Article
Automatic Building Detection for Multi-Aspect SAR Images Based on the Variation Features
Qi Liu 1,2,*, Qiang Li 3, Weidong Yu 1,2 and Wen Hong 2,4

1 The Department of Space Microwave Remote Sensing System, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100194, China; ywd@aircas.ac.cn
2 The School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 101408, China; whong@mail.ie.ac.cn
3 Beijing Institute of Tracking and Telecommunications Technology, Beijing 100094, China; 15101099961@163.com
4 Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China
* Correspondence: liuqi181@mails.ucas.ac.cn

Abstract: Multi-aspect synthetic aperture radar (SAR) images contain more information available for automatic target recognition (ATR) than images from a single view. However, the sensitivity to aspect angles also makes it hard to extract and integrate information from multi-aspect images. In this paper, we propose a novel method based on the variations features to realize automatic building detection in the image level. First, to get a comprehensive description of target variation patterns, statistical characteristic variances are derived from three representative and complementary categories. Then, these obtained features are fused and put in the K-means classifier for prescreening, whose results are used as the training sets in supervised classification later to avoid manual labeling. Second, for more precise detection performance, finer features in vector forms are obtained by principal component analysis (PCA). The variation patterns of these feature vectors are explored in two different manners of correlation and fluctuation analyses and processed by separate support vector machines (SVMs) after fusion. Finally, the independent SVM detection results are fused according to a maximum probability rule. Experiments conducted on two different airborne data sets demonstrate the robustness and effectiveness of the proposed method, in spite of significant target signature variabilities and cluttered background.

Keywords: multi-aspect SAR; building detection; statistical characteristics; PCA; variation patterns; feature fusion; SVM; decision fusion

1. Introduction

SAR is a powerful and potential remote sensing system, which is capable of working in both weak natural light and adverse weather conditions [1]. With the purpose of detecting and classifying targets accurately and efficiently from SAR imagery, ATR is playing a more and more important part in both military and civilian field applications [2]. The designing of ATR includes three steps of detection, discrimination and classification [3]. In this paper, we mainly focus on the fundamental step of detection. Current ATR algorithms can be mainly divided into three types: algorithms based on template, model and deep learning [4]. The template-based ones can hardly meet the real-time requirements, and the emerging deep learning methods require quite abundant samples. Model based algorithms extract and screen image features, and then identify target types by specific classifiers. The more stable and recognizable the features are, the more reliable the recognition results can be [5].

However, because of the backscatter imaging mechanism, features extracted from SAR images are highly sensitive to the SAR acquisition geometry [6]. With a slight change in the target pose or position, the scattering intensity and other characteristics of the same variety of targets can vary quite abruptly. On the other hand, as the aspect angle is limited in the
radar observation at a time, some targets in the scene may be partly or even completely invisible, as the radar cross section (RCS) is partly determined by the corresponding aspect angle [7]. This kind of condition is especially true for man-made targets like the buildings, in which the scattering characteristics of dihedral angles are usually found [8]. In that case, it is commonly reckoned that multi-aspect images of the same scene always contain richer information and can provide better performance than any single of them in target detection tasks [9]. The data acquisition geometry in this paper is shown in Figure 1, where the images are captured by the same airborne sensor at consecutive aspects. Figure 2 takes a building area for example to show its different configurations at different aspects.

Figure 1. Image acquisition geometry. (a) Side view. (b) Front view.

Figure 2. Multi-aspect images of the same building area. The sub-images are taken at the corresponding aspect angles respectively. (a) 10.8°. (b) 12.8°. (c) 16.7°. (d) 20.5°. (e) 30.2°. (f) 39.3°.

Many existing studies have shown the enhancement effects of image combination in the field of target detection [10,11]. Ref. [11] proposes a novel object detection framework, integrating diverse features by jointly considering features in the frequency domain and the original spatial domain. In [12,13], multimodal images are combined via deep learning techniques to show the superiority of diverse data. Multi-aspect SAR images utilization methods can be divided into the following three categories. The first category works on finding the features remain unchanged as the aspect changes. For example, Bhanu et al. [14] compare the positions of strong scattering centers in different images, and select the scattered point pairs roughly stay still as features for model construction. Zhang et al. [15] believe that the intrinsic dimension of the target will always remain the same when the aspect changes within a wide range of degrees. Therefore, man-made targets can be identified by averaging the intrinsic dimensions in the region of interest (ROI) of different
images. The second category aggregates the different performance of the target at different aspects to enrich the referent sample base. Brendel et al. [10] compose one grand image with images in wide angle separations, which is later used as the reference image in a mean squared error (MSE)-template-based ATR system, so that the reference image contains more comprehensive information about the target. The third category pays attention to the inner connections of multi-aspect images. In this strategy, the images are fused through mutual influence, and the internal relevance between them is regarded as an effective criterion for recognition. As an example, Huan et al. [16] put vectors representing different images into the same matrix, which is then dealt with PCA and wavelet fusion methods. The resulting vectors separated from the processed matrix are used as features for classifiers. Zhang et al. [8] take advantage of sparse representation classification (SRC) among multiple aspects for a single joint recognition decision. Its ability to describe each sample precisely under the inner correlation constraints among samples brings it wide acceptance. The deep learning methods applied in multi-aspect SAR are usually based on the connection analyzing as well. Pei et al. [17] propose the multi-aspect deep convolutional neural network (MVDCNN), where they compare images from adjacent aspects step by step with a parallel network topology. Relationship exploration is completed progressively in different network layers. Zhang et al. [18] propose a deep neural network that containing Bi-LSTM model, so they can learn the connections of the training samples in both forward and backward directions independently. In the above literatures, the utilization of multi-aspect images have been demonstrated to be a remarkable improvement compared with single aspect methods. However, there are still some limitations in their practical applications. The first category has strict requirements on the interval and quantity of image samples in each class. The interval is usually recommended to be one degree, and no missing aspects in a wide range is recommended. In the second category, not many variations are allowed in both the target itself and the surrounding environment. When there are not enough training samples, targets in the interval aspect positions are still hard to be identified. The third category emphasizes the internal relationship between the images, but it may not work well when the relationship happened to be weak, especially when the aspects are quite separated.

