Wuhan, in 2016. He is currently pursuing the Ph.D. degree in armament science and technology with Space Engineering University, Beijing, China.

His research interests include computer vision and remote sensing image processing, especially on objection detection and instance segmentation.

• • •