Letter

A Maximum-Information-Minimum-Redundancy-Based Feature Fusion Framework for Ship Classification in Moderate-Resolution SAR Image

Gaoyu Zhou, Gong Zhang * and Biao Xue

Key Laboratory of Radar Imaging and Microwave Photonics, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China; zhou_gaoyu@nuaa.edu.cn (G.Z.); xuebiao@nuaa.edu.cn (B.X.)

* Correspondence: gzhang@nuaa.edu.cn; Tel.: +86-138-1389-6509

Abstract: High-resolution synthetic aperture radar (SAR) images are mostly used in the current field of ship classification, but in practical applications, moderate-resolution SAR images that can offer wider swath are more suitable for maritime surveillance. The ship targets in moderate-resolution SAR images occupy only a few pixels, and some of them show the shape of bright spots, which brings great difficulty for ship classification. To fully explore the deep-level feature representations of moderate-resolution SAR images and avoid the “dimension disaster”, we innovatively proposed a feature fusion framework based on the classification ability of individual features and the efficiency of overall information representation, called maximum-information-minimum-redundancy (MIMR). First, we applied the Filter method and Kernel Principal Component Analysis (KPCA) method to form two feature subsets representing the best classification ability and the highest information representation efficiency in linear space and nonlinear space. Second, the MIMR feature fusion method is adopted to assign different weights to feature vectors with different physical properties and discriminability. Comprehensive experiments on the open dataset OpenSARShip show that compared with traditional and emerging deep learning methods, the proposed method can effectively fuse non-redundant complementary feature subsets to improve the performance of ship classification in moderate-resolution SAR images.

Keywords: moderate-resolution SAR image; feature fusion; filter method; kernel principal component analysis (KPCA); maximum-information-minimum-redundancy (MIMR); ship classification

Citation: Zhou, G.; Zhang, G.; Xue, B. A Maximum-Information-Minimum-Redundancy-Based Feature Fusion Framework for Ship Classification in Moderate-Resolution SAR Image. *Sensors* **2021**, *21*, 519. <https://doi.org/10.3390/s21020519>

Received: 24 November 2020

Accepted: 11 January 2021

Published: 13 January 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/>).

1. Introduction

Synthetic aperture radar (SAR) is widely used in maritime surveillance and ship monitoring due to its all-time and all-weather observing ability that covers wide areas [1]. Knowledge about the location and type of ship targets can be applicable to multiple applications, such as maritime traffic management, oil spill pollution detection, and illegal smuggling monitoring [2]. Ship classification in SAR images has been studied in depth and has developed into an important part of many operable marine monitoring systems.

Since the swath of SAR images is inversely proportional to image resolution, high-resolution SAR images are usually obtained at the expense of narrow swath. A high-resolution SAR image at the resolution of about 1 m usually corresponds to a swath of 5 km. Such narrow swath may be suitable to inspect specific locations on Earth, but are unsuitable for maritime surveillance. The current satellite-borne SAR can provide swaths of 100–450 km at image resolution of 10–30 m [3], which is more suitable for maritime surveillance in actual situations. In this context, we propose an effective method for ship classification in moderate-resolution SAR images.

Unlike optical images, SAR images are essentially the spatial distribution of electromagnetic wave reflection intensity. The challenges and difficulties of ship target recognition in satellite-borne SAR images are summarized as follows [4]: first, the imaging scene is

large, resulting in densely distributed, numerous types of sea targets; second, the imaging distance is long, resulting in extremely low echo signal-to-noise ratio, so the ship targets are not obvious; third, the imaging environment is complex, which brings about a problem that the imaging quality is significantly reduced under high sea conditions, especially moving targets. The most intuitive result of the above problems is that the ship target occupies only a few pixels in a moderate-resolution SAR image and the visual features are extremely inconspicuous, which brings great difficulties to the maritime target recognition [5,6]. Therefore, how to extract the robust features of the maritime target in the complex sea surface environment is the key to the ship target recognition technology.

The typical SAR image ship target recognition is mainly divided into four stages: pre-processing, target detection, feature extraction, and target classification. Among them, feature extraction is a key factor that affects the ship classification in SAR images, and the selection of features directly determines the accuracy of classification. The widely used features in moderate-resolution SAR images are physical features, such as geometric structure features [7], gray-scale statistical features and so on. Lang et al. [3] proposed naive geometric features (NGFs) for ship classification in moderate-resolution SAR images, which are all derived from two basic features: length and width. It's discovered that when the feature dimension exceeds a certain value, the performance of the classifier will gradually deteriorate, which is called the "curse of dimensionality" [8] in the field of machine learning. M. Dash et al. [9] proposed that a typical feature selection method has four basic steps: generation procedure, evaluation function, stopping criterion, and validation procedure. Kira et al. [10] defined that under ideal conditions, feature selection is to find the minimum feature subset that is necessary and sufficient to identify the target. The existing feature selection methods can be mainly divided into Filter methods and Wrapper methods [11], which have played an important role in the feature selection stage of target classification in SAR images, and have been active for a long time in the field of ship classification [12–15]. The Filter methods evaluate the predictive ability of each feature according to a certain criterion, thereby selecting several "better" features to form a feature subset. Its significant advantage is that it can quickly remove some non-critical noise features and can be used as a feature pre-selector. The Wrapper methods train the subsequent classifiers with the selected feature subset directly, and evaluate the feature subset according to the classification performance of the test set. It relies on subsequent classifiers, and iteratively trains the classifiers to make it computationally expensive, and the computational efficiency is lower than that of the Filter method.

