



Article Multi-Feature Fusion and Adaptive Kernel Combination for SAR Image Classification

Xiaoying Wu^{1,2,3,4}, Xianbin Wen^{1,3,4,*}, Haixia Xu^{1,3,4}, Liming Yuan^{1,3,4,*} and Changlun Guo^{1,3,4}

- ¹ School of Computer Science and Engineering, Tianjin University of Technology, Tianjin 300384, China; xiaoyingwu@hist.edu.cn (X.W.); xuhaixia@email.tjut.edu.cn (H.X.); 183128306@stud.tjut.edu.cn (C.G.)
- ² School of Information Engineering, Henan Institute of Science and Technology, Xinxiang 453003, China
- Key Laboratory of Computer Vision and System, Ministry of Education, Tianjin 300384, China
- ⁴ Tianjin Key Laboratory of Intelligence Computing and Novel Software Technology, Tianjin 300384, China
- Correspondence: xbwen@email.tjut.edu.cn (X.W.); yuanliming@tjut.edu.cn (L.Y.); Tel.: +86-186-2205-6806 (L.Y.)

Abstract: Synthetic aperture radar (SAR) image classification is an important task in remote sensing applications. However, it is challenging due to the speckle embedding in SAR imaging, which significantly degrades the classification performance. To address this issue, a new SAR image classification framework based on multi-feature fusion and adaptive kernel combination is proposed in this paper. Expressing pixel similarity by non-negative logarithmic likelihood difference, the generalized neighborhoods are newly defined. The adaptive kernel combination is designed on them to dynamically explore multi-feature information that is robust to speckle noise. Then, local consistency optimization is further applied to enhance label spatial smoothness during classification. By simultaneously utilizing adaptive kernel combination and local consistency optimization for the first time, the texture feature information, context information within features, generalized spatial information between features, and complementary information among features is fully integrated to ensure accurate and smooth classification. Compared with several state-of-the-art methods on synthetic and real SAR images, the proposed method demonstrates better performance in visual effects and classification quality, as the image edges and details are better preserved according to the experimental results.

Keywords: multi-feature; adaptive; kernel combination; SAR; image classification

1. Introduction

Synthetic aperture radar (SAR) can acquire images of different land covers through clouds and rain in all weather conditions and times, and it has a certain surface penetrating capability. Therefore, SAR images are widely used in disaster monitoring, environment mapping, urban planning, military reconnaissance, and crop yield estimation [1]. SAR image classification is one of the core contents of SAR image interpretation. However, with SAR image classification it is not easy to obtain satisfactory results, due to the speckle embedding in SAR imaging. This remains a challenging task to be resolved. In recent years, the research on SAR image classification technology has mainly focused on two crucial aspects: more effective representation of feature information, and enhancing the local consistency on pixel labels.

To date, various methods have been proposed for feature information representation in classification. The most basic feature information representation is various feature extraction, which includes gray-level co-occurrence matrix (GLCM), Gabor filters (GFs), wavelet (WL), attribute profiles (APs), etc. GLCM was employed for retrieval of sea surface wind direction from SAR images [2], and GFs were used for SAR images to detect change [3]. Discrete wavelet transform processing on coherent sea-ice textures, while being decomposed, has been utilized to perform fractal analysis of SAR images [4]. AP



Citation: Wu, X.; Wen, X.; Xu, H.; Yuan, L.; Guo, C. Multi-Feature Fusion and Adaptive Kernel Combination for SAR Image Classification. *Appl. Sci.* **2021**, *11*, 1603. https://doi.org/10.3390/app 11041603

Academic Editor: Hyung-Sup Jung

Received: 7 December 2020 Accepted: 4 February 2021 Published: 10 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). was used to distinguish heterogeneous and homogeneous image contents in SAR image segmentation by Boldt et al. [5], while Tombak et al. [6] investigated AP for pixel-based SAR image classification. These features always retain image detail features, which do not fully exploit the contextual and spatial information to suppress the speckle effect.

In recent years, the representation of feature information has gradually developed to mine context information. The Markov random field (MRF) has been widely used in SAR image classification to exploit the context information [7]. The Markov random field model was used to cluster the SAR image, and according to the centroid of the region, the region category was located with complete edge [8]. However, this method always considers the neighborhood to be usually in the small fixed-size, which is not very useful to preserve the structures. In addition, the composite kernel support vector machine (SVM) is utilized to exploit the context information [9,10], such as the correlation between pixels, when it classifies SAR images. According to experience set parameters, a composite kernel was constructed, and it explores both the two elements of information in the adaptive region: the texture feature and the context [9]. Moreover, a composite kernel method was proposed to combine the polarimetric and morphological feature fusion [10] with adjustment parameter of experience. However, they both set the kernel parameters artificially in the construction of the composite kernel, the contribution of different features is balanced by experience, and the feature difference cannot be reflected autonomously, so they cannot accurately express the image information.

In order to exploit the information in different aspects, multiple features are further joined for representation. Multi-feature was fused in low-rank representation for SAR target recognition [11], and it was utilized to classify the SAR image [12], improving the classification accuracy of the SAR image. However, the multiple features are united in the simple stacking way, and the difference between the multiple features cannot be taken into account. The method cannot fully mine the feature information, and the essence of the image cannot be better highlighted. Although the approaches above can effectively improve the classification accuracy, "salt and pepper" noise often appears in their classification results due to multiplicative speckle. A good way to solve this problem is to introduce the feature space and combine features dynamically according to difference between features, which can effectively suppress the noise of classification maps and make full use of feature information.

Another aspect is to improve the local consistency of pixel labels. Typical methods include over-segmentation techniques used in the preprocessing step and optimization based on probability theory performed as a postprocessing procedure. The over-segmentation technique divides the whole SAR image into small uniform patches or superpixels, which is an indispensable part of the subsequent classifier learning and labeling process [13–15]. In the segmentation process, pixels in each patch share the same label. Optimization based on probability theory integrates prior label smoothing into the segmentation task to solve the largest probability [16-18]. Dempster–Shafer evidence theory makes up for the deficiency caused by the incomplete, partially inaccurate, or uncertain information, and the hypothesis with the maximum credibility and likelihood degree is selected as the fusion result [16], but it is sensitive to basic reliability allocation. Graph optimization integrates prior information into graph construction, solves the maximum posterior probability to enhance the local consistency of labels, and obtains segmentation results with higher classification accuracy and better spatial connectivity [17]. However, the construction of graph structure takes more time and requires high computational cost. Majority vote algorithms are simple to calculate, flexible in resolving conflicts [18,19], and do not require the allocation of confidence.

Both the above two aspects can effectively promote the performance of SAR image classification. However, most of the existing SAR image classification methods usually consider only one of them. In this paper, to address the aforementioned problem for SAR image classification, we propose a classification framework of multi-feature fusion and adaptive kernel combination (MAKC), which constructs the feature space to fully mine

multi-feature information while using optimization to ensure the local consistency of labels. The main contributions are summarized as follows: (1) Constructing three-dimensional feature space of generalized spatial information for fusion, meaning space information is utilized more deeply than ever before, instead of only plane contextual information. (2) Adaptive kernel combination reflects the feature difference to mine the spatial information, with dynamic weights according to image content rather than artificial setting according to experience. (3) Different from the original over-segmentation algorithm, it combines the Gamma distribution of SAR images and introduces non-negative logarithmic likelihood value to evaluate pixel similarity, which makes the algorithm more suitable for SAR images. (4) Compared with the previous postprocessing optimization procedure that only considers confidence level, a conflict resolution rule that can comprehensively consider the quality of features enhances the local consistency of pixel labels to fuse features, and also changes the traditional stacking way of feature fusion.

The rest of this paper is organized as follows. In Section 2, we review the notation and background materials. In Section 3, the proposed method is detailed. In Section 4, the experimental results are given and compared with several state-of-the-art classification methods. Then, in Section 5, the effectiveness of the method step and the adjustment of parameters are discussed. Finally, in Section 6, the paper is summarized and future works are suggested.

