

## Research Article

# SAR Image Target Recognition Based on Monogenic Signal and Sparse Representation

XiuXia Ji <sup>1</sup> and Yinan Sun<sup>2</sup>

<sup>1</sup>Nanjing Vocational College of Information Technology, Nanjing 210023, China

<sup>2</sup>Wuhan University, School Electronic Information, Wuhan 430072, China

Correspondence should be addressed to XiuXia Ji; [jixx@njcit.cn](mailto:jixx@njcit.cn)

Received 4 November 2020; Revised 27 December 2020; Accepted 31 December 2020; Published 18 January 2021

Academic Editor: Fangqing Wen

Copyright © 2021 XiuXia Ji and Yinan Sun. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

It is necessary to recognize the target in the situation of military battlefield monitoring and civilian real-time monitoring. Sparse representation-based SAR image target recognition method uses training samples or feature information to construct an overcomplete dictionary, which will inevitably affect the recognition speed. In this paper, a method based on monogenic signal and sparse representation is presented for SAR image target recognition. In this method, the extended maximum average correlation height filter is used to train the samples and generate the templates. The monogenic features of the templates are extracted to construct subdictionaries, and the subdictionaries are combined to construct a cascade dictionary. Sparse representation coefficients of the testing samples over the cascade dictionary are calculated by the orthogonal matching tracking algorithm, and recognition is realized according to the energy of the sparse coefficients and voting recognition. The experimental results suggest that the new approach has good results in terms of recognition accuracy and recognition time.

## 1. Introduction

As a new kind of reconnaissance remote sensing device, SAR is widely used in aerial and space reconnaissance, monitoring, and intelligent tracking of moving aerial targets [1–3]. UAVs are widely used in military surveillance, smart home monitoring, and target tracking. UAV-borne SAR has become an important development direction of UAV remote sensing earth observation technology. In the situation of military battlefield monitoring and smart cities monitoring, it is necessary to classify and identify the target. SAR image target recognition mainly refers to radar detection of targets, processing of echo information, and determination of target attributes, categories, or types. The spectral property for SAR images is determined by the back-scattered signal that is simply the back reflected part of the microwaves scattered from the land cover. Since the back-scattered signal is very weak, it is very difficult to distinguish different types of targets in SAR images. At the same time, the inherent speckle

noise will play a vital role in information extraction from SAR images. Being affected by the inherent speckle noise, SAR images are inferior in readability. In addition, the image features change tremendously as slight fluctuations of imaging parameters or the variation of surroundings which will affect the accuracy and speed of SAR image target recognition.

SAR image target recognition mainly consists of three stages: image processing, feature extraction, and classifier design. The purpose of image processing is to remove speckle noise, segment SAR image, and make it easy to extract features and recognize the image target. Feature extraction directly affects the accuracy of SAR image target recognition. Features used for image classification usually include principal component analysis (PCA), generalized 2-dimensional principal component analysis (G2DPCA), independent component analysis (ICA), and wavelet. For two-dimensional images, monogenic signal perfectly reproduces the monogenic amplitude of the signal energy, monogenic phase of

the signal structure information, and monogenic orientation of the signal geometry information, which has been widely used in the field of image processing [4–6].

Similarly, there are many classifiers for SAR image recognition, such as support vector machine (SVM) [7],  $k$ -nearest neighbor (KNN) [8], and neural network (NN) [9]. In the above recognition algorithms, in order to ensure the recognition performance, KNN classifier theoretically requires an infinite number of training samples, which is obviously difficult to meet in practical application. SVM classifier transforms the linear inseparable problem into a linear divisible problem by using spatial projection. The huge amount of computation caused by training seriously affects the recognition speed of SAR image targets. NN classifier uses the training samples to learn the parameters and weights of training networks. When the categories and numbers of training samples are large, the corresponding computation is also very large, which will lead to the lack of convergence in the training process. Therefore, the research of SAR image target recognition algorithm urgently needs to inject new elements.

In recent years, sparse representation of image signals has been widely concerned in the field of pattern recognition. Wright J first proposed a sparse representation-based classifier (SRC); it constructed an overcomplete dictionary with multiple kinds of training samples with label information and classified by using sparse representation coefficients of testing samples on the dictionary [10]. Now, sparse representation has been widely applied to face recognition [11–13], the direction of arrival estimation, tensor modeling and angle estimation [14–16], and SAR image target recognition [17, 18]. Extended maximum average correlation height (EMACH) filter is actually a kind of template filter; it is widely applied to the recognition of specific military targets because of its high matching ability and strong antinoise ability. EMACH combined with exponential wavelet fractal feature and G2DPCA feature is applied to target detection of SAR image, and EMACH combined with G2DPCA feature is used to complete SAR image target recognition.

SAR image target recognition algorithm based on sparse representation is designed from two aspects. The one is using dictionary learning to complete recognition. There are two main methods; one is to train dictionaries directly to make them discriminant. That is to say, the dictionary should be designed according to the training sample and has certain adaptability. If the features of training samples are directly extracted to construct overcomplete dictionary, the dictionary dimension is high and the redundancy is large, which will affect the speed of testing samples recognition. The second is to study and optimize the dictionary in order to improve its recognition ability. For example, Literature [19] used discriminant KSVD dictionary learning method to complete SAR image target recognition. The other one is using sparse coefficient to complete recognition. Unlike training multicategory dictionaries, this approach only requires training a whole dictionary without paying attention to the category of each dictionary. The recognition algorithm based on this method usually needs to consider adding the recognition error to the cost function when learning the dictionary, so that the trained dictionary has good recognition ability.

In this paper, SAR image target recognition is also studied from the point of dictionary design and sparse coefficient solving. EMACH filter is used to train samples and generate template samples; the monogenic features of all templates are extracted; three subdictionaries are generated according to the monogenic amplitude, the monogenic phase, and the monogenic orientation; the sparse coefficients of the testing samples' monogenic feature in each level dictionary are solved; and the category of samples to be tested is determined by the image reconstruction error.

