# Ship Detection in Panchromatic Optical Remote Sensing Images Based on Visual Saliency and Multi-Dimensional Feature Description

**Ting Nie [1,*], Xiyu Han [1,2], Bin He [1], Xiansheng Li [1], Hongxing Liu [1] and Guoling Bi [1]**

[1]   Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China; hanxiyu15@mails.ucas.ac.cn (X.H.); Heb@ciomp.ac.cn (B.H.); lixiansheng@ciomp.ac.cn (X.L.); liuhongxing@ciomp.ac.cn (H.L.); biguoling@ciomp.ac.cn (G.B.)
[2]   Department of Mechatronic Engineering, University of Chinese Academy of Sciences, Beijing 100049, China
*   Correspondence: nieting@ciomp.ac.cn; Tel.: +86-0431-8617-6885

check for updates

**Abstract:** Ship detection in panchromatic optical remote sensing images is faced with two major challenges, locating candidate regions from complex backgrounds quickly and describing ships effectively to reduce false alarms. Here, a practical method was proposed to solve these issues. Firstly, we constructed a novel visual saliency detection method based on a hyper-complex Fourier transform of a quaternion to locate regions of interest (ROIs), which can improve the accuracy of the subsequent discrimination process for panchromatic images, compared with the phase spectrum quaternary Fourier transform (PQFT) method. In addition, the Gaussian filtering of different scales was performed on the transformed result to synthesize the best saliency map. An adaptive method based on GrabCut was then used for binary segmentation to extract candidate positions. With respect to the discrimination stage, a rotation-invariant modified local binary pattern (LBP) description was achieved by combining shape, texture, and moment invariant features to describe the ship targets more powerfully. Finally, the false alarms were eliminated through SVM training. The experimental results on panchromatic optical remote sensing images demonstrated that the presented saliency model under various indicators is superior, and the proposed ship detection method is accurate and fast with high robustness, based on detailed comparisons to existing efforts.

**Keywords:** ship detection; hyper-complex Fourier transform; visual saliency; panchromatic optical remote sensing images; rotation-invariant modified LBP

## 1. Introduction

Ships are important targets of real-time monitoring and wartime attacks at sea, and their accurate and fast detection can play a key role in the analysis of enemy situations, precision guidance, and military mapping. Ships also play an irreplaceable role in rescue, the safety management of fishing vessels, and so on. However, automatic detection is prone to false alarm and missed detection due to complex interference such as shooting weather, sea surface clutter, cloud, fog occlusion, and uneven illumination. How ships can be quickly and reliably detected in and extracted from panchromatic remote sensing images (with one channel) has become a critical issue [1–3].

At present, the ship detection method mainly includes three main stages [4–7]: the separation of sea and land, the region of interest (ROI) location, and target feature description and false alarm elimination, and the last two steps have received an increasing amount of attention. In the stage of locating ROIs (our interest areas of the suspicious ship in the image), gray statistical features, edge detection, and visual saliency are the most popular tools. The first two kinds of algorithm are simple to implement, but when sea conditions become complicated because the wind and waves are strong, the contrast of

targets and backgrounds is low, and the position effect is poor. Even a ship target could be submerged into the background under these circumstances. In recent years, methods based on visual saliency models have been successfully applied in target localization, and these methods simulate the human visual system and quickly focus on the ROIs in the image. Visual saliency models can be mainly divided into two types: top-down and bottom-up models. The top-down models, which use cognitive factors such as pre-knowledge, context information, expectations, and motivations to perform a visual search, are related to specific tasks and goals. This model belongs to an advanced cognitive process to consciously calculate the characteristics according to the task, and the existing top-down models usually bring in high computational costs and do not have a generic model. Therefore, the top-down models are more complicated and are rarely used in engineering projects. Most of the saliency detection models, including the spatial domain and transform domain, are bottom-up. In [8], ITTI constructed a saliency map based on intensity, color, and orientation features. This method has a good implementation; however, the model performed in the spatial domain is relatively complicated in structure and intensive in computation. Guo proposed a phase spectrum quaternary Fourier transform (PQFT) method for color images, and the edge detection performance was satisfactory, but the continuity of the segmentation region was poor with respect to panchromatic images [9]. Hou proposed the spectral residual method (SR), which is a significant detection model based on phase spectrum analysis in the frequency domain [10]. Li proposed the hyper-complex frequency domain transform model (HFT) with higher significant detection accuracy in simple scenes [11]. Dong proposed a novel visual saliency detection method based on differences in statistical characteristics to locate ROIs [12]. Xu presented a global saliency model based on high-frequency coefficients of multi-scale and multi-direction wavelet decomposition for ship detection [13]. Generally speaking, the processing results of the aforesaid algorithms were unsatisfactory with respect to panchromatic images. Specifically, these algorithms were not only sensitive to the boundaries of the target areas, but the detected target area was also incomplete in complex backgrounds.

With respect to the feature description and false alarm elimination phase [14–16], there are two representative categories. One is the method based on the statistical analysis of targets and false alarm characteristics. This kind of method requires the artificial design of multi-dimensional description features and then classification by machine learning. For instance, Yang combined the LBP descriptor of the image texture with the ship structure and used the adaptive boosting (AdaBoost) algorithm to generate hypotheses for ships. The ideal detection results were obtained for large-scale ships in high resolution images of calm sea surfaces; however, there was no difference in LBP distribution between each part of the ship structure at low resolution and in complex backgrounds, and the performance was degraded [17]. Qi designed a ship histogram of an oriented gradient (S-HOG) descriptor to obtain rotation invariance and applied principal component analysis (PCA) to compute the orientation. However, the complexity of the algorithm was high [18]. Shi used the learning algorithm based on AdaBoost classification to extract histograms of an oriented gradient (HOG) for distinguishing targets and false alarms [19]. Yang employed the compactness and length–width ratio to remove false alarms. The two kinds of features related to shape description, and the method lacked sufficient strong features [7]. It is important to design and select distinguished features. The other method can detect features automatically via deep learning models. Liu used a convolutional neural network (CNN) based on GPU acceleration to design feature layers of different depths, and completed real-time detection for $704 \times 704$ images [20]. Wang proposed a CNN-based renormalization method to improve the quality of object proposal with respect to very high resolution (VHR) remote sensing images [21]. Nie used a ship detection algorithm based on a transfer-learned single shot multi-box detector (SSD). The SSD fully utilizes the feature expression of each convolution layer [22]. Deep learning has been proven to fit big data and object detection extremely well in such databases as ImageNet (a large-scale visual database for visual object recognition). Moreover, large-scale weight parameters in the network can consume a great deal of computing and memory resources; for example, the AlexNet model (a deep convolution neural network proposed by Alex) exceeds 200 MB, and the VGG-16 (the visual geometry

group model) goes beyond 500 MB. The hardware migration of this kind of method is complicated and is not suitable for real-time processing onboard. Hence, it is crucial to conduct in-depth research on large-field and panchromatic remote sensing images in various complex conditions for fast and effective ship detection.

Aiming at these addressed problems, a novel approach based on visual saliency and multi-dimensional descriptor is proposed. The three stages in the framework are similar to those in [23]: sea-land segmentation, ROI extraction, and target discrimination, as shown in Figure 1. However, the last two steps have been improved. In the second stage, a visual saliency model based on an improved PQFT algorithm is constructed. Firstly, the image spatial correlation feature, the contrast feature, and the multi-direction feature are considered to establish the quaternion. Secondly, a hyper-complex Fourier transform is executed. Gaussian filtering of different scales is performed on the transformed result to synthesize the best saliency map. Finally, an adaptive binary segmentation of ROIs based on GrabCut (interactive foreground extraction using iterated graph cuts) is used to locate candidates. The presented saliency model has a high detection accuracy and obtains more complete target contours, regardless of the variety of scenes. In the last stage, the rotation-invariant modified LBP feature is constructed to make up for the poor discrimination of different regions caused by the lack of contrast description in the original LBP. At the same time, shape, texture, and moment-invariant features are extracted and concatenated as a feature description for SVM classification to eliminate false alarms and identify targets.
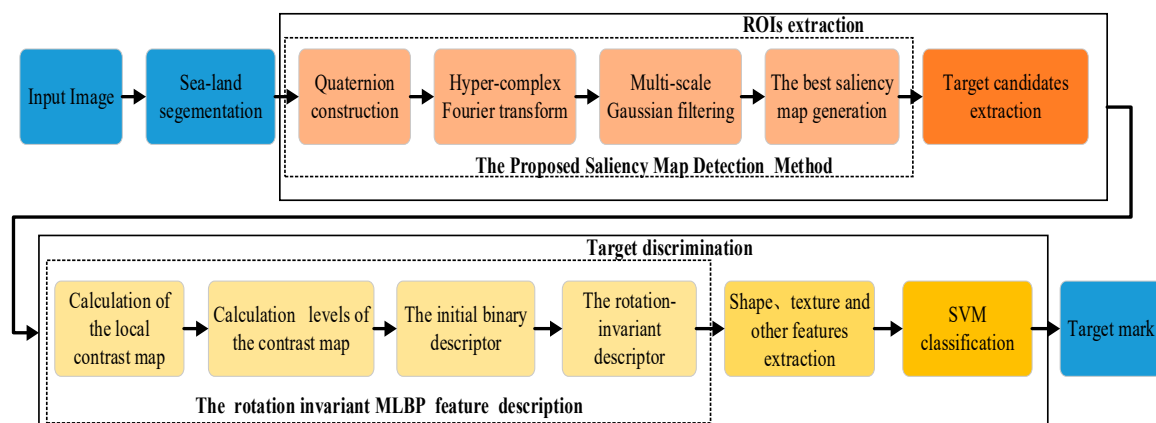


**Figure 1.** Diagram of the proposed hierarchical ship detection scheme.

The novel method makes the following contribution:

(1) A novel saliency map is proposed based on multiple features of quaternion transform, and it is shown to obtain more complete ROIs under complex sea backgrounds. The gray distribution of the saliency map is more uniform, which improves the accuracy of the detection algorithm for panchromatic optical remote sensing images.

(2) In the ROI extraction stage, we propose an adaptive segmentation algorithm based on GrabCut to obtain a more accurate binary region.

(3) The rotation-invariant modified LBP (MLBP) feature is employed to make up for the contrast description of different regions.

(4) Using shape, texture, and moment-invariant features to construct a strong target description operator achieves better performance in both detection accuracy and efficiency.

(5) It is an efficient and convenient model for hardware transplantation and engineering applications on the basis of ensuring detection accuracy.

The rest of this paper is structured as follows: Section 2 mainly introduces the method of extracting ROIs based on saliency map detection. Section 3 introduces the rotation-invariant MLBP features and

other combined features for false alarm elimination. Experimental results and analysis are provided in Section 4. Section 5 reports the conclusion and possible extensions.

## 2. ROI Extraction

It is well known that the size of one panchromatic optical remote sensing image is very large, and the number and area of ships in the image are lower, resulting in redundant background information. Moreover, ship targets can be described by high-level semantics such as texture, shape, direction, and other features. Therefore, ships can be considered as significant targets on the image and use saliency map extraction to locate them in a large image. In this section, we will introduce the two stages of ROI extraction: the detection of saliency maps and the method of obtaining target candidates by binary segmentation. In the first step, a multi-frequency domain significance map model based on the improved hyper-complex frequency domain and quaternion Fourier transform based on PQFT is designed. In the second step, an adaptive segmentation is proposed to obtain binary images of saliency maps to extract some basic parameters of the ROIs and determine potential targets.