2. Background

In this section, we briefly review texture information representations of SAR images, which are gray-level co-occurrence matrix (GLCM), wavelet (WL), and attribute profile (AP). In addition, the information entropy, image entropy, and entropy rate superpixel methods are introduced. SAR image size is $I_1 \times I_2$. A gray-level co-occurrence matrix (GLCM) can express the imaging mechanism and the statistical characteristics of SAR images [20,21]. Derived from the co-occurrence matrix on four directions, 0°, 45°, 90°, and 135°, six measures named contrast, correlation, homogeneity, energy, mean value, and median value are selected, then the average value of four directions on each measure is computed to represent the image pixel, and the feature vector of each pixel is represented as

$$F_{ij}^{G} = (f_{ij1}^{G}, f_{ij2}^{G}, f_{ij3}^{G}, f_{ij4}^{G}, f_{ij5}^{G}, f_{ij6}^{G}), 1 \le i \le I_{1}, 1 \le j \le I_{2}.$$
(1)

Then, the 3D GLCM feature tensor of $I_1 \times I_2 \times 6$ size is obtained for the SAR image, which is

$$F_{fea}^{G} = \begin{bmatrix} F_{11}^{G} & F_{12}^{G} & \cdots & F_{1I_{2}}^{G} \\ F_{21}^{G} & F_{22}^{G} & \cdots & F_{2I_{2}}^{G} \\ \cdots & \cdots & \cdots & \cdots \\ F_{I_{1}1}^{G} & F_{I_{1}2}^{G} & \cdots & F_{I_{1}I_{2}}^{G} \end{bmatrix}.$$
 (2)

The GLCM feature is very good for contrast enhancement and despeckling, and effectively describes spatial information about the relative position of pixels, but does not involve structural relationships and weak the target signal or edge information. The characteristics of texture and noise are not essentially distinguished. Fortunately, WL and AP can make up for GLCM.

WL transform provides frequency information through multi-resolution analysis [22], which can essentially distinguish the characteristics of texture and noise. Inspired by reference [23], the two-level four-tap Daubechies wavelet decomposition is performed over a 9×9 window for each pixel of the SAR image in this section. The wavelet energy is extracted from seven channels, and the feature vector of each pixel is denoted as

$$F_{ij}^{W} = (f_{ij1}^{W}, f_{ij2}^{W}, f_{ij3}^{W}, f_{ij4}^{W}, f_{ij5}^{W}, f_{ij6}^{W}, f_{ij7}^{W}), 1 \le i \le I_{1}, 1 \le j \le I_{2}.$$
(3)

Then, a 3D WL feature tensor of $I_1 \times I_2 \times 7$ size can be obtained [4,24], and is represented as

$$F_{fea}^{W} = \begin{bmatrix} F_{11} & F_{12} & \cdots & F_{1L} \\ F_{21}^{W} & F_{22}^{W} & \cdots & F_{2L_2}^{W} \\ \cdots & \cdots & \cdots & \cdots \\ F_{I_1 1}^{W} & F_{I_1 2}^{W} & \cdots & F_{I_1 I_2}^{W} \end{bmatrix}.$$
 (4)

AP provides a multi-level characterization of an image, which is created by the sequential application of morphological attribute filters (AFs). AFs are the operators of the connection, which only consider the connected components of the image. Thus, they do not distort or insert new edges when processing images, but simply merge existing flat areas. So, AP can be used to model different kinds of structural information of the scene in order to increase the effectiveness of classification. The four SAR attributes of area, the diagonal, moment of inertia, and standard deviation are applied [5,6]. There are three thresholds for each attribute, and the feature vector of each pixel is represented as

$$F_{ij}^{A} = (f_{ij1}^{A}, f_{ij2}^{A}, f_{ij3}^{A}, \cdots f_{ij23}^{A}, f_{ij24}^{A}, f_{ij25}^{A}), 1 \le i \le I_{1}, 1 \le j \le I_{2}.$$
(5)

Then, a 3D attribute profile tensor of $I_1 \times I_2 \times 25$ size is obtained on the SAR image according to the thickening and thinning operation, which is

$$F_{fea}^{A} = \begin{bmatrix} F_{11}^{A} & F_{12}^{A} & \cdots & F_{1I_{2}}^{A} \\ F_{21}^{A} & F_{22}^{A} & \cdots & F_{2I_{2}}^{A} \\ \cdots & \cdots & \cdots & \cdots \\ F_{I_{1}1}^{A} & F_{I_{1}2}^{A} & \cdots & F_{I_{1}I_{2}}^{A} \end{bmatrix}.$$
 (6)

Information entropy was proposed by Shannon. Shannon et al. postulated that the amount of information could be expressed in terms of the amount of eliminated uncertainty [25,26]. According to information theory, information entropy is considered from the statistical property of the whole information source. For a particular information source, information entropy is only one. Suppose the random event set is $\{X_i, i = 1, 2, \dots, N\}$, the probability of its occurrence is p_i , and the condition $\sum_{i=1}^{N} p_i = 1, 0 \le p_i \le 1, i = 1, 2, \dots, N$ is satisfied: then define information entropy $H(X) = -\sum_{i=1}^{N} p_i \log p_i$.

An SAR image is composed of pixels. Pixels of different intensities occupy different areas in the image, making the image show different shapes, and different shape image areas contain different information. Because the image distribution has block structures, there is a position correlation between each pixel. Therefore, on the basis of the onedimensional entropy, the image entropy, which can reflect the spatial property of gray-scale distribution, is introduced to form the two-dimensional entropy, considering the two-dimensional spatial property of the SAR image. Assuming an SAR image is $I_1 \times I_2$ size, and the image data have a non-negative value that is $f(i, j) \ge 0$, the image entropy E(f) is defined as

$$p_{ij} = f(i,j) / \sum_{i=1}^{I_1} \sum_{j=1}^{I_2} f(i,j), E(f) = -\sum_{i=1}^{I_1} \sum_{j=1}^{I_2} p_{ij} \log p_{ij}$$
(7)

where f(i, j) is the gray-scale value of the pixel whose coordinate position is (i, j) in the image, and p_{ij} is the probability of pixel gray-scale value f(i, j) appearing in the image.

The entropy rate superpixel method is an over-segmentation technique that can improve the local consistency of pixel labels. Among the existing superpixel algorithms, the entropy rate method can not only obtain superpixels of adaptive size from the global perspective, but also has the adaptive ability in the heterogeneous region and the ability to preserve local details for optical images [27,28]. For optical images, the distance between

pairwise neighboring pixels can be used to generate superpixels by the entropy rate method. Definition of pixel similarity between the two in terms of distance is as follows:

$$d(x,y) = \|x - y\|_{2} = \sqrt{(x - y)^{T}(x - y)}$$
(8)

where x, y are two pixel values in the image, and the similarity between the two pixels is calculated by Formula (8). An image is mapped to a graph G = (V, E) with vertices denoting the pixels and the edge weights denoting the pairwise similarities given in the form of a similarity matrix. The entropy rate of the random walk on the constructed graph is a criterion to obtain compact and homogeneous clusters, and a balancing function encourages clusters with similar sizes.

However, the original entropy rate method is not completely suitable for SAR images due to the fact that speckle noise will severely affect the distance directly calculated between individual pixels, so an accurate and robust distance measure between neighboring pixels should be derived.

3. Methodology

In this section, the implementation of the proposed method is illustrated concretely. The overall framework of the proposed method is shown in Figure 1. The figure includes a preprocessing step and three main parts from left to right.



Figure 1. Framework of the proposed method.

The preprocessing takes the SAR image as the starting point, solves the pixel similarity based on Gamma distribution by the non-negative logarithmic likelihood difference, and generates the superpixel mapping, and the specific description is shown in Section 3.1.

The three bodies are the feature space construction, adaptive kernel combination, and conflict resolution. Feature space construction is carried out on three group texture information feature tensors from the SAR image, respectively, which are gray-level co-occurrence matrix (GLCM), wavelet (WL), and attribute profile (AP) from top to bottom. Each feature tensor is clustered through the superpixel mapping to obtain 3D adaptive blocks. Based on these blocks, the internal context information feature space and the generalized space information feature space are constructed, as detailed in Section 3.2.

In the adaptive kernel combination stage, the texture feature kernel, contextual kernel and spatial kernel corresponding to three feature spaces in one group are adaptively combined through dynamic weight to obtain a composite kernel, which realizes information fusion within features of one group. Dynamic weights according to image content rather than artificial setting combine feature kernels, which reflect the feature difference to mine the spatial information. The stage is described in detail in Section 3.3.

Conflict resolution is the decision embodiment of local consistency optimization, and the classification maps generated by three composite kernels accordingly are considered to enhance the local consistency smoothing labels, which realizes complementary information fusion among features in different groups, and the final classification result is obtained. See Section 3.4 for detailed description.