In all the presented methods, it is always the major target signature variations among different aspects that cause trouble for detection. In this paper, we propose a new method for building detection with multi-aspect SAR images. With this method we process these variations into recognizable and essential features in the detection procedure, instead of avoiding them by requiring small aspect separation or stable environment conditions. We have noticed that as the aspect changes, some statistic characteristics of the background tend to stay relatively steady, while the same characteristics would vary sharply in building areas in the same scene. The different variation patterns between target and background can contribute to target discrimination in the complexity of disturbance in urban areas. In our method, the holistic scene to be detected is partitioned into a fixed number of grids, and their respective local variation patterns are taken for discrimination. As a single feature has only limited potentials, we adopt five indexes derived from three complementary characteristics to get a comprehensive description. By calculating and integrating variances from different indexes, we are able to put the grids into a K-means classifier for prescreening. After that, in order to reduce the information loss when the statistical histograms drop directly to one dimension of variance, we recalculate two variation patterns in vector forms based on PCA via correlation and fluctuation analyses. Separate SVM classifiers work independently under the resulting two variation patterns, whose training sets are provided by modified K-means clustering results instead of manual labeling. At last, the SVM detection results are fused according to a maximum probability rule. Experiments show that the method has good adaptability to significant target signature variabilities and has no strict requirements on the number and intervals of images.

The remaining part of the paper is structured as follows: we first introduce the common difficulties in multi-aspect target detection in Section 2. Then, in Section 3, the
proposed method for building detection is presented. Extensive experiments are conducted on airborne SAR images in Section 4. Finally, conclusions are drawn in Section 5.

2. Significant Target Signature Variabilities in Multi-Aspect Images

In addition to target deformations like affine transformation caused by radar perspective conversion naturally, there are also some significant target signature variabilities in the multi-aspect image sequence [19,20]. These variabilities include target scintillation, both intentional and unintentional target obscuration, changing background surfaces caused by inherent speckle noise and shadowing, etc. [21]. In the following part, we would illustrate these variabilities with specific examples.

The stated variabilities make the images in the time series carry discrepant information to some extent. As a result, the targets become harder to discriminate or fit into uniform descriptions. Because of the existence of the variabilities, we have decided not to search for stable features or fixed association relationships between all aspects, but simply focus on describing the variation patterns contained in the image sequence. By discriminating the targets with the difference of variation patterns, we can ensure the robustness of the algorithm in images cluttered or fuzzy.

2.1. Target Scintillation

In SAR images, flat surfaces such as building roofs in urban areas are often shown as dark areas in many aspects because of their surface scattering properties. They are only highlighted in some specific aspects, depending mainly on their incline angles to the ground and positions relative to the radar platform. As an example, Figure 3 shows three different scattering conditions of the same group of buildings in different aspects. In Figure 3a, a large part of the building group is highlighted, but the remaining parts are still more ambiguous and weaker than the surroundings. In Figure 3b, the buildings are partly shown, and these parts are almost complementary to Figure 3a. In Figure 3c, the buildings are almost invisible and hard to recognize.

![Figure 3. Target scintillation in different aspects. (a) Buildings mostly visible. (b) Buildings partially visible. (c) Buildings almost invisible.](image)

2.2. Target Obscuration

Radar detection has the ability of penetration. This ability is generally related to wavelength and polarization mode used by the detector, but also related to the aspect angle of the current image inevitably. In the image in Figure 4a, the buildings are obscured by the trees nearby, while in Figure 4b,c, parts of the buildings under the trees are visible. The appearance and disappearance of the obscurations are also responsible for the variabilities of the targets.
2.3. Background Changing: Speckle Noise and Shadowing

Speckle noise cannot be completely eliminated from SAR images and will always cause trouble in SAR target detection. However, when the variation features are taken as the detection criteria, the problem of speckle noise can be avoided to a large extent. Speckle noise usually has a relatively uniform distribution in the whole scene and hence little influence on regional statistical characteristics. It has even less influence on variation features as it just changes randomly with aspects, which is very different from the changing patterns of targets in the same scene.

In images change significantly with aspects, the reliability of some traditional methods tends to be greatly affected. Shadowing happens to be one of the main factors that cause this degree of change. The change in shadow with aspects is immediate and noticeable, and can bring unavoidable interference to the work of target detection. For instance, geometrical properties are commonly used in building detection methods [22]. However, when the targets are partly shadowed by urban greening vegetation or other buildings nearby, their areas, contours, shapes and connectivity can be affected a lot. The presence of complex objects in the background will present great challenges for the detection. Figure 5 shows the influence of shadows at different aspects on building forms in SAR images. Therefore, a more robust approach that is not sensible to these factors is needed.

Figure 4. Target variation caused by ambient obscurations. (a) Buildings completely obscured. (b) Buildings partially obscured. (c) Buildings partially obscured and complementary to (b).

Figure 5. Shadows change with aspects. The sub-images are taken at the corresponding aspect angles respectively. (a) 11.0°. (b) 22.6°. (c) 40.8°.
3. Proposed Method of Multi-Aspect Building Detection

3.1. Multi-Aspect Building Detection Framework

There are three steps contained in our method. First, we quantify the variations of 5 indexes from three different categories, analyze them to roughly define the areas where targets are likely to appear; then the features from these categories are refined in two different ways and put in the SVM classifier, respectively, to determine the exact building locations. At last, the results obtained from SVM classifiers are fused at the decision level to get our final detection results. The block diagram of the algorithm is shown in Figure 6.

3.2. Variances Derived from Statistic Characteristics as Prescreening Features

To achieve fully automatic target detection, we need to address the problem that unsupervised learning can fail to meet the accuracy requirements while supervised learning needs mass work in manual sample labeling. In that case, we have decided to take a step of prescreening with K-means to roughly define the target area locations, whose results are later taken as training sets in SVM classifiers with some proper modifications. In the process of prescreening, we tend to prioritize strict constraint conditions to ensure the correctness of the results. The utilization of one single feature has only limited constraint effect, for better performance we need to seek the fusion approaches for multiple features.