These algorithms are intuitive and efficient, and they do not change the essential attributes of features, but rely on the accuracy and stability of feature evaluation criteria. Therefore, more nonlinear dimensionality reduction methods have emerged, such as the kernel principal component analysis (KPCA) method [16]. KPCA is one approach of generalizing linear Principal Component Analysis (PCA) into a nonlinear case using the kernel method and it has shown a significant performance in SAR image classification [17,18]. KPCA utilizes the potential nonlinear characteristics of the data in the feature space by introducing an appropriate kernel function in the vector inner product calculation process, and the specific representation of nonlinear mapping does not need to be known, which greatly reduces the amount and complexity of calculation.

With the rapid development of artificial intelligence technology, numerous deep learning (DL) methods have been developed for SAR target recognition. The most commonly used is convolutional neural network (CNN), but it heavily relies on the quality and quantity of training data [19], which is not suitable for our research topic of moderate-resolution SAR image classification in this article. In [19], CNN and metric learning are combined to perform SAR image classification tasks and get a recognition accuracy of 83.67% on the OpenSARShip dataset. Shao et al. [20] proposed a channel-wise and spatial attention (CSA) block which introduces channel attention and spatial attention mechanisms at the same time to enhance the feature extraction ability, and achieved an accuracy of 84% on the OpenSARShip dataset. Recently, a DL algorithm that does not have high requirements for the

amount of training data has emerged, that is, the Stacked Autoencoder (SAE). It is a stack of the shallow autoencoder (AE) model, and each layer is based on the expression features of the previous layer. It is an unsupervised learning framework that uses a backpropagation algorithm to reconstruct the input at the output while minimizing the reconstruction error, and performs better than AE. Chen et al. [21] proposed a deep feature extraction method based on SAE and perform classification on hyperspectral data sets, achieving an average accuracy of 94%. Gadhiya et al. [22] applied a SAE to reduce the dimensionality of input feature vector while retaining useful features of the input for multi-frequency PolSAR image. The lowest classification accuracy can reach 91% when using the softmax classifier, which also proves the effectiveness and superiority of SAE in SAR image classification. In this article, the SAE network will be used for comparative experiments.

Based on the above analysis, we consider the feature extraction problem from two perspectives, one is to extract the most separable features, and the other is to extract the most robust and effective features. Therefore, the Filter method is applied to evaluate features and the features with better classification capabilities are selected into the feature subset. And the KPCA method is applied to retain the principal components of the original features and maximize the information representation efficiency of the feature subset to the target. Then, the maximum-information-minimum-redundancy (MIMR) method is utilized to perform reliable feature fusion to maximize information representation and minimize feature redundancy, and different classifiers are selected for ship classification. In Section 2, we will introduce the formation process of two different feature subsets after dimensionality reduction, including the Filter method and KPCA method. The feature fusion framework based on the MIMR method is presented in Section 3. In Section 4, we will first introduce our experimental dataset, and we will present the results of experiments. And Section 5 concludes our work and puts forward future research directions.

2. Materials and Methods

The overview of the basic framework is shown in Figure 1. It mainly includes three parts: feature extraction, feature dimensionality reduction, and feature fusion. For the feature extraction part, taking into account the characteristics of the experimental data in this article, that is the lower resolution results in the inconspicuous structural details of the ship, and faces the problem of imbalance in the data set among various categories, we extract the following features to form the original feature set $F_0 = \{x_1, x_2, x_3, \dots, x_{23}\}$ after preprocessing the SAR images and extracting the region of interest (ROI).

Figure 1. Overall block diagram based on the maximum-information-minimum-redundancy (MIMR) multi-feature fusion framework.

- Geometric structure features [23]: area x_1 , perimeter x_2 , shape complexity x_3 , length x_4 , width x_5 , aspect ratio x_6 , the maximum distance from the center of mass to the target pixel x_7 , and the average distance from the center of mass to the target pixel x_8 .
- Brightness features [23]: quality x_9 , mean x_{10} , and standard deviation x_{11} .
- Texture features [23]: energy x_{12} , entropy x_{13} , moment of inertia x_{14} , correlation x_{15} , and moment of inverse difference x_{16} .
- Moment invariant features [24]: 7 Hu moment invariant features with translation, rotation and scale invariance $x_{17} \sim x_{23}$.

Generally, the original feature value has a large range. For example, x_8 may be as high as 10^4 , while the variation range of x_{12} is (0,1). In order to eliminate the negative impact of the feature range on the classification effect, we first normalize the extracted features. Then, we adopted two feature dimensionality reduction methods to obtain two feature subsets in linear space and non-linear space, F_{Filter} and F_{KPCA} , to reduce the redundancy between features and ensure the maximum efficiency of the information representation of the target.

2.1. Filter Method

In the Filter method, evaluation criteria based on distance measurement, information measurement, and dependency measurement are applied to feature evaluation. In this article, we evaluate original features by three evaluation criteria and the comprehensive ranking method is utilized to select several better features as a candidate feature subset.