### 2.1. The PQFT Model

Guo proposed the PQFT in the hyper-complex frequency domain [9]. The generalized coordinated color features are used to construct the quaternion. The hyper-complex quaternion matrix is represented as follows:

$$q(x, y) = f_0 + f_1\mu_1 + f_2\mu_2 + f_3\mu_3 \tag{1}$$

where $\mu_1$, $\mu_2$, and $\mu_3$ are the imaginary units, $\mu_1^2 = \mu_2^2 = \mu_3^2 = \mu_1\mu_2\mu_3 = -1$, and $f_0$ is the real part. If the value of the real part is 0, it is called a pure quaternion.

By combining the multi-feature components of the color image, the hyper-complex quaternion is formed as follows:

$$f_0 = I(x, y, t) - I(x, y, t - 1) \tag{2}$$

$$f_1 = (r + g + b)/3 \tag{3}$$

$$f_2 = RG = R - G \tag{4}$$

$$f_3 = BY = B - Y \tag{5}$$

where $R = r - (g + b)/2$, $G = g - (r + b)/2$, $B = b - (r + g)/2$, and $Y = (r + g)/2 - |r - g|/2 - b$. r, g, and b are the three channels of the color image. $f_0$ is the motion characteristic component and is used to express the brightness difference between two adjacent frames. $I(x, y, t)$ and $I(x, y, t - 1)$ represent the current frame and the previous frame image in the video. $f_0 = 0$ when it is used for a single-frame static image.

The hyper-complex Fourier transform $Q[u, v]$ of the quaternion $q(x, y)$ of the input image is computed, and the polar coordinate transformation form is described as follows:

$$Q[u, v] = \|Q[u, v]\|e\mu^{\Phi(u,v)} \tag{6}$$

where $\|.\|$ is the module of each element in the hyper- complex matrix, and $Q[u, v]$ is the frequency domain expression of $q(x, y)$.

The amplitude spectrum $A(u, v)$, phase spectrum $P(u, v)$, and Eigen spectrum $\chi(u, v)$ are calculated as follows:

$$A(u, v) = \|Q[u, v]\| \tag{7}$$

$$P(u, v) = \tan^{-1}\left(\frac{\|V(Q[u, v])\|}{S(Q[u, v])}\right) \tag{8}$$

$$\chi(u, v) = \frac{V(Q[u, v])}{\|V(Q[u, v])\|} \tag{9}$$

where $S$ and $V$ represent the real and imaginary part of the $Q[u,v]$, respectively.

The phase spectrum is preserved, the inverse quaternion Fourier transform is carried out, and the saliency map of PQFT is obtained:

$$S = g * \left\| Q^{-1}\left\{A(u,v)e^{XP(u,v)}\right\} \right\|^2 \tag{10}$$

where $g$ represents the Gaussian filtering function, and $Q^{-1}$ is the inverse quaternion Fourier transform. $S$ is the saliency map.

By the above calculation, the PQFT model can quickly achieve a saliency map of color video images and obtain good detection results. The PQFT is established by using two kinds of features (color and motion information) to construct the quaternion in the RGB or CIE Lab mode. However, the panchromatic remote sensing image has only one gray channel, and a single image has no motion information, which can result in a great reduction in the extraction accuracy for PQFT. The method adopts single-scale filtering and does not take into account the multi-scale phenomenon of different target sizes.

### 2.2. The Proposed Saliency Map Detection Method

Aiming at these addressed problems, a novel saliency map detection method is proposed based on the improved PQFT algorithm, called the modified PQFT (MPQFT). The method improves the PQFT in two aspects: one regards how the quaternion is constructed for the panchromatic remote sensing image to obtain uniform and complete target areas; the other regards how the saliency map is generated to ensure that differently sized targets can be detected at the same time. As shown in Figure 2, the proposed algorithm in the process of the MPQFT mainly consists of four steps: constructing the quaternion using different feature maps, carrying out hyper-complex Fourier transform to the quaternion, calculating multi-scale saliency maps, and generating the best saliency map. The process of the MPQFT is shown in Figure 2.
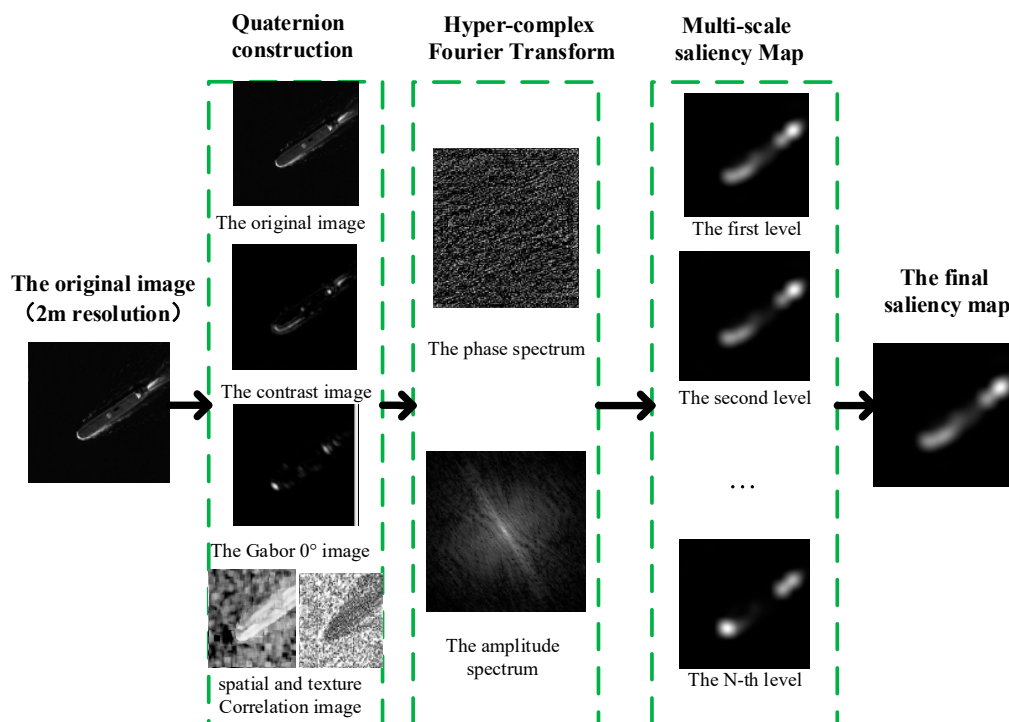


**Figure 2.** Procedure for the modified phase spectrum quaternary Fourier transform (MPQFT) algorithm.

For the ship target itself, it usually has a large change in intensity and obvious boundaries compared with clouds, fog, islands, and other scenes. In addition, texture and spatial similarity of ships are also conducive information in the complex background. Combining the two aspects would benefit the extraction process. Therefore, in the process of constructing the quaternion, we adopted the contrast map, the Gabor filtering map, and the fusion feature map that is established through combining spatial correlation with texture feature. The construction process of these feature maps is shown in detail below.

Firstly, the contrast value of one pixel is measured by the Euclidean distance of regions R1 and R2, and the calculation formula is as follows:

$$c(i,j) = D\left[\frac{1}{N1}\sum_{k1=1}^{N1} p_{k1}, \frac{1}{N2}\sum_{k2=1}^{N2} p_{k2}\right] \tag{11}$$

where $p_{k1}$ and $p_{k2}$ are gray values of pixels in R1 and R2, respectively. $D[.]$ indicates the Euclidean distance. R1 and R2 are the square regions centered on the same pixel $(i,j)$. i and j are the position of the row and column in the image. N1 and N2 are the number of pixels in R1 and R2, respectively. The size of regions R1 and R2 are $5 \times 5$ and $41 \times 41$ pixels, respectively. Both of them are empirical parameters. We c obtain the contrast map of the whole image by Equation (11).

Secondly, the Gabor filtering map that is sensitive to the scale and direction of the images is introduced [24,25]. The Gabor kernel function is multiplied by a Gaussian and sine function as follows:

$$g(x,y,\lambda,\theta,\sigma,\varphi,\gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \exp(i(2\pi x'/\lambda + \varphi)) \tag{12}$$

where $x' = x\cos\theta + y\sin\theta$, $y' = -x\sin\theta + y\cos\theta$. $\lambda$ is expressed in pixels when participating in the calculation. In general, it is less than one-fifth of the size of the input image and greater than or equal to 2. $\theta$ represents the direction of Gabor filter fringes, and ranges from $0°$ to $360°$. $\varphi$ is the phase offset that ranges from $-180°$ to $180°$. $\gamma$ is used to adjust the elliptic aspect ratio after the Gabor transform and when $\gamma$ equals 1, and the shape is approximately circular. $\sigma$ is the standard deviation of the Gaussian function in the Gabor function. We selected the following parameters: $\lambda = 2$, $\theta = 0°$, $\gamma = 0.5$, $\varphi = 0°$, and $\sigma = 0.5$. As shown in Figure 3, one example with different parameters is given.



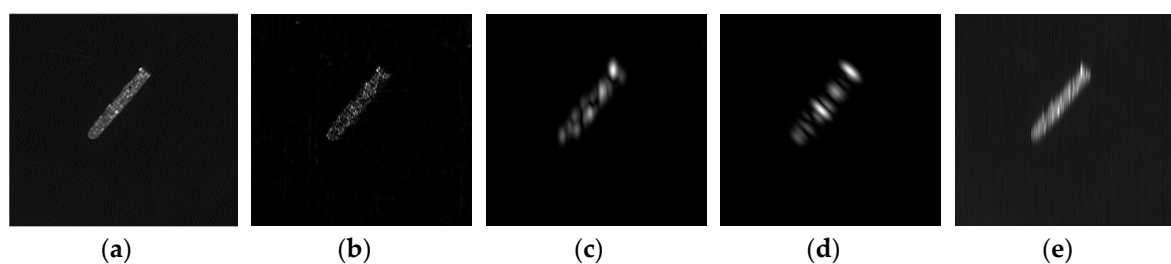| (a) | (b) | (c) | (d) | (e) |

**Figure 3.** One Gabor filter example with different parameters. (**a**) The original image. (**b**) The result of $\lambda = 2$, $\theta = 0°$, $\gamma = 0.5$, $\varphi = 0°$, and $\sigma = 0.5$. (**c**) The result of $\lambda = 2$, $\theta = 45°$, $\gamma = 0.5$, $\varphi = 0°$, and $\sigma = 0.5$. (**d**) The result of $\lambda = 10$, $\theta = 45°$, $\gamma = 0.5$, $\varphi = 0°$, and $\sigma = 0.5$. (**e**) The result of $\lambda = 2$, $\theta = 0°$, $\gamma = 0.1$, $\varphi = 0°$, and $\sigma = 0.5$.