The main contributions of this paper are summarized as follows:

- (1) The proposed SAR image target recognition method uses sparse representation to identify SAR image target, which eliminates the need for additional suppression of speckle noise, reduces the steps of SAR image processing and saves time
- (2) The proposed SAR image target recognition method uses three subdictionaries to construct an overcomplete cascade dictionary, and it obviously improves the recognition speed in the precondition of ensuring the recognition accuracy

The remainder of this paper is organized as follows. Section 2 gives a brief survey of the related works about SAR image target recognition method based on sparse representation. Section 3 describes the motivation and design of the proposed SAR image recognition method which is the major contribution of this paper. Experimental results and discussion are shown in Section 4. Finally, the concluding remarks are given in Section 5.

## 2. SAR Image Recognition Based on Sparse Representation

When sparse representation is originally used in face recognition, an overcomplete dictionary is mainly constructed by the training samples; the testing sample is expressed as a linear combination of atoms in the overcomplete dictionary. The face image can be accurately recovered according to the sparse representation coefficients, and the recognition of the target is realized according to the distance metric between the reconstructed image and the testing sample image.

There are similarities and differences between SAR image target recognition and face recognition. They are same in constructing the overcomplete dictionary with the training samples. If the testing sample can be represented linearly by the atoms of the overcomplete dictionary, and the coefficient of the corresponding target category in all the sparse representation coefficients is large, then the representation of the testing sample in the dictionary is sparse, and the recognition discrimination is completed according to the energy characteristics of the sparse representation coefficient. SAR image target recognition method based on sparse representation omits the reconstruction process in face recognition method. The framework of the specific recognition method is shown in Figure 1.

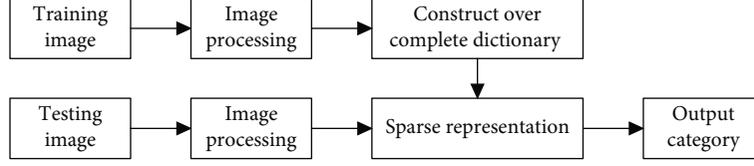


FIGURE 1: The framework of SAR image target recognition based on sparse representation.

In SAR image target recognition, it is assumed that there are  $k$  class samples; column vector set  $D_i = [v_{i,1}, v_{i,2}, \dots, v_{i,n_i}] \in R^{m \times n_i}$  of matrix is composed of  $n_i$  training samples of class  $i$  target. Any kind of testing sample  $y \in R^m$  can be expressed as a linear combination of the training samples.

$$y = \alpha_{i,1} v_{i,1} + \alpha_{i,2} v_{i,2} + \dots + \alpha_{i,n_i} v_{i,n_i}, \quad (1)$$

where  $\alpha_{i,j} \in R (j = 1, 2, \dots, n_i)$  is the coefficient of linear representation of the testing sample in the dictionary.

If the class  $i$  of the testing sample  $y$  is unknown, all  $k$  class  $n$  training sample sets are formed into matrix  $D$ , that is,

$$D = [D_1, D_2, \dots, D_k] = [v_{1,1}, v_{1,2}, \dots, v_{k,n_k}]. \quad (2)$$

The linear representation of test sample  $y$  under all training samples is as follows.

$$y = \alpha_{1,1} v_{1,1} + \alpha_{1,2} v_{1,2} + \dots + \alpha_{k,n_k} v_{k,n_k}, \quad (3)$$

$$y = D\hat{x}, \quad (4)$$

where  $\hat{x} = [0, \dots, 0, \dots, \alpha_{i,1}, \alpha_{i,2}, \dots, \alpha_{i,n_i}, \dots, 0, \dots, 0]^T \in R^n$  is a coefficient vector.

In an ideal case, only the same type of training sample factor as the testing sample may be nonzero in the  $\hat{x}$ ; the corresponding coefficient for the other class samples shall be 0. Since the sparse coefficient  $\hat{x}$  contains the information of the target category, the recognition of the target can be realized by the solution of the formula  $y = D\hat{x}$ .

In the SAR image target recognition method based on sparse representation, the structure of overcomplete dictionary is very critical. The dictionary must be of low dimension, and the atoms of the dictionary should correspond to the properties of the SAR image. At the same time, sparse representation coefficients of different classes of targets over the dictionary must be distinguishable.

Considering the above points, if the overcomplete dictionary is composed of SAR image pixels, that is, the overcomplete dictionary is constructed directly by stretching the central region of the image into a column vector. Because of the high dimension of the dictionary, it will directly affect the speed of solving the subsequent sparse representation coefficients. The overcomplete dictionary can be constructed by using the representation vector which can describe the characteristics of SAR image target. Thus, it can reduce the

dimension of dictionary atoms and improve the speed of sparse solution. In [20], G2DPCA features are used to construct the overcomplete dictionary, which reduces the dimension of extracted features, improves the recognition performance, and has good robustness to the change of target azimuth.

### 3. The Proposed SAR Image Target Recognition Method

The above SAR image target recognition method based on sparse representation takes the feature information extracted by all categories of the training samples or the training samples as the atoms of the overcomplete dictionary. When the number of the training samples is too large, the dictionaries generated by these two methods are too large, which is bound to affect the speed of recognition. In this paper, a cascade dictionary is adopted, that is, subdictionaries are generated according to the monogenic feature of the training samples and the category of the sample to be determined by the reconstruction error. For simplicity, the system framework is divided into the following four stages: image processing, EMACH train, feature extraction, and target recognition. Explanation of each step is described below.

**3.1. Image Processing.** First of all, we extract the region containing the target in the image and generate a new target image based on the biggest scattering point of the target section as the sample image. Since sparse representation can effectively remove noise [21], there is no need for additional noise removal of SAR images in this paper. Take T7 tank image ( $138 \times 138$ ) with angle as a sample. Select the target region of  $45 \times 45$  contained in the center of the image as the sample image (as shown in Figure 2).

It can be seen from the images before and after the processing that the sample image after processing aims at description on the target, the detail information of the target has been enhanced, and the influence of the surrounding clutters on the target is reduced.

**3.2. EMACH Train.** EMACH obtains a two-dimensional function by training the sample image, and then, the correlation response of the image is obtained after doing the correlation operation with the image to be detected of the same size, and the target is judged according to the intensity of the response.