3.1. Segmentation of Non-Negative Logarithmic Likelihood Difference Based on *Gamma Distribution*

To enhance local consistency, a 3D adaptive block is obtained by an improved entropy rate superpixel method. Under speckle noise, pairwise similarities cannot be calculated directly. In order to reduce speckle noise impact, the Gamma distribution statistical characteristics of the SAR image are taken into account. For the intensity image, the probability density function of the Gamma distribution is defined as

$$P(x) = \frac{\beta^{\gamma} x^{\gamma - 1}}{\Gamma(\gamma)} e^{-\beta x}$$
(9)

where γ is the shape parameter, β is the scale parameter, and $\Gamma(\cdot)$ is the Gamma function. The non-negative logarithmic likelihood difference value [29] is adopted to represent the intensity similarity between pixels based on Gamma distribution, which is defined as

$$d(i,j) = \left|\log P(i) - \log P(j)\right| \tag{10}$$

where *i* and *j* are intensity values of two pixels, and P(i) is the value obtained by substituting *i* into the probability density function (9) of the Gamma distribution, which reflects the statistical characteristic on the local neighborhood of the pixel with *i* value and suppresses the effect of speckle noise. The logarithmic transformation of P(i) is performed to improve the sensitivity to the difference in likelihood probability values without changing the monotonicity of the original data. The absolute value of the difference between $\log P(i)$ and $\log P(j)$ represents the similarity between two pixels with *i* and *j* values, replacing (8) in the original method to reduce the impact of speckle noise, so that pixel similarity is not represented by pixel gray value difference directly.

Using Formula (10), the superpixel results by the improved superpixel algorithm are shown in Figure 2a. The region-based statistical characteristic can reflect the change rule of most data in the image, so it can suppress the noise. From Figure 2, the superpixels obtained by the improved method have adaptive sizes and shapes. In particular, the region in the red rectangle box is significantly more attached to the edge than the previous shown in Figure 2b. The red rectangle area is zoomed in Figure 2c,d, from top to bottom; it contains dark gray crop features, black vegetation features, white bare land features, and dark black river features. As can be seen from the two figures, the superpixel edge is better attached to the boundary of the black vegetation area in Figure 2c, while in Figure 2d, the superpixel edge is far away from the boundary of the vegetation area. Then, the shape and size of each block according to local structure can be adaptively adjusted, instead of the boundary being far away from the shape contour of the local object.



Figure 2. Comparison of superpixel results: (a) the improved algorithm, (b) the previous algorithm, (c) the zoom part of (a), and (d) the zoom part of (b).

3.2. Feature Space Construction

The above superpixel results for the SAR image are mapped to each 3D feature. So, the adaptive 3D feature blocks, where S_i^{flag} ($i = 1, \dots, num_{block}, num_{block}$ denotes the number of blocks, and *flag* is a member of {*G*, *A*, *W*}), are generated on each 3D feature tensor. Moreover, each pixel of each 3D feature tensor is clustered according to the superpixel label at the corresponding position. The 3D feature blocks can preserve the structure of the object comprehensively due to it being multi-layer, adhering to edges, and of unfixed size. Hence, they can explore the contextual and spatial information more accurately. The generalized neighborhood of each block S_i^{flag} includes the neighborhood blocks $S_{i,j}^{flag}$ ($j = 1, \dots, num_{nei}, num_{nei}$ is the number of block S_i^{flag} neighborhood) in the current layer and other layers, as shown in Figure 3b.



Figure 3. (a) Adaptive feature blocks and (b) generalized neighborhood of an adaptive block S_i .

The internal context information and generalized spatial information need to be extracted from each shape adaptive feature block and neighborhood. Because each shape-adaptive 3D feature block S_i^{flag} includes a group of neighbor F_{ij}^{flag} vectors, F_{ij}^{flag} is the expression of feature information which is GLCM, AP, and WL extracted above, and the number of neighbor vectors is denoted as num_{pix} , so the internal context information of each 3D feature block S_i^{flag} is expressed as the average of each feature F_{ij}^{flag} values in S_i^{flag} , which is defined as

$$S_i^{mean,flag} = \left(\sum_{k=1}^{num_{pix}} F_{ij}^{flag}\right) / num_{pix}.$$
(11)

The average value vector $S_i^{mean,flag}$ is assigned to each pixel in the shape-adaptive block S_i^{flag} to replace the original feature value F_{ij}^{flag} . Then, the internal context information feature F_{mean}^{flag} can be constructed by performing the same filtering operation on all shape-adaptive blocks, which is

$$F_{mean}^{flag} = \left\{ S_1^{mean,flag}, S_2^{mean,flag}, \cdots, S_i^{mean,flag}, \cdots, S_{num_{block}}^{mean,flag}, \cdots, S_1^{mean,flag}, S_2^{mean,flag}, \cdots, S_i^{mean,flag}, \cdots, S_{num_{block}}^{mean,flag}, \cdots, S_i^{mean,flag}, \cdots, S_i^{$$

where num_{block} is the number of all S_i^{flag} shape-adaptive blocks on one layer, and the number of $S_i^{mean,flag}$ elements in the feature set F_{mean}^{flag} is $I_1 \times I_2$. Then, 3D internal context information feature space F_{mean}^{flag} is constructed by $S_i^{mean,flag}$ vectors.

Considering that the neighboring shape-adaptive blocks have similar spatial information on different layers in 3D space, the generalized spatial information is represented by spatial distance metric similarity, which is expressed as

$$S_i^{spa,flag} = \sum_{j=1}^{num_{nei}} w_{i,j} \times S_{i,j}^{mean,flag},$$
(13)

$$w_{i,j} = \left\| S_{i,j}^{mean,flag} - S_i^{mean,flag} \right\|_2^{-1} / \left(\sum_{j=1}^{num_{nei}} \left\| S_{i,j}^{mean,flag} - S_i^{mean,flag} \right\|_2^{-1} \right), \tag{14}$$

where $S_i^{spa,flag}$ is the generalized spatial information value of the shape-adaptive block S_i^{flag} , $S_{i,j}^{mean,flag}$ is the mean of the *j*th neighborhood of block S_i^{flag} , and num_{nei} is the number of neighboring S_i^{flag} shape-adaptive blocks. $\|S_{i,j}^{mean,flag} - S_i^{mean,flag}\|_2$ is the Euclidean distance that measures the dissimilarity between neighboring block and center block [30,31]. $w_{i,j}$ is the weight, which is inversely proportional to the Euclidean distance. The larger the Euclidean distance and the smaller the weight, then the less impact the neighborhood value has on the new generalized spatial information value. The vector $S_i^{spa,flag}$ is assigned to each pixel in the shape-adaptive block to replace $S_i^{mean,flag}$. The generalized spatial information feature F_{spa}^{flag} is constructed by performing the same filtering operation on all shape-adaptive blocks, which is described as

$$F_{spa}^{flag} = \left\{ S_1^{spa,flag}, S_2^{spa,flag}, \cdots, S_i^{spa,flag}, \cdots, S_{num_{block}}^{spa,flag}, \cdots, S_1^{spa,flag}, S_2^{spa,flag}, \cdots, S_i^{spa,flag}, \cdots, S_{num_{block}}^{spa,flag} \right\}_{I_1 \times I_2}.$$
 (15)

 num_{block} is the number of all shape-adaptive blocks of one layer, and the number of $S_i^{spa,flag}$ elements in the feature set F_{spa}^{flag} is $I_1 \times I_2$. Then, 3D generalized spatial information feature space F_{spa}^{flag} is constructed by $S_i^{spa,flag}$ vectors.