We consider the comparison between a group of multi-aspect sequential images a kind of time domain analysis for a fixed scene. In order to achieve a comprehensive description of the targets, it is essential to find more characteristics covering spatial and time-frequency domain analyses in the image level. For this purpose, we choose characteristics of three categories by experimental investigation, with the aim to ensure that they are aspect-sensitive, complementary to each other and easy to acquire and store. Five specific indexes are derived from three characteristics, that is, mean amplitudes and highlighted pixel proportions derived from intensity, regional homogeneity and dissimilarity from texture and $l_1, l_2$ norm of low frequency components in the wavelet decomposition. For a certain index, the variance among multi-aspect images is calculated as a feature value and different features are combined to form the criterion for prescreening. As we can see, in the target and non-target regions, there is not much difference in the average and range of the indexes, but unignorable difference in their variances.
3.2.1. Intensity Variance

The intensity of pixels is the most intuitionistic feature of SAR images. The signature variabilities in multi-aspect images have great influence on intensity of the targets. So, we have to examine the variances of indexes derived from intensity, and look for the difference of their representation forms between building areas and background. We first divide the holistic scene into $n \times n$ grids, for each grid the intensity histograms from different aspects are obtained. Then, the variances of mean values and bright pixel proportions are calculated, respectively, from different aspect histograms. By now, each grid has got two scalar feature values under the same category of intensity:

$$m_j = \frac{1}{N} \sum_{i=1}^{N} x_i^j$$  \hspace{1cm} (1)

$$m_{\text{ave}} = \frac{1}{P_n} \sum_{j=1}^{P_n} m_j$$  \hspace{1cm} (2)

$$V_m = \frac{1}{P_n} \sum_{j=1}^{P_n} (m_j - m_{\text{ave}})^2$$  \hspace{1cm} (3)

$$r_j = \frac{\sum_{i'} x_{i'}^j}{\sum_{i=1}^{N} x_i^j}$$  \hspace{1cm} (4)

$$r_{\text{ave}} = \frac{1}{P_n} \sum_{j=1}^{P_n} r_j$$  \hspace{1cm} (5)

$$V_r = \frac{1}{P_n} \sum_{j=1}^{P_n} (r_j - r_{\text{ave}})^2$$  \hspace{1cm} (6)

where $i$ is the sequence number of the bins in the histogram, $j$ is the sequence number of multi-aspect images. $N$ is the total number of bins in the histogram, $P_n$ is number of images involved. $x_i^j$ is the amplitude of the $ith$ bin in the $jth$ histogram. $T_r$ is the threshold set to distinguish bright pixels from others. $m_j$ is the mean value index of the $jth$ image, $a_j$ is the highlighted pixel proportion index of the $jth$ image. $V_m$ is the variance of the mean values, $V_r$ is the variance of bright pixel proportions. $V_m$ and $V_r$ are the two features both derived from the characteristic of intensity. In Figure 7a, which shows the mean intensity index in different aspects, the diagram on the left comes from a grid in the background area. We can see that it experiences a slow change as the aspect changes. The diagram on the right shows how the same index changes sharply in a grid of building area. In addition, we can see that the mean values of the two grids are quite close, indicating that there is no obvious difference based on the index amplitude alone. Figure 7b shows the highlighted proportions are of the same conditions.

3.2.2. Texture Variance

Texture reflects the different organization forms of the pixels within different parts of the images. The gray level co-occurrence matrix (GLCM) is generally used to describe image texture by studying the spatial correlation of the pixels [23]. To use the GLCM principle, we first convert the radar image to a gray level image by grading the pixel intensity into $L$ levels. Then, the occurrence frequency of pixel pairs at each grade level is counted according to specified direction and distance. At last, the co-occurrence matrixes
\( P_\theta \) obtained in different directions are averaged to serve the subsequent feature extraction steps. The final co-occurrence matrix \( P \) of pixel \((x, y)\) is shown as:

\[
P(\theta) = \frac{1}{N_{\theta}} \sum_\theta P_\theta \quad \text{s.t.} \quad p, q = 1, 2, \ldots, L; \quad \theta = 0, 45, 90, 135
\]

where \( d_r \) and \( d_c \) are the specified displacement of a pixel pair at the row and column directions, \( L \) is the general grades of gray levels, and \( \theta \) is the direction of counted pixels.

\[
P_\theta(p, q|d, \theta) = \{(x, y), (x + dr, y + dc)| I(x, y) = p; I(x + dr, y + dc) = q\}
\]

\[
P = \frac{1}{N_{\theta}} \sum_\theta P_\theta \quad \text{s.t.} \quad p, q = 1, 2, \ldots, L; \quad \theta = 0, 45, 90, 135
\]

\( \text{Figure 7. Normalized variations of intensity as aspect changes in background and target areas.} \)

(a) The intensity mean values vary with aspects. (b) The proportions of the highlighted pixels vary with aspects.

Our purpose is to obtain the texture of the grids in general for comparison between different images, instead of the elaborating characteristics of a certain image. In this condition, GLCM is only formed at the central pixel in each grid to represent the grid’s texture characteristic. Of all the texture values calculated from the co-occurrence matrix, we find that the indexes of homogeneity and dissimilarity can lead to the best distinction results via experiments. Figure 8 shows the normalized texture variations in multi-aspect images. Figure 8a compares the homogeneity variances in target and background grids, while Figure 8b compares the dissimilarity variances in the same conditions. The variances in different images are calculated as follows, where \( V_h \) and \( V_d \) are the indexes derived from the characteristic of texture.

\[
\text{hom} = \sum_{p=0}^{L-1} \sum_{q=0}^{L-1} \frac{P(p, q)}{1 + |p - q|}
\]

\[
\text{hom}_\text{ave} = \frac{1}{P_n} \sum_{j=1}^{P_n} \text{hom}_j
\]

\[
V_h = \frac{1}{P_n} \sum_{j=1}^{P_n} (\text{hom}_j - \text{hom}_\text{ave})^2
\]

\[
\text{dis} = \sum_{p=0}^{L-1} \sum_{q=0}^{L-1} (p - q)^2 P(p, q)
\]

\[
\text{dis}_\text{ave} = \frac{1}{P_n} \sum_{j=1}^{P_n} \text{dis}_j
\]

\[
V_d = \frac{1}{P_n} \sum_{j=1}^{P_n} (\text{dis}_j - \text{dis}_\text{ave})^2
\]
3.2.3. Variance of Wavelet Low Frequency Components

Wavelet decomposition extracts features in the image domain through time frequency analysis. In wavelet decomposition, the low frequency wavelet components are not sensitive to insignificant disturbance and can reflect the intrinsic signatures of an image [16]. In this paper, we perform 2D wavelet decomposition at 3 levels to each divided grid as shown in Figure 9. Figure 9d shows the decomposition results of Figure 9a in principle, where $LL_k$ denotes the low frequency component of the $k$th level decomposition while $LH_k$, $HL_k$ and $HH_k$ denote high frequency components.

![Figure 8](image-url)

**Figure 8.** Normalized variations of texture as aspect changes in background and target areas. (a) The regional homogeneity varies with aspects. (b) The regional dissimilarity varies with aspects.

![Figure 9](image-url)

**Figure 9.** Three-level wavelet decomposition. (a) Original image to show principle of wavelet decomposition. (b) Original image of target region. (c) Original image of background region. (d) Schematic diagram of decomposition result. (e) Decomposition result of target region. (f) Decomposition result of background region.