First input images of size  $45 \times 45$  extend the pixel to a one-dimensional vector  $x_i$  with the length of  $45 \times 45$  line by line from left to right and from top to bottom. Define  $h$  as

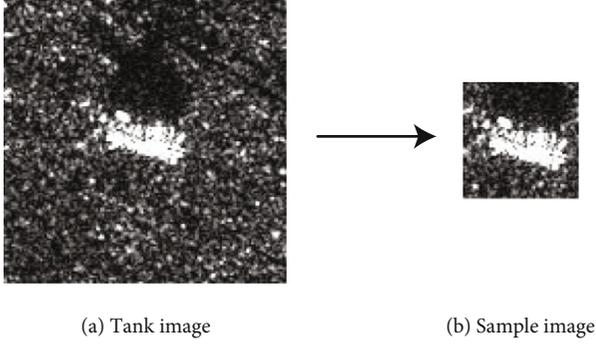


FIGURE 2: Image processing.

the filter of EMACH. FFT2 () represents Fourier operation and assign  $\beta \in (0, 1)$ ,  $M = \text{FFT}(m)$ ,  $X_i = \text{FFT}(x_i)$ , and  $i = 1, 2, \dots, N$ .

$$C_x^\beta = \frac{1}{N} \sum_{i=1}^N [x_i - (1 - \beta)m]^+ [x_i - (1 - \beta)m], \quad (5)$$

$$S_x^\beta = \frac{1}{N} \sum_{i=1}^N [X_i - (1 - \beta)M]^+ [X_i - (1 - \beta)M], \quad (6)$$

where  $m$  is the mean value of  $x_i$ . The symbol of “+” represents matrix transposition.

$$J(h) = \frac{h^+ C_x^\beta h}{h^+ (I + S_x^\beta) h}. \quad (7)$$

When the value of formula (7) is maximum,  $h$  is the EMACH filter.

In this paper, the sample image in MSTAR database is processed. The templates are obtained using EMACH, and all the images within the range of  $12^\circ$  azimuth are selected to train a template, and 30 templates are trained for each type of image, for a total of 150 templates. Figure 3 shows all EMACH template images.

**3.3. Feature Extraction.** The traditional Gabor filter uses an adjustable filter to filter amplitude and phase in different directions and scale. Riesz transform is introduced in the analysis of monogenic signal. The expression of the Riesz transform kernel in the spatial domain is defined as

$$\eta_z = \left( \eta_x, \eta_y \right) = \left( \frac{x}{2\pi \|z\|^3}, \frac{y}{2\pi \|z\|^3} \right). \quad (8)$$

Suppose  $w = (u, v)$ ,  $u, v$  represent two coordinates in the frequency domain, the frequency domain response of Log-Gabor filter is

$$g_{LG}(w) = \exp \left\{ \frac{-[\log(w/\gamma_0)]^2}{2[\log(\sigma/\gamma_0)]^2} \right\}, \quad (9)$$

where  $\gamma_0$  is the central frequency and  $\sigma$  is the scale of Log-Gabor filter. Log-Gabor filters of different scales can be obtained by modify  $\sigma$ .

The band communication number generated by the filter of 2D image  $f$  can be expressed as

$$h = f * F^{-1}(g_{LG}(w)), \quad (10)$$

where  $*$  is the convolution operator and  $F^{-1}$  represents the inverse Fourier transform.

$$h_x = h * \eta_x, \quad (11)$$

$$h_y = h * \eta_y, \quad (12)$$

where  $h$  is the real part of monogenic transformation and  $h_x$  and  $h_y$  are the two imaginary parts.

For a given image  $f$ , the monogenic amplitude  $A$ , the monogenic phase  $\phi$ , and the monogenic orientation  $\theta$  can be calculated by the following formula:

$$A = \sqrt{h^2 + h_x^2 + h_y^2}, \quad (13)$$

$$\phi = \arctan \left( \frac{\sqrt{h_x^2 + h_y^2}}{h} \right), \quad (14)$$

$$\theta = \arctan \left( \frac{h_y}{h_x} \right). \quad (15)$$

Obtain the monogenic feature of the ind template image with  $s$  scale of log-Gabor filter; it can be described as

$$\left\{ A^{i,1}, \phi^{i,1}, \theta^{i,1}, A^{i,2}, \phi^{i,2}, \theta^{i,2}, \dots, A^{i,S}, \phi^{i,S}, \theta^{i,S} \right\}. \quad (16)$$

Pull the monogenic feature into one-dimensional vector and normalize it:

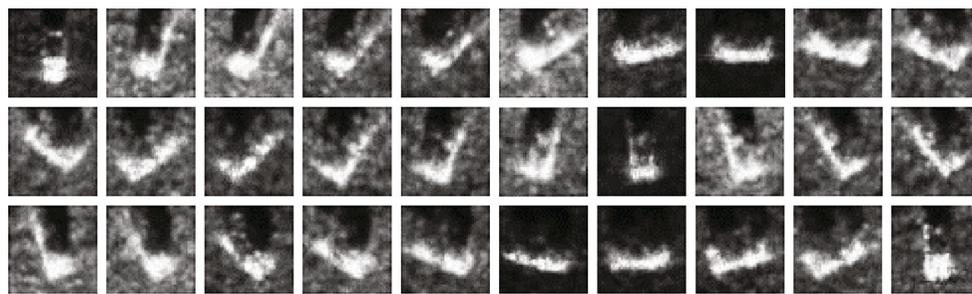
$$D^i = \left\{ \chi_A^{i,1}, \chi_A^{i,2}, \dots, \chi_A^{i,S}, \chi_\phi^{i,1}, \chi_\phi^{i,2}, \dots, \chi_\phi^{i,S}, \chi_\theta^{i,1}, \chi_\theta^{i,2}, \dots, \chi_\theta^{i,S} \right\}. \quad (17)$$

Assuming the total number of the training samples is  $n$ , and the subdictionary  $D_k$  can be expressed as:

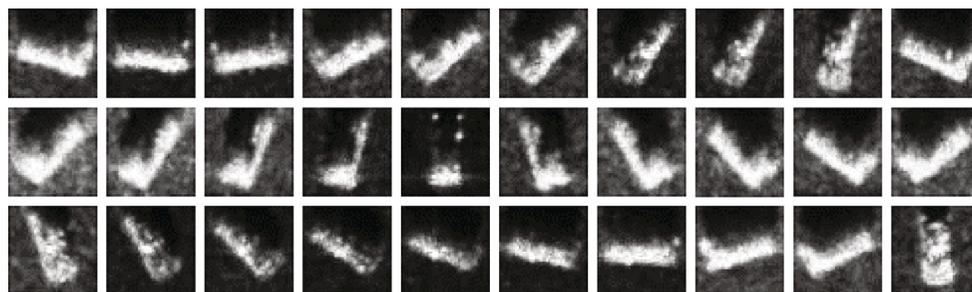
$$D_1 = \left\{ \chi_A^{1,1}, \chi_A^{1,2}, \dots, \chi_A^{1,S}, \chi_\phi^{1,1}, \chi_\phi^{1,2}, \dots, \chi_\phi^{1,S}, \chi_\theta^{1,1}, \chi_\theta^{1,2}, \dots, \chi_\theta^{1,S} \right\}, \quad (18)$$

$$D_2 = \left\{ \chi_\phi^{2,1}, \chi_\phi^{2,2}, \dots, \chi_\phi^{2,S}, \chi_\theta^{2,1}, \chi_\theta^{2,2}, \dots, \chi_\theta^{2,S} \right\}, \quad (19)$$

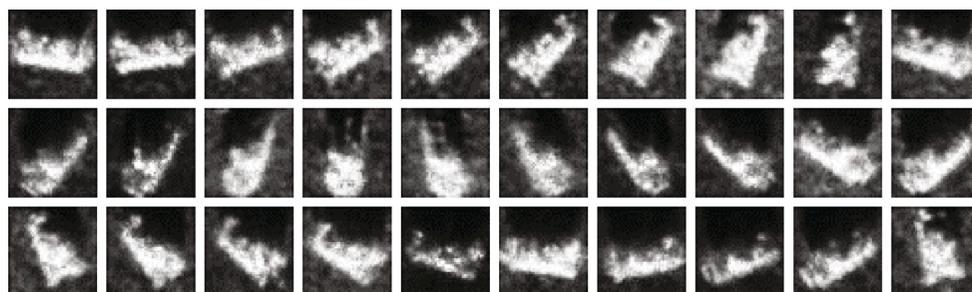
$$D_3 = \left\{ \chi_\theta^{3,1}, \chi_\theta^{3,2}, \dots, \chi_\theta^{3,S}, \chi_\theta^{n,1}, \chi_\theta^{n,2}, \dots, \chi_\theta^{n,S} \right\}. \quad (20)$$



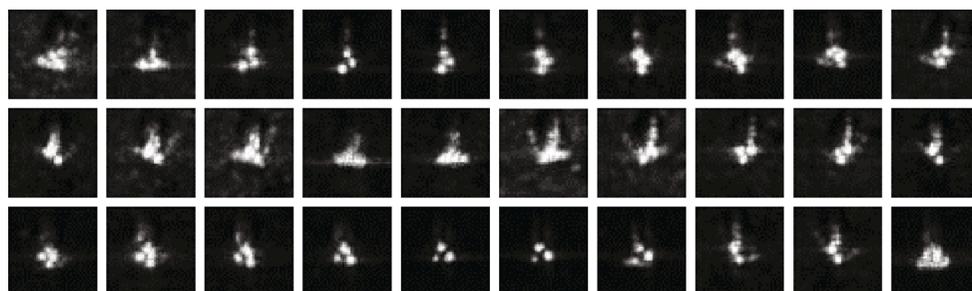
(a) Template images of BRDM2



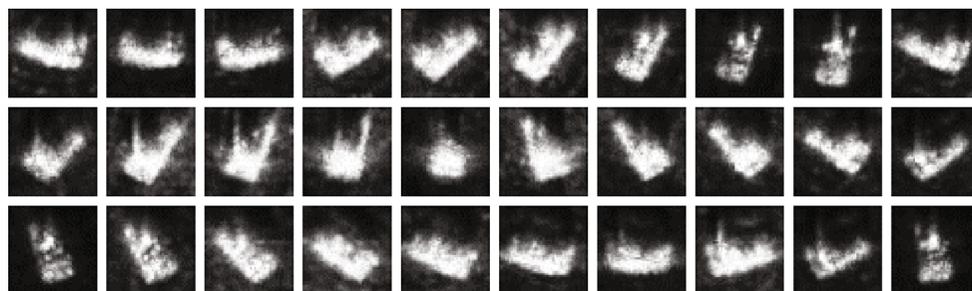
(b) Template images of 2S1



(c) Template images of T72



(d) Template images of SLICY



(e) Template images of ZSU234

FIGURE 3: Template images.

Each subdictionary can be treated as a binary classifier, cascading all subdictionaries into a cascading dictionary.

**3.4. Target Recognition.** In early research, ensembles were shown empirically and theoretically to possess better accuracy than any single component classifier [22]. Since each classifier can only classify the samples of corresponding category, it can be considered a weak classifier. However, a plurality of weak classifiers are cascaded to form a strong classifier with strong recognition capability, and the cascade classifier designed by the hierarchical structure can make most of the samples of the previous categories be recognized directly in the previous classifiers, and only a small part of samples of the latter categories or the missed samples will pass all the subclassifiers, and the recognition time can be obviously reduced.

As stated earlier, the objective of this paper is to develop a cascade classifier that will improve SAR image target recognition accuracy and reduce recognition time. For this purpose, we trained three subdictionaries and then combined them into a cascade dictionary. Because the final classifier uses the cascade structure, the recognition performance of each subclassifier is very important. In this paper, the subclassifiers are designed based on the monogenic features, each monogenic feature generates a subdictionary, the sparse coefficient of the testing sample under each subdictionary is obtained, the testing sample is reconstructed by the coefficient, and the sample category is determined by the reconstruction error. The selection of reconstruction error has a great influence on the recognition performance of subclassifiers. In general, a larger reconstruction error threshold results in a shorter recognition time, but a higher recognition error rate is also generated. The reconstruction error threshold selected in this paper is large. In this way, although there are some samples missed recognition, the last voting mechanism guarantees the correct recognition of this part of samples.

The system framework of the proposed SAR image target recognition method based on monogenic signal and sparse representation is shown in Figure 4.

