# Multi-Layer Model Based on Multi-Scale and Multi-Feature Fusion for SAR Images

**Aobo Zhai [1,2,\*]** (ID)**, Xianbin Wen [1,2,\*], Haixia Xu [1,2], Liming Yuan [1,2] and Qingxia Meng [3]**

[1] Key Laboratory of Computer Vision and System, Ministry of Education, Tianjin University of Technology, Tianjin 300384, China; xuhaixiaxhx@163.com (H.X.); yuanleeming@163.com (L.Y.)

[2] Tianjin Key Laboratory of Intelligence Computing and Novel Software Technology, Tianjin University of Technology, Tianjin 300384, China

[3] School of Computer Science and Technology, Tianjin University, Tianjin 300072, China; 1mengqingxia@163.com

\* Correspondence: aobozhai1@163.com (A.Z.); xbwen317@163.com (X.W.); Tel.: +86-22-6021-6808 (A.Z. & X.W.)

**Abstract:** A multi-layer classification approach based on multi-scales and multi-features (ML–MFM) for synthetic aperture radar (SAR) images is proposed in this paper. Firstly, the SAR image is partitioned into superpixels, which are local, coherent regions that preserve most of the characteristics necessary for extracting image information. Following this, a new sparse representation-based classification is used to express sparse multiple features of the superpixels. Moreover, a multi-scale fusion strategy is introduced into ML–MFM to construct the dictionary, which allows complementation between sample information. Finally, the multi-layer operation is used to refine the classification results of superpixels by adding a threshold decision condition to sparse representation classification (SRC) in an iterative way. Compared with traditional SRC and other existing methods, the experimental results of both synthetic and real SAR images have shown that the proposed method not only shows good performance in quantitative evaluation, but can also obtain satisfactory and cogent visualization of classification results.

**Keywords:** sparse representation classification (SRC); multi-layer structure; multi-feature fusion; multi-scale; SAR image

## 1. Introduction

Synthetic aperture radars (SAR) can obtain stable image data as we are observing the planet Earth. It is not affected by light conditions and can be used day and night under various conditions [1,2]. In recent years, SAR image classification has received more attention as an important part of image understanding and interpretation. A considerable number of image classification algorithms have been proposed, such as support vector machine (SVM) [3], neural network (NN) [4], wavelet decomposition, and sparse representation classification (SRC) [5], etc. Among these existing methods, the traditional SVM and NN methods show high reliability in pattern recognition. However, the relevant computation cost is expensive, and they are easily affected by the selection of features. SRC, which is based on sparse representation and was proposed by Mallat and Zhang [6], has been proven to be an extremely powerful tool in image processing and can obtain good performance in the final processing results [7–15].

The basic ideas of SRC are the linear description hypothesis and spatial joint representation mechanism. This is based on the minimum residual between the original and the reconstruction signal. The sparse coefficients associated with the different classes are selected to reconstruct the original signal. Actually, SRC cannot be directly applied to SAR image classification due to the imaging mechanisms of SAR being different to those of nature imagery. However, if an SAR image is transformed into a

specific feature space, the SRC can be efficiently used in SAR image classification. A joint sparsity model (JSRM) is proposed based on SRC [16], in which the small neighborhood around the test pixel are represented by linear combinations of a few common training samples. Furthermore, the features cannot be represented well on a single scale, which results in the low accuracy of classification results. Neighboring regions of different scales correspond to the same test pixel and they should offer complementary and correlated information for classification. Different sizes of textures in an image have different performance in different scales. The hierarchical sparse representation-based classification (HSRC) [17] can solve the problem in a previous reference [16] to a certain extent, but the HSRC belongs to classification based on each pixel, which only depends on the selection of features in the spatial domain and the selected scale for each layer. This may lead to a loss or misrepresentation of information, resulting in poor classification accuracy and time-consuming training requirements.

In this paper, aiming to overcome the above-mentioned problems, we proposed a novel approach, which is called the multi-layer with multi-scale and multi-feature fusion model (ML–MFM), for SAR image classification. This maintains high accuracy and robustness in addition to having reduced time requirements. Firstly, in order to fix the deficiency of using a single feature and to provide more textural and gray statistical level information [5,12,16], we extracted three types of features of a SAR image for different classes and different scales, which are respectively the gray-level histogram, gray-level co-occurrence matrix (GLCM), and Gabor filter [18–21]. In other words, a discriminative feature vector is composed of the gray-level histogram, GLCM and Gabor filter for each class, while the feature matrix is constructed by the column vector composed of discriminative feature vectors of all classes and row vectors composed of discriminative feature vectors of all scales. Moreover, motivated by the fusion of characteristics from multiple frames in reference [22], a multi-scale fusion strategy was used to construct the dictionary. Thus, the extracted features under different scales can be merged together to construct the column vectors of the dictionary (see Figure 1), which can allow complementation between sample information and reduce the time complexity. Following this, we should segment an SAR image into a host of homogeneous regions called superpixels, with the structural information captured by a discriminative feature vector extraction for each superpixel. Finally, inspired by the idea of layers in the spatial pyramid in reference [21], the multi-layer operation is utilized to refine the classification results by adding a threshold decision condition to SRC in an iterative way. If a superpixel meets the condition as the new atoms in the dictionary, the category is recorded. Otherwise, it will be used as the testing sample for the next layer (Figure 3 depicts the above-mentioned basic framework). Compared with other methods, the final classification results of the proposed method have higher accuracy.
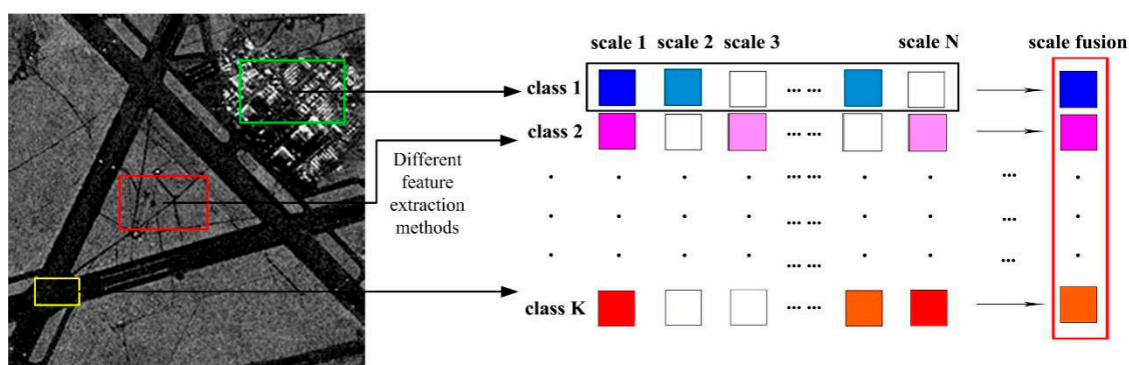


**Figure 1.** The model of multi-scale fusion strategy.

The remainder of this paper is organized as follows. In Section 2, we briefly review the SRC, while the procedure of our novel model regarding the use of ML–MFM for SAR image classification is explored. The experimental results for synthetic and real SAR images are presented and compared

with others in Section 3. Comparison with the HSRC [17] and the major innovation points are provided in Section 4. Finally, conclusions are drawn and future research directions are described in Section 5.