- Separability

The separability metric uses the distance between samples to measure the separability of features for different types of samples. The most commonly used distance measures are intra-class distance and inter-class distance. The definition of these two distance measures is expressed as [10]

$$S_{W_i} = \sum_{\omega=1}^3 \frac{N_{\omega}}{N} \left(\frac{1}{N_{\omega}} \sum_{s=1}^{N_{\omega}} (x_i - E(x_i))^2 \right) \quad (1)$$

$$S_{B_i} = \sum_{\omega=1}^3 \frac{N_{\omega}}{N} \left((E(x_i) - E_{x_i})^2 \right) \quad (2)$$

where i is the feature label and ω is the type of ships. $E_{x_i} = \frac{1}{N} \sum_{\omega=1}^3 \sum_{s=1}^{N_{\omega}} x_i$ is the mean vector of all the samples; N_{ω} is the number of the ω_{th} type samples and N is the number of all the samples. The intra-class distance and inter-class distance measures reflect the structural information spread by the sample from different levels, and the most commonly used digital feature for feature selection is the ratio of the intra-class distance to the inter-class distance, which is defined as [25]:

$$J_i = \frac{S_{B_i}}{S_{W_i}} \quad (3)$$

We hope that features can effectively distinguish different target categories, that is, the smaller the intra-class distance is, the better, and the larger the inter-class distance is, the better. Therefore, the larger J_i of the feature is, the better its separability is.

- Stability

A subset of features for which a classifier obtains the best results may not necessarily be applicable to other classifiers. Therefore, the stability of the feature itself is also a factor that must be paid attention to in feature selection. The normalized coefficient of variance is used to measure the stability of the target feature, which is defined as:

$$ST_i = \frac{E(x_i^2) - E^2(x_i)}{E(x_i^2)} \quad (4)$$

where $E(x_i^2)$ is the mean square of the feature and $E^2(x_i)$ is the square of the mean of the feature. The smaller the normalized variance coefficient of the feature ST_i is, the more stable the feature is [25].

- Pearson correlation coefficient

In statistics, the Pearson correlation coefficient is often used to measure the degree of linear correlation of variables, which is defined as the ratio of covariance to standard deviation. The greater the absolute value of the correlation coefficient, the stronger the correlation, that is, the closer the correlation coefficient is to 1 or -1 , the stronger the correlation; the closer the correlation coefficient is to 0, the weaker the correlation, which is defined as:

$$\rho_{i,j} = \frac{\sum x_i x_j - \frac{\sum x_i \sum x_j}{N}}{\sqrt{\left(\sum x_i^2 - \frac{(\sum x_i)^2}{N}\right) \left(\sum x_j^2 - \frac{(\sum x_j)^2}{N}\right)}} \quad (5)$$

Obviously, the lower the $\rho_{i,j}$ is, the smaller the redundancy between features x_i and x_j is, and the more they should be selected into the candidate feature subset.

After feature extraction, we first analyze the separability, stability and Pearson correlation coefficient of features. Then we comprehensively evaluate the features according to their separability and stability, and subtract the features when the separability is lower than a certain threshold and the stability is higher than a certain threshold. Next we analyze the correlation coefficients of the remaining features, and select the features with lower correlation with other features into the feature subset. Finally, a feature vector is constructed for each ship chip: $F_{Filter} = \{x_1^{(1)}, x_2^{(1)}, x_3^{(1)}, \dots, x_{M0}^{(1)}\}$.

2.2. KPCA Method

An initial motivation of KPCA is to perform PCA in the high-dimensional feature space. As we know, PCA minimizes the distance between the data and the linear principal component; while KPCA first selects a suitable nonlinear mapping, and then maps the original data into a high-dimensional nonlinear feature space F , to minimize the distance between the data and the nonlinear principal component [16]. KPCA constructs an optimal plane in this high-dimensional space to make the mapped data separable. Then PCA is performed on the data mapped to the high-dimensional space to obtain a better classification and recognition effect.

Given the sample set $X = \{x_1, x_2, \dots, x_i, \dots, x_L\}$ ($i = 1, 2, \dots, L$), where $x_i \in R^n$ is a vector of n dimension, R^n is the put data space and L is the total number of samples. The pre-selected nonlinear mapping is Φ , and the high-dimensional nonlinear mapping space after mapping is F , so the samples in the feature space is donated by:

$$\Phi(X) = \{\Phi(x_1), \Phi(x_2), \dots, \Phi(x_L)\} \quad (6)$$

Kernel techniques are introduced to avoid huge calculations. K is a $L \times L$ kernel matrix which is defined by the inner product:

$$K_{ij} = K(x_i, x_j) = \Phi^T(x_i) \Phi(x_j) \quad (7)$$

The covariance matrix on the feature space F can be constructed by:

$$C = \frac{1}{L} \sum_{i=1}^L \Phi(x_i) \Phi(x_i)^T \quad (8)$$

The corresponding eigenvalue equation is:

$$Cv = \lambda v \quad (9)$$

The principal components are decided by the eigenvalue λ and the eigenvector v of the covariance matrix C . According to the reproductive kernel theorem, that is, any vector in the space (even the base vector) can be linearly represented by all samples in the space. So the eigenvector v can be expressed as:

$$v = \sum_{i=1}^L \alpha_i \Phi(x_i) \quad (10)$$

Therefore, Equation (9) can be transformed to:

$$K\alpha = \lambda' \alpha \quad (11)$$

where $\lambda' = L\lambda$ are the eigenvalues. The eigenvalues are obtained and sorted in descending order $\lambda_1' \geq \lambda_2' \geq \dots \geq \lambda_n'$, and the corresponding eigenvectors are $\alpha_1, \alpha_2, \dots, \alpha_n$. We select the first P_0 principal components $\alpha_1, \alpha_2, \dots, \alpha_{P_0}$ according to the cumulative contribution rate of the eigenvalues. Then we calculate the projection $Y = K\alpha$ of the sample X , and Y is the original data after dimensionality reduction, that is, the reduced feature vector $F_{KPCA} = \{y_1^{(2)}, y_2^{(2)}, y_3^{(2)}, \dots, y_{P_0}^{(2)}\}$ for each ship-chip.