3.3. Adaptive Kernel Combination

Three different feature kernels are expressed based on the above feature information 3D space. From feature tensors information F_{fea}^{flag} , where flag is a member of $\{G, A, W\}$, a set of feature pixels is randomly selected as training samples, which is denoted as $\{F_1^{fea,flag}, F_2^{fea,flag}, \cdots, F_n^{fea,flag}\}$, where *n* is the number of the training samples. Spatial pixels from F_{mean}^{flag} and F_{spa}^{flag} are extracted, respectively, corresponding to the position indexes of selected feature pixels, which are denoted as $S_1^{mean,flag}, \cdots, S_n^{mean,flag}, S_1^{spa,flag}, \cdots, S_n^{spa,flag}$, respectively. After that, three different kernels use the radial basis function as kernel function, and are created based on the three kinds of training samples, that is

$$K_{fea}^{flag}\left(F_{i}^{fea,flag},F_{j}^{fea,flag}\right) = \exp\left(-\left\|F_{i}^{fea,flag}-F_{j}^{fea,flag}\right\|/2\delta^{2}\right),\tag{16}$$

$$K_{mean}^{flag}\left(S_{i}^{mean,flag},S_{j}^{mean,flag}\right) = \exp\left(-\left\|S_{i}^{mean,flag}-S_{j}^{mean,flag}\right\|/2\delta^{2}\right),\tag{17}$$

$$K_{spa}^{flag}\left(S_{i}^{spa,flag},S_{j}^{spa,flag}\right) = \exp\left(-\left\|S_{i}^{spa,flag}-S_{j}^{spa,flag}\right\|/2\delta^{2}\right).$$
(18)

By adaptive kernel combination of the above three kernels, the composite kernel can be constructed. To determine the dynamic weight, principal component analysis [32] is used on F_{fea}^{flag} , F_{mean}^{flag} , and F_{spa}^{flag} respectively to obtain the first three principal components (PCs) in each feature space, which can reduce the computational cost. Canny filter is used to detect the edge on PCs to obtain the binary image of the contour. The image entropy [25,26] is used to measure the amount of structure information in the binary image. If image entropy is high, this means the structural information contained in the image is rich. The image entropies are calculated as follows:

$$E_{fea}^{flag} = -\sum_{i=0}^{1} P_i^{fea,flag} \log P_i^{fea,flag}, E_{mean}^{flag} = -\sum_{i=0}^{1} P_i^{mean,flag} \log P_i^{mean,flag},$$

$$E_{spa}^{flag} = -\sum_{i=0}^{1} P_i^{spa,flag} \log P_i^{spa,flag}.$$
(19)

 $P_i^{fea,flag}$ is the probability that the pixel gray value is *i* in the binary image of F_{fea}^{flag} , *i* is equal to 0 or 1. Similarly, $P_i^{mean,flag}$ and $P_i^{spa,flag}$ are the probability in the different features F_{mean}^{flag} and F_{spa}^{flag} , respectively. E_{fea}^{flag} , E_{mean}^{flag} , and E_{spa}^{flag} are image entropy. The obtained image entropy ratios are set as the dynamic weight of the corresponding kernels, and the composite kernel is defined as

$$\begin{split} W_{fea}^{flag} &= \frac{E_{fea}^{flag}}{E_{sum}^{slag}}, W_{mean}^{flag} &= \frac{E_{mean}^{flag}}{E_{sum}^{slag}}, W_{spa}^{flag} &= \frac{E_{spa}^{flag}}{E_{sum}^{slag}}, E_{sum}^{flag} &= E_{fea}^{flag} + E_{mean}^{flag} + E_{spa}^{flag}, \\ K_{com}^{flag} \left(F_i^{com, flag}, F_j^{com, flag} \right) \\ &= W_{fea}^{flag} K_{fea}^{flag} \left(F_i^{fea, flag}, F_j^{fea, flag} \right) + W_{mean}^{flag} K_{mean}^{flag} \left(S_i^{mean, flag}, S_j^{mean, flag} \right) + W_{spa}^{flag} K_{spa}^{flag} \left(S_i^{spa, flag}, S_j^{spa, flag} \right). \end{split}$$

The adaptive kernel combination obtains composite kernels, which are incorporated into the SVM classifier to implement the SAR classification. Each group of feature information generates a composite kernel and obtains a SAR classification result. The below three classification maps are fused to obtain SAR image final classification.

3.4. Conflict Resolution of Local Consistency Optimization

For enforcing the local consistency on pixel labels, optimization by probability theory is used to smooth labels based on the above three classification maps. Conflict resolution is used to determine the class of pixel p(i, j) in conflict labels as follows on the SAR image, and $1 \le i \le I_1, 1 \le j \le I_2$.

In no-conflict situations, $cou_k^{i,j}$ denotes the classification times that each pixel p(i,j) is classified as class k in three classification maps, m is the class number, and then final class label l can be obtained by the maximum count value, that is

$$Lab_{i,j}^{nocon} = \left\{ l \middle| cou_l^{i,j} = \max(cou_1^{i,j}, \cdots, cou_k^{i,j}, \cdots, cou_m^{i,j}), 0 \le cou_k^{i,j} \le 3. \right\}.$$
(21)

In conflict situations, where a pixel has three different classification maps' labels, it is hard to choose the right map's label as the final result. To address this issue, a probability is solved to determine the label of a pixel based on image feature information quality. The equivalent number of looks (ENL) [33] is an index to measure the smoothness of uniform regions. The higher the value, the less the noise affects the feature information and the better the feature quality. In order to further improve the reliability of labels, confidence level is combined with ENL to make decisions. A confidence level is defined with the initiation value zero for each map, and the valid decision times of each F_{fea}^{flag} feature information are counted according to classification map $Lab_{i,j}^{nocon}$ in no-conflict situations. The label of largest probability is selected, the probability is defined as

$$Pro_{flag} = \frac{CL_{flag}}{\sum\limits_{flag=G}^{W} CL_{flag}} + \frac{M_{flag}}{\sum\limits_{flag=G}^{W} M_{flag}},$$
(22)

where
$$M_{flag} = \delta_{flag}^2 / \mu_{flag}^2$$
 (23)

$$CL_{flag} = \text{Countif}\left(classmap_{i,j}(F_{fea}^{flag}) = Lab_{i,j}^{nocon}\right).$$
(24)

The ENL values M_{flag} are denoted as M_G , M_A , M_W , flag is a member of $\{G, A, W\}$. They equal to the square ratio of the standard deviation σ_{flag} and the expectation μ_{flag} of intensity in SAR image area. Formula (24) calculates the confidence level of each classification map. If $classmap_{i,j}(F_{fea}^{flag})$, that is the pixel label at positions (i, j) in classification map of feature F_{fea}^{flag} , is equal to the same position pixel label $Lab_{i,j}^{nocon}$, which is final label result in the no-conflict case, then the confidence level CL_{flag} of feature F_{fea}^{flag} is increased by 1, that is, count the number of labels in each feature classification map that are equal to the final label results in the no-conflict case. CL_G , CL_A , CL_W count the number of times that each feature F_{fea}^{flag} classification map is consistent with the result of $Lab_{i,j}^{nocon}$ in no-conflict situations obtained by (21). Finally, the undetermined test pixel sample p(i, j) label is the label in F_{fea}^{flagx} classification map with the highest probability value Pro_{flagx} in conflict situations, which is

$$Lab_{i,j}^{conflict} = \left\{ l \middle| Pro_{flagx} = \max(Pro_G, Pro_A, Pro_W), \text{ then } l = \text{classmap}_{i,j}(F_{fea}^{flagx}) \right\}.$$
(25)

In the end, the label matrix *Lab* of classification result is obtained. The procedure of multi-feature fusion and adaptive kernel combination algorithm is outlined in Algorithm 1.

Algorithm 1 Multi-feature fusion and adaptive kernel combination

INPUT: Original SAR image $X \in I_1 \times I_2$, the training samples and their corresponding labels, and the training samples number of each class is n_i , the number of superpixels is num_{block} . **OUTPUT**: The label matrix *Lab* of classification result on the SAR image.

Step 1: The SAR image *X* is represented as three texture information tensors: GLCM feature tensor F_{fea}^G , WL feature tensor F_{fea}^W , and AP feature tensor F_{fea}^A .

Step 2: The SAR image X is segmented according to the number of superpixels num_{block} and Formula (10), and obtain S^G , S^W , S^A so that the superpixels result is mapped to the above three texture information tensors F^G_{fear} , F^W_{fear} , F^A_{fear} .

Step 3: The internal context information feature spaces F_{mean}^{G} , F_{mean}^{W} , F_{mean}^{A} are constructed on S^{G} , S^{W} , S^{A} according to Formula (12).

Step 4: The generalized spatial information feature spaces F_{spa}^{G} , F_{spa}^{W} , F_{spa}^{A} are constructed on F_{mean}^{G} , F_{mean}^{W} , F_{mean}^{A} according to Formula (15).