## 2. Materials and Methods

### 2.1. SRC

We assume an SAR image contains $K$ classes. $D_K$ is the $k$th class of the sub-dictionary constructed by concatenating feature vectors of the $k$th class. We can define a dictionary $D$ constructed by $[D_1, D_2, \ldots, D_K]$ for an SAR image. The testing sample $y$ can be formulated by a series of training samples as follows:

$$y = \psi(y) = Dx \in R^M \tag{1}$$

where $\psi(\cdot)$ is an Eigen function which can be used to realize the transformation from pixel to feature space and $x = [0, \ldots, 0, x_{i,1}, \ldots, x_{i,n_i}, 0, \ldots, 0] \in R^n$ is a sparse coefficient vector whose entries are zeros except those associated with the $i$th class. A sparse coefficient $x$ indicates that it will be easier to estimate the identity of the testing sample $y$. A sparse coefficient $x$ can be obtained by solving the following error-constrained Equation (2) or the sparsity-constrained Equation (3):

$$\hat{x} = \text{argmin} \|x\|_0 \text{ subject to } \|y - Dx\|_2 \leq \sigma \tag{2}$$

$$\hat{x} = \text{argmin} \|y - Dx\|_2 \text{ subject to } \|x\|_0 \leq sl \tag{3}$$

where $\sigma$ is the error tolerant limit and $sl$ is the sparsity level which can represent the maximum number of selected atoms in the dictionary. Moreover, $\|\cdot\|_0$ and $\|\cdot\|_2$ denote $l_0$ and $l_2$ norms, respectively. Usually, the problem of solving sparse coefficients can be performed using the orthogonal matching pursuit (OMP) method [23].

After obtaining the sparse coefficient $\hat{x}$, the class label $\hat{k}$ of the test pixel $y$ can be determined by the minimal error between $y$ and its approximation from the sub-dictionary of each class:

$$\hat{k} = \underset{k}{\text{argmin}} \|y - D\hat{x}_k\|_2, \ k = 1, \ldots, K \tag{4}$$

where $\hat{x}_k$ represents the coefficients in $\hat{x}$ belonging to the $k$th class. In order to demonstrate the drawback of the SRC algorithm clearly, a simple experiment was performed on two real SAR images (the original SAR image is in Figures 10a and 12a), observed in Figure 2. We can see that the final SRC results are unacceptable from Figure 2a,b. This is mainly because the SRC algorithm extracts features of SAR images only by using the pixel-by-pixel method, resulting in a lack of complementation between sample information. Therefore, to solve this problem, superpixels and more complete features need to be taken into account.
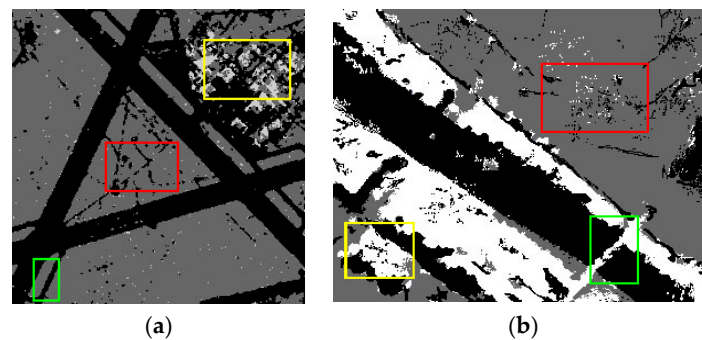


(**a**)　　　　　　　　　(**b**)

**Figure 2.** Classification results with the sparse representation classification (SRC) algorithm of a previous study [5] on (**a**) SAR1 and (**b**) SAR2.

*2.2. Proposed Multi-Layer and Multi-Feature Model (ML–MFM)*

In this section, the multi-layers and multi-feature model (ML–MFM) based on the SRC algorithm is proposed. The basic framework of the proposed method is shown in Figure 3.
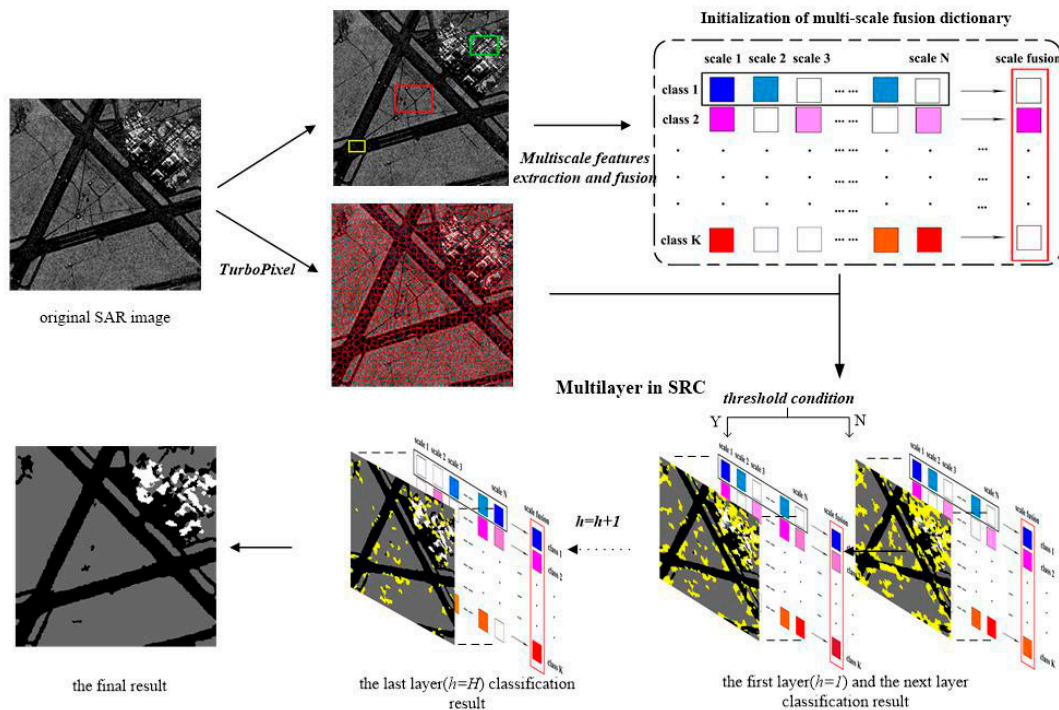


**Figure 3.** The basic framework of multi-layers and multi-feature model (ML–MFM).

Figure 3 can be understood in four parts: superpixel generation, multi-feature extraction, multi-scale with fusion strategy and multi-layer sparse representation classification. In the first stage, the over-segment algorithm is used. The initialization dictionary is subsequently used in different scales, before being fused by the fusion strategy, which is introduced in Section 2.2.2. The first classification is performed using the initialization dictionary and the superpixels. Finally, SRC is used in multi-layers through iterations to obtain the final result.

2.2.1. Multilayer SRC

There are some difference between SAR image classification and face recognition, the training samples of which can be controlled by a human in a standard data set. SAR images contain various complex terrains. It is difficult to guarantee enough training samples to represent each pixel. To deal with the challenge, reference [17] developed a hierarchical sparse representation classifier to improve the classification of the SAR image, which we called the multi-layer SRC.

In this classifier, $h$ $(1 \leq h \leq H)$ represents the layer of classification. Sparse representation is used for each layer. Thus, we performed sparse representation $h$ times. As the number of layers $h$ $(1 \leq h \leq H)$ increases, the classification map becomes closer to the final result. Finally, the number of $h$ is analyzed in Section 3.1.

2.2.2. Multiscale Fusion-Based Dictionary

In the application of sparse representation, a dictionary is first constructed. To counter the existence of speckle and the complex appearance in the SAR image, we transform the pixel value space into a feature space, which reduces the computational complexity and extracts the discriminative features from the SAR image. The gray-level histogram, GLCM and Gabor filter features are extracted

to capture statistical properties in the SAR image. We concentrated on these three types of features to form a feature vector for representing each pixel or superpixel. This non-linear feature will provide competitive performance by representing statistical information and capturing texture information in adjacent areas.

Moreover, different texture features of images show different performances in different scales [24]. In the classification of the image, many experiments have proved that different scales correspond to the same test sample $y$. The information from different scales complements each other, which is useful to classify each pixel.