The kernel function provides a connection from linear to nonlinear and any algorithm that can represent the dot product between two vectors. The specific representation of mapping Φ does not need to be known, so the amount of calculation is greatly reduced, and the complexity is also eased. Here we choose the Gaussian kernel function with better learning ability.

3. MIMR Feature Fusion Method

In addition to using a certain feature alone, the rational use of different types of features combined with each other can achieve complementary effects. The most important problem to be solved is how to effectively combine different types of features, which is called feature-level fusion [26]. Feature-level fusion can not only increase the feature information of the image, but also effectively integrate the advantages between different features and improve the adaptability of the algorithm. At present, the existing feature fusion algorithms are mainly divided into three categories: feature combination, feature selection, and feature conversion [27,28]. A simple feature combination may result in useless redundant information between various dimensions, which will affect the performance of the classifier. Therefore, we need to consider that the number of feature subsets is as low as possible to efficiently represent the target, while taking into account the low redundancy between features. Inspired by the minimal-redundancy-maximal-relevance (mRMR) method [25], this paper proposes a maximum-information-minimum-redundancy (MIMR) feature fusion method.

The idea of the MIMR method is to consider the information representation efficiency of the target feature subsets with different physical attributes while minimizing the redundancy between feature vectors F_{Filter} and F_{KPCA} . This coincides with our original intention, which is to reduce the redundancy of different feature vectors while ensuring classification accuracy. Various features are normalized before fusion, which is of great significance to the improvement of the overall classification effect.

Given the feature set $F_{con} = [F_{Filter}, F_{KPCA}]$, and the target type ω , the feature selection problem is to find a feature set $S = [\lambda F_{Filter}, \mu F_{KPCA}]$ that “optimally” characterizes ω . The optimal characterization condition often means the maximum percentage of correct classification (PCC), which usually requires the feature set S for classification to contain as much information of the target as possible.

Furthermore, it has been recognized that the combinations of individually good features do not necessarily lead to good classification performance. In other words, there may be redundancy between “good” features. Here we utilize mutual information (MI) to represent redundancy, and minimum redundancy represents the smallest correlation

between features, that is, the smallest MI between features. Given two random variables X and Y , their MI is defined in terms of their probabilistic density $p(x)$, $p(y)$, and $p(x, y)$:

$$I(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \quad (12)$$

Therefore, the optimization model can be established as

$$\begin{aligned} \min_{\lambda_i, \mu_j} R &= \sum_{i=1}^M \sum_{j=1}^P I(\lambda_i x_i, \mu_j y_j) \\ \text{s.t.} &\begin{cases} \sum_{i=1}^M \lambda_i = 1 \\ \sum_{j=1}^P \mu_j = 1 \end{cases} \end{aligned} \quad (13)$$

where $\{x_i, i = 1, 2, \dots, M\}$ and $\{y_j, j = 1, 2, \dots, P\}$ are the feature vectors in F_{Filter} and F_{KPCA} . $\{\lambda_i, i = 1, 2, \dots, M\}$ and $\{\mu_j, j = 1, 2, \dots, P\}$ are the weight coefficients of a single feature vector in F_{Filter} and F_{KPCA} . The requirement to solve this optimization model is actually to solve a constrained nonlinear multivariable function [29]. The idea of solving this optimization model is to find a set of optimal feature weight vectors $\{\lambda_1, \lambda_2, \dots, \lambda_M, \mu_1, \mu_2, \dots, \mu_P\}$ through iterative optimization, so that the MI, that is, the redundancy between feature vectors is minimized while ensuring PCC. Finally, an optimal fused feature vector will be constructed by $S = \{\lambda_1 x_1, \lambda_2 x_2, \dots, \lambda_M x_M, \mu_1 y_1, \mu_2 y_2, \dots, \mu_P y_P\}$.

To further improve efficiency, we first adopted the Fisher Score method [29] to make a preliminary judgment on the concatenated features F_{con} . According to the Fisher criterion, when the Fisher Score of the selected feature is higher, it means that the feature makes the distance between different categories in the sample points farther, the closer the distance between samples of the same category, and the classification ability of the feature is better. The Fisher Score of the i_{th} feature is defined as [30]:

$$FS_i = \frac{\sum_{\omega=1}^3 N_{\omega} (E(x_i) - E_{x_i})^2}{\sum_{\omega=1}^3 N_{\omega} \sigma_i^2} \quad (14)$$

where i is the feature label and ω is the type of ships. E_{x_i} is the mean vector of all the samples, and σ_i^2 represents the variance of the ω_{th} type sample corresponding to the i_{th} feature variable. When the Fisher Score of a feature is extremely low (lower than the threshold), the feature will be subtracted. Then we perform the MIMR feature fusion method on the remaining features.