By column-stacking the $LL_3$ from different aspect images, we get a matrix $M$ represents the wavelet low frequency components. We calculate the $l_{1,2}$ mixed norm of $M$ by calculating $l_2$ norm of each row in $M$ and $l_1$ of the resulting vector afterwards. The value of $M_{1,2}$ is taken as the variance of wavelet components for each grid, in order to properly reflect the variation relationship among the components [22]. In the following formulas, $w_i^j$ is the amplitude of the $ith$ bin of the histogram from the $jth$ aspect. $V_{io}$ stands for the $M_{1,2}$...
we used. For intuitive observation, the mean value of each low frequency component in different images is shown in Figure 10.

\[
\begin{align*}
  w_{\text{ave}}^i &= \frac{1}{P_n} \sum_{j=1}^{P_n} w_i^j \\
  w_v^i &= \frac{1}{N} \sum_{i=1}^{N} (w_i - w_{\text{ave}}^i)^2 \\
  V_w &= \sum_{i=1}^{N} w_v^i
\end{align*}
\]

Figure 10. Normalized variations of low frequency component mean value in background and target areas.

3.3. Prescreening Based on Fused Features by K-Means

After we have got the variances of characteristics on image intensity, texture and wavelet, we integrate them into a vector for each grid and put it in the K-means classifier to determine preliminarily whether the grid belongs to target areas or not. K-means is one of the most widely used unsupervised classifiers who can make full use of existing features to give effective predictions. This procedure is regarded as prescreening in our works.

Because we have considered the image features from quite comprehensive perspectives, the results of the prescreening are also proved to be of low false alarm rate. Still, we cannot be entirely sure about the correctness of the results offered by K-means. In the procedure of variance calculation, the indexes transform from multidimensional vectors directly to scalars, and the information loss thus becomes unignorable. Therefore, in the following steps, the features will be refined and the detected results will be used as training sets for SVM classifiers for finer discrimination.

However, the mistakes in the training set are more likely to be enlarged in the SVM classification outcomes. To address this problem, we would modify the training samples based on the areas and aggregation conditions reflected in the relative positions of the
detected regions, as a supplement to further ensure the reliability of the samples. Taking into account the characters of buildings, we would delete fifteen percent of the isolated small areas in the detected regions, as buildings are more likely to appear in the form of large areas of connectivity in principle. The method of modification can be described as the following steps:

1. Binarization. The pixel judged as targets in prescreening are set to 1, while pixels judged as background set to 0;
2. Count the area of all the connected areas in the scene, and arrange them from smallest to largest;
3. Find the smallest 30% of the area and calculate the Sum of Euclidean distances from them to the center of the mass of the largest 30% of the area;
4. In the 30% of the smallest area, the half with the greater sum of distances is discarded to be background after modification.

This move may cause miscalculation, but it will do more good than harm in the long run.

3.4. Refining Features for Accuracy Improvement

In this part, to locate the targets more precisely based on the prescreening results, we would first provide finer features than scalar variables for each grid. Back to Section 3.2, when we first got the histograms of intensity, GLCM texture and wavelet low frequency components as statistical characteristics, instead of deriving scalar indexes from them, we would use them as vectors directly to explore the variation patterns among different aspects.

For each grid, if we join histograms of different characteristics from end to end, the new constructed feature vector will be detailed but easily redundant. With these vectors not further optimized, too much calculation will be required due to too high feature dimensions. Moreover, these histograms are naturally of different dimensions, which will eventually lead to unnecessary difference in their weights and influence when put together.

In this condition, we have decided to use the PCA method for feature selection. This method has the effects of unifying and reducing dimensions of these histograms and retaining decisive features with appropriate dimensions. It could concentrate feature energies and extract features through selecting appropriate basis function in low dimension space. Take the characteristics of intensity for an example, for a certain grid, we arrange the multi-aspect histograms as column vectors into a \( N \times P \) matrix \( H \). Then the correlation matrix \( C \) of \( H \) is calculated and its eigenvalue equation is solved as shown in (18–19).

The solved eigenvectors \( \xi \) corresponding to the maximum \( p \) eigenvalues \( \lambda \) are taken to form an orthogonal vector basis \( W \), which is used as the transformation matrix to perform dimension reduction. The vectors are thus reduced to \( p \) dimensions from \( N \) in the resulting matrix \( S \). Set \( p \) a constant quantity for all the features, we can assure them the same dimension and importance.

\[
\begin{align*}
C &= E \left[ H_{N \times P} H_{N \times P}^T \right] \\
\lambda \xi &= C \xi \\
W &= [\xi_1, \xi_2, \ldots, \xi_p]^T \\
S &= WH
\end{align*}
\]

After all the features of reduced dimensions are obtained by PCA for each grid, we can use them as materials to study the variation patterns along with aspects. A proper definition is needed here to explicitly describe the variation relationship among the feature vectors from different aspects. There are two options both proved to be effective and complementary to each other in this situation. One of them focuses on the correlation of
the vectors and the other analyzes the fluctuation between the vectors. For the first one, we calculate the covariance matrix of $S$:

$$S = (S_1, S_2, \ldots, S_P)$$  \hspace{1cm} (22)

$$\text{Cov}(S_i, S_j) = \frac{\sum_{k=1}^{P} (S_{ik} - \bar{S}_i)(S_{jk} - \bar{S}_j)}{P - 1}$$  \hspace{1cm} (23)

$$c_{ij} = \text{Cov}(S_i, S_j), \text{ s.t. } i, j = 1, 2, \ldots, P$$  \hspace{1cm} (24)

$$D(S) = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1P} \\ c_{21} & c_{22} & \cdots & c_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ c_{P1} & c_{P2} & \cdots & c_{PP} \end{bmatrix}$$  \hspace{1cm} (25)

As we can see in (23), the covariance matrix has a definition similar to the variance of scalars. It is an extension of variance in multi-dimensional cases. In fact, the diagonal elements in $D(S)$ represent the variances of the column vectors respectively in $S$, while the non-diagonal elements reflect the degree of correlations among the columns. Apparently, the latter are negatively correlated with the variation amplitudes of the columns, so we would make them into one of the criteria we are looking for. Because of the symmetry of the covariance matrix, in order to avoid repetition, we take the upper triangle elements of $D(S)$ to form a new vector representing the correlation variation pattern for a certain feature.