We selected $n_i$ vectors of the $l$th scale training samples from the $i$th class in the $h$th layer as columns to construct a matrix $A_{i,l}^h = [f_{i,l,1}^h, f_{i,l,2}^h, \ldots, f_{i,l,n_i}^h] \in R^{m \times n_i}$, where $m$ denotes the dimension of the extracted feature vector. Scale $l$ means the scale sizes $(2l + 1) \times (2l + 1)$ of the extracted features. Therefore, $l = 1$ is the finest scale and $l = L$ is coarsest scale, which depends on the resolution of the image. This can be calculated by $L = floor(resolution * 3/2)$. We constructed a matrix by the concatenation of the $n_i$ training sample vectors of all $K$ defined classes and all $L$ scales in the fixed $h$th layer as follows:

$$A^h = \begin{bmatrix} A_{1,1}^h & A_{1,2}^h & \cdots & A_{1,L}^h \\ A_{2,1}^h & A_{2,2}^h & \cdots & A_{2,L}^h \\ \vdots & \vdots & \cdots & \vdots \\ A_{K,1}^h & A_{K,2}^h & \cdots & A_{K,L}^h \end{bmatrix} \tag{5}$$

Following the dictionary $D^h$ in the fixed $h$th layer is defined by row element using the average fusion strategy in matrix $A^h$. This is shown as follows:

$$D^h = \begin{bmatrix} D_{1,fusion}^h \\ D_{2,fusion}^h \\ \vdots \\ D_{K,fusion}^h \end{bmatrix} \xleftarrow{D_{i,fusion}^h = \frac{1}{n_1+n_2+\cdots+n_L} \sum\limits_{l=1}^{L} \sum\limits_{j=1}^{n_L} f_{i,l,j}^h} \begin{bmatrix} A_{1,1}^h & A_{1,2}^h & \cdots & A_{1,L}^h \\ A_{2,1}^h & A_{2,2}^h & \cdots & A_{2,L}^h \\ \vdots & \vdots & \cdots & \vdots \\ A_{K,1}^h & A_{K,2}^h & \cdots & A_{K,L}^h \end{bmatrix} = A^h \tag{6}$$

### 2.2.3. Multi-Layer and Multi-Feature Model

It is well-known that pixels are not natural entities but a result of the discrete representation of an image, with structural information captured in a region rather than a pixel. Furthermore, the computational complexity increases rapidly with an increase in the scale of pixels used. Therefore, we first divided the SAR image into superpixels to integrate the contextual information of neighboring pixels and to reduce computational complexity. In our method, the operation of superpixel generation uses the TurboPixel algorithm [25]. Furthermore, in order to encode gray, textural and spatial information into superpixels, we described each superpixel $sp_t \in sp = \{sp_t | 1 \leq t \leq N\}$ by a $m$ dimensional feature vector $f_{sp_t} = [f_{t,1}, f_{t,2}, \ldots, f_{t,m}]$, in which $N$ is the maximum number of superpixels.

Unlike other recognition methods, SAR image classification lacks training samples to effectively represent each pixel. To deal with the challenge, inspired by the idea of the hierarchical sparse representation [17], we proposed a multi-layer operation based on SRC and the multi-fusion scale dictionary for SAR image classification.

Based on dictionary construction and the superpixel $sp_t$, we used Equation (2) with the $l_0$ norm to ensure a sparse solution. The sparse coefficient can be solved by OMP to obtain $\hat{x}$. We define the $h$th layer of the minimum residual error and class as Equations (7) and (8), respectively.

$$r^h = res_{min}^h(\psi(sp_t)) = \min_{i=1,\ldots,K} res_i^h(\psi(sp_t)) \tag{7}$$

$$\hat{k}^h = \underset{i}{argmin} \|\psi(sp_t) - D^h \hat{x}_i\|_2, i = 1, 2, \ldots, K \tag{8}$$

where the $res_i^h(\psi(sp_t))$ is the residual error in the $h$th layer under the fusion scale; $r^h$ is the minimum residual error; and $c^h$ is the category of $sp_t$. Different from Equation (4), we introduce a parameter $\Delta T$ as the threshold value to limit the superpixel and judge whether it belongs to this class instead. If the $r^h$ is within the specified tolerance limit, the pixel belongs to the current class $i$. Otherwise, the uncertain samples are classified again in the next layer. Following this, the class of superpixels in $h$th ($1 \leq h < H - 1$, where $H$ is layer number) layer are expressed as:

$$label(sp_t) = \begin{cases} \underset{i}{\arg\min} \|\psi(sp_t) - D^h \hat{x}_i\|_2, i = 1, 2, \ldots, K; r^h \leq \Delta T \\ uncertain, \qquad\qquad\qquad otherwise \end{cases} \tag{9}$$

where $r^h \leq \Delta T$ is a restricted condition to ensure that the superpixel belongs to the category. In fact, the choice of the threshold value $\Delta T$ will influence our final results to some extent and it will be further discussed in Section 3.1.

The uncertain superpixels in the $(h + 1)$th $A_i^h = [f_{i,1}^h, f_{i,2}^h, \ldots, f_{i,n_i}^h] \in R^{m \times n_i}$ ($1 \leq h < H$) layer will be classified by Equations (7)–(9). We selected superpixels from each class, which are labeled as the new training samples in the $h$th layer. Following this, we extract the feature vector at the $h$th layer. Arranging these vectors as the columns vector $A_{sp}^h = [A_1^h, A_2^h, \ldots, A_K^h]$, we define the dictionary $D^{(h+1)}$ in the fixed $(h + 1)$th layer based on the dictionary $D^h$ in the fixed $h$th layer.

$$D^{(h+1)} = \begin{bmatrix} D_{1,fusion}^{(h+1)} \\ D_{2,fusion}^{(h+1)} \\ \vdots \\ D_{K,fusion}^{(h+1)} \end{bmatrix} = Average\left(\begin{bmatrix} D_{1,fusion}^h \\ D_{2,fusion}^h \\ \vdots \\ D_{K,fusion}^h \end{bmatrix}, \begin{bmatrix} A_1^h \\ A_2^h \\ \vdots \\ A_K^h \end{bmatrix}\right) = Average(D^h, A_{sp}^h) \tag{10}$$

In fact, the uncertain superpixels decrease with an increase of layers. When $h = H$, there is still a small number of uncertain points (marked by yellow in Figure 3). However, the dictionary $D^H$ is modified by Equation (10), before the remaining pixels will be classified by a traditional sparse repreentation classifier until it outputs the final result. The whole ML–MFM for the SAR image classification Algorithm is summarized as follows:

---

**Algorithm 1.** ML–MFM for synthetic aperture radar (SAR) Image Classification

---

**Input:** SAR image, threshold $\Delta T$, class number $K$, layer $H$.
**Output:** classification map.
1. Segment the SAR image into superpixels by [25].
2. Construct the initial fusion dictionary $D^1$ by Equations (5) and (6), while the fusion dictionary contains $K$ class, $D^1 = [D_{1,fusion}^1, D_{2,fusion}^1, \ldots, D_{K,fusion}^1]$. Choosing a specified number of pixels $n_i$ from the original SAR image as the samples, each sample can be represented by the $m$ dimension extracted variety of features.
3. Multi-layer SRC and dictionary in layers are constructed.