In terms of classifier selection, we consider several mature classifiers including: k-Nearest Neighbor (KNN) classifier based on Euclidean distance, support vector machine (SVM) classifier based on kernel function, and neural network classifier based on artificial intelligence [31–33]. KNN classifier is simple to calculate and easy to execute, making it the most commonly used classifier in target recognition; SVM classifier is a new machine learning method aiming at finding the optimal classification surface. Due to its excellent learning ability, it has also been widely used in SAR image target classification and recognition tasks; neural network classifier for ship classification in SAR images has become a boom, but the lack of a sufficient number of labeled class-balanced databases limits its application and development.

Considering that the number of samples between the categories of experimental data in this article is extremely unbalanced, it is difficult for neural network classifiers to train a good network model, so here we choose KNN classifier and SVM classifier to ensure the classification accuracy and the speed of the algorithm, and we use LibSVM [34] implementation.

The detailed process to obtain the weight coefficients $\{\lambda_1, \lambda_2, \dots, \lambda_m, \mu_1, \mu_2, \dots, \mu_p\}$ of fused feature vector by the MIMR method is shown in Algorithm 1.

Algorithm 1 The process of the MIMR algorithm

Step 1. After normalizing the feature vectors F_{Filter} and F_{KPCA} , perform feature combination to obtain a feature subset $F_{con} = [F_{Filter}, F_{KPCA}]$.

Step 2. Utilize the Fisher Score method to evaluate the $M_0 + P_0$ features in F_{con} , we get $\{FS_1, FS_2, \dots, FS_K\}$. Sets the threshold ε , when $FS_k < \varepsilon$, the feature x_k are subtracted.

Step 3. For the remaining $M + P$ features, initialize the weight vectors of a single feature vector $\sum_{i=1}^M \lambda_i^0 = 1, \sum_{j=1}^P \mu_j^0 = 1$ (constraints).

Step 4. Calculate the mutual information between the M features of F_{Filter} and P features of F_{KPCA} , we will get a $M \times P$ MI matrix $I(\lambda_i x_i, \mu_j y_j)$. Then calculate the objective function

$$R = \sum_{i=1}^M \sum_{j=1}^P I(\lambda_i x_i, \mu_j y_j).$$

Step 5. Utilize iterative optimization method to find the local minimum that satisfies the constraints. Compare the objective function R with the default value of the optimality tolerance ξ . When $R > \xi$, return to step 3; when $R < \xi$, go to step 6.

Step 6. When objective function R is non-decreasing in feasible directions, that is, reaches minimum, and constraints are satisfied to within the default value of the constraint tolerance, the optimization is completed and $\{\lambda_1, \lambda_2, \dots, \lambda_M, \mu_1, \mu_2, \dots, \mu_P\}$ is obtained.

4. Experiment

4.1. Dataset

In this article, the experimental dataset selected is the OpenSARShip [35] dataset, which comes from the dual-polarized satellite-borne SAR data detected by the European Space Agency's Sentinel-1 satellite. It has 11,346 ships and their corresponding Automatic Identification System (AIS) information. The OpenSARShip provides two available products of the interferometric wide swath mode (IW) mode: the single look complex (SLC) and ground range detected (GRD) products, with VV-VH dual polarization. The image spatial resolution in GRD mode is $20 \text{ m} \times 22 \text{ m}$, which fits well with our research topic of moderate-resolution SAR image, so we choose images in GRD mode for the experiments. Figure 2 shows the type distribution of data sets in GRD mode. Figure 3 is an example of the data set. Since many ship chips in the dataset are extremely small in size (smaller than 30×30) and are not suitable for classification, we selected three types of ship chips with a suitable size (greater than 70×70) for experiment, including 250 Cargo ships, 240 Container ships and 134 Tankers.

Figure 2. The statistical results of the types of ships in the GRD mode of the dataset.

Figure 3. Example of the data set. **Top to bottom:** cargo ship; container ship; tanker.

4.2. Results and Discussion

After feature extraction and analysis, the separability, stability, and Pearson correlation coefficient analysis result are shown in Figure 4.

Figure 4. Feature analysis result. (a) separability analysis; (b) stability analysis; (c) Pearson correlation analysis.

From the result, it can be seen intuitively that different features have different properties and distinguishability. For example, the width x_5 has good separability (as shown in Figure 4a), but it performs poorly in stability on cargo ships (as shown in Figure 4b). In addition, there is a strong correlation between x_{20} , x_{21} , x_{22} and x_{23} in the Hu moment invariant features (as shown in Figure 4c although they are relatively stable (as shown in Figure 4b), this is the very reason that we need to perform feature selection and fusion. Firstly, $\{x_2, x_6, x_8, x_9, x_{18}, x_{20}, x_{21}, x_{22}, x_{23}\}$ are subtracted due to their low separability. Secondly, $\{x_3, x_{10}, x_{11}, x_{13}, x_{14}, x_{15}\}$ are subtracted due to their poor stability. Finally, we evaluate the correlation coefficient between the remaining features and the feature vector after screening is $F_{Filter} = \{x_1, x_4, x_5, x_7, x_{12}, x_{16}, x_{17}, x_{19}\}$. Then we combine the feature vector F_{KPCA} obtained by the KPCA method to perform MIMR-based feature fusion for experiments.

In each experiment, we randomly selected 200 container ships, 200 bulk carriers, and 100 tankers as the training set, and the rest as the test set. In order to eliminate the bias caused by random sampling to a single experimental data, the above experiments were carried out 50 times. The performance of the classifier is finally reported as the average classification accuracy. KNN classifier and SVM classifier are selected to perform

classification tasks to comprehensively evaluate the method. The recognition result is shown in Table 1.