The principle of the second way is more straightforward in comparison. The variation of different columns is a direct combination of the variances from each row in matrix $S$:

$$S_o = \frac{1}{P_n-1} \sum_{i=1}^{P_n} (S_i - \bar{S}) \circ (S_i - \bar{S})$$  \hspace{1cm} (26)

After that, for the same characteristic, feature vectors from different aspects are summarized as an available variation vector by their variation relationship. The feature vectors from different characteristics are then connected end to end, forming the new criterion that will be adopted by SVM.

3.5. SVM as Classifier for Accuracy Improvement

Corresponding to the above two criteria from different variation pattern analytical perspectives, two separate classifiers are adopted to form independent classification results. These results will be then fused to make the final detection decisions. When it comes to the classifier types, it is commonly agreed that supervised classifiers can achieve better performance with proper samples. SVM is a binary classifier widely used in SAR classification due to its conspicuous performance in feature learning and class separating [24–26]. The basic principle of SVM can be stated as follows [27]: SVM first transforms its samples into a high-dimensional Euclidean space, and then separates them with a decision surface found in this new space with its kernel function.

$$f(x) = \text{sgn}\left\{ \sum_{i=1}^{I} a_i y_i K(x_i, x) + b \right\}$$  \hspace{1cm} (27)

$$K(x_i, x) = \exp\left\{ -\frac{(x - x_j)^2}{\sigma^2} \right\}$$  \hspace{1cm} (28)

where $x_i$ is support vector, $y_i$ is class label of $x_i$, $a_i$ is Lagrange multiplier of $x_i$, $b$ is the threshold used in this classification, $K$ is the Gaussian kernel function and $f(x)$ stands for the final classification results by SVM.

As mentioned in Section 3.3, SVM classifier use the detected regions as the samples of the training set to avoid manual labeling. Besides, we also delineate several random
size districts in the same scene with no presence of any targets as control terms. These districts have been dealt with the same dividing and feature extracting process as above, and the obtained grids are added into the training set as negative samples. The grids not considered as targets in the prescreening step are all put into the test set and reclassified. The detection results coming from different SVM classifiers are combined in the following part according to a maximum probability rule.

3.6. Fusion Strategy for SVM

By now, we have adopted two different methods to calculate the variations of the same set of feature vectors and thus got the detection results from two separate classifiers. In this part, we fuse these results at the decision level according to a maximum probability rule [4], in which the proposition with the highest probability is obtained. The probabilities are provided by SVM classifiers. The detection results obtained by different classifiers are formed into a set $T$:

$$ T = \{T_1, T_2, \ldots, T_H\} $$

$$ T_h(r,c) = \{T^1_h(r,c), T^2_h(r,c)\} $$

$$ t_1(r,c) = \max\{T^1_h(r,c)\} $$

$$ t(r,c) = \arg\max\{T^1_h(r,c)\} $$

$$ t_2(r,c) = \frac{1}{H-1} \sum_{h=1, h \neq t}^H T^2_h(r,c) $$

where $H$ is the number of used classifiers, $T^1_h(r,c)$ is the probability that the classifier $h$ regards the grid at the position of row $r$ and column $c$ as a target region. $T^2_h(r,c)$ is the probability that the same grid regarded as background by $h$. $t_1(r,c)$ is the maximum probability that the grid $(r,c)$ considered to be target by all the classifiers. $t_2(r,c)$ is the probability that the same grid considered to be background by the union of the rest of the classifiers except the one contributes to $t_1$. All the grids satisfying $t_1(r,c) > t_2(r,c)$ constitute the final decided target regions.

4. Experiments and Discussion

4.1. Dataset

We use images from two different experiments carried out in different time and places to verify the validity of the proposed method. The first experiment was performed in Yangjiang City, Guangdong Province, in which the airborne SAR system worked in X-band at an altitude of 3600 m. The images were obtained at a depression angle of 65.5 degree through spotlight imaging mode. The resolution of the images is 0.05 m. The second experiment was performed in Zhoushan City, Zhejiang Province. The working band of its radar platform was 9.6 GHz while the flight altitude was 7000 m. The depression angle was 55.0 degree. The resolution of the images is 0.3 m. In terms of time, the first experiment took place in the year of 2019 and the second in 2017. The $\omega K$ imaging algorithm and motion compensation measures for the images can be found in [28].

Detection algorithms based on feature extraction are inevitably sensitive to location shift, rotation, and non-uniform illumination in multi-aspect SAR images [29]. Measures would be taken to improve these phenomena in the preprocessing stage. However, as the contents contained in different images are indeed distinctive ones, and the proposed method advocates taking advantage of the variations, we find it unnecessary to pursue the strict per-pixel registration. Therefore, we chose to ensure that the key points of as many of building structures as possible stay basically the same locations among images. The preprocessing steps are as arranged follows: Before registration, all the images used have been fixed to the right size, adjusted to the same contrast using histogram equalization, normalized to avoid unnecessary differences. The registration is realized roughly by SAR-
SIFT algorithm. The key points found by SAR-SIFT are mostly likely to appear in the man-made target areas naturally [30]. We select key points that are centrally contained within the main high-lighted regions and calculate the summary of the distance of the selected matching point pairs. The high-lighted areas are determined by the OSTU method. Transformations including translation, rotation and scaling are then used to make the summary of matching point distance minimum.

In the following verifying experiments, we use at most 6 images obtained in different aspects for each scene. The details of aspect information in both of the experiments are described in Table 1. Take one of the scenes of Experiment1 for an example, the multi-aspect SAR images after preprocessing along with the corresponding optical image are shown in Figure 11. The SAR flight experiment was done in 2019.06 as mentioned, while the optical image was filmed in 2016.10, so there may be some difference in their surface objects. Figure 12 shows some of the background areas used as negative samples in SVM training set in one of the aspects.

Table 1. Aspect information of images available in the experiments.

| Aspect Sequence | Aspect in Experiment1 (°) | Aspect Interval in Experiment1 (°) | Aspect in Experiment2 (°) | Aspect Interval in Experiment2 (°) |
|-----------------|--------------------------|----------------------------------|--------------------------|----------------------------------|
| Aspect1         | −25.4                    | 0                                | 10.8                     | 0                                |
| Aspect2         | −14.7                    | 11.3                             | 20.5                     | 9.7                              |
| Aspect3         | −1.7                     | 12.4                             | 30.2                     | 9.7                              |
| Aspect4         | 11.0                     | 12.7                             | 31.2                     | 1.0                              |
| Aspect5         | 22.6                     | 11.6                             | 39.3                     | 8.1                              |
| Aspect6         | 40.8                     | 18.2                             | 41.5                     | 2.2                              |

Figure 11. Images at different aspects and the optical image of the same scene.
4.2. Performance of the Proposed Method

4.2.1. Basic Performance Verification

We analyze the effects of the proposed method according to four indicators: precision, accuracy, miss rate and false alarm rate. Specially, as the progress of building detection is based on the divided grids all the time in the proposed method, the calculation of indicators is also based on the grids. For example, when the number of grids correctly detected in the target regions is $N_a$, grids effectively judged to be non-target in the background regions is $N_b$, the accuracy of the detection results in the scene is calculated by:

$$\text{accuracy} = \frac{N_t + N_b}{n^2}$$

The same goes for other indicators.