- Classify all superpixels in the first layer by Equations (7) and (8) with orthogonal matching pursuit (OMP);

- Find the best representative atom's label by Equation (9);

    **while** $1 \leq h \leq H$
    **if** $r^h \leq \Delta T$
        $\underset{i}{\min} res_i^h(\psi(sp_t))$;
        $label(sp_t) \leftarrow i$;
        updating dictionary by (10)
        $D^{(h+1)} = Average(D^h, A_{sp})$;
    **else** $uncertain(sp_t^h) \leftarrow sp_t = \psi(sp_t)$;
        $h \leftarrow h + 1$;
    **end while**

---

## 3. Results

In this section, the proposed model is now applied in the classification of synthetic and real SAR images. To validate the performance of the proposed method, we use both types of images in quantitative evaluation and visualization results. We mainly compare our results with the results of previous studies [3,5,16,17], in which their parameters are tuned to obtain the best results. Figures 8a, 9a and 10a are the synthetic SAR images, which are from the Brodatz database. These synthetic SAR images have three, four and five types of different textural regions, while the size of each image is $512 \times 512$, $335 \times 335$, and $512 \times 512$, respectively. The test images are named Syn1, Syn2 and Syn3, respectively. In addition, three real SAR images (SAR1, SAR2 and SAR3) were tested in experiments. SAR1 has a size of $256 \times 256$, which covers the China Lake Airport, California, with a Ku-Band radar with a 3-m resolution. SAR2 has a size of $321 \times 258$, which covers the pipeline over the Rio Grande river near Albuquerque, New Mexico, with a Ku-Band radar with 1-m resolution. SAR3 has a size of $284 \times 284$, which covers the X-Band radar with 3-m resolution. The central processing unit time was obtained by running the Matlab code on a DELL computer with Inter (R) Core (TM) i7CPU, 3.4 GHz, 16 GB RAM with MATLAB 2014(a) on Windows 10 (64-bit operating system) in our experiment.

### *3.1. Experimental Settings*

In the experiment, we used the TurboPixel [25] algorithm to over-segment the original image into homogeneous regions and to obtain the superpixels. As each superpixel has different sizes, the features of each superpixel need to be processed so that the fusion features of all superpixels have the same dimensions (i.e., m = 60 in our method). Sixteen effective distribution features and four statistical features suggested by a previous study [17] were used. Thus, the features extracted by gray-level histogram and GLCM have 16 dimensions and four dimensions for each superpixel, respectively. After calculating the convolution of the initial bank of Gabor filters, which consists of 40 filters with five scales and eight orientations, the mean value of the filter response corresponding to each superpixel was computed for every filter. Therefore, the Gabor feature of each superpixel was a 40-dimension vector corresponding to 40 Gabor filters with the total number of dimensions being 60. In addition, the ranges of the radial basis function kernel width and penalty coefficients are (0.0001, 0.001, 0.01, 0.1, 1, 10) and (0.1, 1, 10, 100, 500, 1000) respectively.

#### 3.1.1. Influence of Parameters

It is necessary to set ideal parameters to obtain satisfactory results. There are two main parameters that need to be set in our model, namely $H$ and $\Delta T$ (threshold). Based on plenty of experimental data and the analysis of results, each parameter should satisfy the following condition: $1 \leq H \leq 6$, $0.07 \leq \Delta T \leq 0.24$. We noted that the parameter $H$ is influenced by the resolution of the SAR images (similar to a previous reference [17]), as mentioned in Section 2.2. We used an experiment to show the influence of $H$, which is depicted in Figure 4. The horizontal axis represents the layer $h$, while the vertical axis represents a certain superpixel number. Here, we artificially set the total superpixel number to 1000 of SAR1. From Figure 4, it is more intuitive to find the most suitable layer range. This is because when $H > 6$, the speed of the growth of certain points slow significantly in histogram and line chart. In many experiments, if we set $H = 6$ as the maximum number of layers, it is the best choice with regards to time and precision. In addition, it is worth noting that $\Delta T$ is the threshold to control the categories of accuracy (blue solid line) and the kappa coefficient (green dotted line) with our proposed method. From Figure 5, we can see that when $\Delta T$ is too small, there are many uncertain superpixels until $h = H$. However, when $\Delta T$ is too large, the finer areas cannot be placed into classes. Therefore, appropriate parameter selection is very important.
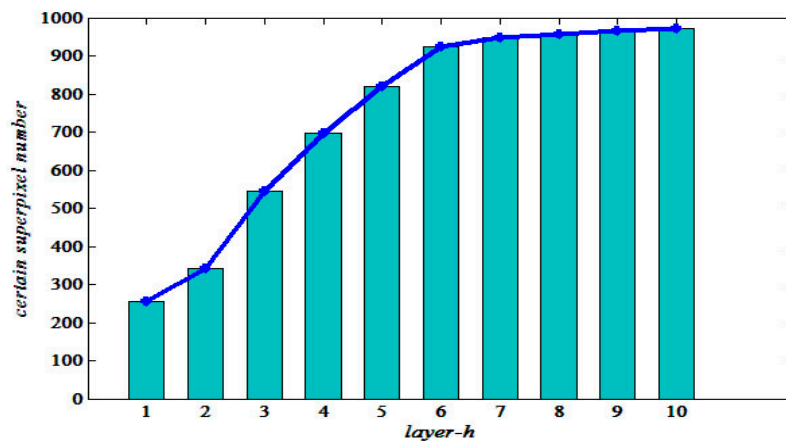
**Figure 4.** Illustration of certain and uncertain superpixels with different layers corresponding to SAR1 (Figure 7a) in our method (the total number of superpixels is 1000).
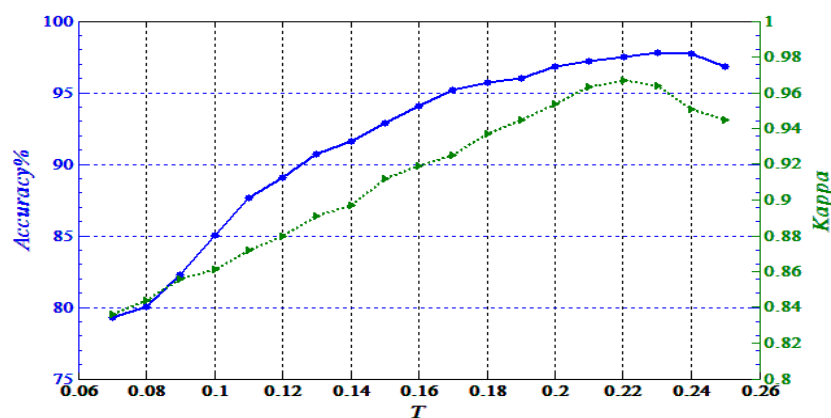


**Figure 5.** Influence of threshold $\Delta T$ in classification accuracy (blue solid line) and Kappa coefficient (green dotted line) corresponding to SAR1 (Figure 7a).