The main steps of the SAR image target recognition method based on monogenic signal and cascade dictionary are described below.

*Step 1.* Process the training samples.

*Step 2.* Train the samples with the EMACH filter and generate the template samples.

*Step 3.* Extract the monogenic features of the template samples, and generate subdictionaries  $D_i$  ( $i = 1, 2, 3$ ) according to the monogenic amplitude, the monogenic phase, and the monogenic orientation features.

*Step 4.* For any testing sample  $y \in R^m$ , the monogenic feature is extracted as  $X = \{A, \varphi, \theta\}$ . Repeat the following steps until the condition is met.

- (i) For signal  $X_i$  ( $i = 1, 2, 3$ ), use the improved orthogonal matching pursuit algorithm to solve underdetermined linear equations  $X_i = D_i \alpha$ , and find out the most sparse coefficient  $\alpha$

$$\min_{\alpha} \|\alpha\|_1 \text{ s.t. } \|y - D_i \alpha\|_2 < \varepsilon. \quad (21)$$

- (ii) Since all the atoms in the dictionary have labels, the sum of coefficients of the  $k$  class can be calculated:

$$r_k(y) = \text{sum}(|\alpha_{k,i}|), i = 1, 2, \dots, n_k. \quad (22)$$

- (iii) Determine the category of testing sample according to the sum of coefficients

$$\text{identity}(y) = \arg \max_k (r_k(y)). \quad (23)$$

- (iv) If the category is the same as the previous category, directly output category  $k$  and skip to Step 5; otherwise,  $i = i + 1$ ; if  $i > 3$ , skip to Step 5; otherwise, Step 4 is repeated

*Step 5.* If the testing sample falls into different categories under the three subdictionaries, the category is determined by the voting recognition according to the reconstruction error.

$$\text{identity}(y) = \arg \min_k (|X_i - D_i \alpha_k|). \quad (24)$$

## 4. Experimental Results and Analysis

In this section, we describe and discuss the experimental results obtained on the study sites introduced in Section 3. We carried out several experiments with the aim to supply a complete analysis of the performance of the proposed SAR image target recognition method. We investigate different aspects: (i) we perform an experiment to evaluate the recognition accuracy of the proposed recognition method and (ii) we perform an in-depth comparative analysis of the performance of the proposed method with respect to the other recognition methods.

**4.1. Data Set Used for Experiments.** The experimental data used herein is from the SAR ground still data of the SAR ground target high resolution provided by the Working Group of the United States DARPA/AFRL MSTAR. As a result of the SAR imaging, even if the target is same, the azimuth difference can cause the difference of the characteristic information, and the difference of the azimuth angle can

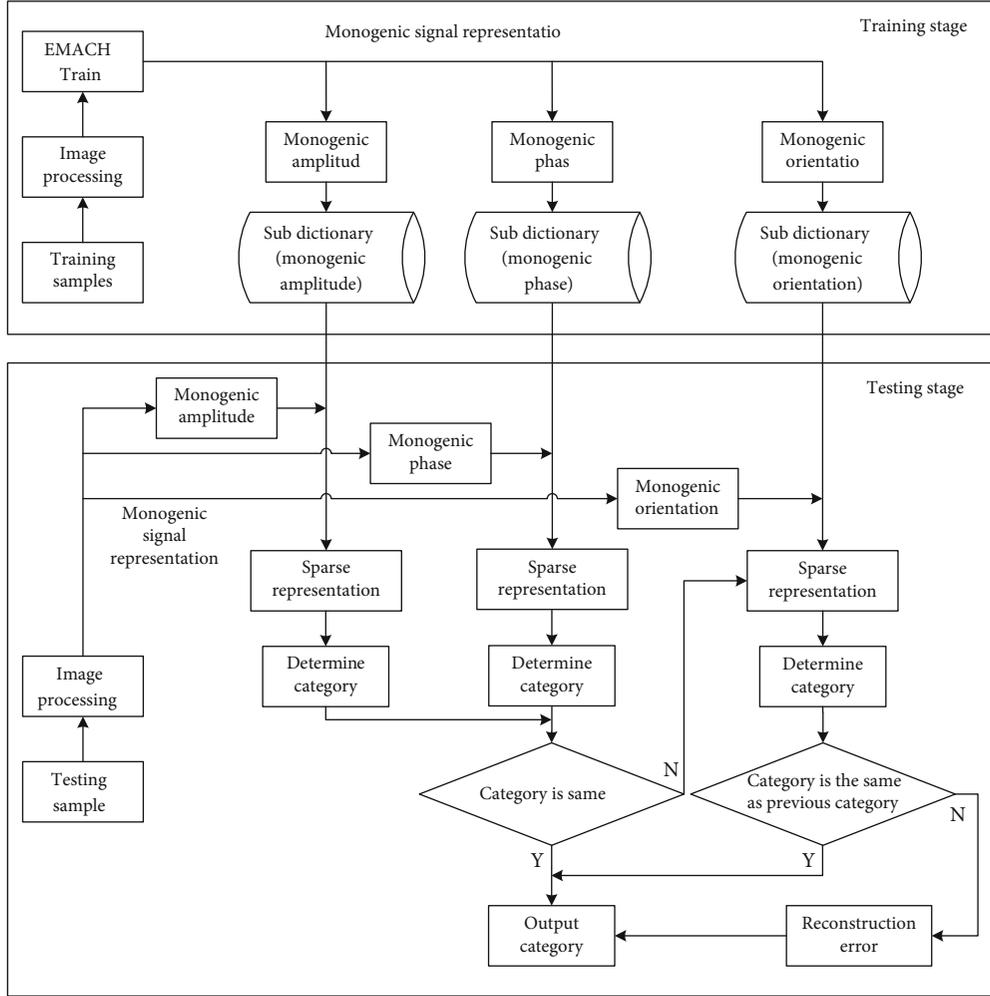


FIGURE 4: The system framework of cascade structure classifier.

cause the difference between the postimaging targets, because the training samples are required to contain all the imaging data at different angles. The SAR image data in the MSTAR database is comprehensive; the azimuth coverage of each target is from  $0^\circ$  to  $360^\circ$ . The experiment selects a subset of MSTAR database, including five types of SAR images: BRDM2, 2S1, T72, SLICY, and ZSU234. The imaging resolution of the image is  $0.3\text{ m} \times 0.3\text{ m}$ , and the azimuth range is from  $0^\circ$  to  $360^\circ$ . In the experiments, the imaging data at  $17^\circ$  are selected as the training samples, and the imaging data when the angle is  $15^\circ$  are used as the testing samples. Since some of the SAR images are noise polluted seriously from certain angle, the amount of images sample for every single target can be used is different. The amount of training samples and testing samples used in this experiment is shown in Table 1.

