remote sensing

MDPI

Article

Ship Detection in Optical Remote Sensing Images Based on Saliency and a Rotation-Invariant Descriptor

Chao Dong [1,2], Jinghong Liu [1,*] and Fang Xu [1]

[1] Key Laboratory of Airborne Optical Imaging and Measurement, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China; dongchao315@mails.ucas.ac.cn (C.D.); xufang59@126.com (F.X.)

[2] University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: liujinghong@ciomp.ac.cn; Tel.: +86-182-0431-9940

Received: 8 January 2018; Accepted: 1 March 2018; Published: 5 March 2018

Abstract: Major challenges for automatic ship detection in optical remote sensing (ORS) images include cloud, wave, island, wake clutters, and even the high variability of targets. This paper presents a practical ship detection scheme to resolve these existing issues. The scheme contains two main coarse-to-fine stages: prescreening and discrimination. In the prescreening stage, we construct a novel visual saliency detection method according to the difference of statistical characteristics between highly non-uniform regions which allude to regions of interest (ROIs) and homogeneous backgrounds. It can serve as a guide for locating candidate regions. In this way, not only can the targets be precisely detected, but false alarms are also significantly reduced. In the discrimination stage, to get a better representation of the target, both shape and texture features characterizing the ship target are extracted and concatenated as a feature vector for subsequent classification. Moreover, the combined feature is invariant to the rotation. Finally, a trainable Gaussian support vector machine (SVM) classifier is performed to validate real ships out of ship candidates. We demonstrate the superior performance of the proposed hierarchical detection method with detailed comparisons to existing efforts.

Keywords: remote sensing; visual saliency; radial gradient transform; covariance matrix; Gaussian SVM

1. Introduction

Sea target detection has been a standing topic in the field of remote sensing image processing for several decades due to the wide range of applications, such as fishery management, vessel traffic services, illegal oil spills, naval warfare, and maritime activities, etc. From the perspective of data sources, ship detection can be roughly classified into three domains: synthetic aperture radar (SAR) images, infrared (IR) images, and visible remote sensing (VRS) images. Because the synthetic aperture radar (SAR) method has the capacity to image day and night regardless of weather conditions, all SAR-based methods expend greatly, achieving impressive performance. However, the invisibility of small and wooden boats in SAR images may result in detection failures. Besides, lack of color and texture features makes SAR imagery unsuitable for recognizing the ship targets. IR images are employed to enhance the vision effect in weak light conditions but they also have some drawbacks, such as poor signal-to-noise ratio, insufficient structure information, and varied gray levels [1]. Compared with SAR and IR images, the VRS images investigated in this paper are more intuitive and capture more details and complex structures of an observed scene, which can be further used in target recognition. However, the above-mentioned facts about VRS images complicate the background and pose three main challenges to ship detection:

- The high variability of targets caused by the viewpoint variation, imaging sensor parameters, occlusion, ship wakes, color, speed, and material of ships, etc.

- High false alarm rate due to islands, heavy clouds, ocean waves, and the various and uncertain sea state conditions, like partial cloud cover, fog, wind, and swell.
- The third issue is the computation burden. Most detection methods have high computational cost. Hence, reducing computational cost is considered to be a key issue for the large-scale remote sensing images.

In consideration of these challenges mentioned above, we believe that a practical ship detection method should meet two requirements: it should be robust to the interference of the high variability of targets and background clutter such as waves, islands, clouds, and so forth. Of equal importance, with the purpose of the engineering applications, it should have lower calculation complexity and satisfy the requirements of real-time processing.

We have performed a thorough investigation into the existing approaches. Unfortunately, to our best knowledge, most existing ship detection methods are only efficient under certain conditions and are unable to satisfy all these goals simultaneously. For instance, some studies focused on discriminating targets from their surroundings according to the difference of intensity contrast or statistical distribution (e.g., [2,3]), but they are not suitable for the situation where the target intensity is similar to the background. The method in [4] employed the Bayesian decision theory, which was only efficient for detecting some small ship targets. With the development of machine learning, many researchers approached object detection through feature extraction and two-class classification operations. For example, Zhu [2] combined local multiple patterns with the shape and the texture features to enhance the discriminative ability of the feature set, and then a semi-supervised classification was adopted to remove the false alarms. However, the pre-detection algorithm only worked well on the images that had a quiet sea background. Shi [5] employed a hyperspectral algorithm to extract candidate regions and a local feature descriptor combined with AdaBoost classifier for discrimination. Nevertheless, they needed to generate four classifiers to solve the variation of ship direction. Han [6] developed an algorithm by combining weakly supervised learning (WSL) and high-level feature learning, which can reduce human labor for annotating training data, but the multi-scale sliding window adopted to handle the different size of the targets is time-consuming. In conventional machine learning-based methods, feature extraction is quite important for high-performance object detection systems. However, the selection of distinguished features is still a challenging problem. To alleviate this problem, deep learning, which can automatically learn features from data, has been attempted for the recognition of the ship targets [7–9]. Zhou [7] designed a ship detection method based on convolutional neural networks and a singular value decompensation algorithm. Tang [8] exploited deep neural network for high-level feature representation and ship classification. Lin [9] proposed a fully convolutional network to label every pixel of the input image into three classes: land, sea and ships. However, such methods have complex training phases and the complex steps also make the implementation difficult.

Since the ships in a VRS image of the sea are salient objects, they are usually sparsely distributed and can easily be identified by the human visual attention system. Thus, the saliency models are introduced to identify attention-grabbing regions which may contain salient objects. The saliency model can be mainly divided into two types: spatial domain saliency, and transform domain saliency. For the former, one of the earliest spatial domain saliency models was proposed by Itti [10]. The algorithm constructed the final saliency map based on intensity, color and orientation features. Harel [11] defined a computational saliency model based on Markov chains and treated the equilibrium distribution over map locations as saliency values. With respect to the latter values, some studies tried to obtain the saliency map in the transform domain, which played an important role in the ship detection. For instance, Bi [12] extracted the target candidate regions by using a bottom-up and multiscale visual attention mechanism. In a similar fashion, Guo [13] employed the SR model (Spectral Residual) [14] to obtain initial target curve; Qi [15] applied the PFT model (Phase Spectrum of Fourier Transform) followed by a homogeneous filter to extract the candidate regions; and Xu [16] constructed a combined saliency model with self-adaptive weights to prescreen the ship candidates. These saliency models calculated in the frequency domain mentioned above have better performance (especially

in a highly cluttered backgrounds) as compared with the spatial domain saliency methods [10,11]. However, they also have some drawbacks, such as the low resolution of saliency map, low target integrity, and blurring of the target boundary. The frequency-tuned saliency detection method [17], which can obtain full-resolution saliency maps and well-defined boundaries of objects, was applied by Wang [18] to extract the target regions. Despite the fact that these models have a low missed detection rate, they still suffer from the interference mentioned earlier and cause a high number of the false alarms. Therefore, it is still vital to go deeply into the study of fast and efficient ship detection methods which can pop out the targets and suppress the distractors under complex uncertain situations.

In order to solve the problem in the ship detection, the first requirement is an efficient pre-detection model accelerating the prescreening process and decreasing the false alarms. Furthermore, a robust feature set is also required to discriminate the ships from non-ship targets. To meet the two requirements mentioned above, a practical ship detection scheme is presented in this paper. The workflow of our detection algorithm is given in Figure 1.

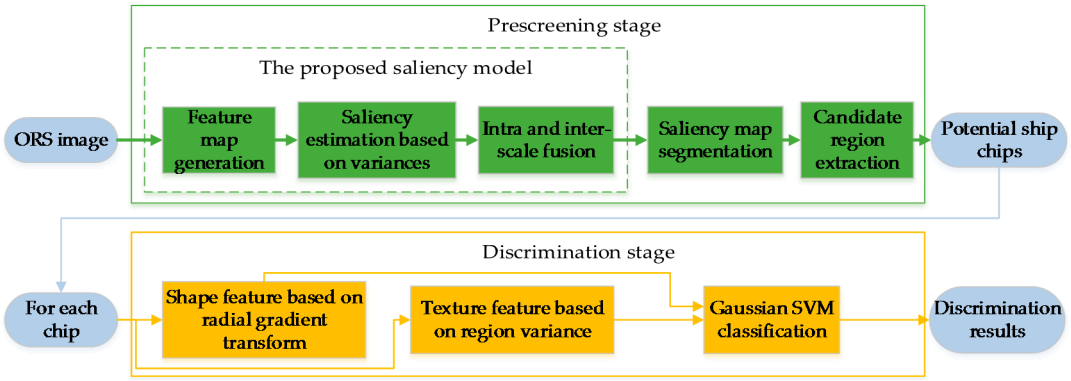


Figure 1. Diagram of the proposed hierarchical target detection scheme. The prescreening stage consists of saliency detection and binary segmentation. The discrimination stage includes feature extraction (radial gradient descriptor [19] and sigma set descriptor [20]) and classification by using Gaussian support vector machine (SVM) [21]. ORS: optical remote sensing.