### 3.1.2. Analysis of Multi-Feature Fusion and Multi-Scale Fusion

Multi-Feature Fusion

In this part, the multi-feature fusion is analyzed to verify its effectiveness in obtaining satisfactory results. In our paper, the fusion strategy is introduced to construct the dictionary. We perform an experiment on the original SAR1 (Figure 6a) to show the influence of multi-feature fusion on the dictionary and the impact of classification results. The rectangular areas of Figure 6a–e are marked by red, yellow and green, respectively. Figure 6b shows the results of the method with the gray-level histogram; Figure 6c shows the results of the GLCM; Figure 6d shows the results of the Gabor method; and Figure 6e shows the results of the multi-features method. We can see that Figure 6e has fewer miscellaneous points than Figure 6b–d. The reason is that the fusion features includes distribution features and four statistical features. Therefore, the dictionary $D^h$ includes more information to obtain better results, which is an advantage that is absent in the method with single features. Therefore, the multi-feature fusion strategy is important.
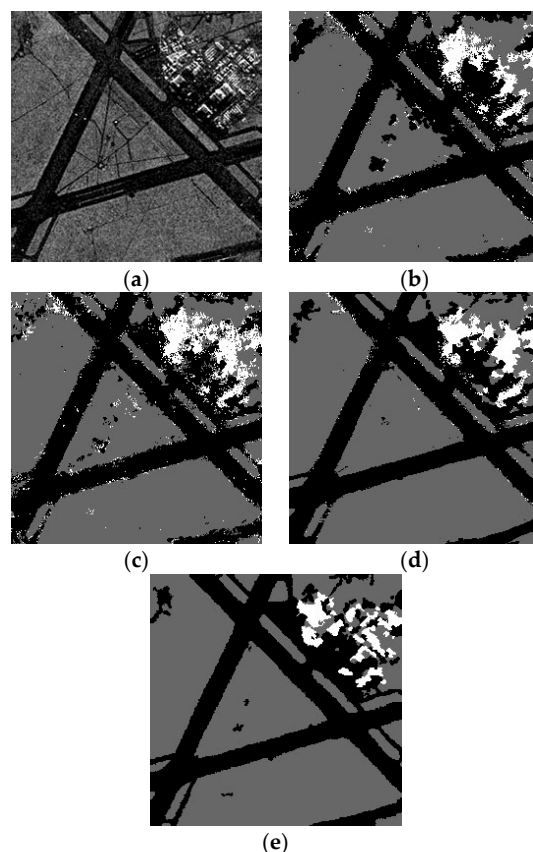
**Figure 6.** Comparison of (**a**) the original SAR1 with (**b**) gray-level histogram; (**c**) gray-level co-occurrence matrix (GLCM); (**d**) with Gabor; (**e**) with multi-features.

Multi-Scale Fusion

In this part, multi-scale fusion is analyzed to verify its effectiveness in obtaining satisfactory results. In our paper, the fusion strategy is introduced to construct the dictionary. We perform an experiment on the original SAR1 (Figure 7a) to show the influence of the fusion strategy on the dictionary after merging features under different scales. The rectangular areas of Figure 7b,c are both marked by red, yellow and green. Figure 7b shows the results method with the fusion strategy; Figure 7c shows the results method without the fusion strategy. We can see that Figure 7b has fewer miscellaneous points than Figure 7c. The reason for this is that the fusion dictionary $D^h$ includes each scale information (homogeneous and marginal regions). This has similar effects on the dictionary under multi-scales, which are absent in the method without the fusion dictionary and are important to SAR image processing. Therefore, the multi-scale fusion strategy is important.
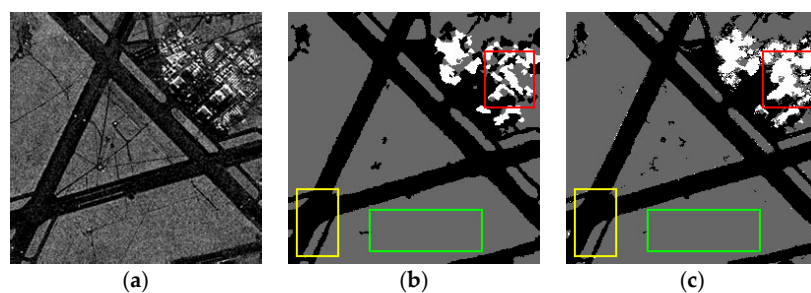


**Figure 7.** Comparison of (**a**) the original SAR1 (**b**) with fusion strategy; and (**c**) without fusion strategy.

## 3.2. Results on Synthetic SAR Images

In this section, we test the capability of the proposed algorithm by applying it to the synthetic SAR images Syn1, Syn2, and Syn3. The superpixels of Syn1, Syn2, and Syn3 are 2800, 1500 and 2800, respectively. In our method, $H = 6$ and $\Delta T$ is 0.221. The scale (patch) size in the support vector machine (SVM) [3], SRC [5] and JSRM [16] is fixed and we set it to be $3 \times 3$. The ground truth was used to calculate the accuracy of the classification results to evaluate the contrast algorithms. We can see that the proposed method can obtain a higher accuracy of classification than the results of previous studies [3,5,16] and can reduce the processing time found in reference [17]. Moreover, as shown Figures 8–10, as well as Table 1, the proposed method can keep the details (edges) in a similar way to reference [17] in the visual representation. However, the results of the other methods in finer textural regions (marked with pink and yellow), such as Figure 9e–g, have significantly different degrees of error, which is caused by the lack of samples. Although our method has no significant improvement in accuracy compared to the method in reference [17], there are benefits to not requiring an extensive amount of time in pixel-by-pixel training and having less miscellaneous points existing in the final classification.
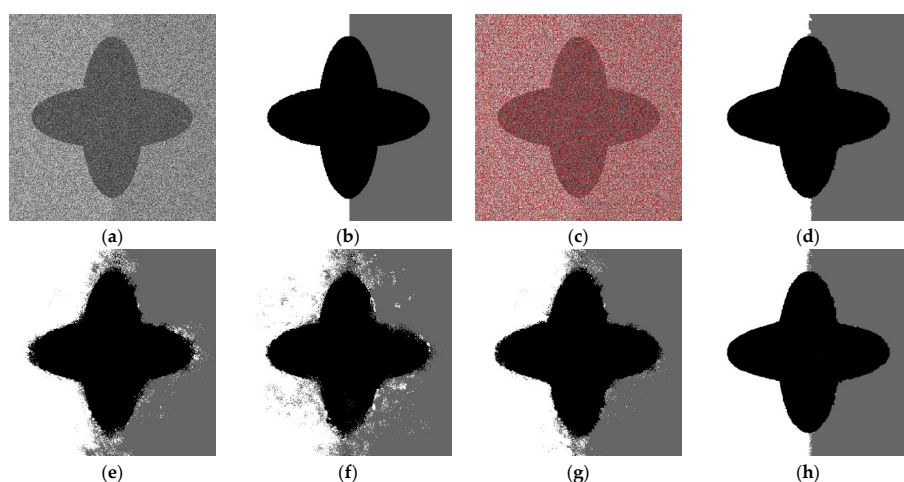


**Figure 8.** Results of different methods in Syn1: (**a**) Synthetic SAR images; (**b**) Ground truth; (**c**) Superpixels map; (**d**) Proposed method; (**e**) support vector machine (SVM) [3]; (**f**) SRC [5]; (**g**) joint sparsity model (JSRM) [16]; (**h**) hierarchical sparse representation-based classification (HSRC) [17].
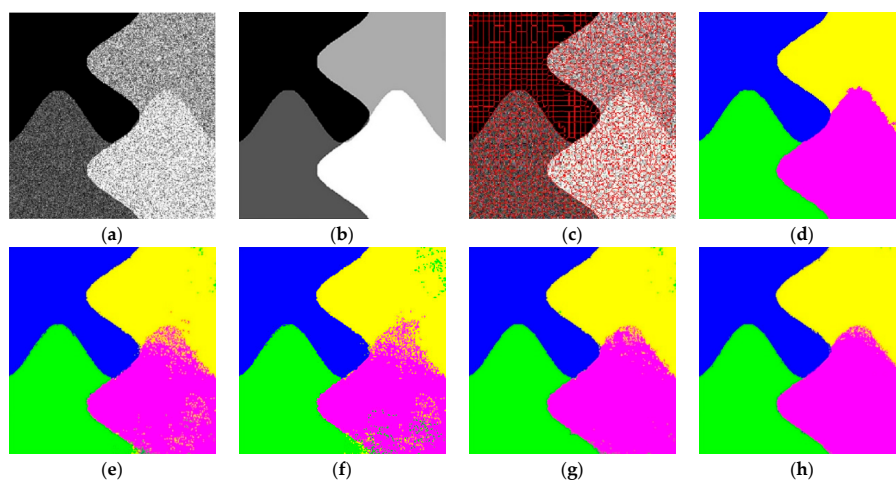


**Figure 9.** Results of different methods in Syn2: (**a**) Synthetic SAR images; (**b**) Ground truth; (**c**) Superpixels map; (**d**) Proposed method; (**e**) support vector machine (SVM) [3]; (**f**) SRC [5]; (**g**) joint sparsity model (JSRM) [16]; (**h**) hierarchical sparse representation-based classification (HSRC) [17].