To assure the results reliable, we conduct the experiments in different scenes and take an average of their indicators to show the overall performance of the proposed method. Among the scenes, Scene1–4 are from Experiment 1 and Scene5–6 are from Experiment 2. The performance of the method shows no significant variations in completely different scenes, proving our method steady and widely applicable. In the experiments, we set the dimension of PCA $p_{\text{can}} = 10$, the image segmentation parameter $n = 24$, and the number of bins in histograms of features $N = 36$.

Limited to the article length, Figure 13 shows the detection results in two different scenes and the corresponding optical images, as well as true values manually labeled. The two scenes are taken from the two experiments, respectively. Table 2 displays the indicators in all the scenes and their average.

Figure 14 compares the detection results provided by K-means with the final detection results and the ground truth manually labeled in Scene3. The indicators before and after SVM classification are listed in Table 3. As we can see, K-means has a quite low degree of false alarm rate, but a high degree of miss rate, while SVM can lead to significant improvements in accuracy and address the problem of high miss rate of K-means.

Figure 15 displays different instances detected effectively corresponding to different types of target signature variabilities shown in Section 2. In region 1, the grids experience severe target scintillation are plotted. In region 2, the targets are distracted by both factors of target scintillation and obscuration. In region 3, more and more parts of the building area gradually become shadowed by nearby trees as the aspect changes. In other scenes, interference such as trees, blocks of green space, roads, stacked building materials and water bodies are also common in urban context.
Table 2. Detection results in different scenes.

| Scene Sequence | Precision (%) | Accuracy (%) | Miss Rate (%) | False Alarm Rate (%) |
|----------------|---------------|--------------|---------------|----------------------|
| Scene1         | 87.4          | 92.2         | 18.2          | 12.6                 |
| Scene2         | 80.6          | 89.9         | 17.9          | 19.4                 |
| Scene3         | 87.3          | 90.1         | 21.1          | 12.7                 |
| Scene4         | 79.3          | 89.3         | 18.1          | 20.7                 |
| Scene5         | 70.7          | 84.9         | 22.6          | 29.3                 |
| Scene6         | 82.6          | 87.7         | 20.3          | 24.7                 |
| Average        | 81.4          | 89.0         | 19.7          | 19.9                 |

Table 3. Detection results before and after SVM.

| Step            | Precision (%) | Accuracy (%) | Miss Rate (%) | False Alarm Rate (%) |
|-----------------|---------------|--------------|---------------|----------------------|
| K-means         | 92.2          | 84.5         | 46.3          | 7.8                  |
| K-means+SVM     | 87.3          | 90.1         | 21.1          | 12.7                 |

Figure 13. Cont.
Figure 13. Detection results in different scenes and labeled true values. (a) Detection and labeled results in Scene1 in Experiment1. (b) Detection and labeled results in Scene5 in Experiment2.

Figure 14. Comparison of K-means and SVM detected results. (a) Detection results by K-means. (b) Detection results by SVM. (c) Manually labeled results. (d) Optical image of the same scene.
Figure 15. Detected regions with significant target signature variabilities.

4.2.2. Comparison with Single Aspect Detection

Table 4 takes Scene1 as an example to show its advantages over the detection results of any individual aspect. We can see from the table that some images are quite unrecognizable when standalone, but can work much better together with our proposed method. Traditionally, when it comes to the integration of different images, the most immediate thought comes to our mind should be a superposition of the grayscale in each image. But this can lead to a significant overlay of noise. A direct improvement should be picking up the maximum pixels in each image. However, the detection results of this maximum intensity image show that it is far inferior to our method. Its false alarm index can be much higher than normal as the dark areas have no significant superior over noise. Figure 16 presents the results in the maximum intensity image and the single aspect images with the precision and accuracy in the top two. The detection works done on the isolated images are also conducted by the same types of features and classifiers in our method, only that there is no variation accessible, so the features themselves are use in its place instead.

Table 4. The comparison between proposed method and isolated image.

| Detected Image   | Precision (%) | Accuracy (%) | Miss Rate (%) | False Alarm Rate (%) |
|------------------|---------------|--------------|---------------|----------------------|
| Proposed Method  | 87.3          | 90.1         | 21.1          | 12.7                 |
| Aspect1          | 29.1          | 61.8         | 50.0          | 70.9                 |
| Aspect2          | 32.6          | 63.2         | 38.3          | 67.4                 |
| Aspect3          | 66.9          | 86.1         | 25.8          | 33.1                 |
| Aspect4          | 56.8          | 80.7         | 44.5          | 43.2                 |
| Aspect5          | 30.8          | 61.3         | 40.6          | 69.2                 |
| Aspect6          | 28.2          | 58.7         | 44.5          | 71.6                 |
| Maximum Intensity| 60.8          | 84.2         | 18.8          | 39.2                 |
4.2.3. Comparison with Other Existing Methods

Table 5 compares the detection effects of some typical recognition methods for multi-aspect images with the proposed method in the same scene. These previously developed methods have covered a variety of features, including intensity features, time-frequency features and transformation features [22], but the variation patterns of the features have not been considered yet. Nilubol and Pham [31] perform Radon transform to multi-aspect images, and generate features from Fourier transformations of intensity statistical variables. Hidden Markov models are used for classification. Wang et al. [32] combine wavelet moment and entropy features in the step of feature extraction, and the resulting feature vectors are put into a SVM classifier. Huan et al. [33] combine PCA, ICA and Gabor wavelet features via decision fusion method, and achieve classification with SVM classifier. We compare the detection results of the proposed method with those of the above methods in the same scenes.

In the contrast experiments, the images from six aspects of Scene1 are manually labeled as the training set. Other scenes except Scene1 are used for detection in the test set. It can be seen that the Probability of Correct Classification (PCC) of some methods are relatively lower than their original proposed values in the references. That is mainly because there is more interference in the background of the data set used in this paper compared with the MSTAR database used in the original experiments of those methods. Another reason for this phenomenon is that the aspect intervals of the images applied are also wider than MSTAR.