The scheme contains two main coarse-to-fine stages: prescreening and discrimination. According to the difference of the region characteristics between regions of interest (ROIs) and natural background, a novel and practical ship candidate detection scheme based on region variance is proposed in the prescreening stage. Firstly, our model decomposes an input image into non-overlapping regions of square blocks and estimates their variances of simple features. Secondly, the information entropy is introduced to adaptively tune the relative weight of the saliency maps estimated by variances of the different feature. Finally, the inter-scale fusion is performed to increase the contrast between the salient and non-salient patches. In this way, not only can the targets be precisely detected, but the false alarms are also significantly reduced. After obtaining the saliency map, binary segmentation is operated to extract the candidate regions. In the discrimination stage, taking advantage of the symmetrical shape of the ships, the radial gradient histogram [19] is applied for guaranteeing the rotation invariance. Additionally, the region covariance descriptor [20], which is robust to large rotations and illumination changes, is utilized to describe the texture feature of the targets. Both shape and texture features are extracted and concatenated as a feature vector for subsequent classification, and then a trainable Gaussian support vector machine (SVM) classifier is performed to further remove the false alarms and maintain the real ship targets. Compared with the previous works [2,15], our detection model can achieve better performance in terms of both detection accuracy and running time. As a result, it is potentially of great benefit in the complex task of ship detection.

The rest of this paper is organized as follows. Section 2 introduces the framework of the visual saliency detection. Section 3 describes the discrimination stage, including the combined rotation-invariant

descriptor as well as the Gaussian SVM. Experimental results are provided in Section 4. We then briefly conclude on the method, performances, and future work in Section 5.

2. Ship Candidate Extraction Based on Saliency

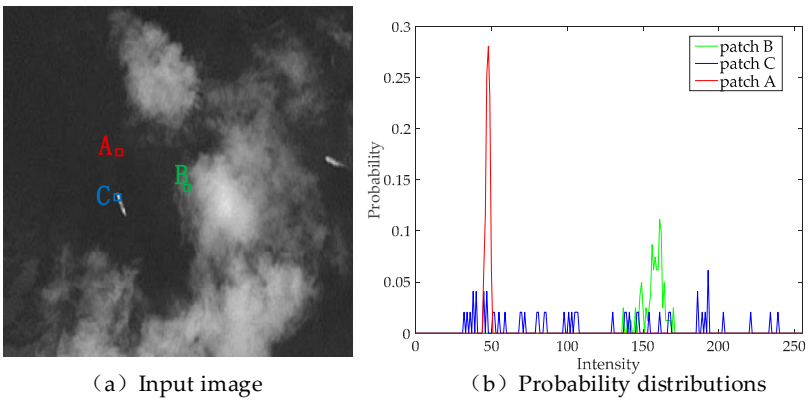
In the prescreening stage, the saliency value of each region is determined by quantifying its variance. The attended regions are detected to highlight potential objects by performing the fast and efficient saliency detection. Secondly, the ship candidates are extracted from the segmented binary image.

2.1. The Proposed Saliency Model

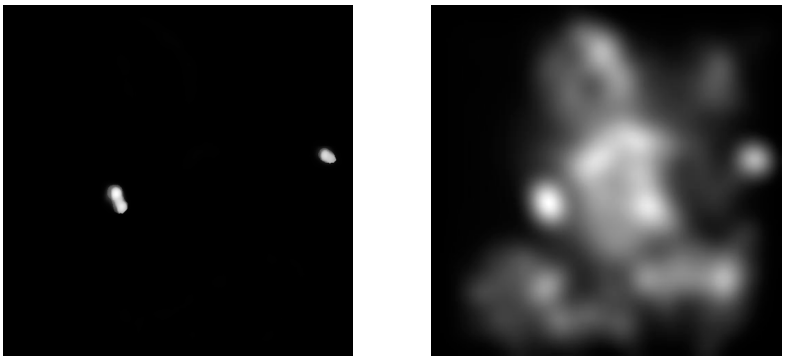
The ship targets in a VRS image of the sea are more salient than the background because the pixels of the targets are variable while those of the background have great similarities. Then, if we extract different low-level features from the images, the feature set of the ships will be very distinct from those of the sea backgrounds. It can be also concluded that a patch which contains a part of the target has more complex information compared with the one which only contains the similar background. To describe the distinction mentioned above, statistical characteristics have been investigated and proved to be powerful descriptions in remote sensing image processing. For instance, the variance weighted information entropy (WIE) was applied to detect target both in infrared and SAR images and achieved impressive performance [22,23]. In this paper, the value of region variances from optical remote sensing images in their uniform areas, the area including false alarms, and the partial area of ship target is primarily tested.

To illustrate the general idea, consider that patch A, patch B and patch C shown in Figure 2a,b present their corresponding gray level distributions. It can be observed that the gray level distributions of patch A (red line) and patch B (green line) are very different from that of patch C (blue line). Due to the wide range gray level distribution of patch C, the region variance values of highly non-uniform areas (patch C) which allude to ROIs are usually greater than those of the homogeneous backgrounds (patches A and B). In other words, region variance, as a basic regional statistical characteristic, can measure the complexity of a given patch to some extent. The similar conclusion could also be found in another model [24]. Therefore, it is reasonable to connect the saliency of a region with its variance. We constructed a novel saliency model based on this fact. As shown in Figure 2c, if our model is performed, the ship targets pop out from the background. We also compare our results with the seminal model of Itti in Figure 2d. The Itti model computes intensity, color, and orientation maps for a given input image base on a center-surround operation. The resulting feature maps are combined into the saliency map using a winner-takes-all network and an inhibition of the return mechanism. As shown in the second raw of Figure 2, compared with the Itti model, the proposed model is more effective in suppressing the background interference.

Next, we explain the proposed saliency model in detail. Figure 3 shows its schematic diagram. There are four main steps. First, we extract pixel amplitude and amplitude derivative features. Secondly, the rarity values for each scale are estimated based on the region variance of the different feature map. Afterwards, selection algorithm and intra-scale fusion are applied based on the information entropy. Finally, we obtain the final saliency map by performing the multi-layer cellular automata [25]. A detailed description of the proposed saliency model is provided hereinafter.



（a）Input image （b）Probability distributions



（c）Our result （d）Itti’s result

Figure 2. The proposed saliency model based on region variance analysis. (**a**) the original image; (**b**) probability distributions of patches A–C. The region variance values in terms of intensity information for A–C are 2.8, 87.6 and 1.8 × 10^3, respectively, and the sizes of A–C are all 8 × 8 pixels. For contrast, our and Itti's results are shown in (**c**,**d**), respectively.

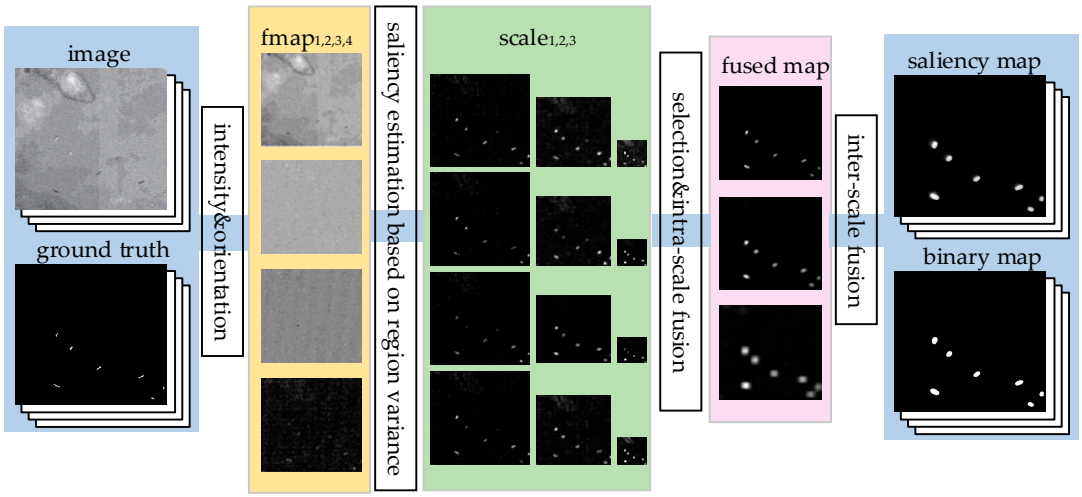


Figure 3. Proposed algorithm from input image (**left**) to saliency map (**right**). For visibility, feature maps, rarity maps, and saliency maps are normalized to [0,1].