Next, a fusion feature map is established through combining spatial correlation with texture to enhance the quaternion. The lower the spatial correlation, the more likely a candidate it is. The spatial correlation of the region x to the domain $X'$ for the pixel (i, j) is defined as follows:

$$Correlation(i,j) = \frac{cov(x,x')}{\sigma_x * \sigma_{x'}} = \frac{E(x \odot x') - E(x)E(x')}{\sqrt{E(x) - E^2(x)}\sqrt{E(x') - E^2(x')}} \tag{13}$$

where x and X′ are the square regions centered on the pixel p(i, j) and p(i, j + N), respectively. I and j are the row and column coordinates of the image. N is set to 5 in this paper. x and X′ are areas represented by the black and blue borders in Figure 4. The sizes of them are both k × k pixels.$\sigma_x$ and $\sigma_{x'}$ are standard deviations of the corresponding areas. E(x) and E(x′) are the mean value of the region x and X′, respectively. x ⊙ x′ represents the dot product of the gray value in the corresponding position of the two regions, and the size of x ⊙ x′ is k × k pixels. E(x ⊙ x′) is the mean value of the x ⊙ x′.
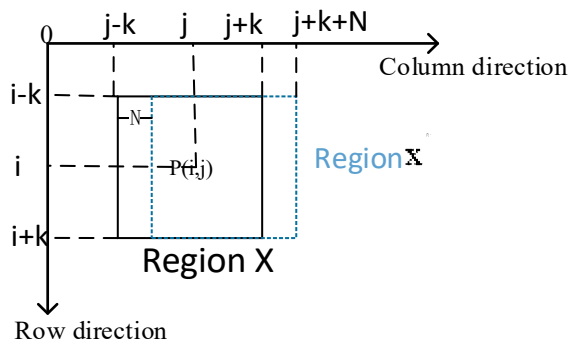


**Figure 4.** Schematic diagram of the spatial correlation regions.

We then calculate the unified LBP [17] to reflect texture properties of points, line corners, and flat areas in the image.

$$\text{LBP}_{N,R}^{riu2} = \begin{cases} \sum_{n=0}^{N-1} s(g_n - g_c) & if\,(U(\text{LBP}_{N,R}) <= 2) \\ N+1 & otherwise \end{cases} \tag{14}$$

where $s(.)$ is a binary symbol, $s(x) = 1, x > 0$, and $s(x) = 0, x \leq 0$. $U(LBP_{N,R})$ represents the number of exchanges between 0 and 1 on the N-bit binary number [17], $g_c$ and $g_n$ are the gray values of the central and the domain pixel, respectively. It was found that the LBP value of the background was higher than the target. The LBP map is processed as follows.

$$\text{LBP}(i, j)_{coeff} = \left| \frac{\overline{LBP} - \text{LBP}(i,j)}{\overline{LBP}} \right| \tag{15}$$

where $\overline{LBP}$ is the mean value of the LBP map, and |.| represents the absolute value. The final fusion feature is obtained by multiplying the correlation value Corre*lation*(i, j) and the coefficient in the LBP$_{coeff}$(i, j) in the corresponding position:

$$\text{Fusion}_{map} = \text{Corre}lation(i, j) \odot \text{LBP}_{coeff}(i, j) \tag{16}$$

Through the above steps, we completed the construction of the feature maps. $f_0 \sim f_3$ in Equation (1) can be assigned: $f_0$ is the original image, $f_1$ is a contrast map by Equation (11), $f_2$ is a Gabor transform map based on Equation (12), and $f_3$ is a blend feature map combining spatial correlation with the texture feature based on Equation (16). Therefore, the quaternion is q(x, y) = $f_0 + f_1\mu_1 + f_2\mu_2 + f_3\mu_3$.

$Q[u, v]$ is obtained by performing the hyper-complex Fourier transform in Equation (6). The amplitude spectrum A( u, v ) and phase spectrum P( u, v ) of Fourier transform is calculated by Equations (7) and (8). At the same time, in order to enhance the significant part, the Gaussian kernel function with different scales is used to obtain the scale space of the amplitude spectrum.

$$\overline{A_i(u, v)} = g_i(u, v) * A(u, v) \tag{17}$$

$$g_i(u, v) = \frac{1}{\sqrt{2\pi}2^{k-1}t_0} e^{-\frac{u^2+v^2}{2^{2k-1}t_0^2}} \tag{18}$$

where k is a spatial scale parameter, k = 1, ... , K, $K = \lceil \log_2 \min\{H, W\} \rceil + 1$, and H and W are the height and width of the image, respectively, $t_0 = 0.5$. We built a series of saliency maps at different scales as follows:

$$S_k = g * \|Q^{-1}\{\overline{A}_i(u,v)e^{XP(u,v)}\}\|^2 \tag{19}$$

The best saliency map is selected by methods according to the theory of minimum information entropy, and other saliency maps are discarded [26]. However, the discarded saliency maps also contained significant information at different scales for differently sized ships. Therefore, we linearly synthesized the saliency maps according to the entropy:

$$S = S_{\min} + \frac{1}{H_k(x)} S_k \tag{20}$$

$$H_k(x) = -\sum_{i=1}^{n} p_i \log p_i \tag{21}$$

where $H_k(x)$ represents the information entropy of one saliency map. $S_{\min}$ is the saliency map that the entropy is the minimum of $S_k$. $S$ is the saliency map we obtained.

We provide one example to show the process of the quaternion construction and the result of saliency map. The image comes from the satellite GF-2 and has a 2 m resolution. We can see from Figure 5 that the proposed method achieves a better locating result when it is subjected to complex interference and that each feature map suppresses the background interference effectively. Moreover, the gray value distribution of the saliency map is relatively uniform, so the binary value leads to a more complete target contour in the process of binary segmentation. The color images should be converted into grayscale images, and the following processing is similar to gray images.
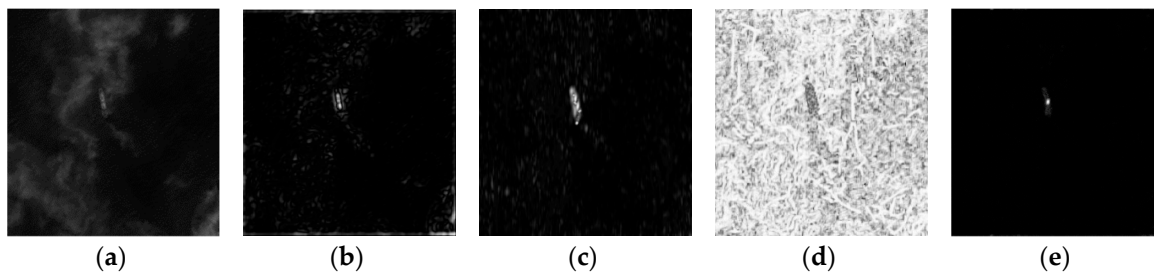


|     (a)      |     (b)      |     (c)      |     (d)      |     (e)      |

**Figure 5.** The process of the quaternion construction, and the result of the proposed saliency model. (**a**) The original image covered by thick cloud. (**b**) The contrast map generated by Equation (11). (**c**) The Gabor transform map generated by Equation (12). (**d**) The blend feature map combining spatial correlation with texture feature generated by Equations (13)–(16). (**e**) The final saliency map based on Equation (20).

### 2.3. Target Candidates Extraction

After the detection of saliency maps, saliency regions are enhanced, and the background is suppressed. In order to locate positions of candidate targets in the image, binary segmentation is needed. Furthermore, these areas include ships and false alarms. We can eliminate some simple false alarms by characteristic parameters and obtain ROI candidates. In order to extract the characteristic parameters of saliency regions, such as length and width, it is necessary to obtain binary images of saliency maps. Therefore, an adaptive segmentation based on the Otsu method and GrabCut [27,28] was executed to extract candidate target regions.

The original GrabCut algorithm is an interactive segmentation model with high precision. In this algorithm, the Gaussian mixture model (GMM) is used to model the foreground and background region of the image through annotating artificially. The foreground and background areas of the model need to be defined in advance. The foreground should include the complete target as much as possible.

The rest of the image will be the background. Any point in the image corresponds to a Gaussian component of the target or background. This obviously requires too many manual inputs in advance. Otsu is a classical threshold segmentation algorithm that is simple to calculate, but its accuracy is low in complex scenes. Therefore, we propose an adaptive segmentation method combining the two algorithms to obtain more accurate binary regions.

(1) Initial binary segmentation: Slices of ROIs are segmented to obtain binary images by the Otsu algorithm. One example (2 m resolution) is displayed in Figure 6 to show how the algorithm is executed. Figure 6b is the binary result of the Otsu method. We can see that the target region is not complete.

(2) External moment calculation: The external moment of the binary image is computed. The upper left point is $(x_{upper}, y_{upper})$, and the lower right point is $(x_{lower}, y_{lower})$. The region of the red border is the external moment. The upper left point (pink point) and lower right point (blue point) can be obtained, as shown in Figure 6c.

(3) As regards the definition of foreground and background regions, the external moment is extended 10 pixels to establish the foreground and background regions. The inside of the white solid line is the foreground, and the other part is the background, as shown in Figure 6d. In this way, the foreground and the background can be obtained by expanding the external moment without manual input.

(4) As regards GrabCut calculations, the GrabCut model iterates the energy model according to the input foreground and background region until the energy tends to be stable [27].

(5) As regards the final binary segmentation, the final binary segmentation image is obtained by a GarbCut model. In the saliency map, the region above the threshold is defined as the ROIs, while the other is considered to be the background. The external moment is recalculated to facilitate the subsequent application.
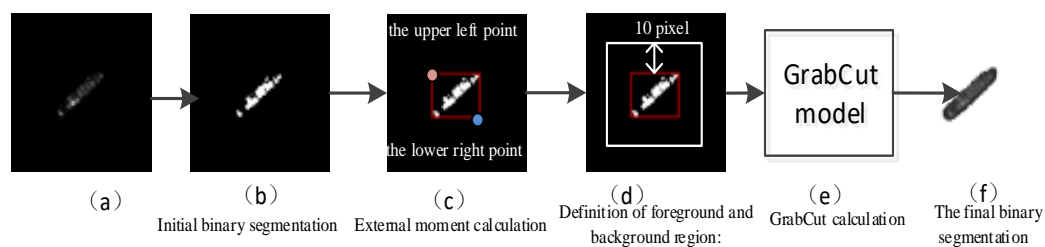


**Figure 6.** The adaptive segmentation proposed by our method.

Figure 7 shows the ROI extraction of images in complex backgrounds with a 2 m resolution. We can obtain many ROIs with different sizes, including ship targets, broken cloud blocks, sea waves, and other false alarms, as shown on the second column in Figure 7. The third column is the segmentation results obtained by the adaptive segmentation. Some false alarms can be eliminated by relevant simple shape features including length, width, area, aspect ratio, and tightness. By applying simple shape analysis, large clouds, islands, and small waves can be ruled out.
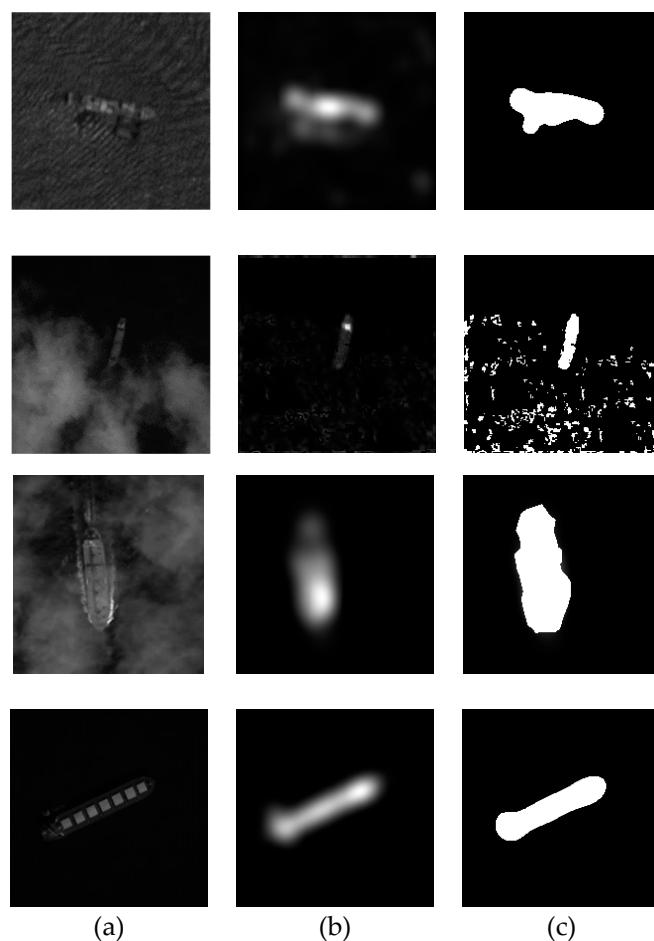
**Figure 7.** The regions of interest (ROIs) are extracted by the proposed method. (**a**) Test images. (**b**) Saliency maps generated by our MPQFT in Section 2.2. (**c**) Image segmentation by the method in Section 2.3.