From Table 5, we conclude the PCC obtained by the proposed method is significantly higher than that obtained by other methods listed. It shows that our method has good adaptability in complex scenes and limited available sample conditions.
Table 5. PCC of some detection methods.

| Detection Method                                                       | Average PCC (%) |
|-----------------------------------------------------------------------|-----------------|
| Intensity statistical features+HMM                                    | 72.9            |
| Grayscale wavelet moment and entropy+SVM                             | 71.5            |
| PAC, ICA and Gabor wavelet components+SVM                            | 79.0            |
| Proposed method in this paper                                        | 89.0            |

4.2.4. Detection under Different Division Parameters

In the proposed method, the mesh density in the step of image segmentation is one of the critical parameters we must choose carefully. As the basis of feature obtaining, the area of the grids will certainly affect the actual accuracy of the experimental results. In order to provide insights into the determination of the grid number $n \times n$, we carry out experiments with different values of $n$. The range of $n$ is set between 18 and 36, as we all know that when the value of $n$ is too small, the presented accuracy value might be high, but it will not make any sense. From Figure 17, we can see the accuracy peaks at $n = 24$, which we choose to use in the experiments. When the partition is further refined, the calculation amount of the algorithm increases quickly, but the precision and accuracy decrease slowly. Besides, in both cases of false alarm and miss rate, $n = 24$ is in a trough position. The PCC value in Figure 17 are taken from the average of different scenes.

4.2.5. Influence of Image Number and Robustness to Image Interval

During the procedure of the experiments, we reckon that the quality of the detection results is closely related to the number of images we have got. In this part we would analyze the suitable image number for detection. Additionally, our images are generally evenly separated (see Table 1), when we reduce the image number gradually, we can observe the different effects caused by the change of aspect intervals.
For the six images of Scene3, keeping all the other variables constant, we change the image number to observe the consequences. It is proved in Figure 18 that the proposed method is insensitive to the number and interval of images, but it is still beneficial to increase the number appropriately, even if no new images are actually added in this process. That is to say, the reuse of existing images can improve the detection performance effectively, especially when the image number is limited. Through multiple experiments we also find that the maximum intensity image can be also used as material of the repetition. However, excessive repetition on limited images does not lead to optimization of results.

In order to achieve the desired effect with the least amount of computation, the image selection cannot be just random. While there are no specific requirements for the aspects in which the images are generated, the quality of images should be checked in advance. For example, the mean value, variance and entropy of images from Aspect1 and Aspect6 are all relatively lower, so they should be eliminated whenever only part of the images are recommended to be used.

To further observe the influence of image interval on the detection results, we fix the number of images in Figure 19 and expand the range of aspects. We can see that the PCC has not changed much under the affection.

![Figure 18. Detected results change with image number.](image-url)
5. Conclusions

Most of the existing multi-aspect detection methods are designed for isolated targets with relatively simple background. The proposed method provides a new choice in the image level for complex application scenarios. Based on the variations between different images, it can work effectively in the presence of diverse information, and thus be applied in cluttered backgrounds like urban areas for their monitoring and planning.

Our method contains three steps: Firstly, we calculate the variances of different indexes educed from different characteristics, and integrate the variances as criteria for prescreening. Secondly, we remodel the variations of the same indexes into vectors for finer feature fusion. The vectors are then put into two SVM classifiers, respectively, according to two different variation pattern definitions. Thirdly, the independent results of the SVMs are fused at decision level for final judgment. It is not necessary to know the aspect of each image in advance in the proposed method. There are also no strict restrictions on the number of images and their aspect intervals. The method may be improved from several aspects in the future: new registration methods specifically developed for multi-aspect images may be beneficial for the subsequent detection steps. Different feature screening methods or attempts with other emerging classification algorithms could provide additional performance improvement. Further measures can be taken in the processing of target area boundaries. Finally, it is expected to combine multi-aspect SAR images and optical images for multi-modal applications.

Author Contributions: Conceptualization, W.Y. and W.H.; methodology, Q.L. (Qi Liu); software, Q.L. (Qi Liu); validation, Q.L. (Qi Liu); formal analysis, Q.L. (Qi Liu); investigation, Q.L. (Qi Liu); resources, Q.L. (Qi Liu) and W.Y.; data curation, W.Y.; writing—original draft preparation, Q.L. (Qi Liu); writing—review and editing, Q.L. (Qi Liu) and W.H.; visualization, Q.L. (Qi Liu); supervision, W.Y. and W.H.; project administration, Q.L. (Qiang Li) and W.Y.; funding acquisition, W.H. All authors have read and agreed to the published version of the manuscript.
Funding: This research was funded by the National Natural Science Foundation of China, grant number 61860206013.

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

References

1. Zhang, F.; Yao, X.; Tang, H.; Yin, Q.; Hu, Y.; Lei, B. Multiple Mode SAR Raw Data Simulation and Parallel Acceleration for Gaofen-3 Mission. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 2018, 11, 2115–2126. [CrossRef]

2. Pei, J.; Huang, Y.; Sun, Z.; Zhang, Y.; Yang, J.; Yeo, T.S. Multiview Synthetic Aperture Radar Automatic Target Recognition Optimization: Modeling and Implementation. IEEE Trans. Geo. Sci. Remote Sens. 2018, 56, 6425–6439. [CrossRef]

3. Song, S.; Xu, B.; Yang, J. SAR Target Recognition via Supervised Discriminative Dictionary Learning and Sparse Representation of the SAR-HOG Feature. Remote Sens. 2016, 8, 683. [CrossRef]

4. Liu, M.; Wu, Y.; Zhao, W.; Zhang, Q.; Li, M.; Liao, G. Dempster Shafer Fusion of Multiple Sparse Representation and Statistical Property for SAR Target Configuration Recognition. IEEE Geosci. Remote Sens. Lett. 2014, 11, 1066–1101. [CrossRef]

5. Novak, L.; Owirka, G.; Brower, W.; Weaver, A. The Automatic Target Recognition System in SAIP. Lins. Lab. J. 1997, 391, 187–202. [CrossRef]

6. Brown, M.Z. Analysis of multiple-view Bayesian classification for SAR ATR. Proc. SPIE—Int. Soc. Opt. Eng. 2003, 5095, 265–274. [CrossRef]

7. Tria, M.; Ovarlez, J.P.; Vignaud, L.; Castelli, J.C.; Benidir, M. Discriminating Real Objects in Radar Imaging by Exploiting The Squared Modulus of The Continuous Wavelet Transform. IET Radar Sonar Navig. 2007, 1, 27–37. [CrossRef]