Given an H × W image I, it is observed that the targets of interest usually have great intensity fluctuations and obvious edges [26,27], and we extract the pixel intensity and intensity derivatives to define the four-dimensional feature vector $\mathbf{f_k}$ for the kth pixel in I.

$$\mathbf{f_k} = \left[I_k \left| \frac{\partial I_k}{\partial x} \right| \left| \frac{\partial I_k}{\partial y} \right| \left| \frac{\partial I_k}{\partial x \partial y} \right| \right]^{\mathrm{T}} \tag{1}$$

where I_k is the intensity of the kth pixel, and the image derivatives are calculated through the filter $[-1\ 0\ 1]^T$. To obtain a maximum features decorrelation, we transform the feature map into four linearly uncorrelated maps by performing PCA decomposition (Principal Component Analysis) [28]. The resulting four feature maps after PCA transformation are denoted as $fmap^j$ and shown in the second column of Figure 3. The resulting feature vector of the kth pixel is redefined as $\left[f_k^1\ f_k^2\ f_k^3\ f_k^4 \right]^T$ where $k = 1, 2, \ldots, \mathrm{H} \times \mathrm{W}$. Non-overlapping patches with the size of $n \times n$ pixels are drawn from each feature map. A patch of each feature map is denoted as p_i^j, and the region variance of p_i^j can be expressed as

$$\mathrm{var}_i^j = \frac{1}{n^2 - 1} \sum_{\mathrm{k} \in p_i^j} \left(f_k^j - \overline{f_i^j} \right)^2 \tag{2}$$

where $\overline{f_i^j}$ denotes the mean value of the feature points in p_i^j. Then, the rarity value of patch is defined as

$$\mathrm{r_i^j} = 1 - \exp\left(\frac{-\mathrm{Var_i^j}}{\mathrm{Z}}\right) \tag{3}$$

where Z is a normalization factor equal to $\max_{i \in fmap^j} \left(\mathrm{var}_i^j \right)$. We obtain a set of four maps called rarity maps as shown in Figure 3. To integrate data information together, a selection algorithm is applied to the rarity maps. The first step is to compute for each rarity map an efficiency coefficient (EC_j), which is estimated by the information entropy. We can obtain the entropy value by considering the rarity map as a probability map. Image entropy can reflect the degree of difference in the gray values of pixels. According to the definition of entropy, the stronger the discriminative ability of the rarity map is, the smaller the entropy is. Then, the EC_j is defined as

$$EC_j = \frac{1}{H_j} = \frac{1}{-\sum_{i=1}^{n} p_i \log(p_i)} \tag{4}$$

where P_i represents the probability of gray level i in the image, and H_j denotes the information entropy of the rarity map j. When EC_j is greater, the rarity map is more efficient. We sort the rarity maps based on each map efficiency coefficient EC_j. $\mathbf{r_1}$ is the most efficient map, and $\mathbf{r_4}$ is the least efficient one. Finally, we eliminate $\mathbf{r_4}$, the fusion is then the sum of the rest maps weighted by EC_j:

$$\mathbf{s} = \sum_{j=1}^{3} EC_j \times \mathbf{r_j} \tag{5}$$

Note that the patch size $n \times n$ specifies the resolution of the saliency map and affects the performance of the algorithm. There are different outputs at different scales. The saliency map with small scales (small patch size $n \times n$) may tend to favor the boundaries rather than the entire body of a big ship target. In other words, it only focuses on the edges of targets and may introduce inner holes to the detection results. On the contrary, the small target boundary in the saliency map with large scales would be blurry. Furthermore, if the distance between the ship targets is too small, the ship candidates will be detected as a whole, the number of the ships cannot be distinguished. The situation becomes complex when different sizes of targets occur in the VRS images. To overcome this issue, we obtain multi-scale saliency

maps by changing the patch size $n \times n$ and perform the inter-scale fusion to produce a better saliency map. In this step, the multi-layer cellular automata [25] is introduced to integrate multi-scale saliency maps and improve the contrast between salient and non-salient patches. Pixels which have the same coordinates in different saliency maps are neighbors in the multi-layer cellular automata. It can enhance saliency consistency among similar regions by exploiting the intrinsic relationship in the neighborhood. Consider the scales $N = \{n_1, n_2, \cdots, n_M\}$. The saliency map at each scale is resized to the scale of the original image and denoted as $\{\mathbf{S_1}, \mathbf{S_2}, \cdots, \mathbf{S_M}\}$, and then the multi-layer cellular automata is expressed as:

$$\ln(\mathbf{S_m^{t+1}}) = \ln(\mathbf{S_m^t}) + \sum_{\substack{i=1 \\ i \neq m}}^{M} sign(\mathbf{S_i^t} - \gamma_i \cdot \mathbf{1}) \cdot \ln(\frac{\lambda}{1-\lambda}) \tag{6}$$

where $\mathbf{S_m^t} = \left[S_{m1}^t, \cdots, S_{mP}^t\right]^T$ denotes the saliency value of all pixels on the m-th map at time t, and P is the total number of pixels. The length of the vector $\mathbf{1} = [1, 1, \cdots, 1]^T$ is P. γ_i denotes the threshold of the i-th saliency map generated by Otsu [29]. We empirically set $\ln(\frac{\lambda}{1-\lambda}) = 0.5$ based on the analysis of [25]. After T time steps, the final saliency map $\mathbf{S}^T$ is defined as

$$\mathbf{S}^T = \frac{1}{M} \sum_{m=1}^{M} \mathbf{S_m^T} \tag{7}$$

In inter-scale fusion step, the number of time steps T is determined by the convergence time. We set T = 10. We still need to further investigate the appropriate set of scale parameters. The input images with increased sizes of the targets, from top to bottom, are shown in the first column of Figure 4. For small ship targets, edges are blurred with a large scale in accordance with the aforementioned discussion. When the patch size is 4×4, the middle areas of bigger ships have low salient values and only the edges are preserved. When the patch size goes up to 8×8 or 16×16, the performance gets better. To sum up, scale parameters that are too large or too small could cause poor performance. After several experiments, the scale parameter is fixed as N = {4, 8, 16} for better performance, and thus the number of scales is M = 3. Then single-scale saliency model can be easily extended to operate on multiple scales. Via performing the multi-scale saliency and selecting the appropriate set of scale parameters, our model is insensitive to the variation in target size. As shown in Figures 3 and 4, the output saliency map is now unique and the ship targets can be detected accurately even in a highly cluttered background.

2.2. Target Candidates Extraction

The final saliency map needs to be segmented to extract the candidate regions. In this step, we use the optimal threshold generated by the Otsu algorithm [29] to acquire the binary map. The optimal threshold is determined by the integration of the histogram and is selected automatically. The pixels with larger saliency values than the obtained threshold are defined as targets, while the rest of the pixels in the image are treated as backgrounds. Then, we define the smallest rectangle containing the connected region as ship candidates. There are two types of test images with complex backgrounds as shown in Figure 5. One set of images is covered by the clouds, and the other is disturbed by the islands. The first column presents the test VRS images, the second column presents their corresponding saliency maps, and the binary maps and prescreening stage results are shown in the third and the fourth column, respectively. As shown in Figure 5d, after saliency detection, segmentation, masking and extraction processing, the ship candidates are cut from the input image according to the location of each detected region in the binary image and marked with red boxes.