## 3. Target Discrimination

Although some sea surface background interference can be suppressed after the ROI extraction stage, there are still some false alarms, such as coastal buildings and thick clouds, due to the complexity of the sea surface. Therefore, it is necessary for us to use the feature extraction and machine learning technique to confirm real ships. In this section, we introduce the description operator MLBP combined with SVM training to eliminate false alarms and confirm real targets.

### 3.1. The MLBP Feature Description

The LBP operator is gray rotation-invariant [29]. However, the LBP method only considers the size relationship between the central and neighboring pixel, with no contrast relationship. Therefore, different contrast distributions have the same LBP value as shown in Figure 8.

The modified LBP feature (MLBP) is proposed in this paper to solve this problem. The following process is for any pixel (i, j) in the image. Firstly, the local contrast between the central pixel and neighbor pixels is calculated:

$$g_p = I_p - I_{(i,j)} \tag{22}$$

where $I_{(i,j)}$ denotes the gray value of the i-th row and the j-th column of the image. $I_p$ is the gray value of the neighbor pixel centered in the pixel (i, j), and p = 1, 2, . . . N. N represents the number of pixels in the neighbor field, and it is set to 8 in this paper. $g_p$ is a contrast value between the central pixel and one neighbor pixel.
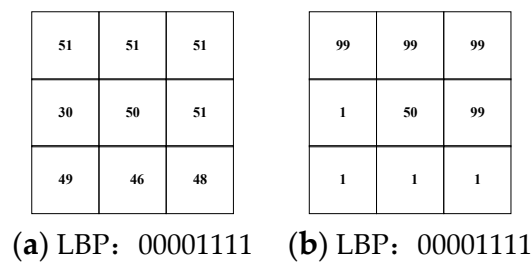
| 51 | 51 | 51 |
|---|---|---|
| 30 | 50 | 51 |
| 49 | 46 | 48 |

| 99 | 99 | 99 |
|---|---|---|
| 1 | 50 | 99 |
| 1 | 1 | 1 |

(**a**) LBP：00001111    (**b**) LBP：00001111

**Figure 8.** Example of local binary pattern (LBP). (**a**) The contrast between the central pixel and neighboring pixels is very small. (**b**) The contrast between the central pixel and neighboring pixels is very high.

Secondly, the maximum (maxC) and minimum (minC) of $g_p$(p = 1, 2, ... N) can be found. The range of values between maxC and minC is then divided into L levels. Therefore, each contrast value corresponds to a certain level. The level of $g_p$ is calculated as follows:

$$l_p = \left\lceil \frac{g_p - \min C}{(\max C - \min C)/L} \right\rceil \tag{23}$$

where p = 1, 2, ... N. $\lceil . \rceil$ is a symbol of rounding in the small direction, maxC and minC represent the maximum and minimum contrast values, respectively. L represents the number of levels and is set to 4 based on a large number of experiments. If $l_p > L$, then $l_p = L$.

The contrast level $l_p$ is transformed in to a binary description of 0 or 1:

$$S_p = \begin{cases} 1, l_p = L \\ 0, l_p \neq L \end{cases} \tag{24}$$

For the pixel (i, j), the binary descriptor MLBP is a series of numbers of 0 and 1 constructed with $S_p$, $MLBP = \{S_p\}$, and $p = 1, 2, \ldots, N$. In order to obtain rotation invariance, the number of exchanges between 0 and 1 is used to describe it again, and the rotation-invariant MLBP proposed in this paper is finally obtained as follows:

$$MLBP^{riu} = \begin{cases} \sum_{p=1}^{N} S_p, if(U(MLBP) <= 2) \\ N+1 \end{cases} \tag{25}$$

$$U(MLBP) = \sum_{p=2}^{N} \left| f(s_p - s_{p-1}) \right| + \left| f(s_N) - f(s_1) \right| \tag{26}$$

$$f(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases} \tag{27}$$

where $U(MLBP)$ represents the number of exchanges between 0 and 1 for the series. $MLBP^{riu}$ is a rotation invariance binary descriptor.

One example of the MLBP description process is given in Figure 9. The value of the central pixel is 120, N = 8, L= 4, and the size of the neighbor is 8 pixels, displayed in blue in Figure 9. The surrounding pixel values are compared with the central value to obtain the contrast map as shown in Figure 9b. The level of each neighbor pixel is calculated according to Equation (22) as shown in Figure 9c. The initial binary description of this area is 00000101, according to Equation (23), and the number of exchanges between 0 and 1 is 4. The rotation invariance binary descriptor of this area is 9.
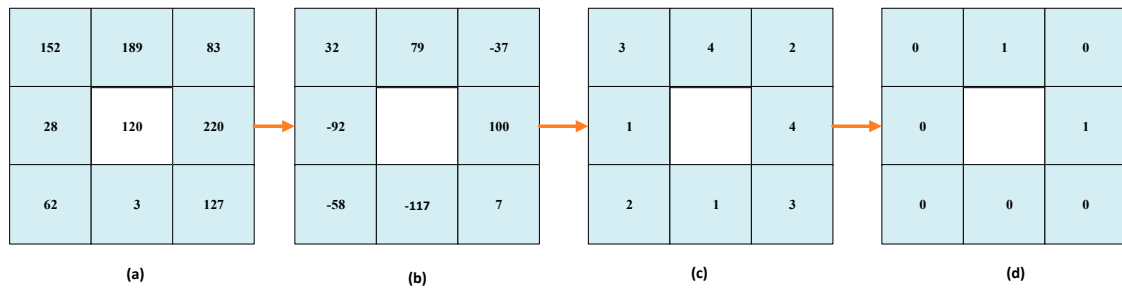
**Figure 9.** One example of the MLBP description process. (**a**) The original image. (**b**) A contrast map calculated using Equation (22) (**c**) The contrast level calculated using Equation (23). (**d**) Initial binary description 00000101.

The square neighbor area is used as an example above. In the actual application, a circular area can be adopted, and the bilinear interpolation method is applied for the estimation of the value that is not at the center of the pixel. According to the method proposed in this paper, the binary description of area (a) and (b) in Figure 8 are 11101111 and 00001111, respectively. The rotation-invariant MLBP is 1 and 2. Obviously, the MLBP of different contrast distributions is different.

### 3.2. SVM Training

The main aim of the SVM classification is to discriminate the real ship targets from candidates based on different features. We have already obtained the texture description in Section 3.1. In addition, we use the other 10 feature descriptions, including shape description, contrast and texture description, and invariant moment features. The shape feature includes length, width, area, and the perimeter of the minimum external moment. Furthermore, we selected three morphological features to eliminate false alarms, including length–width ratio, compactness, and rectangularity, which are calculated as follows:

$$RH = H/W \tag{28}$$

$$RT = P^2/S \tag{29}$$

$$RR = S/(H \times W) \tag{30}$$

where H and W are the length and width of the minimum external moment in pixels; P is the perimeter of the contour in pixels; S is the number of the connected area points in pixels.

At the same time, we analyzed the texture characteristics of a large number of ships and non-ships. The texture of waves and clouds changed slowly, and that of ships varied greatly. Therefore, we introduced the histogram variance of MLBP (proposed by Section 3.1) and the correlation and contrast of the grey-level co-occurrence matrix (GLCM) [23] to strengthen the ability to distinguish ships and non-ships. The contrast and correlation of the GLCM are calculated as follows.

$$M\_contrast = \sum_i \sum_j (i-j)^2 P_{ij} \tag{31}$$

$$M\_correlation = \sum_i \sum_j P_{ij} \frac{(i-\mu)(j-\mu)}{\sigma^2} \tag{32}$$

where $P_{ij}$ is an element of GLCM; $\mu$ is the average of GLCM, $\mu = \sum_i \sum_j i \cdot P_{ij}$; $\sigma^2$ is the standard deviation, $\sigma^2 = \sum_i \sum_j (i-\mu)^2 P_{ij}$. The contrast reflected the degree of the depth of the texture. The lighter the texture groove, the lower the contrast, and the more blurred the visual effect is. The correlation measured the degree of similarity between the spatial GLCM elements in the row or column direction. If that value of the matrix is equal in size, the value of the correlation value is small.