8. Zhang, H.; Nasrabadi, N.M.; Zhang, Y.; Huang, T.S. Multi-View Automatic Target Recognition using Joint Sparse Representation. IEEE Trans. Aerosp. Electron. Syst. 2012, 48, 2481–2497. [CrossRef]

9. Ding, J.; Chen, B.; Liu, H.; Huang, M. Convolutional Neural Network With Data Augmentation for SAR Target Recognition. IEEE Geosci. Remote Sens. Lett. 2016, 13, 364–368. [CrossRef]

10. Brendel, G.F.; Horowitz, L.L. Benefits of aspect diversity for SAR ATR: Fundamental and experimental results. Proc. SPIE—Int. Soc. Opt. Eng. 2000, 4053, 567–578. [CrossRef]

11. Wu, X.; Hong, D.; Tian, J.; Chanussot, J.; Li, W.; Tao, R. ORSIm Detector A Novel Object Detection Framework in Optical Remote Sensing Imagery Using Spatial-Frequency Channel Features. IEEE Trans. Geo. Sci. Remote Sens. 2019, 57, 5146–5158. [CrossRef]

12. Hong, D.; Yao, J.; Meng, D.; Xu, Z.; Chanussot, J. Multimodal GANs Toward Crossmodal Hyperspectral-Multispectral Image Segmentation. IEEE Trans. Geo. Sci. Remote Sens. 2020, 59, 5103–5113. [CrossRef]

13. Hong, D.; Gao, L.; Yokoya, N.; Yao, J.; Chanussot, J.; Du, Q.; Zhang, B. More Diverse Means Better Multimodal Deep Learning Meets Remote-Sensing Imagery Classification. IEEE Trans. Geo. Sci. Remote Sens. 2021, 59, 4340–4354. [CrossRef]

14. Bhanu, B.; Jones, G. Exploiting azimuthal variance of scatterers for multiple-look SAR recognition. Proc. SPIE—Int. Soc. Opt. Eng. 2002, 4727, 290–298. [CrossRef]

15. Zhang, Z.; Hong, L. Man-made targets detection based on intrinsic dimension of SAR image samples. Electron. Meas. Technol. 2016, 39, 34–39. [CrossRef]

16. Huan, R.; Pan, Y. Target recognition for multi-aspect SAR images with fusion strategies. Prog. Electromagn. Res. 2013, 134, 267–288. [CrossRef]

17. Pei, J.; Huang, Y.; Huo, W.; Zhang, Y.; Yang, J.; Yeo, T.S. SAR Automatic Target Recognition Based on Multiview Deep Learning Framework. IEEE Trans. Geo. Sci. Remote Sens. 2017, 56, 2196–2210. [CrossRef]

18. Zhang, F.; Fu, Z.; Zhou, Y.; Hu, W.; Hong, W. Multi-aspect SAR target recognition based on space-fixed and space-varying scattering feature joint learning. Remote Sens. Lett. 2019, 10, 998–1007. [CrossRef]

19. Moshing, J.C.; Ross, T.D.; Bradley, J. An Evaluation of SAR ATR Algorithm Performance Sensitivity to MSTAR Extended Operating Conditions. Proc. SPIE—Int. Soc. Opt. Eng. 2019, 13370, 13. [CrossRef]

20. Ross, T.D.; Bradley, J.J.; O’Connor, M.P. SAR ATR: So what’s the problem? An MSTAR perspective. Proc. SPIE—Int. Soc. Opt. Eng. 1999, 3721, 606–610. [CrossRef]

21. Knee, P.; Thiagarajan, J.J.; Ramamurthy, K.N.; Spanias, A. SAR target classification using sparse representations and spatial pyramids. IEEE RadarCon 2018, 5, 294–298. [CrossRef]

22. Chen, L.; Zhan, P.; Cao, L.; Li, X. Discrimination and Correlation Analysis of Multiview SAR Images with Application to Target Recognition. Sci. Program. 2021, 2021, 1–9. [CrossRef]

23. Wei, L.; Wang, K.; Lu, Q.; Liang, Y.; Li, H.; Wang, Z.; Wang, R.; Cao, L. Crops Fine Classification in Airborne Hyperspectral Imagery Based on Multi-Feature Fusion and Deep Learning. Remote Sens. 2021, 13, 2917. [CrossRef]

24. Shan, C.; Huang, B.; Li, M. Binary Morphological Filtering of Dominant Scattering Area Residues for SAR Target Recognition. Comput. Intell. Neurosci. 2018, 2018, 1–15. [CrossRef] [PubMed]

25. Target recognition in SAR images using radial Chebyshev moments. Signal Image Video Process. 2017, 11, 1033–1040. [CrossRef]

26. Cui, Z.; Cao, Z.; Yang, J.; Feng, J.; Ren, H. Target recognition in synthetic aperture radar images via non-negative matrix factorisation. IET Radar Sonar Navig. 2015, 9, 1376–1385. [CrossRef]

27. Zhao, Q.; Princep, J.C. Support vector machines for SAR automatic target recognition. IEEE Trans. Aerosp. Electron. Syst. 2001, 37, 643–654. [CrossRef]

28. Ian, G.; Cumming, F.H.W. Digital Processing of Synthetic Aperture Radar Data: Algorithm and Implementation; Publishing House of Electronics Industry: Beijing, China, 2007.
29. Sandirasegaram, N.; English, R. Comparative Analysis of feature extraction (2D FFT and Wavelet) and classification (Lp metric distances, MLP NN, and HNeT) algorithms for SAR imagery. *Proc. SPIE—Int. Soc. Opt. Eng.* 2005, 5808, 314–325. [CrossRef]

30. Dellinger, F.; Delon, J.; Gousseau, Y.; Michel, J.; Tupin, F. Sar-sift: A sift-like algorithm for sar images. *IEEE Trans. Geosci. Remote Sens.* 2015, 53, 453–466. [CrossRef]

31. Nilubol, C.; Pham, Q.; Mersereau, R.; Smith, M.; Clements, M. Hidden Markov modelling for SAR automatic target recognition. In Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP’98 (Cat. No. 98CH36181), Seattle, WA, USA, 12–15 May 1998.

32. Wang, H.; Sun, F.; Zhao, Z.; Cai, Y. SAR image ATR using SVM with a low dimensional combined feature. *Autom. Target Recognit. Image Anal. Multispectral Image Acquis.* 2007, 6786, 67862J. [CrossRef]

33. Huan, R.; Pan, Y.; Zhang, P. SAR target recognition using PCA, ICA and Gabor wavelet decision fusion. *J. Remote Sens.* 2012, 16, 262.