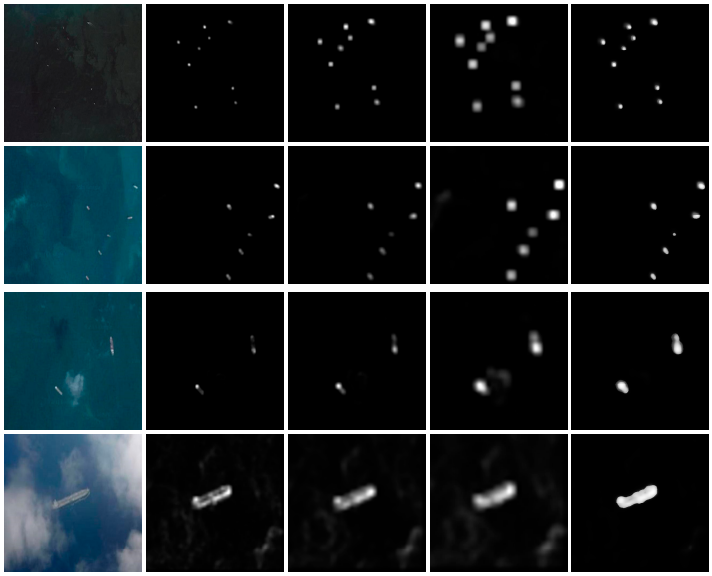


Figure 4. (**a**) A set of images containing the different size of targets; (**b**–**d**) Estimated saliency maps obtained at different scales (patch size parameter n = 4, 8, 16 respectively); (**e**) Final saliency map based on the multi-layer cellular automata described in the text. By performing multi-scale saliency, our model is robust in the presence of target size changes. In addition, it can pop out the targets accurately and suppress the cluttered background effectively.

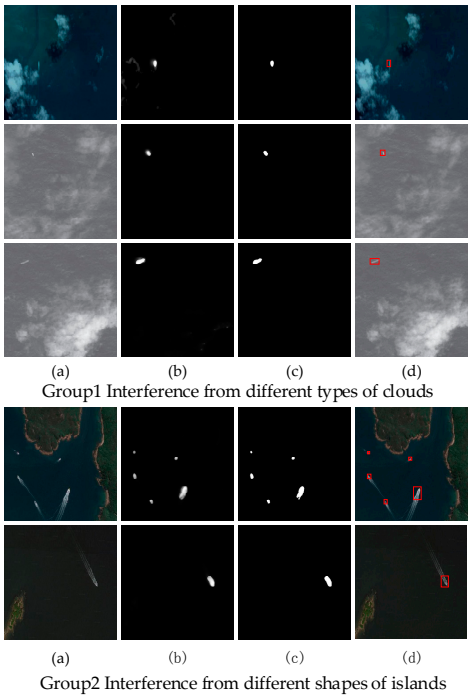


Figure 5. Candidate region detection results in complex backgrounds. Group 1 consists of images interfered by the cloud cover and group 2 consists of images interfered by islands. (**a**) Test image; (**b**) Saliency map generated by our visual attention model; (**c**) Image segmentation with Otsu algorithm; (**d**) Candidate region detection results.

3. Ship Discrimination

These attended regions acquired by visual saliency model could correspond to either ship objects in the image or false alarms. The discrimination process is performed to further remove pseudo-targets and confirm whether they are real ship targets. Therefore, a two-step solution is adopted to identify real ships, namely feature extraction and machine learning techniques. Feature extraction is conductive to the subsequent classification. An effective and robust descriptor characterizing the ship target is the key of the final discrimination. Considering the fact that arbitrary direction of the ship candidates brings difficulty to target detection, investigating a robust descriptor that allows the ship to be well recognized without the influence of direction is critically needed. In our approach, the rotation-invariant features describing the shape and the texture information of the targets are extracted and concatenated as a feature vector for subsequent classification. Finally, a trainable Gaussian support vector machine (SVM) classifier is performed to further remove the false alarms and maintain the real ship targets. The discrimination stage is described briefly as followed.

3.1. Rotation-Invariant Global Gradient Descriptor

Taking advantage of the symmetrical shape of the ships, the histogram of oriented gradients (HOG) descriptor [30] is introduced to distinguish between the ships and non-ship targets [15,16]. Note that the HOG descriptor usually samples cells on grids to describe objects, thus it is clearly not rotation-invariant and not applicable to directly describe targets because the direction of the ship in chips is arbitrary. To make up for the deficiency, Qi [14] performed the PCA transform to obtain the direction of the main axis and rotated the ship candidates to the vertical direction before extracting HOG feature from the ship candidates. In a similar manner, Xu [15] performed the segmentation algorithm and radon transform to estimate the ship target heading. Considering that the estimation of the principal axis direction is time-consuming and not always accurate enough, we introduce the radial gradient transform (RGT) [19] which can eliminate the computation of estimating an orientation to guarantee the rotation invariance. Moreover, the RGT descriptor, which was initially developed for real-time tracking, is faster compared with the other rotation-invariant descriptor [31,32].

The specific process of the RGT transform is shown in Figure 6. Two orthogonal basis vectors, r and t, denote the radial and tangential direction at a point p, and point c is the center of the chips. By projecting onto r and t, the gradient g is reformulated as $(g^T r)r + (g^T t)t$. The rotation matrix for some angle θ is denoted as R_θ. If we rotate the patch about its center by the angle θ, a new local coordinate system and gradient will be expressed as:

$$R_\theta p = p', R_\theta r = r', R_\theta t = t', R_\theta g = g' \tag{8}$$

and the radial gradient after the rotation can be expressed as $\left(g'^T r', g'^T t'\right)$. It is easy to verify that the coordinates of the gradient in the local frame are invariant to the rotation by:

$$\begin{aligned} \left(g'^T r', g'^T t'\right) &= \left((R_\theta g)^T R_\theta r, (R_\theta g)^T R_\theta t\right) \\ &= \left(g^T {R_\theta}^T R_\theta r, g^T {R_\theta}^T R_\theta t\right) \\ &= \left(g^T r, g^T t\right) \end{aligned} \tag{9}$$

Then, the radial gradient direction can be calculated by the formula:

$$\theta_{RGT} = \arctan \frac{g^T t}{g^T r} \tag{10}$$

and the magnitude is given by

$$g_{RGT} = \sqrt{\left(g^T r\right)^2 + \left(g^T t\right)^2} \tag{11}$$

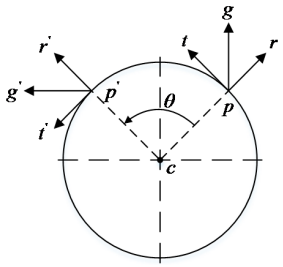
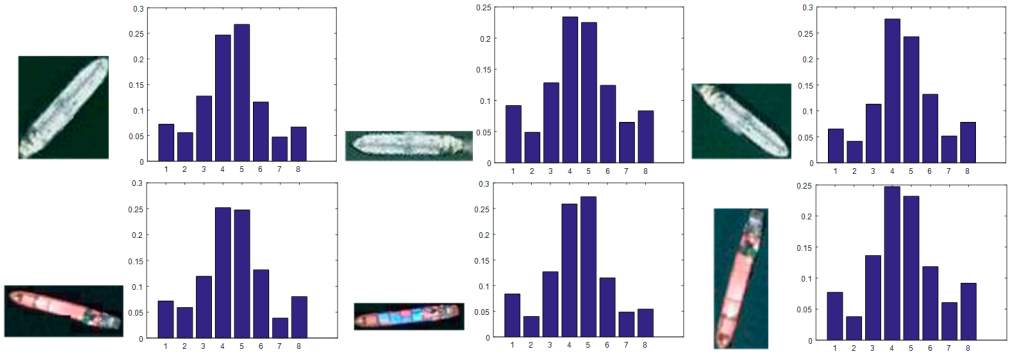
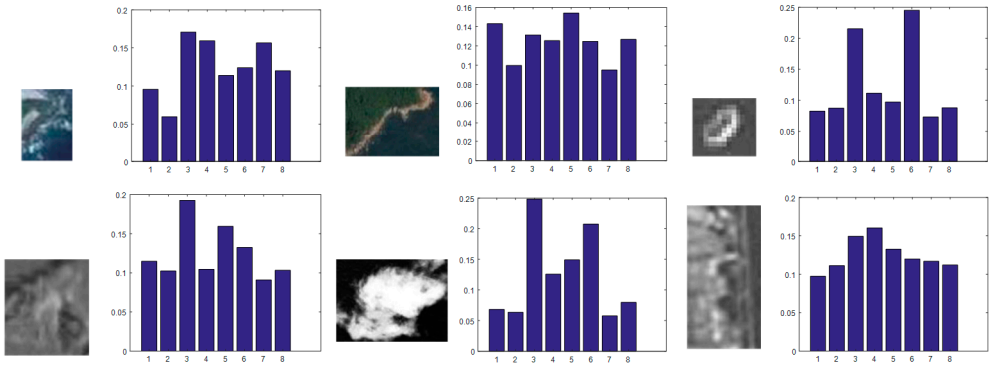


Figure 6. Illustration of radial gradient transform. For the given radial coordinate system (r, t), when the chip is rotated by θ, the projections of the gradient in (r, t) remain the same.