In order to compare efficiency of proposed method, we establish two characteristics as a comparison basis:

(i) *Recognition accuracy*: proportion of the testing samples correctly recognized by the algorithm.

(ii) *Recognition time*: the time it takes an algorithm to complete a given task.

The experiment was conducted with Matlab-2012b 64bit, installed on windows 10 professional 64bit with an Intel Core i7 Processor (8MCache, up to 3.90 GHz), and 16 GB of RAM.

**4.2. Recognition Performance Analysis.** The monogenic features of each kind of sample images are extracted and drawn into column vectors to form three subdictionaries  $D_i (i = 1, 2, 3)$ , concatenate the subdictionaries, and form a cascade classifier shown in Figure 4. For the testing samples, the region with  $45 \times 45$  size of the target in the center of the image is also selected to realize the segmentation, and the recognition accuracy of each class of testing samples with the proposed method in this paper is shown in Table 2.

The correct sample of primary recognition refers to the samples that do not need to participate in the final voting mechanism recognition, and the correct sample of voting recognition refers to the samples that need to participate in the final voting mechanism recognition, that is, the samples that are missed in cascade recognition. Suppose the total number of samples is  $N$ , correct sample number of primary

TABLE 1: Experimental sample data set.

Target type	Training sample	Testing sample
BRDM2	298	274
2S1	299	274
T72	232	196
SLICY	299	288
ZSU234	299	274

recognition is  $n_1$ , and correct sample number of voting recognition is  $n_2$ .

$$\text{Primary recognition accuracy} = \frac{n_1}{N} \times 100\%, \quad (25)$$

$$\text{Voting recognition accuracy} = \frac{n_2}{N} \times 100\%, \quad (26)$$

$$\text{Recognition accuracy} = \frac{n_1 + n_2}{N} \times 100\%. \quad (27)$$

As can be seen from Table 2, in the design of the classifier, because the reconstruction error threshold is large, about 10% of the samples are missed in the first recognition. However, because of the existence of voting mechanism, this part of the missed samples has been correctly recognized at the time of the final vote.

Additionally, in order to further assess the effectiveness of the proposed recognition method, this paper compared it with the other four methods including the traditional classifiers, such as the SVM recognition method (method 1), the KNN recognition method (method 2), the method with the dictionary directly generated by image pixels (method 3), and the method with the dictionary generated by G2DPCA features in [20] (method 4).

In method 1, SVM is a small sample learning classification method, which has strong generalization ability. Nonlinear processing can be easily realized by introducing kernel function mapping. Suppose monogenic feature is  $x = \{x_1, x_2, \dots, x_M\}$ , where  $M$  is the number of categories of the samples. For the tagged training sets  $(x_i, y_i)$ ,  $i = 1, 2, \dots, M$ ,  $y_i \in \{1, -1\}$ , SVM is to solve the following optimization problems.

$$\min_{w, b, \eta} \frac{1}{2} w^T w + C \sum_{i=1}^K \eta_i \text{ subject to } y_i (w^T f(x_i) + b) \geq 1 - \eta_i, \eta_i \geq 0, \quad (28)$$

where  $C$  is the penalty factor of the error term.

Usually, RBF kernel function is used to map the training vector to high dimensional space.

$$K(x_k, x_g) = \varphi(x_k)^T \varphi(x_g) = \exp(-\gamma \|x_k - x_g\|^2) > 0, \quad (29)$$

where  $\gamma$  is a nuclear factor. In this paper, the kernel function selected is  $k(u, v) = \exp(-\gamma \|u - v\|)$ ,  $\gamma = 5$ .

In method 2, KNN is a simple and effective technique for objects classification according to the closest training examples in the feature space. KNN rates the neighbor of a test sequence among the training sample and uses the class labels of the nearest neighbor to predict the test vector class.

For the tagged training sets  $(x_i, y_i)$ ,  $y_i \in \{1, -1\}$ , the Euclidean distance is often used as the distance metric to measure the similarity between two vectors.

$$d^2(x_i, x_j) = \|x_i - x_j\|^2 = \sum_{k=1}^d (x_{ik} - x_{jk})^2. \quad (30)$$

Parameter  $k$  represents the number of neighbors in a set of training observations which are nearest to the given observation in validation or testing data set. Variation of this parameter will affect the accuracy of each binary classifier inside an expert. In this paper, parameter  $k = 5$ .

In method 3, the dictionary is constructed directly with the image pixels. In method 4, the dictionary is constructed directly with the G2DPCA feature.

Experimental results, for each data set, are represented separately for all methods in Figure 5; the graphical representations of average recognition accuracy and the recognition time are shown in Figures 6 and 7.

As can be seen from Figures 5 to 7, on the premise of the same feature extraction, the recognition accuracy of the SVM classifier is the lowest, and recognition time is the longest. The other two classifiers include the cascade dictionary classifier designed in this paper and the classifier designed to construct the dictionary with the monogenic features of the samples. Their recognition rates are obviously higher than that designed by the dictionary directly generated with the image pixels. The recognition accuracy of the proposed method, method 3, and method 4 are all higher than that of SVM and KNN methods.

A recognition method can not be considered superior to other methods if it requires a great deal of time to yield relatively small improvements. It can be seen from Figures 6 and 7 that by comparing the results of method 4 with that of the proposed method, the recognition accuracy is slightly lower than that of the method 4. However, the average recognition time of method 4 is 19.14 seconds, while that of the proposed method is only 10.75 seconds, the recognition speed is obviously improved. The main reason is that about 86% of the samples are correctly identified at the first recognition and do not need to pass the rest of the dictionaries. In method 4, all the testing samples need to do transvection with all the atoms in the overcomplete dictionary, which obviously slows down the recognition speed.