Step 5: The information feature kernels $K_{fea}^{flag}\left(F_{i}^{fea,flag}, F_{j}^{fea,flag}\right)$, $K_{mean}^{flag}\left(S_{i}^{mean,flag}, S_{j}^{mean,flag}\right)$, and $K_{spa}^{flag}\left(S_{i}^{spa,flag}, S_{j}^{spa,flag}\right)$ are constructed on three feature spaces F_{fea}^{flag} , F_{mean}^{flag} , and F_{spa}^{flag} according to Formulas (16)–(18), and $flag \in \{G, W, A\}$.

according to Formulas (16)–(18), and $flag \in \{G, W, A\}$. **Step 6**: The dynamic weights $W_{fea}^{flag}, W_{mean}^{flag}$ and W_{spa}^{flag} are calculated via Formula (20), and the

adaptive kernel combination gains composite kernels $K_{com}^{flag}\left(F_i^{com,flag}, F_j^{com,flag}\right)$ by $K_{fea}^{flag}\left(F_i^{fea,flag}, F_j^{fea,flag}\right), K_{mean}^{flag}\left(S_i^{mean,flag}, S_j^{mean,flag}\right)$, and $K_{spa}^{flag}\left(S_i^{spa,flag}, S_j^{spa,flag}\right)$; three classification mappings are acquired from these composite kernels with training samples. **Step 7**: According to Formulas (21) and (25), the three classification mappings are fused, and the final classification result *Lab* is obtained by local consistency optimization.

4. Results

4.1. Experimental Setup

In the experiment, three synthetic images and three real images are used to test six methods, namely Syn1, Syn2, Syn3, SAR1, SAR2, and SAR3. In the proposed classification framework MAKC, the numbers of original superpixels are set to 2000, 2800, 4300, 10,500, 64,500, and 4500, respectively. The AP uses four attributes to generate a twenty-five dimensional data set as [50 625 1849], [0.05 0.1 0.15], [58 248 328], [0.5 2 3.5]; the window size of GLCM is 9×9 , the direction step of GLCM is 2, and the window size of WL is 9×9 respectively according to the literature [2,4,5]. The radial basis function kernel parameter σ in (15)–(17) is set to 1 in our method, and the parameters of the SVM classifier in all methods are selected by a fivefold cross-validation on the training set. Fifty training samples are used as testing data. The parameters of the other test methods are set as the defaulted values reported in [9,10,17,34,35], respectively. The experiments were conducted on a laptop computer with an Intel Core i5 2.6-GHz CPU and 16-GB memory; the algorithms were implemented in MATLAB 2016b.

The classification results of the MAKC method are compared with the below methods visually and quantitatively: SVM [34], clone kernel spatial fuzzy c-means clustering (CKS-FCM) [35], composite kernel feature fusion (CKFF) [10], SVM-composite kernel (SVM-CK) [9], and multi-feature weighted sparse graph (MWSG) [17]. The MAKC uses the multi-feature texture information, context information, and generalized spatial information, constructs an adaptive composite kernel for classification with dynamic weights, and performs local consistency optimization to integrate multi-feature information. The SVM classifies images by statistical features without considering spatial information, and represents the classical basic SVM method. For the CKS-FCM method, spatial information is incorporated into the objective function of FCM, and a non-Euclidean distance based on a kernels metric is used in the kernel function. For the CKFF method, the spatial information is exploited by the morphological feature and stacked feature fusion with a composite kernel by manual weights on the SAR image. The SVM-CK method is used to extract context information for spatial information expression, the composite kernel is constructed by manual weights, and multi-feature fusion is not involved. In the MWSG method, multifeature samples distance is expressed by Gaussian kernel distance as weights integrated sparse representation, which constructs an adjacent matrix of a graph framework for SAR image classification; it expresses the overall spatial information. To objectively evaluate the classification results, four metrics of overall accuracy (OA), class-special accuracy (CA), average accuracy (AA), and Kappa coefficient (Kappa) are adopted on real SAR images.

4.2. Results on Synthetic SAR Images

In this section, the experiments performed on three synthetic SAR images are detailed. All of the three synthetic SAR images are corrupted by the complicated simulated noise with three-looks, which follows the multiplicative speckle rule [20]. The test synthetic images are named Syn1, Syn2, and Syn3, and contain different numbers of texture types: 2, 4, and 8, respectively. The size of each image is 512×512 pixels. To evaluate the classification performance, training samples are selected randomly in all of the comparison methods, taking fifty samples for each class. Moreover, the averages of the classification accuracies over ten runs are reported. The classification maps are shown in Figure 4, and the classification results for the overall accuracy and the Kappa coefficient are shown in Tables 1 and 2, respectively.



Figure 4. Cont.



Figure 4. Results of the synthetic SAR images: (a1–a3) ground truths; (b1–b3) synthetic SAR images; (c1–c3) SVM; (d1–d3) CKS-FCM; (e1–e3) CKFF; (f1–f3) SVM-CK; (g1–g3) MWSG; (h1–h3) MAKC, where the numbers represent the sequence index of the images.

Table 1. Overall accuracies of different algorithms on the synthetic SAR images.

| Method | Syn1 OA | Syn2 OA | Syn3 OA |
|---------|---------|---------|---------|
| SVM | 85.85 | 75.75 | 57.37 |
| CKS-FCM | 97.21 | 98.94 | 92.1 |
| CKFF | 98.91 | 92.26 | 92.85 |
| SVM-CK | 98.76 | 96.69 | 95.9 |
| MWSG | 99.31 | 97.66 | 97.14 |
| MAKC | 99.33 | 99.68 | 99.03 |

Table 2. Kappa coefficients of different algorithms on the synthetic SAR images.

| Method | Syn1 Kappa | Syn2 Kappa | Syn3 Kappa |
|---------|------------|------------|------------|
| SVM | 0.72 | 0.68 | 0.51 |
| CKS-FCM | 0.94 | 0.99 | 0.91 |
| CKFF | 0.98 | 0.9 | 0.92 |
| SVM-CK | 0.98 | 0.96 | 0.95 |
| MWSG | 0.99 | 0.97 | 0.97 |
| MAKC | 0.99 | 1 | 0.99 |

For the Syn1, results can be observed in Figure 4a1–h1, which is a two-class synthetic image. Figure 4a1 shows the ground truth image corresponding to Figure 4b1. The classification result of SVM is very poor, as shown in Figure 4c1. This is because SVM does not consider spatial information. Figure 4d1 shows the classification result of CKS-FCM, which has many spots in consistent regions. CKS-FCM combined with non-local spatial information has certain noise suppression, but it is still not ideal. Although Figure 4e1,g1 obtained by CKFF and MWSG, respectively, are better than Figure 4c1,d1, some pixels are obviously misclassified in consistent regions. CKFF and MWSG both take advantage of multi-feature information, CKFF stacks features simply, and MWSG eliminates singular data for computational fusion. More information is utilized than in SVM or CKS-FCM. In Figure 4f1, the SVM-CK obtains classification regions with no noise, but the edge of the segmentation regions is rough. SVM-CK only uses context information without feature fusion to emphasize the details. In contrast, the classification of MAKC shown in Figure 4h1 has good regional consistency and no noise. MAKC not only uses context information,

but also makes use of generalized spatial information, and adjust the details by local consistency optimization of multiple feature fusion.

Figure 4b2 shows a four-class simulated SAR image named Syn2, and the corresponding ground truth is shown in Figure 4a2, which contains four gray levels 0, 77, 179, and 255. The classification results of SVM and CKFF are very poor, as shown in Figure 4c2,e2, respectively. The four regions are not divided by SVM, and CKFF has dense noise points. The four different regions in Figure 4d2,g2 obtained by CKS-FCM and MWSG, respectively, can be separated. However, their classifications have a large number of pixels misclassified. In the classifications of SVM-CK in Figure 4f2, the edge of regions is fuzzy, and there are chaotic points here. The classification result of MAKC has better regional consistency than others, as shown in Figure 4h2. In the Syn2 image, the texture contour is varied, and the type is complex; SVM fails to embody nonlinear information among data, and CKFF does not mine spatial context information and is sensitive to noise under complex data. CKS-FCM only uses non-local spatial information to suppress noise, SVM-CK lacks generalized spatial information and does not use feature fusion, and MWSG has limited expression of local spatial information, and these methods do not enhance local spatial consistency except MAKC.