After obtaining the magnitudes and corresponding radial gradient orientations of the ship candidates, the gradient orientations are divided into eight specific bins in 0–360°. The angle in each bin is 45°, and we will get an eight-dimensional histogram from the gradient image by performing radial gradient transform. As shown in Figure 7, the gradient histogram of targets is basically unchanged even if the ship is rotated with the various angles. For a real ship target, bins 4 and 5 of the histogram have higher statistical quantized values in comparison to the other bins. Theoretically, the target chips share a similar distribution of gradient histograms, which is also illustrated in Figure 7. The obtained global gradient descriptor is robust to the variety of the sizes and rotations, and reliably grasps the shape information of targets. Finally, the magnitudes in bins 1–8 are denoted as a feature vector $f = [f_1, f_2, \cdots, f_8]$.



(a) The radial gradient histograms of ship targets



(b) The radial gradient histograms of false alarms

Figure 7. Radial gradient histogram statistics of the ship candidates. The x-coordinate denotes the eight orientation bins and the y-coordinate denotes the radial gradient statistic information in the gradient histogram. Ships with different orientations, sizes and textures share similar histogram distribution which is different from that of false alarms.

3.2. Region Covariance Descriptor

Second, a region covariance descriptor is applied to describe the texture features of ship candidates. The covariance matrix [33], which was initially proposed for texture classification and object detection, is introduced to characterize the ship targets. The region covariance descriptor is reviewed hereinafter. For a given image patch P, the W × H × d dimensional feature image extracted from P is denoted as F:

$$F(x,y) = \phi(P,x,y) \tag{12}$$

where ϕ denotes the function of the features, such as color, intensity, orientation, filter responses, spatial attributes, etc. Then, the image patch P is represented with the d × d covariance matrix C_p of the feature points.

$$C_p = \frac{1}{n-1}\sum_{i=1}^{n}(f_i - u)(f_i - u)^T \tag{13}$$

where $\{f_i\}_{i=1\ldots n}$ denote the d-dimensional feature points and u is the mean of all points inside P.

We use simple features, namely intensity, color, and the norm of the first and second-order derivatives of the intensity to define the d-dimensional (d = 7) pixel-level feature vector $f(x,y)$:

$$f(x,y) = \left[L(x,y), a(x,y), b(x,y), \frac{\partial P(x,y)}{\partial x}, \frac{\partial P(x,y)}{\partial y}, \frac{\partial^2 P(x,y)}{\partial^2 x}, \frac{\partial^2 P(x,y)}{\partial^2 y}\right] \tag{14}$$

with L, a, and b denoting the color of the pixel in Lab color space. The derivatives are calculated through the filters $[-1\ 0\ 1]^T$ and $[-1\ 2\ -1]^T$, and (x, y) denotes the location information. Hence, the covariance matrix C_p is computed as a 7 × 7 matrix. It has several advantages:

- It provides nonlinear integration of different features through modeling its correlations.
- Due to the low-dimensional representations of the patches, it captures local structures better than linear filters.
- It is insensitive to the large rotations and the illumination changes.

To use C_p as the ship descriptor, the matrix C_p needs to be mapped to a vector. Note that covariance matrices do not lie on the Euclidean space. It is infeasible to change d × d matrix into vector intuitively. To remedy this issue, Hong [20] proposed the sigma point descriptor which can transform covariance matrices on Euclidean vector space by using the Cholesky decomposition. After performing the Cholesky decomposition of C_p, $C_p = LL^T$, we can obtain L, which is a lower triangular matrix. Then the nonzero elements in matrix L can be changed into a $\left(d^2 + d\right)/2$ vector denoted as $f_2 = [L_1, L_2, \ldots, L_{28}]$. Finally, both f_1 and f_2 are concatenated as a feature vector for classification. For the sake of simplicity, we redefine the combined features as $f = [f_1, f_2, \ldots, f_{36}]$.

3.3. Gaussian SVM

The main aim of the classification is to discriminate the real ship targets from the ship candidates based on the obtained features f. The support vector machine (SVM) [21] can non-linearly map the input vector into a very high-dimension feature space. More importantly, the solution of SVM is globally optimal. Due to its high performance in many pattern recognition applications, the SVM is adopted in the discriminative stage. Given a training set of m observations:

$$D = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_m, y_m)\}, y_i \in \{-1, +1\}, \mathbf{x}_i \in R^d. \tag{15}$$

with $\mathbf{x}_i$ denoting the feature vector corresponding to the ith observation labeled, and y_i the input label belonging to -1 and 1, which denote non-ship and ship targets. For non-linear classification problems, to construct a separating hyperplane built in the feature space, the d-dimensional feature

vector $\mathbf{x}$ is first transformed into a D-dimensional feature vector by function ϕ: $\mathbf{x} \in R^d \mapsto \phi(\mathbf{x}) \in R^D$. Then, the sign of the function

$$f(\mathbf{x}) = \mathbf{w} \cdot \phi(\mathbf{x}) + b \tag{16}$$

is taken, where $\mathbf{w}$ and b are to-be-learned parameters, and the optimization problem becomes

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2 \tag{17}$$

subject to

$$y_i\left(w^T\phi(x_i) + b\right) \geq 1, i = 1, 2, \ldots, m \tag{18}$$

Then, the Lagrangian is computed to solve this convex quadratic programming problem and the corresponding dual problem is expressed as

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i\alpha_j y_i y_j \phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j) \tag{19}$$

$$s.t. \sum_{i=1}^{m} \alpha_i y_i = 0 \alpha_i \geq 0, i = 1, 2, \ldots, m. \tag{20}$$

where α_i denotes the Lagrange multiplier. Instead of the explicit computations on $\phi(x_i)^T\phi(x_j)$, the kernel trick is applied and the SVM model for function estimation yields

$$f(\mathbf{x}) = \sum_{i=1}^{m} \alpha_i y_i \kappa(\mathbf{x}, \mathbf{x}_i) + b \tag{21}$$

where $\kappa(\cdot,\cdot)$ is the kernel function. The kernel mapping technique plays an important role in classification performance. One can combine the prior knowledge of the problem at hand through constructing special kernel functions [21]. In our experiment, SVMs with linear, quadratic, cubic, and Gaussian kernels are tested. Finally, the Gaussian SVM is adopted to classifier the ships and non-ship targets. The Gaussian kernel function can be expressed as:

$$\kappa(x_i, x_j) = \exp(-\frac{\|x_i - x_j\|^2}{2\sigma^2}) \tag{22}$$

More details can be found in Section 4.3.

4. Experimental Results and Discussion

We conduct our experiments using a PC equipped with a 3 GHz CPU and 4-GB memory. Firstly, we compare the proposed saliency model both qualitatively and quantitatively with four state-of-art methods in different complex backgrounds (e.g., luminance fluctuation, cloud cover, fog, sea clutter, islands interference). We employ the receiver operating characteristic (ROC) area under the curve (AUC) metric to evaluate the candidate location prediction quantitatively. Secondly, the classification accuracy is adopted to measure the performance of SVMs with different kernel functions, we also compare our combined rotation-invariant feature with S-HOG feature (ship histogram of oriented gradient), single feature f_1 and f_2. Finally, the overall detection performance is compared to further demonstrate the effectiveness and robustness of the proposed scheme.