We randomly extracted 20 samples (as shown in Figure 10) from the dataset for statistical analysis of the correlation and contrast of GLCM and the histogram variance of MLBP, as shown (a), (b), and (c) in Figure 11. The resolution of images from the satellites GF-2 and ZY-3 was 2 m. We can see that the characteristics of ships are distinguished from those of clouds and waves. This is because of the fact that the internal texture of ships is quite different compared with that of clouds and waves.
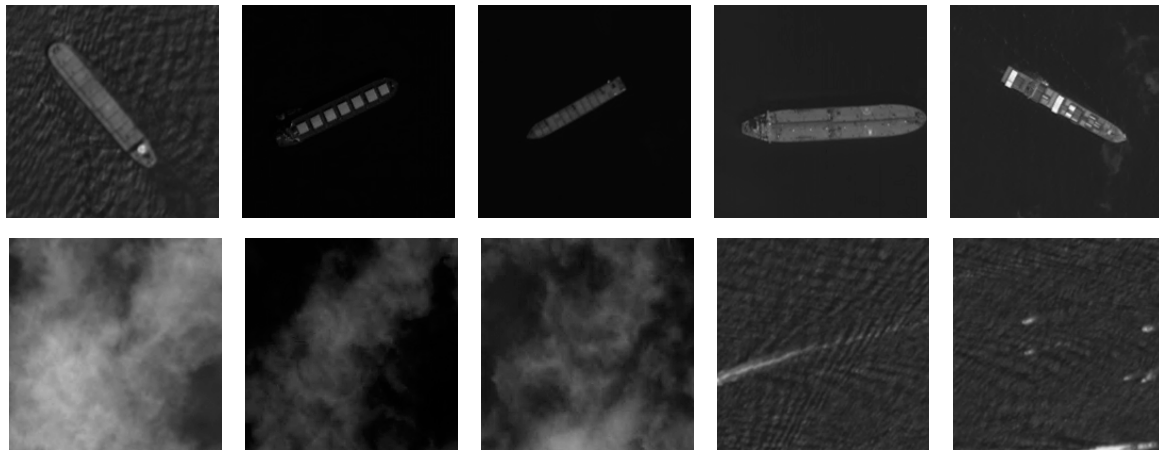
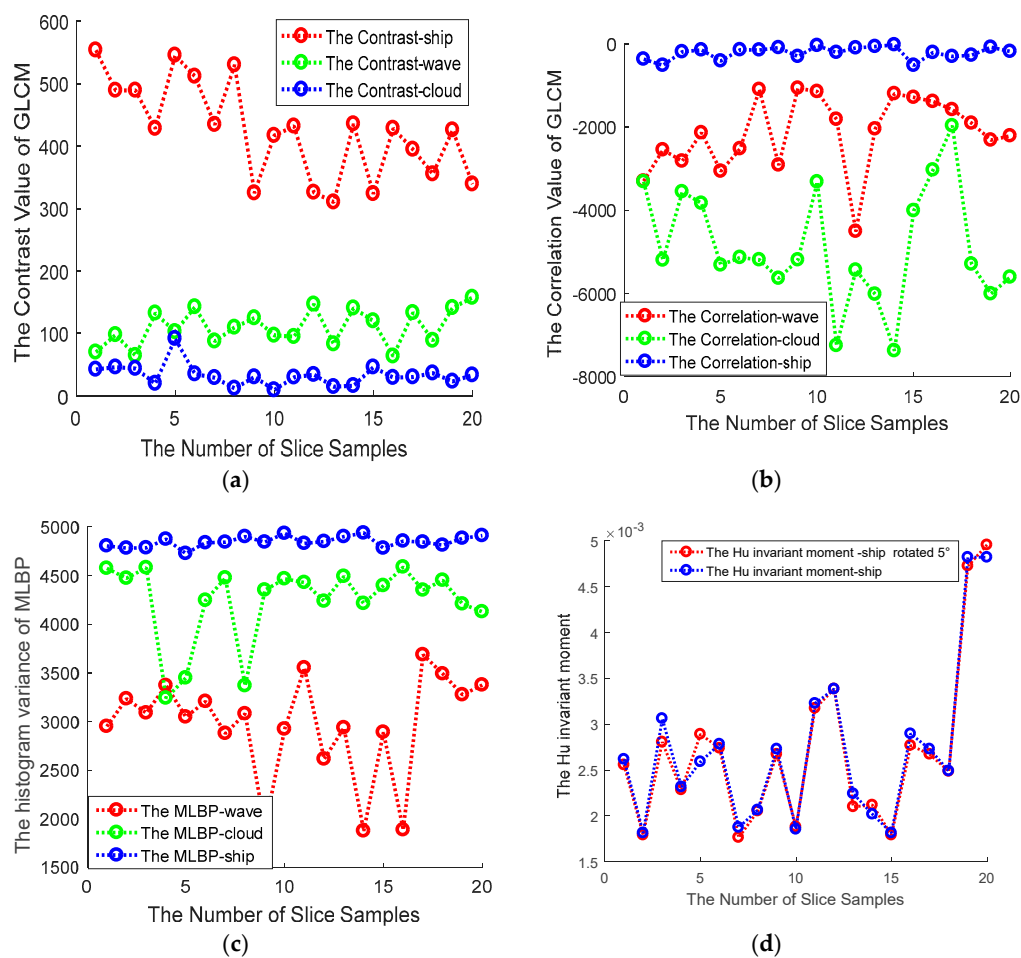**Figure 10.** Some typical ship and non-ship targets involved in the statistics.

**Figure 11.** The results of feature statistical experiments. (**a**) The contrast of grey-level co-occurrence matrix (GLCM) feature statistics. (**b**) The correlation of GLCM feature statistics. (**c**) The histogram variance of MLBP feature statistics. (**d**) The Hu moment feature statistics.

The Hu moment is a highly concentrated image feature with translation, rotation, and scale invariance. The first two moments M1 and M2 of Hu were used as a set of parameters to describe the characteristics of ships. We rotate the 20 ships by 5°, and the second moment of Hu changes little after the rotation, as shown (d) in Figure 11 From the above analysis, it can be seen that the feature description selected in this paper is strong and robust.

We used the SVM to discriminate real ship targets from candidates based on the obtained features. Eleven features should be feature-transformed and mapped to a high-dimensional space. Here, the radial basis function (RBF) is used as the kernel function of the SVM binary classifier, $K(x, x') = \exp(-\|x - x'\|^2/\sigma^2), \sigma > 0$. Before training, each feature is first normalized to the range of 0–1 to reduce the dominant role of a certain dimension feature due to its large magnitude. The RBF function has two key parameters: the penalty factor $c$ and the kernel parameter $\sigma$. We obtained the best parameter through the cross-validation method and set $c = 1$ and $\sigma = 0.7$ [29]. Training images come from the satellites GF-2 and ZY-3 2 m resolution.

## 4. Experimental Results and Discussion

### 4.1. Subjective Visual Evaluation of Saliency Models

In order to test the performance of saliency map extraction proposed in this paper for panchromatic optical remote sensing images, four typical algorithms were compared and analyzed, and they were ITTI, SR, PQFT, and one found in [12]. The ASD and MSRA10K datasets were not used in the evaluation of various algorithms, because these two datasets are mostly color images [30]. Images with a 2 m resolution from the satellites GF-2 and ZY-3 were used to construct 300 test sets under various sea conditions such as different shooting time, conditions, and sea surface false alarms. The slice size is $200 \times 200$ pixels. In addition, in order to evaluate the performance of various algorithms objectively, we extracted precise ground-truth images of contours in advance.

Figure 12 shows the ROI extraction in a set of typical cases, including the calm sea surface, low contrast, obvious sea clutter, and a similar texture.

It is not difficult to find through a large number of experiments that, for panchromatic images, the performance of the frequency domain method is better than the spatial domain method. The spatial domain method cannot suppress the clutter interference and is greatly affected by strong waves and clouds. The SR and PQFT show no significant differences in performance. For the second image in Figure 12c,d, the ship target is completely submerged into the background and cannot be distinguished. The performance of the method found in [12] is better than other frequency domain algorithms, but the detected target area is incomplete, and the target easily melts into the background, as shown in the third, fourth, and fifth images in Figure 12f. The method from [12] only uses gray information, so ROIs cannot be accurately detected. Compared with other algorithms, the method of the saliency map extraction proposed in this paper obtained a better effect, despite a low contrast or a complex background. The detection result was more complete and clearer. Moreover, the brightness value distribution of the target area was uniform, which also helps to detect completed targets in the binary process. This will prevent to some extent incomplete target slices caused by too much brightness or too much darkness.

We selected the PR curve and the comprehensive evaluation index F measure to evaluate the accuracy of various algorithms. Ground-truth images (manual marking) and binary images of saliency maps were recorded as G and M. In the PR curve, P and R refer to the precision and recall of the methods, respectively. Their calculation formulas are as follows:

$$\text{Precision} = \frac{|M \cap G|}{|M|} = \frac{TP}{TP + FP} \tag{33}$$

$$\text{Recall} = \frac{|M \cap G|}{|G|} = \frac{TP}{TP + FN} \tag{34}$$

Among them, the pixels belonging to G and M are simultaneously called TP. The pixels belonging to G and not to M are called FN. The pixels belonging to neither G nor M are called FP.

Recall and precision cannot be discussed in isolation. The F metric is introduced to comprehensively evaluate the detection performance of the saliency model as follows:

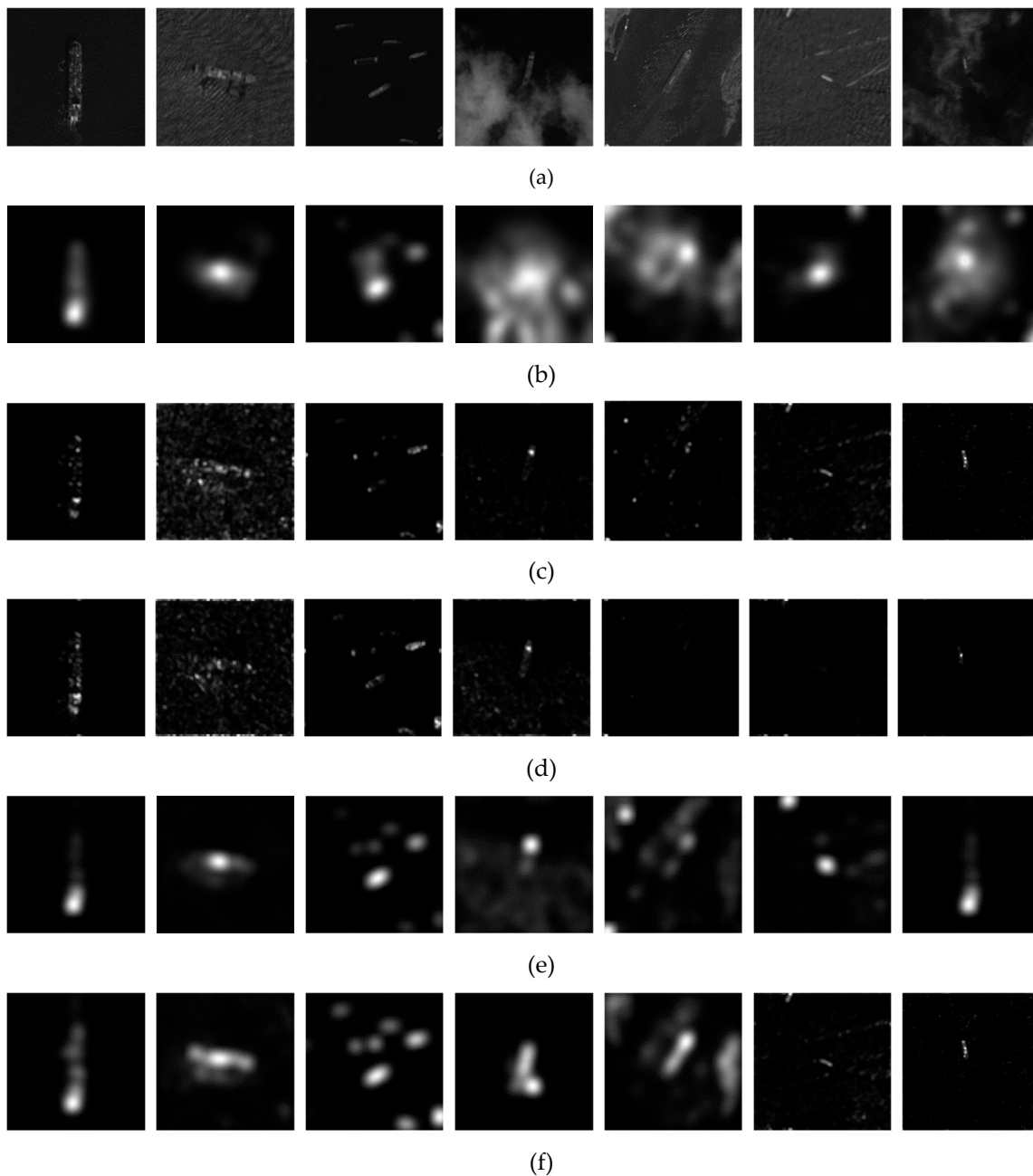$$F_\beta = \frac{(1+\beta^2)\text{Precision} * \text{Re}call}{\beta^2\text{Precision} + \text{Re}call} \tag{35}$$



(a)

(b)

(c)

(d)

(e)

(f)

**Figure 12.** Saliency map extraction of different methods. (**a**) Original images. (**b**) Results of the ITTI method. (**c**) Results of the spectral residual (SR) method. (**d**) Results of the PQFT method. (**e**) Results of a method found in [12]. (**f**) Results of our method.