## 5. Conclusion

In order to solve the problem of higher dimension and larger redundancy in constructing the dictionary with the training samples, this paper proposed a SAR image target recognition method based on monogenic signal and sparse representation. The main innovation points of this method is to train

TABLE 2: The recognition accuracy of the proposed method.

	BRDM2	2S1	T72	SLICY	ZSU234
Total number of samples	274	274	196	288	274
Correct sample number of primary recognition	252	218	165	278	250
Primary recognition accuracy (%)	91.97	79.56	84.18	96.53	91.24
Correct sample number of voting recognition	11	42	20	5	10
Voting recognition accuracy (%)	4.01	15.33	10.20	1.73	3.65
Recognition accuracy (%)	95.98	94.89	94.38	98.26	94.89

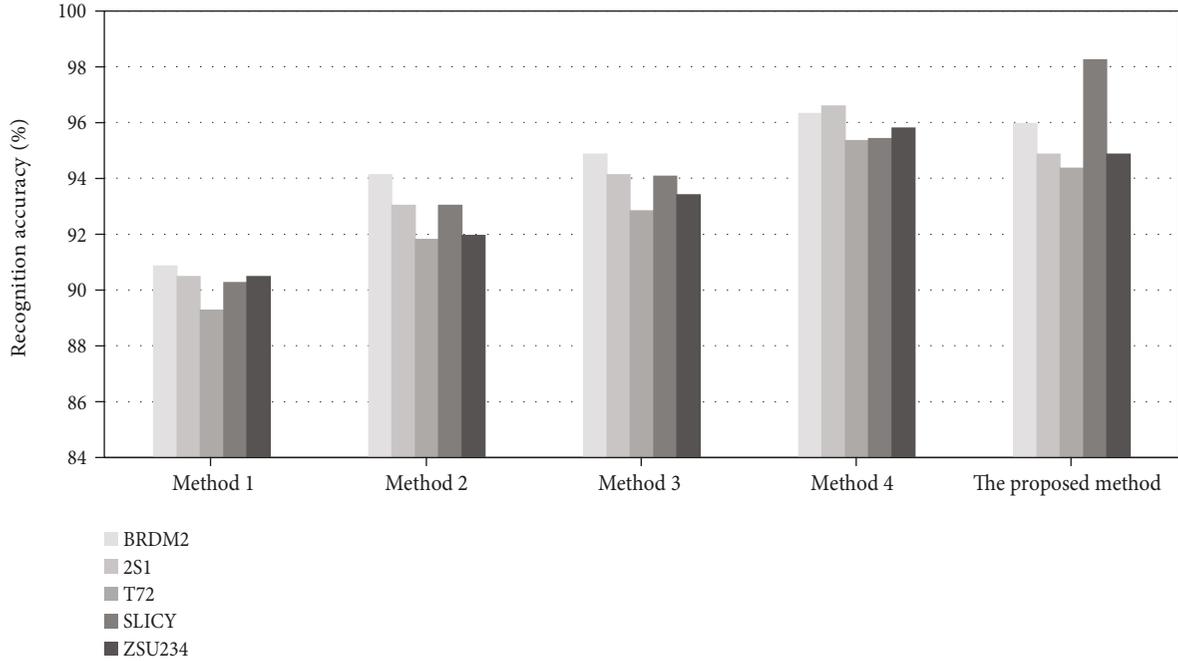


FIGURE 5: Recognition accuracy of different recognition methods.

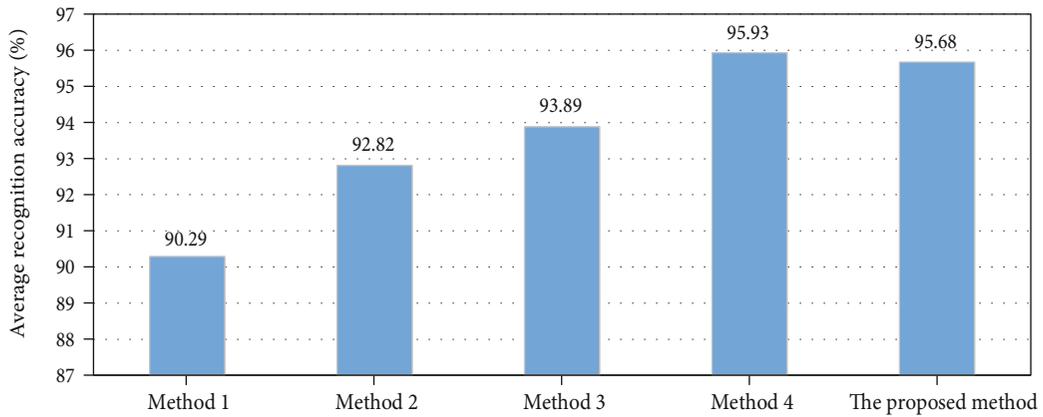


FIGURE 6: Recognition accuracy of different recognition methods.

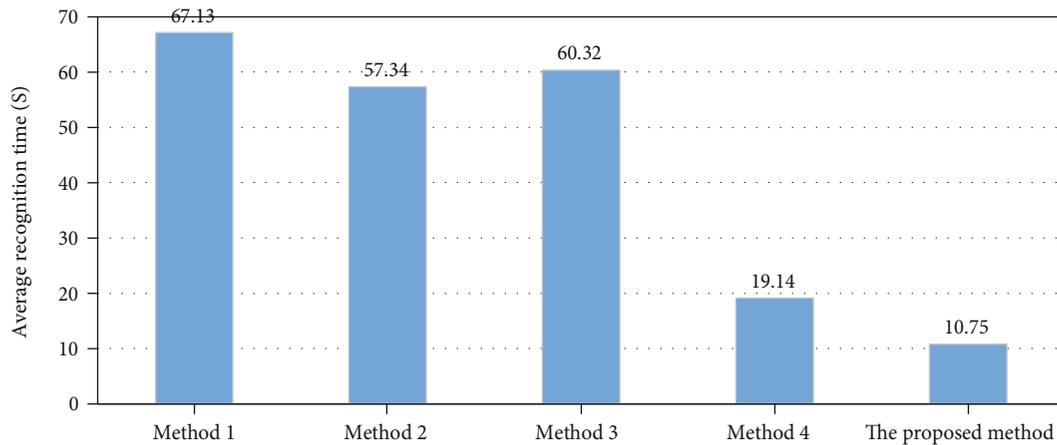


FIGURE 7: Recognition time of different recognition methods.

templates using EMACH filter, generate templates, construct cascade dictionary with the monogenic features of various templates, solve sparse representation coefficients of the testing samples with the proved orthogonal matching pursuit algorithm at each level of dictionary, and determine the category of the testing samples by the reconstruction error of the image. Through experimental results, we have demonstrated that the proposed SAR image target recognition method significantly improved the recognition speed in the precondition of ensuring the recognition rate. Comparison with the traditional dictionary directly generated by G2DPCA feature based on sparse representation indicated that the proposed SAR image recognition method was able to improve the overall accuracy by up to 0.65% and shorten recognition time by up to 43.83%.