4.1. Data Set

All VRS images were collected from Google Earth and were captured under different weather conditions and various viewpoints, the dataset contains 338 ship targets for a total of 162 images of

size 512 × 512 pixels, the corresponding binary maps were manually labeled. The resolution of these images is about 1 m. Sample images of the dataset are listed in the left column of Figure 8.

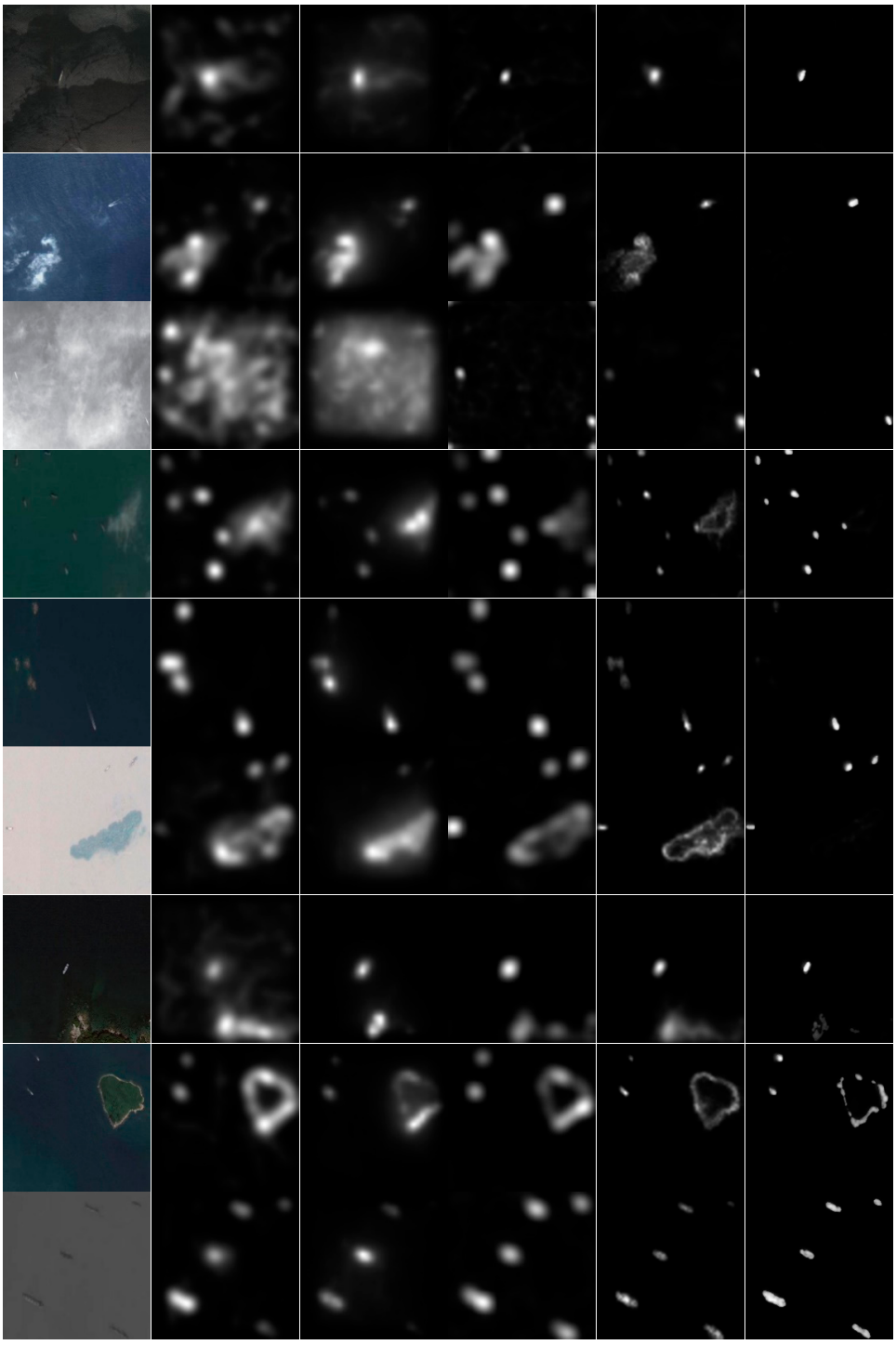


Figure 8. *Cont.*

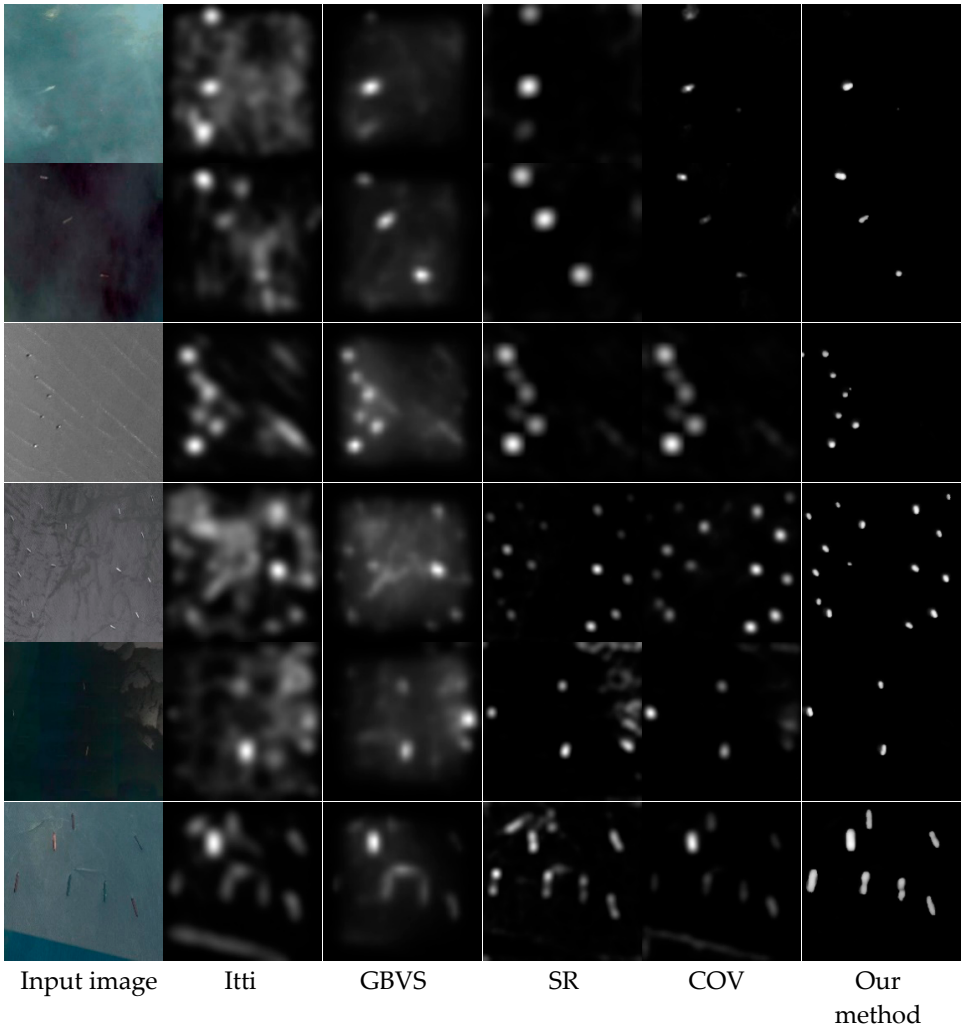


Figure 8. Examples of saliency maps with comparison to the 4 state-of-the-art methods. From left to right: input image, Itti [9], GBVS (Graph-Based Visual saliency) [10], SR (Spectral-Residual) [13], COV (Covariance saliency) [23], and our method. Our method outperforms other typical methods visually.

4.2. Comparison to the State-of-the-Art Saliency Model

Figure 8 presents the results of our saliency approach and other typical models including the Itti [9], GBVS (Graph-Based Visual saliency) [10], SR (Spectral-Residual) [13] and COV (Covariance saliency) [23] methods with respect to some sample images from our dataset. These images can be divided into several types based on the different complex backgrounds, such as thin and thick cloud cover, the interference of islands, fog, sea clutter, etc. The complicated backgrounds make every target detection task unique and challenging. Though it is difficult for all these methods to exactly extract the saliency regions in remote sensing images, our saliency model tends to be less distracted by the cluttered backgrounds in comparison to other methods.