Table 1. The PCC of different feature vectors in different classifiers.

PCC (%)	KNN Classifier	SVM Classifier
F_G	87.39	94.35
F_T	73.29	80.90
F_{Filter}	88.24	95.73
F_{KPCA}	86.04	90.97
F_{SAE}	95.19 (Softmax Classifier)	
F_{mRMR}	83.46	92.19
Feature-fusion (Proposed)	90.48	97.14

Where Feature-fusion is the feature vector $S = \{\lambda_1 x_1, \lambda_2 x_2, \dots, \lambda_M x_M, \mu_1 y_1, \mu_2 y_2, \dots, \mu_P y_P\}$ obtained by the MIMR method proposed in this article. $F_G = \{x_1, x_2, \dots, x_8\}$ and $F_T = \{x_9, x_{10}, \dots, x_{16}\}$ are the geometric structure features and texture features in the original feature set $F_0 = \{x_1, x_2, x_3, \dots, x_{23}\}$. F_{SAE} and F_{mRMR} are the feature vectors obtained by the SAE method and the mRMR method respectively, which are regarded as part of the comparative experiments as well. To understand the impact of features with different physical attributes on classification, we first select feature vectors F_G and F_T to verify the classification ability of the original features we extracted. We can see that F_G have a better classification performance than F_T , this is because texture features reflect the sharpness of the image and the thickness of the texture, which means that it is directly related to the resolution of the image. This also reflects that for the Sentinel-1 products with moderate resolution, the categories of geometric features show better classification performances, which is consistent with our analysis. The feature vector S always has the best performance among all the classifiers, which proves the robustness and stability of our method.

We also found that the same feature vector performs differently on different classifiers. The feature vector F_{SAE} has good performance in Softmax classifier, but it performs poorly in other classifiers. We can conclude that the optimal feature subset may change with the classifier, that is, there may be inherent relevance between the optimal feature subset and the classifier, which can be taken as a future work of our research. After further analyzing the classification results of three different types of ships, we found that the PCC of cargo ships and container ships are higher than that of tankers. This is because the size of tankers varies greatly, which means larger intra-class distance, and tankers account for fewer pixels in ship chips than the other two ship types, which limits the effectiveness of feature extraction and is worthy of our further study.

In addition, we compared the performance of the proposed method with other methods for ship classification in moderate-resolution SAR images, including naive geometric features (NGFs) [3], Geometric features/local radar cross section (LRCS) (three sections) features [34], CNN + Metric [19], Lightweight CNN [20] and Semisupervised learning [36]. The comparison results are shown in Table 2. Compared with the current popular CNN methods, our method does not need to train a complex network framework, nor does it require complex parameter settings. Considering that our experimental data are moderate-resolution SAR images, we pay more attention to geometric structure features and gray-scale statistical features, while discarding those features that require high image resolution. Compared with traditional methods, we choose the Filter method to evaluate features and select “good” features with better classification abilities, and apply the KPCA method to reduce the dimension while retaining the principal components of the original information. Then we propose an MIMR-based feature fusion strategy to minimize the redundancy between features while ensuring the information representation efficiency of the overall feature subset to the target. It can be clearly seen that whether it is the classical machine learning method or the emerging CNN method, the PCC of the proposed method is much higher than them, which proves that our method is simple, feasible, and comparatively

robust. Current satellites capable of SAR imaging mainly include TerraSAR-X [2], Sentinel-1 [35], Radarsat-2 [37], and Gaofen-3 [38], which are operating in different imaging modes with different resolution and swath. With comparable resolutions, we believe that the proposed method should be able to perform well in the SAR images obtained from the current satellite missions.

Table 2. The PCC of the proposed method and the method for comparison in moderate-resolution SAR images.

Methods	Spatial Resolution of Experimental Data	Ship Type of Classification Result	Classifier	PCC (%)
NGFs [3]	15 m × 15 m	Bulk, containers and tankers	KNN	64.6
			SVM	67.6
			MKL	71.0
Geometric/LRCS (3 sections) [34]	20 m × 22 m	Tanker, container ship, and bulk carrier	Geometric-KNN	76.49
			LRCS (3 sections)-KNN	75.69
CNN + Metric [19]	20 m × 22 m	Tanker, container, and bulk carrier	Softmax	83.67
Lightweight CNN [20]	20 m × 22 m	Cargo, tanker, tug and others	Softmax	84
Semisupervised learning [36]	20 m × 22 m	Bulk carrier, container ship and tanker	Softmax	74.96
Proposed method	20 m × 22 m	Cargo, container and tanker	KNN	90.48
			SVM	97.14

5. Conclusions

This article aims to provide a reliable and operable method for ship classification in moderate-resolution SAR images. We extract the motivation of the target classification problem from multi-dimensional feature selection and fusion and correct classification rate. A novel multi-feature fusion framework based on the MIMR method is proposed, which not only considers the classification ability of a single feature, but also considers the information representation efficiency of the overall feature subset to the target, while ensuring that the redundancy is minimized. The results of the comparative experiments show that the proposed method effectively reduces the redundancy between features, and has a fairly good performance for ship targets classification in moderate-resolution SAR images.