In Figure 4b3, the synthetic SAR image Syn3 consists of eight classes, and the graylevels of eight classes in Figure 4a3 come closer to each other, which will increase difficulty for the accurate classification and recognition. It is clear that some local patches are misclassified as other classes in the homogeneous areas in Figure 4c3. In Figure 4d3,e3, the speckle noise in the Syn3 cannot be suppressed by CKS-FCM and CKFF, and the pixels in gray and in white of classification results are mixed together significantly. For the classification results in Figure 4f3,g3, though these two algorithms can divide eight areas roughly, the edges are rugged and some local regions are still not well identified. They cannot produce accurate and clear classification for the Syn3. Fortunately and obviously, the proposed method MAKC in Figure 4h3 achieves the best classification result, not only in region consistency but also in boundary separability. This is because MAKC captures the most information and considers local consistency.

Table 1 lists the average overall accuracy of classification results over ten independent runs, and Table 2 shows the Kappa coefficients of different algorithms on the synthetic SAR images, which is consistent with the visual results. The results show that the proposed MAKC method achieves better classification performance in terms of OA and Kappa than the other state-of-the-art methods. The accuracy of the MAKC method reaches 99% on the synthetic images, especially on Syn2, where the Kappa coefficient reaches 1. The reason for this is that MAKC adopts a variety of features and makes full use of the neighborhood spatial information, the composite kernel fully expresses the nonlinear relationship, and local consistency optimization compensates for feature defects. Similarly, SVM-CK also employs nonlinear and context spatial information but without local consistency optimization and generalized spatial information to implement the classification.

4.3. Results of Real SAR Images

Three real SAR images (SAR1, SAR2, and SAR3) are tested. SAR1 has a size of 475×446 pixels, and was taken on 1 April 1993. It covers the Naval Air Weapons Station China Lake, CA, at the latitude $35^{\circ}41'14.25''$ north, the longitude $117^{\circ}41'29.27''$ west, with Ku-band radar with a 3-m resolution. The image includes three types of ground objects of runway (dark), bare land (gray), and infrastructures (bright). SAR2 has a size of 256×256 pixels, and was taken on 13 August 1997 with C-band radar with a 4-m resolution, and covers a suburb of Beijing, China, about latitude $40^{\circ}04'43.44''$ north, longitude $116^{\circ}11'28.43''$ east. This image includes three types of land cover: water (dark), crop (gray), and vegetation (bright). SAR3 has a size of 505×476 pixels and was taken on 19 June 2007 near to latitude $48^{\circ}31'13.99''$ north, longitude $43^{\circ}28'13.86''$ east. It covers an area of the south Russian steppes northeast of the Black Sea with a 15-m resolution X-band radar. The image consists of four types of land cover: water (dark), vegetation (dark gray), crop (light

gray), and bare land (bright). Figures 5–7 show different classification results obtained by different investigated methods on real SAR images with 50 training samples randomly selected for each class.



Figure 5. Results of a real SAR image: (a) ground truth; (b) real SAR image; (c) SVM; (d) CKS-FCM; (e) CKFF; (f) SVM-CK; (g) MWSG; (h) MAKC.



Figure 6. Results of a real SAR image: (a) ground truth; (b) real SAR image; (c) SVM; (d) CKS-FCM; (e) CKFF; (f) SVM-CK; (g) MWSG; (h) MAKC.



Figure 7. Results of a real SAR image: (a) ground truth; (b) real SAR image; (c) SVM; (d) CKS-FCM; (e) CKFF; (f) SVM-CK; (g) MWSG; (h) MAKC.

For the SAR1, results can be observed from Figure 5b, which includes three types of ground objects of runway, bare land, and airport infrastructures. Figure 5a shows the ground truth image corresponding to Figure 5b. The runway is marked as the first category, and is represented in purple in the classification maps of SAR1. The bare land is identified as the second category, which is indicated by gold in the classification maps of SAR1. The airport infrastructures are the third class, which is shown by orange in the classification map of SAR1. As can be observed, the SVM method of Figure 5c shows a very noisy estimation in its classification map because it only considers the pixel intensity information, with no combination of spatial information. By incorporating the non-local mean to suppress the noise of the SAR image, the CKS-FCM of Figure 5d delivers a smoother visual result. However, it fails to identify the region area although the edge is clear, e.g., the third class, airport infrastructures. By combining nonlinear spatial information, the CKFF in Figure 5e and the SVM-CK in Figure 5f have better visual performance, but still have some incorrect classification labels, e.g., confusion between bare land and airport infrastructures. By expressing the overall spatial information of the SAR image, MWSG in Figure 5g achieves a good result, but fails to identify the pixels in the detailed and edge regions, e.g., the first class, runway. By contrast, the proposed MAKC method in Figure 5h has the best visual classification performance, which not only reduces the speckles greatly but also preserves the meaningful structural information. Because the multi-feature description image is comprehensive, and nonlinear spatial information is mined richly, the local details of the image are considered through consistency enhancement, so the MAKC method achieves convincing results. The runway contour and area classification of the proposed method in the red rectangle are obviously better than those of other methods. Table 3 shows the corresponding classification accuracy of all methods, in which the best results are shown in bold. As can be seen, the proposed MAKC method achieves the highest classification accuracy in terms of AA, OA, and Kappa. The average accuracy is more than 7 percentage points higher than other methods, and the overall accuracy is more than 3 percentage points higher than other methods.

| CLASS | SVM | CKS-FCM | CKFF | SVM-CK | MWSG | MAKC |
|-----------------|-------|---------|-------|--------|-------|-------|
| runway | 56.00 | 88.58 | 82.83 | 85.08 | 92.50 | 88.54 |
| bare land | 60.83 | 61.85 | 91.40 | 87.21 | 91.85 | 96.33 |
| infrastructures | 97.05 | 25.01 | 61.52 | 86.62 | 68.11 | 92.39 |
| AA | 71.29 | 58.48 | 78.58 | 86.30 | 84.15 | 92.42 |
| OA | 60.85 | 65.16 | 89.47 | 86.87 | 91.45 | 95.06 |
| Kappa | 0.25 | 0.35 | 0.69 | 0.64 | 0.74 | 0.84 |

 Table 3. Classification precision results on SAR1.

Figure 6c-h illustrates different classification maps obtained by different test methods on the SAR2 image. As can be seen from Figure 6b, the SAR2 includes three types of land cover: crop, vegetation, and water. In the ground truth Figure 6a, the crop is marked as the first category, and is represented by gold in the classification maps of SAR2. The vegetation is identified as the second category, which is indicated by orange in the classification maps of SAR2. The water is the third class, which is shown by purple in the classification map of SAR2. Observed from Figure 6c,d, the SVM and CKS-FCM classification maps are very noisy because both do not clearly reflect nonlinear information. The CKFF in Figure 6e can deliver a comparatively clear result, but there are noise points in the local area, e.g., second class, vegetation, at the top right of the classification map. CKFF does not extract context spatial information and thus region smoothness is insufficient. Although the SVM-CK and MWSG algorithms of Figure 6f,g show improvements in the reduction of speckles, there are still deviations in local area details, e.g., third class, water. Both do not optimize the local of image and do not utilize generalized spatial information. By contrast, the proposed MAKC method in Figure 6h can provide the best visual performance in these areas, being more accurate in details, with great suppression of noise. This is due to the diversity of information captured by MAKC and the enhancement of local consistency. In the red box, the classification region edge of the proposed method is smoother than other methods and closer to the ground truth, and has less noise points. The corresponding quantitative results are shown in Table 4 with the best results in bold. As can be seen, compared with other methods, the proposed MAKC method exceeds the average of all methods by 4.3%, 4.5%, and 6.8% in terms of AA, OA, and Kappa with the highest values, respectively.