### Data Availability

The experimental data used herein is from the SAR ground still data of the SAR ground target high resolution provided by the Working Group of the United States DARPA/AFRL MSTAR.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### Acknowledgments

This study was jointly supported by the Ph.D. Special Fund Project (YB20180101), the Natural Science Research Project of Jiangsu College Universities (19KJB510043), the National Vocational Education Teachers Teaching Innovation Team Support Project (BZ150706), and Jiangsu Vocational Education Teacher Teaching Innovation Team Support Project (YB2020080102). The authors would like to thank the anonymous reviewers and editor for their critical comments and suggestions to improve the original manuscript.

### References

- [1] X. Wang, M. Huang, and L. Wan, "Joint 2D-DOD and 2D-DOA estimation for co-prime EMVS-MIMO radar," *Circuits, Systems, and Signal Processing*, vol. 1, pp. 1–17, 2021.
- [2] T. Liu, F. Wen, L. Zhang, and K. Wang, "Off-grid DOA estimation for colocated MIMO radar via reduced-complexity sparse Bayesian learning," *IEEE Access*, vol. 7, pp. 99907–99916, 2019.
- [3] L. Wan, X. Kong, and F. Xia, "Joint Range-Doppler-Angle estimation for intelligent tracking of moving aerial targets," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 1625–1636, 2018.
- [4] A. S. Alphonse and D. Dharma, "A novel monogenic directional pattern (MDP) and pseudo-voigt kernel for facilitating the identification of facial emotions," *Journal of Visual Communication and Image Representation*, vol. 49, pp. 459–470, 2017.
- [5] X.-G. Luo, H.-J. Wang, and S. Wang, "Monogenic signal theory based feature similarity index for image quality assessment," *AEU - International Journal of Electronics and Communications*, vol. 69, no. 1, pp. 75–81, 2015.
- [6] S. Gai, B. Zhang, C. Yang, and L. Yu, "Speckle noise reduction in medical ultrasound image using monogenic wavelet and Laplace mixture distribution," *Digital Signal Processing*, vol. 72, pp. 192–207, 2018.
- [7] Z. Sun, Y. Qiao, B. P. F. Lelieveldt, and M. Staring, "Integrating spatial-anatomical regularization and structure sparsity into SVM: Improving interpretation of Alzheimer's disease classification," *NeuroImage*, vol. 178, pp. 445–460, 2018.
- [8] B. Wang, Y. L. Huang, J. Y. Yang, and J. J. Wu, "A Feature extraction method for synthetic aperture radar (SAR) automatic target recognition based on maximum interclass distance," *Science China (Technological Sciences)*, vol. 54, no. 9, pp. 2520–2524, 2011.
- [9] H. Xu and G. Srivastava, "Automatic recognition algorithm of traffic signs based on convolution neural network," *Multimedia Tools and Applications*, vol. 79, no. 17-18, pp. 11551–11565, 2020.
- [10] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.
- [11] Y. Su, X. Gao, and X.-C. Yin, "Fast alignment for sparse representation based face recognition," *Pattern Recognition*, vol. 68, pp. 211–221, 2017.

- [12] M. Khorasani, Brain Engineering Research Center, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran, S. Ghofrani, and M. Hazari, "Sparse representation and face recognition," *International Journal of Image Graphics & Signal Processing*, vol. 10, no. 12, pp. 11–20, 2018.
- [13] M. Liao and X. Gu, "Face recognition approach by subspace extended sparse representation and discriminative feature learning," *Neurocomputing*, vol. 373, no. 15, pp. 35–49, 2020.
- [14] X. Wang, L. Wang, X. Li, and G. Bi, "Nuclear norm minimization framework for DOA estimation in MIMO radar," *Signal Processing*, vol. 135, pp. 147–152, 2017.
- [15] J. Shi, F. Wen, and T. Liu, "Nested MIMO radar: coarrays, tensor modeling and angle estimation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 1, p. 99, 2020.
- [16] F. Wen and J. Shi, "Fast direction finding for bistatic EMVS-MIMO radar without pairing," *Signal Processing*, vol. 173, p. 107512, 2020.
- [17] G. Dong and G. Kuang, "Classification on the monogenic scale space: application to target recognition in SAR image," *IEEE Transactions on Image Processing*, vol. 24, no. 8, pp. 2527–2539, 2015.
- [18] G. Dong, H. Liu, G. Kuang, and J. Chanussot, "Target recognition in SAR images via sparse representation in the frequency domain," *Pattern Recognition*, vol. 96, pp. 1–10, 2019.
- [19] H. J. Qi, Y. H. Wang, and J. Ding, "SAR target recognition based on multi-information dictionary learning and sparse representation," *Systems Engineering and Electronics*, vol. 37, no. 6, pp. 1280–1287, 2015.
- [20] X. X. Ji and G. Zhang, "SAR images target recognition based on sparse representation," *Sensors & Transducers*, vol. 161, no. 12, pp. 67–73, 2013.
- [21] X. Ji and G. Zhang, "Contourlet domain SAR image despeckling via self-snake diffusion and sparse representation," *Multimedia Tools and Applications*, vol. 76, no. 4, pp. 5873–5887, 2017.
- [22] Y. Chen, M. L. Wong, and H. Li, "Applying ant colony optimization to configuring stacking ensembles for data mining," *Expert Systems with Applications*, vol. 41, no. 6, pp. 2688–2702, 2014.