**Figure 10.** Results of different methods in Syn3: (**a**) Synthetic SAR images; (**b**) Ground truth; (**c**) Superpixels map; (**d**) Proposed method; (**e**) support vector machine (SVM) [3]; (**f**) SRC [5]; (**g**) joint sparsity model (JSRM) [16]; (**h**) hierarchical sparse representation-based classification (HSRC) [17].
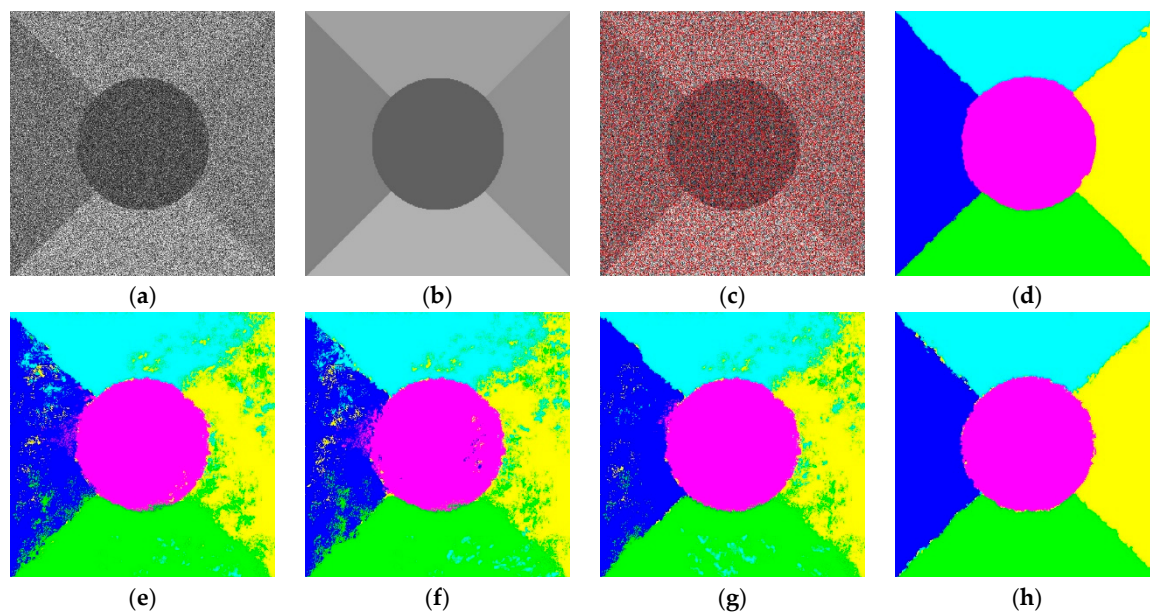
**Table 1.** Comparison of the run times (s) and classification accuracy (%) of different methods.

|  | Proposed Method | | SVM [3] | | SRC [5] | | JSRM [16] | | HSRC [17] | |
|---|---|---|---|---|---|---|---|---|---|---|
| SAR Image | Accuracy | Time | Accuracy | Time | Accuracy | Time (s) | Accuracy | Time | Accuracy | Time |
| Syn1 | 98.79 | 120.32 | 80.35 | 51.88 | 76.38 | 101.37 | 91.73 | 161.37 | 98.89 | 230.59 |
| Syn2 | 97.76 | 103.96 | 88.73 | 54.89 | 80.83 | 85.74 | 94.78 | 137.49 | 98.12 | 201.14 |
| Syn3 | 96.04 | 124.48 | 73.29 | 48.14 | 70.86 | 106.84 | 89.14 | 153.26 | 96.24 | 243.32 |

### 3.3. Results of Real SAR Images

In this section, three real SAR images are used for further analysis. The compared methods are the same as those used on synthetic SAR images. The results are shown in Figures 11–13. These original real images have three, three and four types of different regions as shown in Figures 11b, 12b and 13b, respectively. The superpixels of SAR1, SAR2 and SAR3 are 1000, 1200 and 1100 as shown in Figures 11b, 12b and 13b, respectively. The evaluation of the classification method is based on the visual inspection of the classification and the run time, accuracy, as well as the kappa coefficient. The scale in SVM, SRC and JSRM is set to $7 \times 7$, which represents the best result in the experiments.

From Figure 11c–g, we can see that the proposed method can achieve the classification in different areas and eliminate the influence of shadows, which always leads to categories by mistake. However, when we compare Figure 11c with 11g, it is difficult to know whether our proposed method is better, as it seems that Figure 11g [17] has better visualization results, albeit with some miscellaneous points. Therefore, the accuracy of the quantitative analysis is required for further analysis. From Table 2, it can be seen that the accuracy of the previous study [17] is only slightly higher than our proposed method, but the required running time is too long, as previously seen with synthetic SAR images.
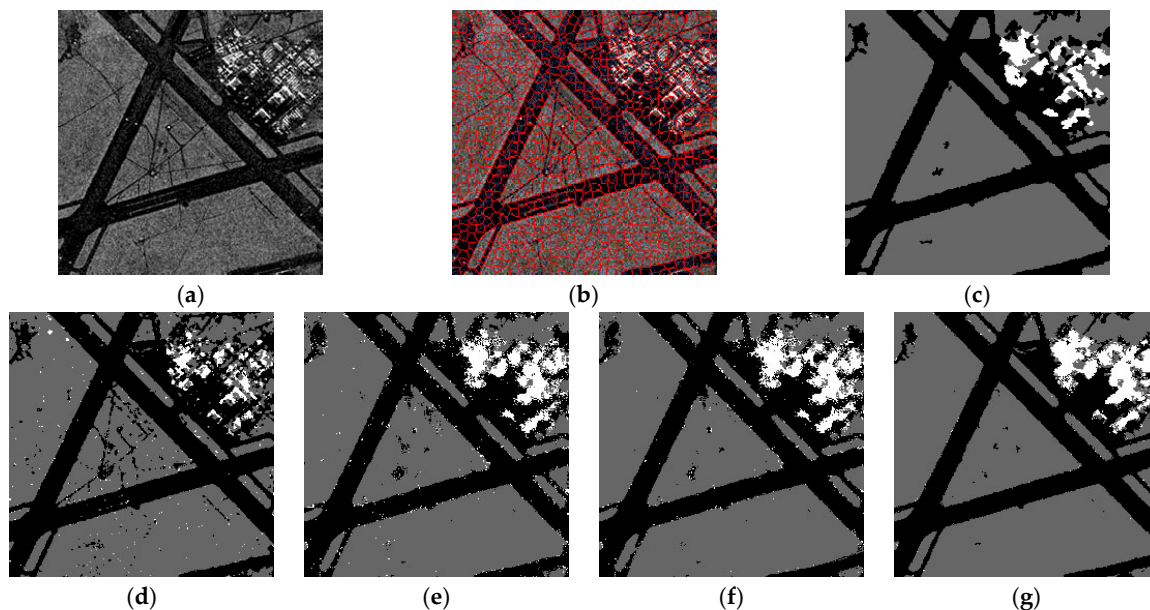
**Figure 11.** Results of different methods in real SAR1: (**a**) Real SAR images; (**b**) Superpixels map; (**c**) Proposed method; (**d**) support vector machine (SVM) [3]; (**e**) SRC [5]; (**f**) joint sparsity model (JSRM) [16]; and (**g**) hierarchical sparse representation-based classification (HSRC) [17].