Among them, the value of $\beta$ is different, and the requirements of recall and precision are different. Combined with actual engineering requirement, this paper pays the same attention to recall and precision, $\beta = 1$.

After experimental analysis, we selected ITTI in the spatial domain and the method in [12] in the frequency domain for comparison with our method. The performance of the two algorithms was relatively good in their domains. Because the test database contains all kinds of complex, typical sea conditions, the threshold T of the binary processing increased from 0 to 255 when the PR curve was drawn. The average value of recall and precision in different cases was obtained as shown below.

It can be seen from the Figure 13 that the PR and F curves of our algorithm are obviously higher than those of the other two methods. From the PR curve, it can be seen that, when the recall rate is 80%, the precision of ITTI and the method in [12] is about 40% and 73%, respectively. The precision of our algorithm is about 84%, which is 11% higher than the method in [12]. Our algorithm has an obvious advantage in the case of complex sea conditions because it has constructed a contrast map, a Gabor map, and a spatial and texture correlation feature to facilitate the integrity of target edge detection. At the same time, MPQFT constructs a multi-scale space, which is helpful for detecting targets of different sizes.
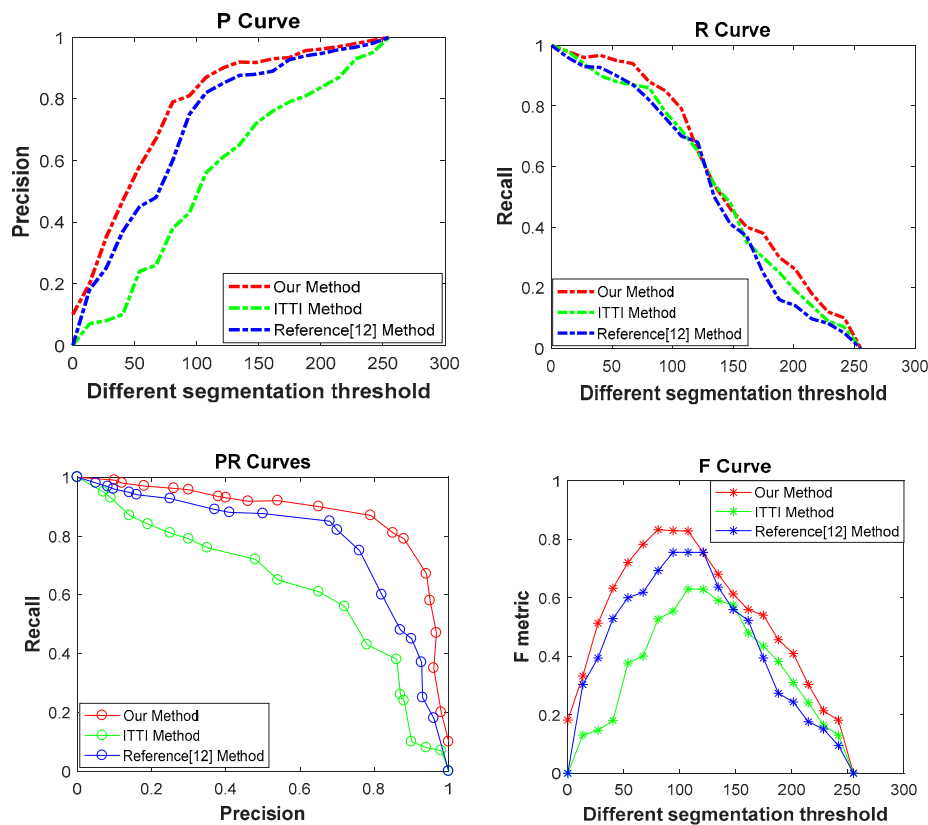


**Figure 13.** Objective comparison of different saliency algorithms.

*4.2. Overall Detection Performances*

Finally, we compared our overall detection method with three typical methods, and the evaluation criteria are defined as follows:

$$Accuracy = \frac{The\ number\ of\ ships\ detected}{The\ number\ of\ all\ targets(\text{ships and non} - \text{ships})\ detected} \tag{36}$$

$$Missed\ alarm\ rate(\text{MA}) = \frac{The\ number\ of\ ships\ judged\ as\ background}{The\ number\ of\ all\ ship\ targets} \tag{37}$$

$$\text{False } alarm \ rate(\text{FA}) = \frac{The \ number \ of \ \text{non}-ships \ judged \ as \ \text{ship targets}}{The \ number \ of \ all \ targets(\text{non}-\text{ships and ships})\text{detected}} \tag{38}$$

We selected 500 typical panchromatic optical remote sensing images with a size of 8192 × 4096 pixels under different imaging conditions from the satellites GF-2 and ZY-3 and with a 2 m resolution. (Some images can be downloaded from the website: www.cresda.com/CN/). In addition, other panchromatic satellite images from various sea surfaces were collected and used as the testing dataset from the publicly available Google Earth service with a resolution of 2 m. These images are very large in size and we have subdivided them into 120 sub-images with 8192 × 4096 pixels. A total of 620 images contain ship targets of different sizes and shapes. We tested our approach on three groups of different sea surfaces: calm sea with little interference, a complex sea surface influenced by clouds and waves, the cloud area percentage of which is about 20%, −30%, and worse imaging conditions, the cloud area percentage of which is more than 50% and as high as 85%. The images contained ships of different types and sizes.

The proposed method was implemented in C++ with Intel (R) Core (TM) i7-4770K CPU at 3.40 GHz and 64.0 GB RAM. The objective evaluation indices of the detection results are listed in Table 1. Because these images are so large in size that they will be compressed in this paper, we provided some local parts of the original images, as shown in Figures 14 and 15. Red borders indicate the ship targets we detected, and the yellow border areas indicate ships we missed with our method. In some cases, we lost some targets, as shown in of Figure 15a,d. In Figure 15a, one ship is almost completely obscured by clouds, making it difficult to locate the target, and another one is lost because the contrast of the ship and the background is too low, making it difficult to distinguish by the human eye. In Figure 15d, more than five ships are connected and docked, making it different from the conventional training model in terms of the shape and area; therefore, the targets were not recognized.

**Table 1.** Detection results of our method in various situations.

| Different Situations | Accuracy | FA | MA |
|---|---|---|---|
| Calm sea | 98.10% | 1.90% | 1.43% |
| Textured sea | 93.40% | 6.60% | 5.43% |
| Clutter sea | 86.80% | 13.20% | 10.88% |

In Table 1, we can see that with the increasing amount of interference, the accuracy of our method decreases slightly from Group 1 to Group 3. As shown in Figures 14h and 15h, our method lost some ships (areas indicated by yellow dotted lines) submerged under clouds, but the average accuracy reached up to 92.8% for different sea surfaces, and the average false alarm rate was 7.2%. Considering the existence of ships with different sizes in an image, the multi-scale saliency map extraction was designed so that the method can accurately identify ships of different sizes simultaneously. Images used in the experiment had a 2 m resolution. For ships smaller than 20 m (10 pixels), detection accuracy decreased, and in this case, targets would be so small that their feature parameters were sometimes inaccurate, affecting detection results.

At the same time, we compared our method with state-of-the-art methods proposed in [7,13,17]. Table 2 was obtained by averaging the evaluation indices under different sea conditions, showing the average performance of the four methods. Table 3 shows a comparison of four methods under different sea conditions. The precision of the methods from [13,17] is high with respect to calm sea surfaces, but for complex sea surfaces, detection performance is greatly reduced, as shown in Table 3. The method proposed in this paper is optimal for detection accuracy. On average, the detection time of our method is 1.6 s. The efficiency of our algorithm is not the highest. Compared with the method from [13], which has comparable accuracy, our efficiency is almost double.
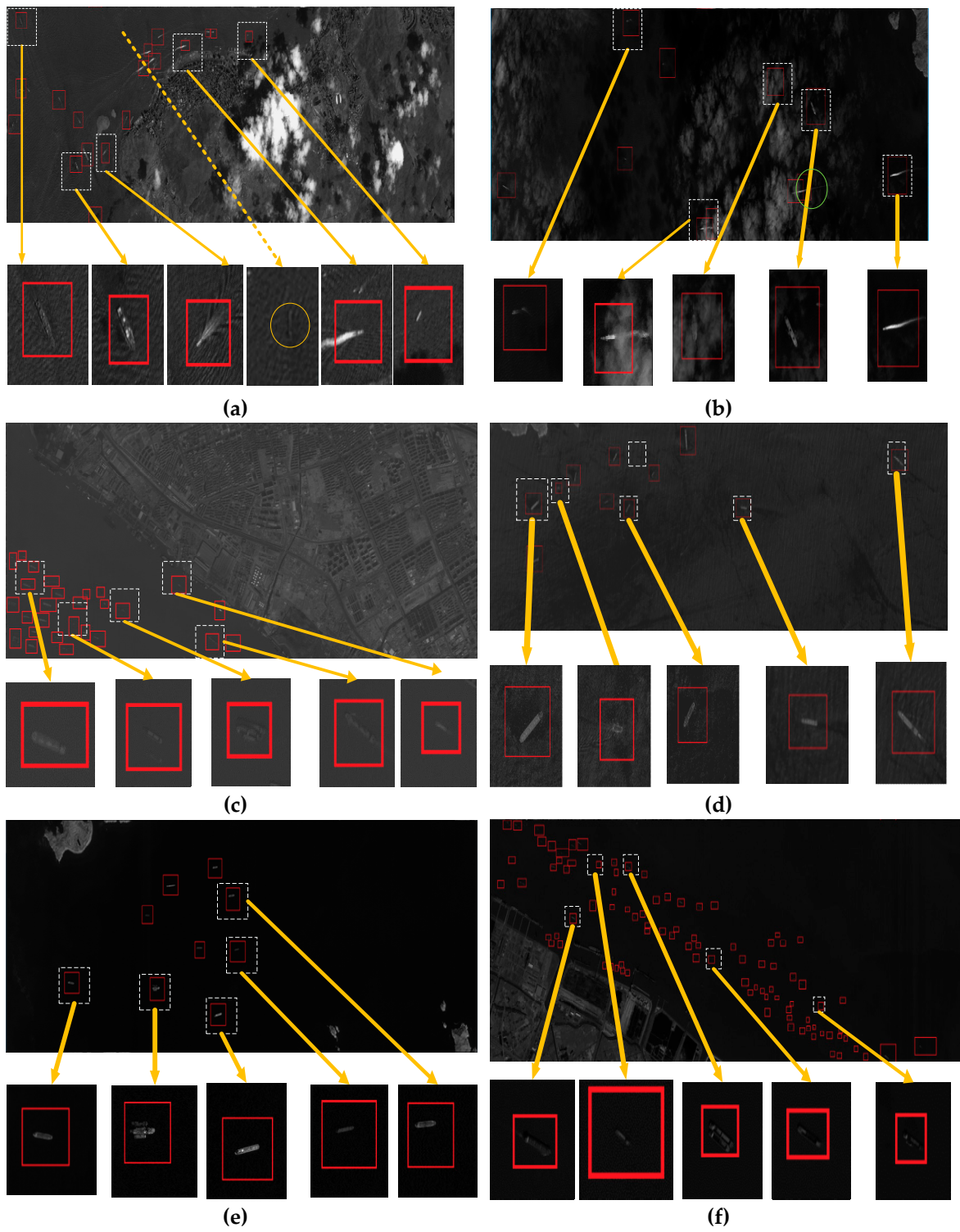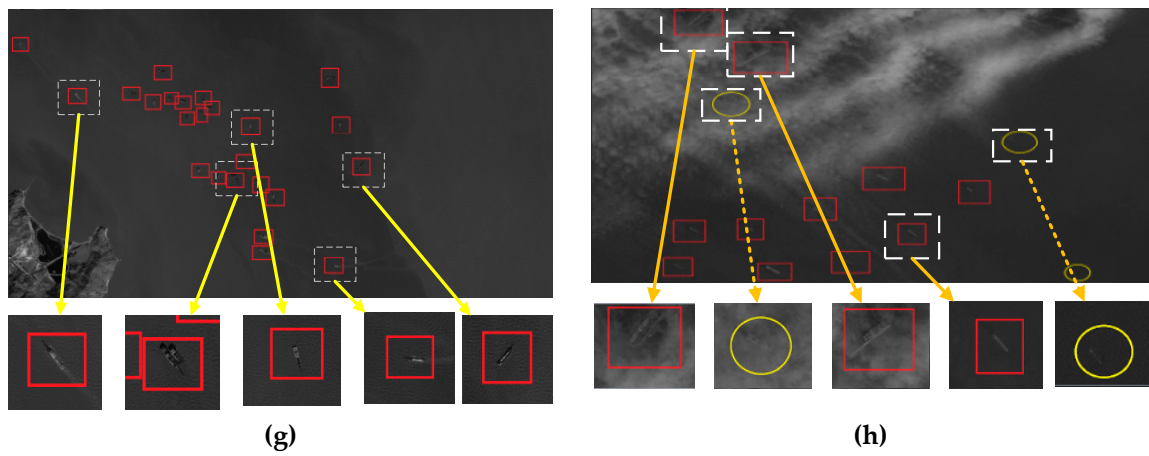
**Figure 14.** *Cont.*