As shown in Figure 8, our proposed saliency model achieves the best results of all saliency models visually both in terms of the accuracy and the integrity of object detection, and has the following advantages:

- Our model can distinguish different ship targets even when they are very close to each other.
- It can identify both large and small ships and highlight the entire ship target regions.
- It can suppress the interference from the complex backgrounds such as cloud, fog and sea clutter.

It is noted that the background suppression abilities of the Itti and GBVS model are weak, especially in the case of the cloud cover. Although the detection results are finer for the SR model,

this model is sensitive to the input image pixels. The COV model is effective for suppressing complex backgrounds, but it is time-consuming and produces more false alarms compared to our model. Overall the proposed saliency model is superior to other typical models and can obtain more accurate shapes and highlight the whole target regions.

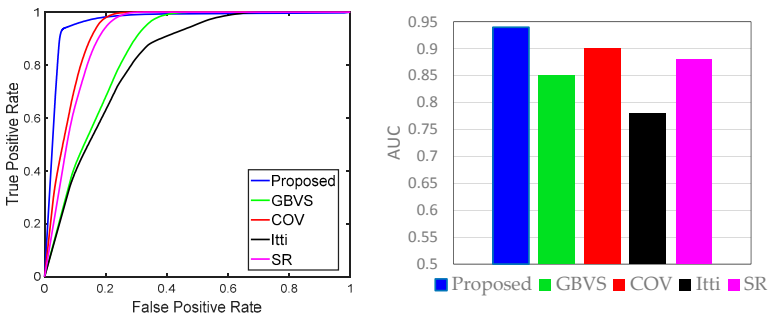
In addition to visual comparisons of saliency maps, we employ the ROC-AUC metric to quantitatively evaluate the performance of the proposed method. Using this metric, the pixels with larger saliency values than a threshold are treated as targets, while the rest of the pixels in the image are treated as backgrounds. Binary maps are used as ground truth. An ROC graph can be drawn by varying the threshold in which the true positive rate (TPR) and the false positive rate (FPR) are plotted on the Y axis and X axis, respectively. The TPR and FPR are expressed as

$$TPR = \frac{tp}{tp + fn} \tag{23}$$

$$FPR = \frac{fp}{fp + tn} \tag{24}$$

where tp is the number of true positives, fp is the number of false positives, tn is the number of true negatives, and fn is the number of false negatives.

The performance in terms of ROC-AUC metric is measured and the results are shown in Figure 9a,b respectively. The ROC curve in the upper-left corner of the graph is best. It can be observed that the proposed saliency model has the highest ROC-AUC performance and outperforms all the other methods in consideration.



(**a**) ROC curve (**b**) AUC

Figure 9. The performance of the saliency models under consideration in terms of receiver operating characteristic area under the curve (ROC-AUC) metric.

We also evaluate the performance of the proposed saliency model in terms of speed with reference to the other methods mentioned above. Table 1 compares the average time taken by each method. Note that SR, COV, and the proposed method are programmed in Matlab, while the codes with regard to Itti and GBVS are quasi Matlab codes which call C++ functions for saving the running time. Nevertheless, a relative overview of the run-time performance of the considered methods is given. It can be observed that COV model has the defect of long running time in despite of good performance. Due to mixed-language programming, the costs of performing Itti and GBVS are relatively low. SR model has the shortest running time because of small calculation efforts. The time complexity of our method is lower than that of other spatial saliency models.

Table 1. The computational run-time(s) of various saliency models under consideration.

Method	Proposed	GBVS	COV	Itti	SR
Time(s)	0.6	1.1	19	0.9	0.08
Code	M	M&C++	M	M&C++	M

4.3. Discrimination Results

There are 543 ship candidates obtained by performing the candidate extraction mechanism. The size of the ship candidate is typically in a range from 11 × 18 to 100 × 92. They are manually classified into 325 ship chips and 218 non-ship chips. They are used to verify the performance of the hierarchical feature extraction as well as the classification approach. Some examples of the extracted ship candidates are shown in Figure 10. Group A and B show the samples of targets and the false alarms, respectively. We randomly select two-thirds of the ship chips and the non-ship chips as the training set. The test set consists of the left chips.

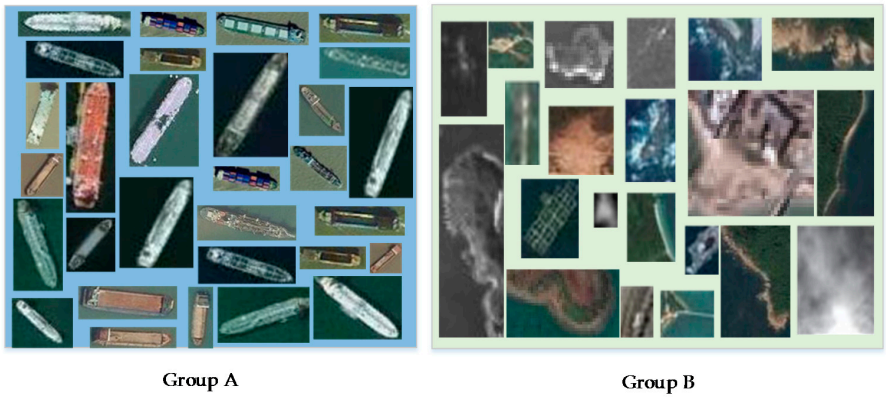


Figure 10. Examples of the extracted ship candidates. Group A corresponds to the ship targets while group B consists of false alarms.

To validate the effectiveness of our combined feature, S-HOG feature (ship histogram of oriented gradient), radial gradient feature, sigma set feature, and the combined feature are separately combined with the classification learner to perform the discrimination. SVM can solve the small sample, nonlinear classification problem and has good generalization performance, which is suitable for the extracted data. Selection of kernel function is a pivotal factor which decides classification accuracy. Based on the above factors, the four different feature sets mentioned above are compared using the SVMs with various kernel functions, namely the linear, quadratic, cubic, and Gaussian functions. The classification accuracy of each method is calculated as:

$$\text{Accuracy} = \frac{\text{Number of correctly classified samples}}{\text{Number of tested samples}} \times 100\% \tag{25}$$

The parameter for the SVM-based classifier is determined by adopting 5-fold cross-validation. The classification accuracy of each feature is shown in Figure 11.

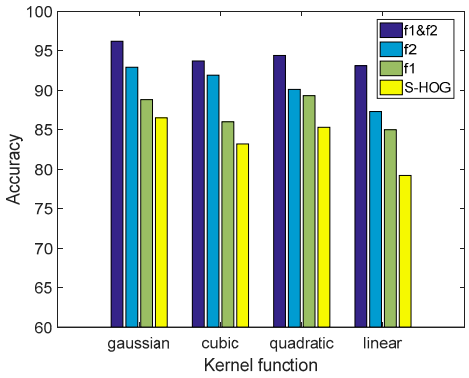


Figure 11. Classification performance of four feature sets with different kernel function.

As can be seen in the comparison shown in Figure 11, for any given kernel function, the ship's classification accuracy based on the combined feature sets is higher than the accuracy based on the single feature set. Note that S-HOG feature has the worst accuracy result; this may be related to low accuracy when estimating the principal axis direction. In addition, for the combined feature f, the accuracy of the SVM with Gaussian kernel is 96.1% which is the highest level of accuracy compared to the others. Therefore, the combined feature set and Gaussian SVM are adopted in the following experiments.

4.4. Comparison of Overall Detection Performances

Finally, we compare our overall detection method with two typical methods. The evaluation criteria are defined as

$$\text{Accuracy} = \frac{\text{Number of correctly detected ships}}{\text{Number of real ships}} \times 100\% \tag{26}$$

$$\text{False ratio} = \frac{\text{Number of detected false alarms}}{\text{Number of detected candidates}} \times 100\% \tag{27}$$

The detection results are listed in Table 2.

Table 2. Target detection results in terms of the accuracy and the false ratio.