From Figure 12c–g, we can see the classification results of different methods, especially in the yellow and red rectangle regions. The yellow and red rectangles of the proposed method in Figure 12c have less miscellaneous points than Figure 12d–f. In general, a smaller number of miscellaneous points indicates a more complete extraction of information and a more stable performance of the algorithm. The different rectangle regions highlight the superiority of the proposed algorithm. However, when we compare Figure 12c with 12g, it is difficult to know whether our proposed method is better, as it seems that Figure 12g [17] has better visualization results, albeit with some miscellaneous points. Therefore, data analysis was used (accuracy, run time and kappa coefficient) for further analysis. From Table 2, it can be seen that the accuracy of the previous method [17] is only slightly higher than our proposed method, but the required running time was too long, as previously seen with synthetic SAR images.
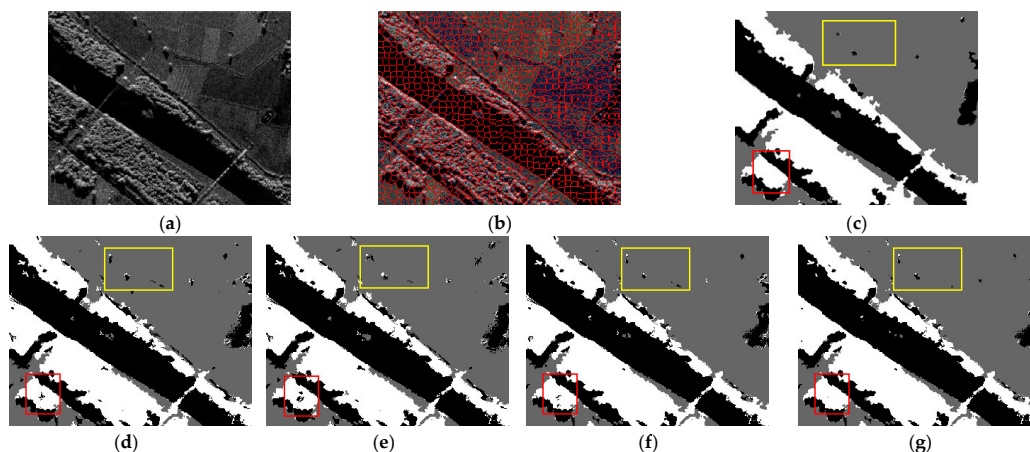


**Figure 12.** Results of different methods in real SAR2: (**a**) Real SAR images; (**b**) Superpixels map; (**c**) Proposed method; (**d**) support vector machine (SVM) [3]; (**e**) SRC [5]; (**f**) joint sparsity model (JSRM) [16]; and (**g**) hierarchical sparse representation-based classification (HSRC) [17].

The analysis of Figure 13 is similar to Figures 11 and 12. From Figures 11–13, the proposed method is shown to be suitable for the SAR image classification and obtains the optimal results. Table 2 provides the quantitative evaluation of different methods. Although the HSRC obtains higher classification accuracy compared with the others, the running time is too long among the different methods. Our method has the absolute advantage in the running time, with competitive accuracy that is only slightly lower than HSRC. Moreover, our method gets the highest robustness, which is reflected by the kappa coefficient. Above all, our method outperforms the others in terms of time consumption and robustness.
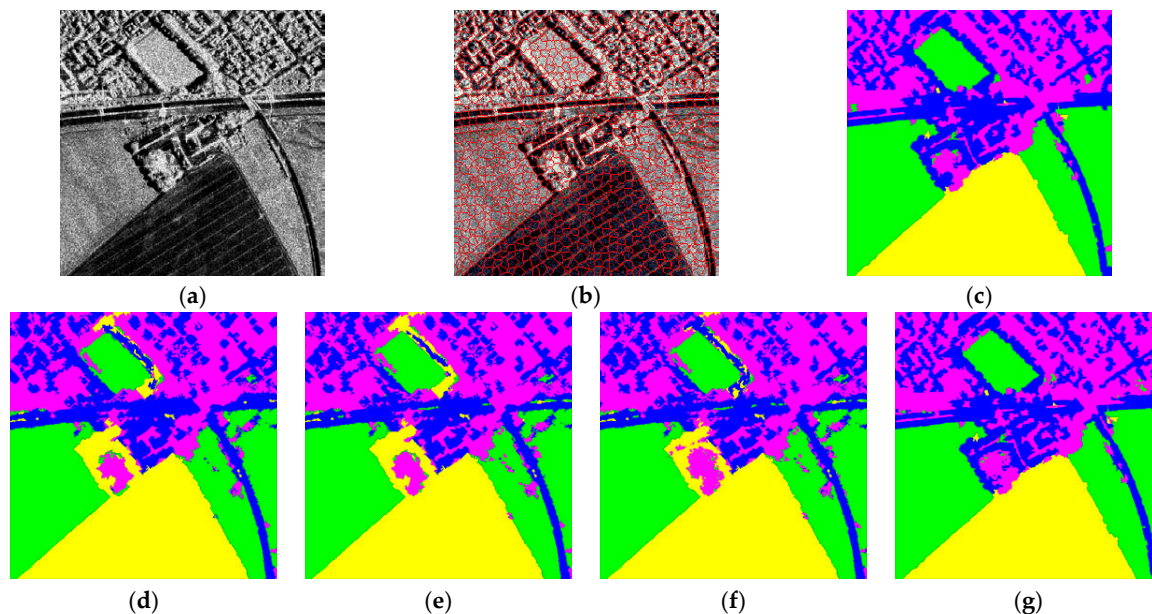


**Figure 13.** Results of different methods in real SAR3: (**a**) Real SAR images; (**b**) Superpixels map; (**c**) Proposed method; (**d**) support vector machine (SVM) [3]; (**e**) SRC [5]; (**f**) joint sparsity model (JSRM) [16]; and (**g**) hierarchical sparse representation-based classification (HSRC) [17].

**Table 2.** Comparison of the average criteria and accuracy (%) of different methods.

| SAR Image | Proposed Method | | SVM [3] | | SRC [5] | | JSRM [16] | | HSRC [17] | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Accuracy | Time | Accuracy | Time | Accuracy | Time | Accuracy | Time | Accuracy | Time |
| SAR1 | 96.18 | 102.38 | 89.79 | 48.68 | 85.38 | 98.67 | 93.67 | 147.73 | 96.20 | 238.95 |
| SAR2 | 97.57 | 121.35 | 87.68 | 43.19 | 87.33 | 99.24 | 92.48 | 139.58 | 97.62 | 253.48 |
| SAR3 | 97.52 | 124.41 | 83.59 | 51.94 | 79.96 | 102.68 | 91.04 | 160.36 | 96.74 | 258.72 |
| AA [1] | 97.42 | | 87.02 | | 84.22 | | 92.40 | | 97.83 | |
| K [2] | 0.961 | | 0.713 | | 0.806 | | 0.941 | | 0.959 | |

[1] AA is average accuracy; [2] K is kappa coefficient.

## 4. Discussion

Traditional SVM [3] is limited by lacking samples, which results in low classification accuracy. For instance, the number of training samples is 300, which is 0.46% of the total samples. Fewer samples affect the selection of optimal parameters by SVM for the testing samples, which will decrease the classification accuracy. In the sparse representation method, the HSRC [13] can solve the problem of reference [16] to a certain extent. It introduces the hierarchical concept and multi-size patch feature to solve the problem of lacking samples. Using SRC in SAR classification for both these methods improves the accuracy and stability. However, HSRC classifies images based on each pixel, which only depends on the selection of features in the spatial domain and the selected scale for each layer. This may lead to the loss or misrepresentation of information, which requires a long time for training.