**(g)**　　　　　　　　　　　　　　　　　　　　**(h)**

**Figure 14.** Detection results of images from the satellites ZY-3 and GY-2. (**a**) Detection results of cloud and wave coexistence. (**b**) Detection results of thick cloud interference. (**c**) Detection results of mist interference. (**d**) Detection results of strong wave interferences. (**e**) Detection results of a calm sea surface. (**f**) Detection results of very low contrast. (**g**) Detection results of middle-size ships with wakes and wave interference. (**h**) Detection results with thick cloud occlusion.
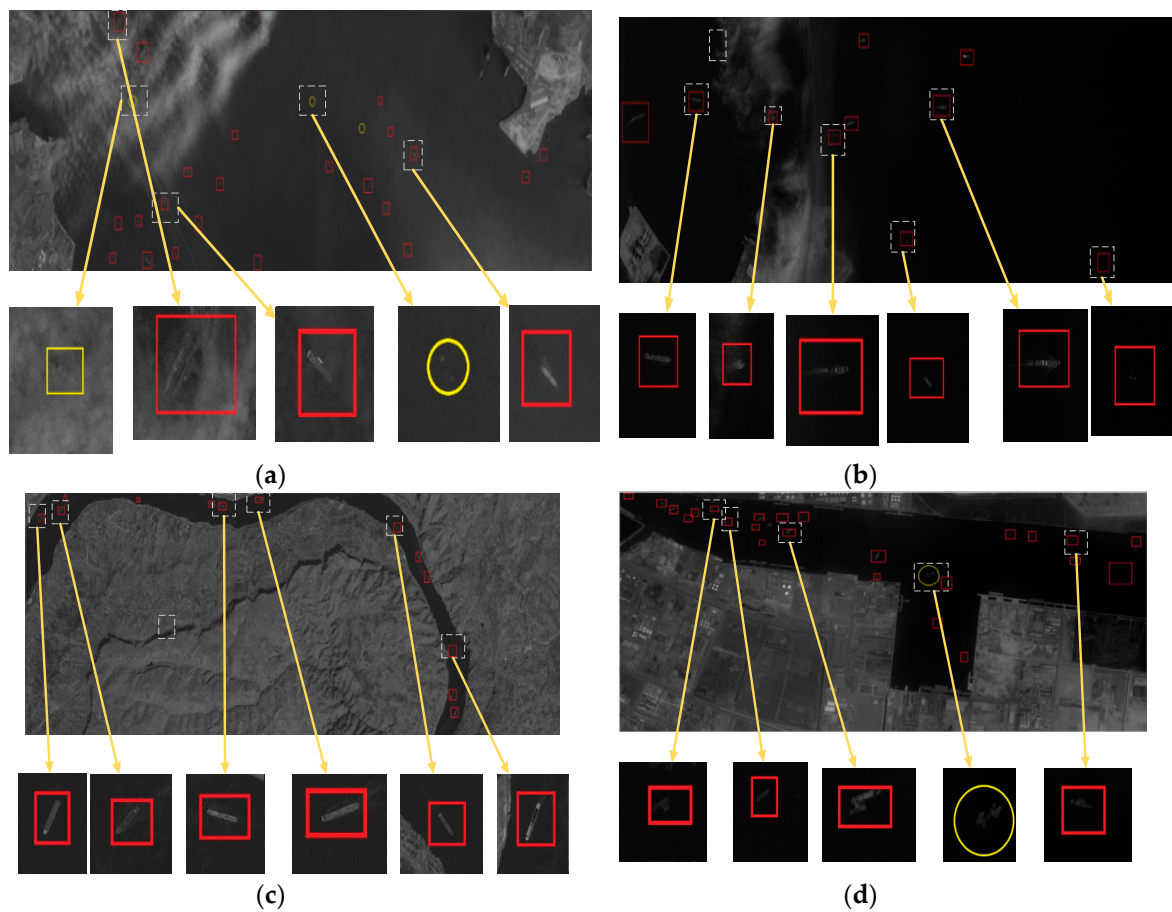


**(a)**　　　　　　　　　　　　　　　　　　　　**(b)**

**(c)**　　　　　　　　　　　　　　　　　　　　**(d)**
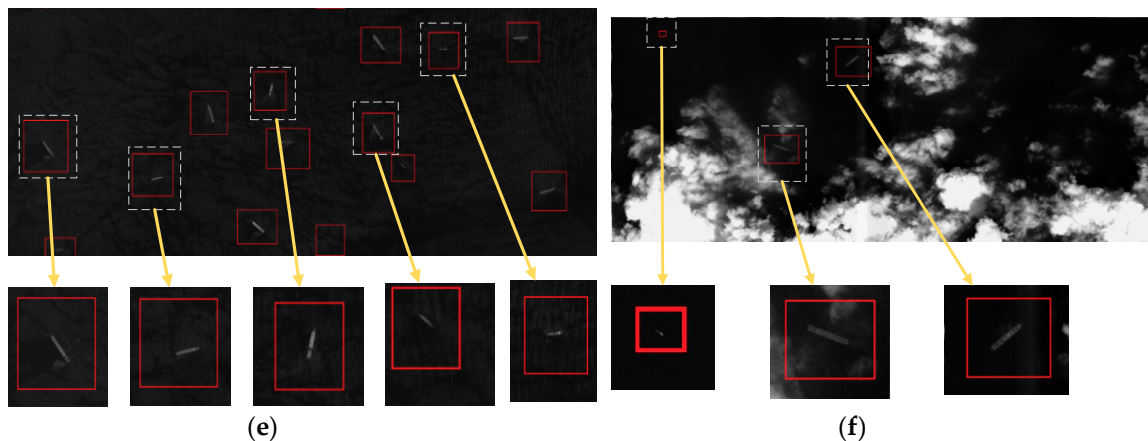
**Figure 15.** *Cont.*

**(e)**    **(f)**

**Figure 15.** Detection results of images from Google Earth. (**a**) Detection results of cloud occlusion. (**b**) Detection results of a little cloud interference. (**c**) Detection results of the inland lake. (**d**) Detection results of the port. (**e**) Detection results of wave interference. (**f**) Detection results of massive cloud occlusion.

**Table 2.** Average detection performance of different methods.

| Method | Accuracy | FA | MA | Time Consumed |
|---|---|---|---|---|
| Method from [7] | 80.2% | 19.8% | 20.3% | 1.2 s |
| Method from [17] | 83.9% | 16.1% | 11.1% | 1.5 s |
| Method from [13] | 88.4% | 11.6% | 7.3% | 3.1 s |
| Our method | 92.8% | 7.2% | 5.9% | 1.6 s |

**Table 3.** Detection results of different methods in various situations.

| Method | Different Situations | Accuracy | FA | MA |
|---|---|---|---|---|
| Method from [7] | Calm sea | 92.20% | 7.80% | 4.40% |
| | Textured sea | 81.30% | 18.70% | 12.40% |
| | Clutter sea | 67.00% | 33.00% | 44.20% |
| Method from [17] | Calm sea | 98.30% | 1.70% | 1.40% |
| | Textured sea | 83.20% | 16.80% | 6.70% |
| | Clutter sea | 70.20% | 29.80% | 25.20% |
| Method from [13] | Calm sea | 97.30% | 2.70% | 1.80% |
| | Textured sea | 85.30% | 14.70% | 5.60% |
| | Clutter sea | 82.60% | 17.40% | 14.60% |
| Our method | Calm sea | 98.10% | 1.90% | 1.43% |
| | Textured sea | 93.40% | 6.60% | 5.43% |
| | Clutter sea | 86.80% | 13.20% | 10.88% |

A linear function combining pixel and region characteristics was employed to select ship candidates in [7]. When the background was covered by clouds, the location of ROIs would fail, resulting in missed alarms. Compactness and the length–width ratio were considered to remove false alarms. The description features were not enough to distinguish between targets and background when the background became complex. The detection performance was greatly reduced. Regarding the calm sea, the method in [17] achieved the best performance. The accuracy was 98.3%, and the false and missed alarm rates were very low. However, this method only adopted intensity distinctness to find ROIs and led to omissions in complex backgrounds. In the false alarm exclusion stage, the LBP histogram features of the bow, stern, left hull, and right hull were used. The algorithm could achieve better results in distinguishing large ships from false alarms, but when the ships were small or the background became complex, the false alarm rate was higher, because there was little difference between each

LBP part of a ship. In [13], the saliency map based on the multi-scale and multi-direction wavelet decomposition was detected to extract ROIs, and the pixel distribution discrimination was designed to eliminate false alarms. Compared with the two previous methods, this method showed a great improvement in accuracy and the false alarm rate. However, this method was designed for color images, and its performance is degraded when applied to panchromatic images. Moreover, the pixel distribution can only remove targets that vary widely in shape. Compared with other algorithms, the performance of our method is a great improvement. In order to improve recall and accuracy, in the extraction of the ROI stage, the MPQFT was designed for panchromatic remote sensing images. The quaternion was constructed with the full consideration of image contrast and edge structure. In addition, in order to find candidates in complex backgrounds, texture and spatial similarity were also adopted. In the latter stage, we proposed the MLBP method combined with 10 other features to reduce false alarms. Figure 14 shows our detection results in some typical sea conditions. On the whole, the algorithm achieved ideal detection results and detected ship targets under complex sea conditions. At the same time, if the sea surface is calm, our method can basically ensure that false alarms and missed detections are less than 5%.