Method	Accuracy	False Ratio
method [2]	85%	10%
method [15]	90%	9%
ours	94%	4%

As can be seen from Table 2, our detection model can obtain higher accuracy and lower false ratio than the other two methods. Note that the method [2] has the worst performance. This is because method [2] generates the candidate regions by image segmentation and uses the simple shape feature to distinguish between the ships and non-ship targets. While our model and method [15] extract the ship candidates by using the visual attention mechanism, this can obtain few false alarms and low missing rate. Besides, the improved HOG feature can describe target shape information efficiently. In addition, compared to method [15], we extract not only shape features but also texture features. This is beneficial to further removing false alarms, which have similar shapes to the targets. Through the analysis above, it can be concluded that our model is effective for eliminating the false alarms and preserving the real targets and outperforms the other ship detection model considered.

With regard to the time consumption of the overall detection algorithm, compared to slide window algorithm, our saliency detection model can greatly decrease the detection time. The time consumption of saliency model detection is 0.6 s, and the average time consumption of discrimination stage is 0.2 s, which basically meets the needs of the near-real-time tasks.

5. Conclusions

In this paper, we have proposed a hierarchical model to tackle the problem of ship detection in a complex and changing background environment based on optical remote sensing data. The scheme consists of prescreening and discrimination stages. First, a fast and efficient multi-scale saliency model based on region statistical characteristics is performed to locate candidate regions. Through performing saliency detection, our model effectively reduces missed detection and false detection. Second, from a given ship candidate, we extract the combined rotation-invariant feature which offers a more powerful descriptor to capture the shape and texture information of the object. Finally, a trainable Gaussian SVM is employed as the discriminator. Our overall detection model achieves the best performance of 94% in terms of accuracy and 4% in terms of false ratio, outperforming the other typical ship detection model. Experiments on the optical remote sensing data have demonstrated the effectiveness and robustness of the proposed model.

Our future work will focus on two aspects. First, we will build a large dataset including thousands of optical remote sensing images to make sure that the input data of classifier is sufficient. It thus can improve the object detection performance further. Second, more effective features may be further explored and feature selection will be considered. Moreover, the better discriminator will be investigated.

Acknowledgments: The work was supported by the National Defense Pre-Research Foundation of China (Grant No. 402040203) and the Programs Foundation of Key Laboratory of Airborne Optical Imaging and Measurement, Chinese Academy of Science (Grant No. y3hc1sr141).

Author Contributions: Chao Dong and Jinghong Liu designed the proposed detection model. Chao Dong and Fang Xu designed the experiments. Chao Dong drafted the manuscript. Fang Xu edited the manuscript. Jinghong Liu provided guidance to the project, reviewed the manuscript, and obtained funding to support this research.

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

References

1. Mirghasemi, S.; Yazdi, H.S.; Lotfizad, M. A target-based color space for sea target detection. *Appl. Intell.* **2012**, *36*, 960–978. [CrossRef]
2. Zhu, C.R.; Zhou, H.; Wang, R.S.; 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]
3. Yang, G.; Li, B.; Ji, S.F.; Gao, F.; Xu, Q.Z. Ship Detection From Optical Satellite Images Based on Sea Surface Analysis. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 641–645. [CrossRef]
4. 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]
5. Shi, Z.W.; Yu, X.R.; Jiang, Z.G.; 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.
6. Han, J.; Zhang, D.; Cheng, G.; Guo, L.; Ren, J. Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 3325–3337. [CrossRef]
7. Zou, Z.X.; Shi, Z.W. Ship Detection in Spaceborne Optical Image with SVD Networks. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 5832–5845. [CrossRef]
8. Tang, J.X.; Deng, C.W.; Huang, G.B.; Zhao, B.J. Compressed-Domain Ship Detection on Spaceborne Optical Image Using Deep Neural Network and Extreme Learning Machine. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1174–1185. [CrossRef]
9. Lin, H.N.; Shi, Z.W.; Zou, Z.X. Maritime Semantic Labeling of Optical Remote Sensing Images with Multi-Scale Fully Convolutional Network. *Remote Sens.* **2017**, *9*, 480. [CrossRef]
10. 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]
11. Harel, J.; Koch, C.; Perona, P. Graph-based visual saliency. *Adv. Neural Inf. Process. Syst.* **2006**, *19*, 545–552.
12. Bi, F.K.; Zhu, B.C.; Gao, L.N.; Bian, M.M. A Visual Search Inspired Computational Model for Ship Detection in Optical Satellite Images. *IEEE Geosci. Remote Sens. Lett.* **2012**, *9*, 749–753.
13. Guo, W.; Xia, X.; Wang, X. A remote sensing ship recognition method based on dynamic probability generative model. *Expert Syst. Appl.* **2014**, *41*, 6446–6458. [CrossRef]
14. Hou, X.D.; Zhang, L.Q. Saliency Detection: A Spectral Residual Approach. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; pp. 2280–2287.
15. Qi, S.X.; Ma, J.; Lin, J.; Li, Y.S.; Tian, J.W. Unsupervised Ship Detection Based on Saliency and S-HOG Descriptor from Optical Satellite Images. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1451–1455.
16. Xu, F.; Liu, J.H.; Sun, M.C.; Zeng, D.D.; Wang, X. A Hierarchical Maritime Target Detection Method for Optical Remote Sensing Imagery. *Remote Sens.* **2017**, *9*, 280. [CrossRef]
17. Achanta, R.; Hemami, S.; Estrada, F.; Suesstrunk, S. Frequency-Tuned Salient Region Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami Beach, FL, USA, 20–25 June 2009; pp. 1597–1604.
18. Wang, H.L.; Zhu, M.; Lin, C.B.; Chen, D.B. Ship detection in optical remote sensing image based on visual saliency and AdaBoost classifier. *Optoelectron. Lett.* **2017**, *13*, 151–155. [CrossRef]

19. Takacs, G.; Chandrasekhar, V.; Tsai, S.; Chen, D.; Grzeszczuk, R.; Girod, B. Unified Real-Time Tracking and Recognition with Rotation-Invariant Fast Features. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 934–941.
20. Hong, X.P.; Chang, H.; Shan, S.G.; Chen, X.L.; Gao, W. Sigma Set: A Small Second Order Statistical Region Descriptor. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami Beach, FL, USA, 20–25 June 2009; pp. 1802–1809.
21. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
22. Wang, X.L.; Chen, C.X. Adaptive ship detection in SAR images using variance WIE-based method. *Signal Image Video Process.* **2016**, *10*, 1219–1224. [CrossRef]
23. Cao, Z.J.; Ge, Y.C.; Feng, J.L. Fast target detection method for high-resolution SAR images based on variance weighted information entropy. *EURASIP J. Adv. Signal Process.* **2014**, *1*, 45. [CrossRef]
24. Leng, X.G.; Ji, K.F.; Zhou, S.L.; Xing, X.W.; Zou, H.X. An Adaptive Ship Detection Scheme for Spaceborne SAR Imagery. *Sensors* **2016**, *16*, 1345. [CrossRef] [PubMed]
25. Qin, Y.; Lu, H.C.; Xu, Y.Q.; Wang, H. Saliency Detection via Cellular Automata. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 110–119.
26. Wang, Y.H.; Liu, H.W. A Hierarchical Ship Detection Scheme for High-Resolution SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 4173–4184. [CrossRef]
27. Cheng, G.; Han, J.W. A survey on object detection in optical remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2016**, *117*, 11–28. [CrossRef]
28. Wold, S.; Esbensen, K.; Geladi, P. Principal Component Analysis. *Chemometr. Intell. Lab. Syst.* **1987**, *2*, 37–52. [CrossRef]
29. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cyber.* **1979**, *9*, 62–66. [CrossRef]
30. Dalal, N.; Triggs, B. Histograms of Oriented Gradients for Human Detection. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 20–25 June 2005; pp. 886–893.
31. Liu, K.; Skibbe, H.; Schmidt, T.; Blein, T.; Palme, K.; Brox, T.; Ronneberger, O. Rotation-Invariant HOG Descriptors Using Fourier Analysis in Polar and Spherical Coordinates. *Int. J. Comput. Vis.* **2014**, *106*, 342–364. [CrossRef]
32. Cheng, G.; Zhou, P.C.; Han, J.W. Learning Rotation-Invariant Convolutional Neural Networks for Object Detection in VHR Optical Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 7405–7415. [CrossRef]
33. Tuzel, O.; Porikli, F.; Meer, P. Region Covariance: A Fast Descriptor for Detection and Classification. In Proceedings of the European Conference on Computer Vision, Graz, Austria, 7–13 May 2006; pp. 589–600.