The preliminary analysis we presented here can be extended in many ways. We intend to conduct extensive experiments in combination with feature fusion under different polarization channels. In addition, we will continue to study other more advanced classifiers.

Author Contributions: Conceptualization, G.Z. (Gaoyu Zhou) and G.Z. (Gong Zhang); methodology, G.Z. (Gaoyu Zhou); supervision, G.Z. (Gong Zhang); validation, G.Z. (Gaoyu Zhou) and B.X.; investigation, G.Z. (Gaoyu Zhou); writing—original draft preparation, G.Z. (Gaoyu Zhou); funding acquisition, G.Z. (Gong Zhang). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China under grant number 61871218, 61701046 and 61801211, the Fundamental Research Funds for the Central University, China, under grant number 3082019NC2019002, the Open Research Fund of State Key Laboratory of Space-Ground Integrated Information Technology under grant number 2018_SGIIT_KFJJ_AI_03 and the Funding of Key Laboratory of Radar Imaging and Microwave Photonics (Nanjing University of Aeronautics and Astronautics), Ministry of Education.

Data Availability Statement: The data presented in this study are openly available in [35].

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

References

1. Moreira, A.; Prats-Iraola, P.; Younis, M.; Krieger, G.; Hajnsek, I.; Papathanassiou, K.P. A Tutorial on Synthetic Aperture Radar. *IEEE Geosci. Remote. Sens. Mag.* **2013**, *1*, 6–43. [\[CrossRef\]](#)
2. Brusch, S.; Lehner, S.; Fritz, T.; Soccorsi, M.; Soloviev, A.; van Schie, B. Ship Surveillance with TerraSAR-X. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 1092–1103. [\[CrossRef\]](#)
3. Lang, H.; Wu, S. Ship Classification in Moderate-Resolution SAR Image by Naive Geometric Features-Combined Multiple Kernel Learning. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1765–1769. [\[CrossRef\]](#)
4. Ding, H.; Dong, Y.; Liu, N.; Wang, G.; Guan, J. Overview and Prospects of Research on Sea Clutter Property Cognition. *J. Radars* **2016**, *5*, 499–516. [\[CrossRef\]](#)
5. Mou, X.; Chen, X.; Guan, J.; Chen, B.; Dong, Y. Marine Target Detection Based on Improved Faster R-CNN for Navigation Radar PPI Images. In Proceedings of the 2019 International Conference on Control, Automation and Information Sciences (ICCAIS), Chengdu, China, 24–27 October 2019; pp. 1–5. [\[CrossRef\]](#)
6. Chen, X.; Guan, J.; Huang, Y.; Xue, Y.; Liu, N. Radar Signal Processing for Low-Observable Marine Target-Challenges and Solutions. In Proceedings of the 2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP), Chongqing, China, 11–13 December 2019; pp. 1–6. [\[CrossRef\]](#)
7. Clemente, C.; Pallotta, L.; Gaglione, D.; De Maio, A.; Soraghan, J.J. Automatic Target Recognition of Military Vehicles with Krawtchouk Moments. *IEEE Trans. Aerosp. Electron. Syst.* **2017**, *53*, 493–500. [\[CrossRef\]](#)
8. Blum, A.L.; Langley, P. Selection of relevant features and examples in machine learning. *Artif. Intell.* **1997**, *97*, 245–271. [\[CrossRef\]](#)
9. Dash, M.; Liu, H. Feature Selection for Classification. *Intell. Data Anal.* **1997**, *1*, 131–156. [\[CrossRef\]](#)
10. Kira, K.; Rendell, L.A. The Feature Selection Problem: Traditional Methods and a New Algorithm. In *Proceedings of the Tenth National Conference on Artificial Intelligence, San Jose, CA, USA, 12–16 July 1992*; AAAI Press: Menlo Park, CA, USA, 1992.
11. Langley, P. Selection of Relevant Features in Machine Learning. In Proceedings of the AAAI Fall Symposium on Relevance, New Orleans, LA, USA, 4–6 November 1994; pp. 1–5.
12. Lang, H.; Zhang, J.; Zhang, X.; Meng, J. Ship Classification in SAR Image by Joint Feature and Classifier Selection. *IEEE Geosci. Remote Sens. Lett.* **2015**, *13*, 212–216. [\[CrossRef\]](#)
13. Wu, J.; Zhu, Y.; Wang, Z.; Song, Z.; Liu, X.; Wang, W.; Zhang, Z.; Yu, Y.; Xu, Z.; Zhang, T.; et al. A Novel Ship Classification Approach for High Resolution SAR Images Based on the BDA-KELM Classification Model. *Int. J. Remote Sens.* **2017**, *38*, 6457–6476. [\[CrossRef\]](#)
14. Qiu, H.; Wang, X.; Xu, Z.; Zhang, J.; Su, C. Summary of Research Progress on the Recognition of Ship Target from SAR Image. In Proceedings of the 3rd International Conference on Computer Engineering, Information Science & Application Technology (ICCIA 2019), Chongqing, China, 30–31 May 2019. [\[CrossRef\]](#)
15. Mafarja, M.; Heidari, A.A.; Faris, H.; Mirjalili, S.; Aljarah, I. Dragonfly Algorithm: Theory, Literature Review, and Application in Feature Selection. In *Nature-Inspired Optimizers*; Springer: Cham, Switzerland, 2020.
16. Xu, Y.; Zhang, D.; Song, F.; Yang, J.-Y.; Jing, Z.; Li, M. A Method for Speeding Up Feature Extraction Based on KPCA. *Neurocomputing* **2007**, *70*, 1056–1061. [\[CrossRef\]](#)
17. Luo, S.; Sarabandi, K.; Tong, L.; Pierce, L. A LS-SVM-Based Classifier with Fruit Fly Optimization Algorithm for Polarimetric SAR Images. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 1859–1862. [\[CrossRef\]](#)
18. Gokdag, U.; Ustuner, M.; Bilgin, G.; Sanli, F.B. Kernel Extreme Learning Machines for PolSAR Image Classification Using Spatial Features. In Proceedings of the 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, Turkey, 2–5 May 2018.
19. Li, Y.; Li, X.; Sun, Q.; Dong, Q. SAR Image Classification Using CNN Embeddings and Metric Learning. *IEEE Geosci. Remote Sens. Lett.* **2020**, *1*–5. [\[CrossRef\]](#)
20. Shao, J.; Qu, C.; Li, J.; Peng, S. A Lightweight Convolutional Neural Network Based on Visual Attention for SAR Image Target Classification. *Sensors* **2018**, *18*, 3039. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Chen, Y.; Lin, Z.; Zhao, X.; Wang, G.; Gu, Y. Deep Learning-Based Classification of Hyperspectral Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2094–2107. [\[CrossRef\]](#)
22. Gadhiya, T.; Tangirala, S.; Roy, A.K. Stacked Autoencoder Based Feature Extraction and Superpixel Generation for Multifrequency PolSAR Image Classification. In Proceedings of the PReMI 2019 8th International Conference on Pattern Recognition and Machine Intelligence, Tezpur, India, 17–20 December 2019.
23. Zhou, L. *Feature Extraction of Target Chips in SAR Images*; National University of Defense Technology: Changsha, China, 2007.
24. Reiss, T. The Revised Fundamental Theorem of Moment Invariants. *IEEE Trans. Pattern Anal. Mach. Intell.* **1991**, *13*, 830–834. [\[CrossRef\]](#)
25. Peng, H.; Long, F.; Ding, C. Feature Selection Based on Mutual Information Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1226–1238. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Mahmoudi, T.F.; Arabsaedi, A.; Alavipanah, S.K. Feature-Level Fusion of Landsat 8 Data and SAR Texture Images for Urban Land Cover Classification. *J. Indian Soc. Remote Sens.* **2018**, *47*, 479–485. [\[CrossRef\]](#)
27. Sungho, K.; Woo-Jin, S.; So-Hyun, K. Robust Ground Target Detection by SAR and IR Sensor Fusion Using Adaboost-Based Feature Selection. *Sensors* **2016**, *16*, 1117.