Table 4. Classification precision results on SAR2.

| CLASS | SVM | CKS-FCM | CKFF | SVM-CK | MWSG | MAKC |
|------------|-------|---------|-------|--------|-------|-------|
| crop | 92.43 | 83.32 | 89.92 | 93.00 | 88.31 | 95.79 |
| vegetation | 79.29 | 74.28 | 92.23 | 95.56 | 95.22 | 94.85 |
| water | 85.33 | 98.20 | 92.40 | 92.06 | 94.76 | 94.11 |
| AA | 85.68 | 85.27 | 91.51 | 93.54 | 92.76 | 94.92 |
| OA | 84.71 | 83.78 | 91.63 | 93.82 | 93.16 | 94.90 |
| Kappa | 0.77 | 0.75 | 0.87 | 0.91 | 0.89 | 0.92 |

In addition, another experiment is conducted on the SAR3 image in Figure 7b; the image consists of four types of land cover: water, vegetation, crop, and bare land. In ground truth Figure 7a, this water is labeled as the first category, and is represented by blue in classification maps of SAR3. Vegetation belongs to the second category, and in ground truth and in the SAR3 classification maps is purple. Crop is the third category and is orange in ground truth and in classification maps. The fourth category of bare land is in gold in ground truth and the classification maps. The quantitative results are represented in Table 5 with the best results in bold. As can be observed, the proposed MAKC method outperforms all the compared methods in terms of AA, OA, and Kappa when a small number of training samples is available. There are two main reasons for this. On the one hand, multi-feature can well retain the geometric characteristics of SAR images to provide more discriminative information for classification. On the other hand, the discriminative information is sufficiently utilized by mining the abundant spatial information within each feature and fusing the complementary information among features. A visual performance

comparison of different methods is represented in Figure 7c–h. As can be seen, the proposed MAKC method in Figure 7h provides the best regional consistency compared with other test methods. The best consistency of vegetation areas in the red box was the proposed method, while the boundary distinction between bare land and crops was better maintained by the proposed method than the other methods.

| CLASS | SVM | CKS-FCM | CKFF | SVM-CK | MWSG | MAKC |
|------------|-------|---------|-------|--------|-------|-------|
| water | 26.22 | 95.09 | 89.34 | 88.83 | 95.39 | 96.25 |
| vegetation | 0 | 42.42 | 75.60 | 78.90 | 84.26 | 76.41 |
| crop | 99.65 | 39.66 | 83.31 | 86.43 | 81.73 | 88.77 |
| bare land | 36.94 | 74.81 | 77.98 | 86.76 | 81.01 | 92.38 |
| AA | 40.70 | 62.99 | 81.55 | 85.23 | 85.59 | 88.45 |
| OA | 53.49 | 69.55 | 84.31 | 87.22 | 87.11 | 92.15 |
| Kappa | 0.36 | 0.58 | 0.77 | 0.81 | 0.81 | 0.88 |

 Table 5. Classification precision results on SAR3.

5. Discussion

5.1. Effects of Spatial Information on Different Numbers of Training Samples

The step of spatial information extraction within three features is removed, and the three groups complementary feature information is directly classified by SVM to create a fusion result for classification, while other parameters and operations remain unchanged. The corresponding classification performance is compared with the proposed MAKC on real SAR images, which is illustrated in Figure 8. It can be seen that this method has better classification performance for real SAR images, taking into account different numbers of training samples. For the SAR1, the classification accuracy decreases when spatial information extraction is removed. When only ten training samples (TN = 10) are selected for each class, the OA of the proposed MAKC method is 90.1%, while the OA of the MAKC without spatial information is 78.67%, which is shown in Table 6 with the best results in bold. It can be seen that the OA decreases by more than 10%. Because the spatial information is not extracted, the abundant detail information within each feature cannot be well exploited. The same situation can also be observed for the SAR2 and SAR3. Overall, the proposed MAKC can consistently achieve the best results in terms of classification accuracy when compared with the situation without using spatial information within features, so spatial information extraction is effective.

5.2. Effects of Local Consistency Optimization on Different Numbers of Training Samples

Figure 9 shows the classification result comparison between MAKC and without local consistency optimization on different real SAR images, with different numbers of training samples. It can be seen that the classification performance of MAKC outperforms other single features (i.e., AP, GLCM, and WL), because MAKC utilizes the complementary information among different feature groups by local consistency optimization. The advantage of the proposed method becomes more obvious when the available training samples are limited to only TN = 10. For example, as shown in Table 6, for the SAR1 image, the proposed method has the highest classification accuracy with OA = 90.1%, whereas the OAs of other single features (i.e., AP, GLCM, and WL) are 87.2%, 85.68%, and 85.89%, with a decrease of 2.9%, 4.42%, and 4.21%, respectively. The same results can also be observed on the other two data sets (e.g., SAR2 and SAR3). This demonstrates that local consistency optimization is effective in our proposed classification framework, as it assesses information quality and confidence level to fuse different features. Table 6 shows the accuracy of various variation methods and MAKC when the number of training samples is 10 on SAR1. The best results are shown in bold.



Figure 8. Classification result (in OA, AA, and Kappa) comparison between MAKC and MAKC without spatial information, with different numbers of training samples. This is demonstrated on (**a**) SAR1 image, (**b**) SAR2 image, and (**c**) SAR3 image. TN = n means the number of training samples is equal to n.

Table 6. Classification results on SAR1 with TN = 10.

| Method | Non-Spatial | AP | GLCM | WL | MAKC |
|--------|-------------|-------|--------|--------|--------|
| AA | 0.7173 | 0.864 | 0.7081 | 0.7306 | 0.8747 |
| OA | 0.7867 | 0.872 | 0.8568 | 0.8589 | 0.901 |
| Kappa | 0.46 | 0.65 | 0.6 | 0.6 | 0.7 |

5.3. Parameter Discussion of Superpixel Number

In this section, the effect of the number of original superpixels is investigated. The numbers of training and test samples are selected to be the same as in the aforementioned experiments on the SAR1, SAR2, and SAR3. According to the size of the image, the number of original superpixels for SAR1 is selected from 2500 to 18,500 with a step size of 1000. The number of original superpixels for the SAR2 is selected from 59,500 to 67,500 with a step size of 500, while the number of original superpixels on the SAR3 is selected from 1500 to 21,500 with a step size of 1000.

Figure 10 illustrates the OA values of the proposed MAKC method under different original superpixel numbers on all three real SAR images. For SAR1 in Figure 10a, it can be seen that when the number of superpixels is 10,500, the classification accuracy of OA reaches the optimum, marked with a red rectangle, and when the number of original superpixels is selected as other values, OA decreases. This is mainly due to the fact that when the number of original superpixels is fewer than 10,500, the spatial information of different categories may be contained in one superpixel, resulting in a decline in classification accuracy. On the other hand, when the number of original superpixels is higher than 10,500, the size of each superpixel will decrease, so that spatial information in a large homogeneous area of the SAR1 image cannot be fully utilized for classification. In addition, the same situation can also be seen on SAR2 in Figure 10b and SAR3 in Figure 10c with the optimal number of original superpixels being 64,500 and 4500, respectively.



Figure 9. Classification result (in OA, AA, and Kappa) comparison among AP, GLCM, WL, and MAKC, with different numbers of training samples. This is demonstrated on (**a**) SAR1 image, (**b**) SAR2 image, and (**c**) SAR3 image. TN = n means the number of training samples is equal to n.



Figure 10. Effects of the number of original superpixels on (**a**) SAR1 image (TN = 50), (**b**) SAR2 image (TN = 50), and (**c**) SAR3 image (TN = 10, TN = 30, and TN = 50). TN = n means the number of training samples is equal to n.

Note that in Figure 10c, the number of training samples is large, and the accuracy value is stable and high; conversely, the accuracy value fluctuates greatly and is low. The reason for this is that the number of training samples is large, the extracted discrimination features are more, and the recognition ability is improved, so the classification results change little. At the same time, with different numbers of training samples, the best number of original superpixels for the SAR3 image is the same, which demonstrates the robustness of the proposed method.

6. Conclusions

In this paper, a new multi-feature fusion classification framework is proposed to exploit the contextual and generalized spatial information, and dynamically mine nonlinear information by adaptive kernel combination on SAR images, while at the same time enhancing the local consistency, which is termed the MAKC method. First, three discriminant and complementary texture information tensors can be constructed by extracting APs, GLCMs, and WLs from an SAR image. Then, 3D feature blocks are obtained by segmentation mapping based on the gamma distribution and non-negative logarithmic likelihood difference on the SAR image. Feature spaces of context and generalized spatial information are constructed, respectively; adaptive kernel combination explores nonlinear information, and the information within features is fused. Furthermore, local consistency optimization is implemented to fuse complementary information among features. In this way, the different spatial information within each feature group, the abundant texture information of the SAR image, and the complementary information among feature groups are well utilized. The classification results of the proposed MAKC on three synthetic SAR images and three real SAR images are better than several state-of-the-art classification methods in both quantitative and visual performance, which proves the effectiveness of this method.

In the proposed framework, the multi-feature adaptive kernel combination is used to explore the spatial and nonlinear information within the feature tensors. Our future work will take advantage of the framework with other features and combine them with other classifiers to improve classification accuracy, e.g., neural networks. In addition, since the feature information is a tensor, tensor analysis should be studied to extract more discriminant spatial information.