In our paper, we inherit the advantages of reference [17], such as multi-layer. However, the difference is the multi-scale and multi-feature fusion. In the multi-feature fusion stage, we take three different methods to extract the gray and texture characteristics, which are stable in the presence of noise and changes in view, and can enrich the information of images. Moreover, the strategy of the multi-feature fusion was inspired by [23], which fused the different features from multiple layers. We fused the different features from different scales. This reduces the computational time and ensures a rich amount of information. Furthermore, classification based on superpixels can improve the speed of algorithms effectively. Three evaluation metrics (i.e., run time (time), average accuracy (AA) and the Kappa coefficient (K)) are adopted in these experiments to evaluate the quality of classification results. AA represents the mean of the percentage of correctly classified pixels for each class. K estimates the percentage of classified pixels corrected by the number of agreements. We performed comparative experiments with four other methods. The proposed method can solve the time redundancy problem of HSRC, but has its uncertainties. For instance, the uncertain points are always in the process of the algorithm until $h = H$. That is the reason we use the traditional SRC (this step is same as reference [17]) in the last step.

## 5. Conclusions

In this paper, based on superpixels, we presented a new model of classification of SAR images. It validates that adding multiple features, scales and layers can benefit the results of SRC classification and enrich the information of the images. Furthermore, using multiple layers can decrease the computational time due to the use of superpixels. The fusion strategy was introduced to merge each scale together to form a multi-fusion dictionary. With the added benefits, robustness was enhanced and the classification accuracy was improved significantly. The comparison experiments based on synthetic SAR images and real SAR images clearly demonstrate the efficiency and advantages of the proposed classification method. Moreover, the proposed classification method is also able to achieve lower computational costs. These added benefits are general for SAR image classification, and can be suitable for utility in more applications in the area of SAR image classification, as well as in other areas where the SRC method could be applied.

This method provides a slight improvement in calculation time for SAR image classification and application. Moreover, future research will focus on developing more efficient algorithms to cope with the large-scale SAR images.

**Author Contributions:** This work was prepared and accomplished by Ao-bo Zhai, who also wrote the manuscript. Xian-bin Wen outlined the research and supported the analysis. He also revised the work in presenting the technical details. Li-ming Yuan, Hai-xia Xu both suggested the design of comparison experiments and supervised the writing of the manuscript at all stages. Qing-xia Meng provided writing suggestions.

## References

1. Oliver, C.; Quegan, S. *Understanding Synthetic Aperture Radar Images*; SciTech Publishing: Raleigh, NC, USA, 2004.
2. Adragna, F.; Nicolas, J. *Processing of Synthetic Aperture Radar Images*; Wiley: New York, NY, USA, 2010.
3. Akbarizadeh, G. A new statistical-based kurtosis wavelet energy feature for texture recognition of SAR images. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 4358–4368. [CrossRef]
4. Xue, X.R.; Wang, X.J.; Xiang, F.; Wang, H.F. A new method of SAR image segmentation based on the gray level co-occurrence matrix and fuzzy neural network. In Proceedings of the IEEE 6th International Conference Wireless Communications Networking and Mobile Computing, Chengdu, China, 23–25 September 2010.

5.  Wright, J.; Yang, A.Y.; Sastry, S.S.; Ma, Y. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 210–227. [CrossRef] [PubMed]

6.  Mallat, S.G.; Zhang, Z. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process.* **1993**, *41*, 3397–3415. [CrossRef]

7.  Elad, M.; Aharon, M. Image Denoising Via Sparse and Redundant Representations over Learned Dictionaries. *IEEE Trans. Image Process.* **2006**, *15*, 3736–3745. [CrossRef] [PubMed]

8.  Buades, A.; Coll, B.; Morel, J.M. A Non-Local Algorithm for Image Denoising. In Proceedings of the IEEE Conference Computer Vision Pattern Recognition, San Diego, CA, USA, 20–26 June 2005.

9.  Yue, C.; Jiang, W. An algorithm of SAR image denoising in nonsubsampled contourlet transform domain based on maximum a posteriori and non-local restriction. *Remote Sens. Lett.* **2013**, *4*, 270–278. [CrossRef]

10. Fang, L.Y.; Li, S.T.; Mcnabb, R.P.; Nie, Q.; Kuo, A.N.; Toth, C.A.; Izatt, J.A.; Farsiu, S. Fast Acquisition and Reconstruction of Optical Coherence Tomography Images via Sparse Representation. *IEEE Trans. Med. Imaging* **2013**, *32*, 2034–2049. [CrossRef] [PubMed]

11. Li, S.; Yin, H.; Fang, L. Remote Sensing Image Fusion via Sparse Representations over Learned Dictionaries. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 4779–4789. [CrossRef]

12. Xiang, D.; Tang, T.; Hu, C.; Li, Y.; Su, Y. A kernel clustering algorithm with fuzzy factor: Application to SAR image segmentation. *IEEE Geosci. Remote Sens. Lett.* **2011**, *7*, 1290–1294.

13. Wang, W.; Xiang, D.; Ban, Y.; Zhang, J.; Wan, J. Superpixel Segmentation of Polarimetric SAR Images Based on Integrated Distance Measure and Entropy Rate Method. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *99*, 1–14. [CrossRef]

14. Liu, B.; Hu, H.; Wang, H.; Wang, K.; Liu, X.; Yu, W. Superpixel-Based Classification with an Adaptive Number of Classes for Polarimetric SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 907–924. [CrossRef]

15. Zhang, X.; Wen, X.; Xu, H.; Meng, Q. Synthetic aperture radar image segmentation based on edge-region active contour model. *J. Appl. Remote Sens.* **2016**, *10*, 036014. [CrossRef]

16. Chen, Y.; Nasrabadi, N.M.; Tran, T.D. Hyperspectral Image Classification Using Dictionary-Based Sparse Representation. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 3973–3985. [CrossRef]

17. Hou, B.; Ren, B.; Ju, G.; Li, H.; Jiao, L.; Zhao, J. SAR Image Classification via Hierarchical Sparse Representation and Multisize Patch Features. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 33–37. [CrossRef]

18. Swain, M.J.; Ballard, D.H. Color indexing. *Int. J. Comput. Vis.* **1991**, *7*, 11–32. [CrossRef]

19. Haralick, R.M.; Shanmugam, K.; Dinstein, I. Textural Features for Image Classification. *IEEE Trans. Syst. Man Cybern.* **1973**, *3*, 610–621. [CrossRef]

20. Yan, X.; Jiao, L.; Xu, S. SAR image segmentation based on Gabor filters of adaptive window in overcomplete brushlet domain. In Proceedings of the 2nd Asian-Pacific Conference on Synthetic Aperture Radar, Xi'an, China, 26–30 October 2009.

21. Gu, J.; Jiao, L.; Yang, S.; Liu, F.; Hou, B.; Zhao, Z. A Multi-kernel Joint Sparse Graph for SAR Image Segmentation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 1265–1285. [CrossRef]

22. Zhao, S.; Liu, Y.; Han, Y.; Hong, R.; Hu, Q.; Tian, Q. Pooling the Convolutional Layers in Deep ConvNets for Video Action Recognition. *IEEE Trans. Circuits Syst. Video Technol.* **2015**, 1. [CrossRef]

23. Tan, M.; Tsang, I.W.; Wang, L.; Zhang, X. Convex Matching Pursuit for Large-scale Sparse Coding and Subset Selection. In Proceedings of the AAAI Conference on Artificial Intelligence, Toronto, ON, Canada, 22–26 July 2012.

24. Mo, X.; Monga, V.; Bala, R.; Bala, R.; Fan, Z.G. Adaptive Sparse Representations for Video Anomaly Detection. *IEEE Trans. Circuits Syst. Video Technol.* **2013**, *24*, 631–645.

25. Levinshtein, A.; Stere, A.; Kutulakos, K.N.; Fleet, D.J.; Dickinson, S.J.; Siddiqi, K. TurboPixels: Fast Superpixels Using Geometric Flows. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 2290–2297. [CrossRef] [PubMed]