-
28. Zhang, M.; Chen, F.; Tian, B.; Liang, D. Multi-Temporal SAR Image Classification of Coastal Plain Wetlands Using a New Feature Selection Method and Random Forests. *Remote Sens. Lett.* **2019**, *10*, 312–321. [[CrossRef](#)]
 29. El-Farra, N.H.; Christofides, P.D. Bounded Robust Control of Constrained Multivariable Nonlinear Processes. *Chem. Eng. Sci.* **2003**, *58*, 3025–3047. [[CrossRef](#)]
 30. Bolourchi, P.; Moradi, M.; Demirel, H.; Uysal, S. Improved SAR Target Recognition by Selecting Moment Methods Based on Fisher Score. *Signal Image Video Process.* **2019**, *14*, 39–47. [[CrossRef](#)]
 31. Cui, Z.; Li, Q.; Cao, Z.; Liu, N. Dense Attention Pyramid Networks for Multi-Scale Ship Detection in SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 8983–8997. [[CrossRef](#)]
 32. Dong, Y.; Wang, C.; Zhang, H.; Wang, Y.; Zhang, B. Impact Analysis of Incident Angle Factor on High-Resolution Sar Image Ship Classification Based on Deep Learning. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019.
 33. Wu, Y.; Yuan, Y.; Guan, J.; Yin, L.; Chen, J.; Zhang, G.; Feng, P. Joint Convolutional Neural Network for Small-Scale Ship Classification in SAR Images. In Proceedings of the 2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019.
 34. Chang, C.-C.; Lin, C.-J. LIBSVM. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 1–27. [[CrossRef](#)]
 35. Huang, L.; Liu, B.; Li, B.; Guo, W.; Yu, W.; Zhang, Z.; Yu, W. OpenSARShip: A Dataset Dedicated to Sentinel-1 Ship Interpretation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 195–208. [[CrossRef](#)]
 36. Wang, C.; Shi, J.; Zhou, Y.; Yang, X.; Zhou, Z.; Wei, S.; Zhang, X. Semisupervised Learning-Based SAR ATR via Self-Consistent Augmentation. *IEEE Trans. Geosci. Remote Sens.* **2020**, 1–12. [[CrossRef](#)]
 37. Zhang, B.; Wang, C.; Zhang, H.; Wu, F.; Tang, Y.-X. Detectability Analysis of Road Vehicles in Radarsat-2 Fully Polarimetric SAR Images for Traffic Monitoring. *Sensors* **2017**, *17*, 298. [[CrossRef](#)]
 38. Sun, J.; Yu, W.; Deng, Y. The SAR Payload Design and Performance for the GF-3 Mission. *Sensors* **2017**, *17*, 2419. [[CrossRef](#)]