Author Contributions: Methodology, X.W. (Xiaoying Wu) and X.W. (Xianbin Wen); resources, X.W. (Xiaoying Wu); validation, H.X., L.Y., and C.G.; writing—original draft, X.W. (Xiaoying Wu); writing—review & editing, X.W. (Xianbin Wen). All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded in part by the National Natural Science Foundation of China under Grant 61472278, in part by the Major Project of Tianjin under Grant 18ZXZNGX00150, in part by the Natural Science Foundation of Tianjin under Grant 18JCYBJC84800, in part by the Scientific Research Key Project of Tianjin Municipal Education Commission under Grant 2017ZD13, and in part by the Scientific Research Project of Tianjin Municipal Education Commission under Grant 2017KJ255.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Oliver, C.; Quegan, S. Understanding Synthetic Aperture Radar Images; SciTech: West Perth, Australia, 2004; pp. 154–196.
- Gang, Z.; Xiaofeng, L.; Lizhang, Z.; Jingsong, Y.; Lin, R.; Peng, C.; Huaguo, Z.; Xiulin, L. Development of a Gray-Level Co-Occurrence Matrix-Based Texture Orientation Estimation Method and Its Application in Sea Surface Wind Direction Retrieval From SAR Imagery. *IEEE Trans. Geosci. Remote Sens.* 2018, 56, 5244–5260.
- Sumaiya, M.N.; Shantha, S.K.R. Gabor filter based change detection in SAR images by KI thresholding. *Optik* 2017, 130, 114–122. [CrossRef]
- Shahrezaei, H.; Kim, H. Fractal Analysis and Texture Classification of High-Frequency Multiplicative Noise in SAR Sea-Ice Images Based on a Transform- Domain Image Decomposition Method. *IEEE Access* 2020, *8*, 40198–40223. [CrossRef]
- 5. Boldt, M.; Thiele, A.; Schulz, K.; Stefan, H. SAR Image Segmentation Using Morphological Attribute Profiles. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2014**, XL-3, 39–44. [CrossRef]
- Ayşe, T.; İlter, T.; Aptoula, E.; Koray, K. Pixel-Based Classification of SAR Images Using Feature Attribute Profiles. *IEEE Geosci. Remote Sens. Lett.* 2019, 16, 564–567.
- Omati, M.; Sahebi, M.R. Change Detection of Polarimetric SAR Images Based on the Integration of Improved Watershed and MRF Segmentation Approaches. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2018, 11, 4170–4179. [CrossRef]
- 8. Li, X.; Su, J.; Yang, L. Building Detection in SAR Images Based on Bi-Dimensional Empirical Mode Decomposition Algorithm. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 641–645. [CrossRef]
- 9. Guan, D.; Xiang, D.; Dong, G.; Tang, T.; Tang, X.; Kuang, G. SAR Image Classification by Exploiting Adaptive Contextual Information and Composite Kernels. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 1035–1039. [CrossRef]
- Wang, X.; Feng, J.; Cao, Z.; Min, R. Polarimetric-spatial classification of PolSAR images based on composite kernel feature fusion. In Proceedings of the IEEE Radar Conference 2017, Seattle, WA, USA, 8–12 May 2017.
- 11. Zhang, X.; Wang, Y.; Li, D.; Tan, Z.; Liu, S. Fusion of multifeature low-rank representation for synthetic aperture radar target configuration recognition. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 1402–1406. [CrossRef]

- 12. Biao, H.; Bo, R.; Guilin, J.; Huiyan, L.; Jin, Z.; Licheng, J. SAR image classification via hierarchical sparse representation and multisize patch features. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 33–37.
- Zhang, W.; Xiang, D.; Su, Y. Fast Multiscale Superpixel Segmentation for SAR Imagery. IEEE Geosci. Remote Sens. Lett. 2020, 1–5. [CrossRef]
- 14. Shang, R.; Lin, J.; Jiao, L.; Yang, X.; Li, Y. Superpixel Boundary-Based Edge Description Algorithm for SAR Image Segmentation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 1971–1985. [CrossRef]
- 15. Li, T.; Liu, Z.; Xie, R.; Ran, L. An Improved Superpixel-Level CFAR Detection Method for Ship Targets in High-Resolution SAR Images. *IEEE J. Selected Top. Appl. Earth Observ. Remote Sens.* 2018, 11, 184–194. [CrossRef]
- 16. Dong, G.; Kuang, G. Target Recognition via Information Aggregation through Dempster–Shafer's Evidence Theory. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1247–1251. [CrossRef]
- 17. Jing, G.; Licheng, J.; Fang, L.; Xiangrong, Z.; Xu, T.; Puhua, C. Multi-Feature Weighted Sparse Graph for SAR Image Analysis. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 881–891.
- 18. Geng, J.; Wen, J.; Xinyang, D. Multi-scale deep feature learning network with bilateral filtering for SAR image classification. *ISPRS. J. Photogramm.* **2020**, *167*, 201–213. [CrossRef]
- Leyuan, F.; Nanjun, H.; Shutao, L.; Pedram, G.; Jón, A. Extinction Profiles Fusion for Hyperspectral Images Classification. *IEEE Trans. Geosci. Remote Sens.* 2018, 56, 1803–1815.
- Xiaoying, W.; Xianbin, W.; Liming, Y.; Changlun, G.; Haixia, X. Cluster-Based Tensorial Semisupervised Discriminant Analysis for Feature Extraction of SAR Images. *IEEE Access* 2019, 7, 84318–84332.
- Clausi, D.A. Comparison and fusion of co-occurrence, Gabor and MRF texture features for classification of SAR sea-ice imagery. *Atmos. Ocean* 2001, 39, 183–194. [CrossRef]
- 22. Fukuda, S.; Hirosawa, H. Suppression of speckle in synthetic aperture radar images using wavelet. *Int. J. Remote Sens.* **1998**, *19*, 507–519. [CrossRef]
- Rajamani, A.; Krishnaveni, V. Performance Analysis Survey of Various SAR Image Despeckling Techniques. Int. J. Comput. Appl. 2014, 90, 5–17. [CrossRef]
- 24. 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.* 2014, *11*, 1290–1294. [CrossRef]
- 25. Shannon, C.E. A mathematical theory of communication. Bell Syst. Tech. J. 1948, 27, 623–656. [CrossRef]
- Chari, S.K.; Gupta, A.; Gupta, P.; Mohan, J. Threshold selection in image segmentation using parametric entropy measures. In Proceedings of the 2017 Fourth International Conference on Image Information Processing (ICIIP), Shimla, India, 12 March 2018.
- 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. STARS* 2017, 10, 4045–4058. [CrossRef]
- Liu, M.; Tuzel, O.; Ramalingam, S.; Chellappa, R. Entropy rate superpixel segmentation. In Proceedings of the 24th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011, Colorado Springs, CO, USA, 20 June 2011.
- Yoshiyama, F.; Uenohara, S.; Nishijima, K.; Hioka, Y.; Furuya, K. Log-likelihood method to select initial values of multichannel non-negative matrix factorization. In Proceedings of the 2015 IEEE International Conference on Consumer Electronics, Taipei, Taiwan, 6 June 2015.
- Lu, J.; Hu, J.; Zhou, J. Deep Metric Learning for Visual Understanding: An Overview of Recent Advances. *IEEE Signal Proc. Mag.* 2017, 34, 76–84. [CrossRef]
- 31. Jacques, F. Parameter-free fast pixelwise non-local means denoising. Image Process. Line 2014, 4, 300–326.
- 32. Abdi, H.; Williams, L.J. Principal component analysis. Wiley Interdiscip. Rev. Comput. Stat. 2010, 2, 433–459. [CrossRef]
- 33. Luis, G.; Raydonal, O.; Alejandro, F. Unassisted Quantitative Evaluation of Despeckling Filters. Remote Sens. 2017, 9, 389.
- 34. Qu, G.; Yu, Q.; Wang, Y. An improved method for SAR image coastline detection based on despeckling and SVM. In Proceedings of the IET International Radar Conference 2013, Xi'an, China, 27 April 2013.
- 35. Shang, R.; Tian, P.; Jiao, L.; Stolkin, R.; Feng, J.; Hou, B.; Zhang, X. A Spatial Fuzzy Clustering Algorithm with Kernel Metric Based on Immune Clone for SAR Image Segmentation. *IEEE J. STARS* **2016**, *9*, 1640–1652. [CrossRef]