Using the method proposed by this paper, color images should be transformed into gray-scale images for processing. We tested our approach on different sea surfaces in color images. These color images were synthesized from the three spectrums (RGB) from the satellite GF-2 and had an 8 m resolution. Some detection results for color images are shown in Figure 16. The first row contains color images. The second row contains the corresponding gray-scale images and detection results. On the whole, a good performance was achieved. The red border is the target detected by the method. However, it is generally said that the resolution of color images for remote sensing multi-spectral images is much lower than that of panchromatic images. Therefore, the accuracy of this method will be reduced for small ships. As shown in Figure 16, the yellow border is the lost target because the ship is too small in the image with an 8 m resolution.
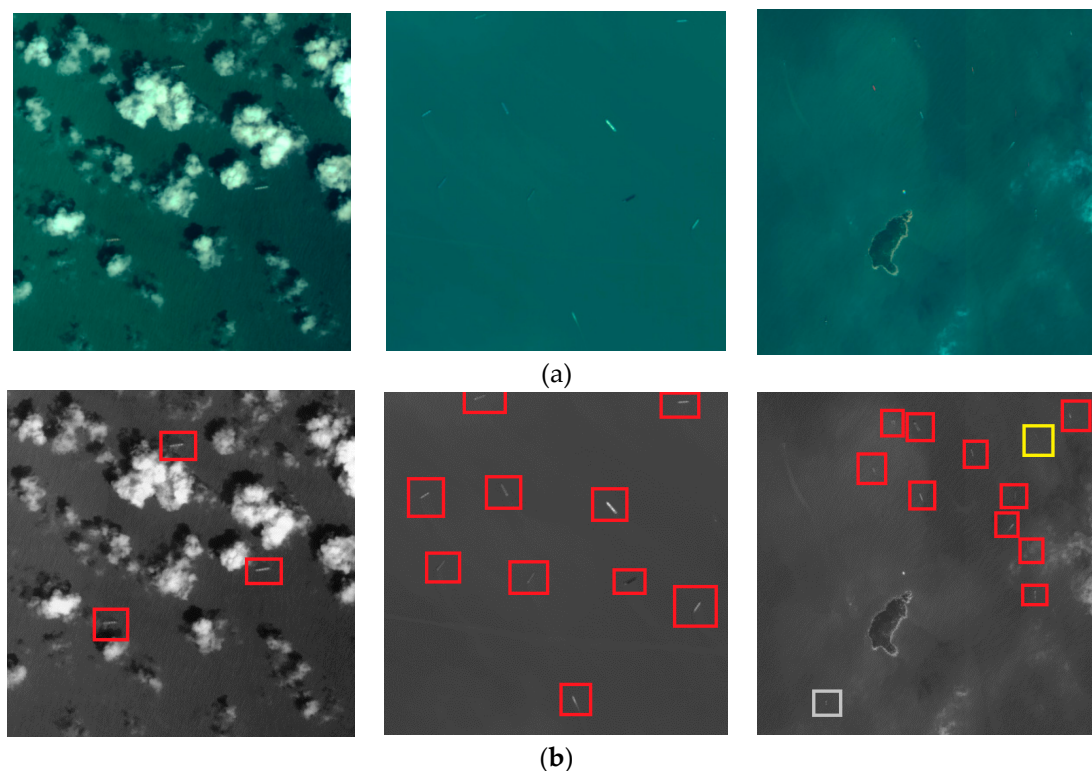


(a)

(b)

**Figure 16.** Detection results of color images. (**a**) Color images. (**b**) The results on the corresponding grayscale images.

A platform of FPGA combined with DSP was built, and the method of ship detection proposed in this paper was optimized and embedded into a hardware platform. DSP used a TMS320C6678 with an 8-core processor. FPGA adopted the VIRTEX-7 chip produced by the company Xilinx. We carried out a large number of hardware tests of large-field optical remote sensing images. The image size was 8192 × 4096 pixels -2 m resolution. The processing time of one image was 242 ms on average. Compared with the performance of the original software method, the efficiency of the method was greatly improved, and the system has strong versatility and expansibility, which lays the foundation for practical engineering applications.

## 5. Conclusions

In this paper, we present a novel ship detection method for panchromatic optical remote sensing images consisting of saliency map extraction and target discrimination in complex backgrounds. We adopted an efficient frequency-domain model based on the hyper-complex Fourier transform of the quaternion to locate candidate regions, which made the brightness distribution of ROIs more uniform and complete and effectively reduced missed detections. Meanwhile, to determine whether the candidate target was a ship or not, multi-dimensional description features were extracted and designed by the characteristics of ships and no-ships. In addition, an improved LBP, which takes into account contrast information between pixels, was presented and provided more powerful descriptions. Finally, we built a database through actual panchromatic remote sensing images, and used SVM training to obtain a more stable model for ship confirmation. The experimental results under various sea backgrounds demonstrate that the proposed ship detection method can obtain high precision and detection robustness.

Although our method has achieved promising results, several issues remain to be further settled. With the development of satellite remote sensing, the hyper-spectral data should be fully used to construct a ship description operator, which can be combined with the advantages of visible and SAR images, making ship detection methods more robust and easily implemented in hardware.

**Conflicts of Interest:** The authors declare that there is no conflict of interest.

## References

1.  Wang, Y.C.; Ning, X.Y.; Leng, B.H.; Fu, H.X. Ship detection based on deep learning. In Proceedings of the IEEE International Conference on Mechatronics and Automation, Tianjin, China, 4–7 August 2019; pp. 275–279.
2.  Ma, J.; Zhou, Z.; Wang, B.; An, Z. Hard ship detection via generative adversarial networks. In Proceedings of the 31th Chinese Control and Decision Conference, Nanchang, China, 3–5 June 2019; pp. 3961–3965.
3.  More, N.; Murugan, G.; Singh, R.P. A survey paper on various inshore ship detection techniques in satellite imagery. In Proceedings of the Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 16–18 August 2018; pp. 1–5.
4.  Corbane, C.; Najman, L.; Pecoul, E.; Demagistri, L.; Petit, M. A complete processing chain for ship detection using optical satellite imagery. *Int. J. Remote. Sens.* **2010**, *31*, 5837–5854. [CrossRef]
5.  Proia, N.; Page, V. Characterization of a Bayesian ship detection method in optical satellite images. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 226–230. [CrossRef]
6.  Xu, Q.; Li, B.; He, Z.; Ma, C. Multiscale contour extraction using a level set method in optical satellite images. *IEEE Geosci. Remote. Sens. Lett.* **2011**, *8*, 854–858. [CrossRef]

7.  Yang, G.; Li, B.; Ji, S.; Gao, F.; Xu, Q. Ship detection from optical satellite images based on sea surface analysis. *IEEE Geosci. Remote. Sens. Lett.* **2014**, *11*, 641–645. [CrossRef]

8.  Itti, L.; Koch, C.; Niebur, E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 1254–1259. [CrossRef]

9.  Guo, C.; Qi, M.; Zhang, L. Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AL, USA, 23–28 June 2008; pp. 1–8.

10. Hou, X.; Zhang, L. Saliency detection: A spectral residual approach. In Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; pp. 1–8.

11. Li, J.; Levine, M.D.; An, X.; Xu, X.; He, H. Visual saliency based on scale-space analysis in the frequency domain. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *35*, 996–1010. [CrossRef] [PubMed]

12. Dong, C.; Liu, J.; Xu, F. Ship Detection in Optical Remote Sensing Images Based on Saliency and a Rotation-Invariant Descriptor. *Remote Sens.* **2018**, *10*, 400. [CrossRef]

13. Xu, F.; Liu, J.; Dong, C.; Wang, X. Ship Detection in Optical Remote Sensing Images Based on Wavelet Transform and Multi-Level False Alarm Identification. *Remote Sens.* **2017**, *9*, 985. [CrossRef]

14. Yang, F.; Xu, Q.; Li, B.; Ji, Y. Ship detection from thermal remote sensing imagery through region-based deep forest. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 449–453. [CrossRef]

15. Sun, Y.; Lei, W.; Hu, Y. Rapid ship detection in remote sensing images based on visual saliency model. *Laser Technol.* **2018**, *42*, 379–384.

16. Zhou, F.; Fan, W.; Sheng, Q.; Tao, M. Ship detection based on deep convolutional neural networks for polar Sar images. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing, Symposium, Valencia, Spain, 22–27 July 2018; pp. 681–684.

17. Yang, F.; Xu, Q.; Li, B. Ship detection from optical satellite images based on saliency segmentation and structure-lbp feature. *IEEE Geosci. Remote. Sens. Lett.* **2017**, *14*, 602–606. [CrossRef]

18. Qi, S.; Ma, J.; Lin, J.; Li, Y.; Tian, J. Unsupervised ship detection based on saliency and s-hog descriptor from optical satellite images. *IEEE Geosci. Remote. Sens. Lett.* **2015**, *12*, 1451–1455.

19. Shi, Z.; Yu, X.; Jiang, Z.; Li, B. Ship detection in high-resolution optical imagery based on anomaly detector and local shape feature. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 4511–4523.

20. Liu, W.; Ma, L.; Chen, H. Arbitrar-oriented ship detection framework in optical remote sensing images. *IEEE Geosci. Remote. Sens. Lett.* **2018**, *15*, 937–941. [CrossRef]

21. Wang, T.F.; Gu, Y.F. CNN based renormalization method for ship detection in VHR remote sensing images. In Proceedings of the IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 1252–1255.

22. Nie, S.; Jiang, Z.; Zhang, H.; Cai, B.; Yao, Y. Inshore ship detection based on mask R-CNN. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 693–696.

23. Nie, T.; He, B.; Bi, G.; Zhang, Y.; Wang, W. A Method of Ship Detection under Complex Background. *ISPRS Int. J. Geo-Information* **2017**, *6*, 159. [CrossRef]

24. Kamarainen, J.K.; Kyrki, V.; Kalviainen, H. Invariance properties of Gabor filter-based features overview and applications. *IEEE Trans. Image. Process.* **2006**, *15*, 1088–1099. [CrossRef] [PubMed]

25. Moreno, P.; Bernardino, A.; Santos-Victor, J. Gabor parameter selection for local feature detection. In Proceedings of the Iberian Conference on Pattern Recognition & Image Analysis, Estoril, Portugal, 7–9 June 2005.

26. Xu, F.; Liu, J.; Sun, M.; Zeng, D.; Wang, X. A Hierarchical Maritime Target Detection Method for Optical Remote Sensing Imagery. *Remote Sens.* **2017**, *9*, 280. [CrossRef]

27. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [CrossRef]

28. Boykov, Y.; Jolly, M. Interaetive graph cuts for optimal boundary & region segmentation of objects in N-D images [C]. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, Kauai, HI, USA, 8–14 December 2001; pp. 731–738.

29. Zhu, C.; Zhou, H.; Wang, R.; Guo, J. A novel hierarchical method of ship detection from spaceborne optical image based on shape and texture features. *IEEE Trans. Geosci. Remote. Sens.* **2010**, *48*, 3446–3456. [CrossRef]

30. Achantay, R.; Hemamiz., S.; Estraday, F. Frequency-tuned salient region detection. In Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 1597–1604.