



Gao, Fei and You, Jialing and Wang, Jun and Sun, Jinping and Yang, Erfu and Zhou, Huiyu (2017) A novel target detection method for SAR images based on shadow proposal and saliency analysis. Neurocomputing, 267. pp. 220-231. ISSN 0925-2312 , <http://dx.doi.org/10.1016/j.neucom.2017.06.004>

This version is available at <https://strathprints.strath.ac.uk/61126/>

Strathprints is designed to allow users to access the research output of the University of Strathclyde. Unless otherwise explicitly stated on the manuscript, Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Please check the manuscript for details of any other licences that may have been applied. You may not engage in further distribution of the material for any profitmaking activities or any commercial gain. You may freely distribute both the url (<https://strathprints.strath.ac.uk/>) and the content of this paper for research or private study, educational, or not-for-profit purposes without prior permission or charge.

Any correspondence concerning this service should be sent to the Strathprints administrator: strathprints@strath.ac.uk

A novel target detection method for SAR images based on shadow proposal and saliency analysis

Gao Fei^a, Jialing You^a, Wang Jun^{a,*}, Jinping Sun^a, Erfu Yang^b, Huiyu Zhou^c

^aElectronic Information Engineering, Beihang University, Beijing 100191, China

^bSpace Mechatronic Systems Technology Laboratory, Department of Design, Manufacture and Engineering Management, University of Strathclyde, Glasgow G1 1XJ, UK

^cSchool of Electronics, Electrical Engineering and Computer Science, Queens University Belfast BT7 1NN, Northern Ireland, UK

Abstract

Conventional synthetic aperture radar (SAR) based target detection methods generally use high intensity pixels in the pre-screening stage while ignoring shadow information. Furthermore, they cannot accurately extract the target area and also have poor performance in cluttered environments. To solve this problem, a novel SAR target detection method which combines shadow proposal and saliency analysis is presented in this paper. The detection process is divided into shadow proposal, saliency detection and One-Class Support Vector Machine (OC-SVM) screening stages. In the shadow proposal stage, localizing targets is performed first with the detected shadow regions to generate proposal chips that may contain potential targets. Then saliency detection is conducted to extract salient regions of the proposal chips using local spatial autocorrelation and significance tests. Afterwards, in the last stage, the OC-SVM is employed to identify the real targets from the salient regions. Experimental results show that the proposed saliency detection method possesses higher detection accuracy than several state of the art methods on SAR images. Furthermore, the proposed SAR target detection method is demonstrated to be robust under different imaging environments.

*Corresponding author at: Electronic Information Engineering, Beihang University, Beijing 100191, China

Email address: wangj203@buaa.edu.cn (Wang Jun)

Keywords: synthetic aperture radar (SAR) target detection, shadow, saliency detection, local spatial autocorrelation, visual features, One-Class SVM (OC-SVM)

1. Introduction

Existing image processing methods cannot meet the demand of the rapid growth of SAR imaging. Effectively addressing the current challenges and enhancing the performance of the state-of-the-art techniques in SAR image interpretation is urgently needed. Target detection has important significance in both civil and military applications. Constant false alarm rate (CFAR) detection forms the basis of conventional SAR target detection methods [1]. Two-parameter CFAR [2] proposed by Lincoln laboratory of MIT marks the progress of SAR target detection. Subsequently, many CFAR detection algorithms from different perspectives emerged, such as Ordered Statistic CFAR (OS-CFAR) [3] and Variability Index CFAR (VI-CFAR) [4]. Nevertheless, due to the dependence of the statistical modeling on the distribution of clutters, CFAR algorithms lack generalisation. In addition, because of ignoring the difference between targets and distracters, CFAR algorithms may introduce a plenty of false alarms in complex environments.

Human visual system can detect and identify SAR targets relatively quickly and accurately in a complicated scene according to their visual features. Inspired by this observation, we aim to establish a novel SAR target detection scheme simulating the human visual attention process. Generally, a visual attention model is constructed adopting object proposals (OP) [5] and saliency detection. OP aims to cover as many objects of interest as possible with as few windows as possible. Compared with traditional detection methods using exhaustive sliding windows to search across images, the application of OP methods in the preprocessing stage allows us to quickly locate the target objects, which greatly improves the computational efficiency. The idea of OP is consistent with human cognitive behaviors where humans scan objects before discriminating them [6].

Approaches to generate object proposals can be divided into two categories. One is grouping proposal methods [7–11], in which an image is segmented into superpixels firstly, and then specific strategies are used to aggregate superpixels into objects. The other one contains window scoring proposal methods, in which candidate windows are directly generated according to the color, edge or other information of the image, and then candidate windows are scored on the basis of how likely it is to contain an object and low scoring windows are filtered out [5, 12–15]. The aforementioned OP methods have achieved favorable experimental results on optical images, while SAR images have some defects such as lack of color information, low resolution, and obscure edge features. Therefore, OP methods are very difficult to obtain satisfactory results when directly applied onto SAR images. Moreover, OP methods can only locate objects roughly in images.

Unlike OP, saliency detection is designed to detect a specific area stimulating the human visual system. Saliency detection can also be divided into two categories: bottom-up (BU) saliency based on scene-driven learning and top-down (TD) saliency based on expectation-driven learning. BU saliency detection plays a dominant role in the building of a visual attention model because of advantages, e.g., fast, involuntary, stimulus-driven, and easily modeled. Depending how it is calculated, BU saliency detection can be classified into three categories: feature based, probability based, and transformation based respectively. In 1998, Itti et al. [16], proposed the first computable bio-inspired saliency detection model, which combines three feature maps including color, intensity, and orientation at different scales. Harel et al. [17] proposed Graph-based visual saliency (GBVS), which uses the characteristics of Markov random field to construct two-dimensional Markov chains to calculate the saliency map on the basis of the ITTI model. Both the ITTI and GBVS models are based on features. The approaches developed by Zhang et al. [18] and Neil Bruce et al. [19], which are respectively based on Bayesian rules and Shannon’s self-information measures, use probability based models. The Spectral Residual [20] and the Phase Spectrum of Quaternion Fourier Transform [21], both of

which are the representatives of the transformation model, analyze and calculate saliency maps using the amplitude or phase spectrum of an image. Besides, a number of newly representative saliency models have emerged [22–25], such as sparse coding based [26, 27], manifold ranking based [28], background hypothesis based [29], etc. Saliency detection has been applied in remote sensing fields [30–32]. However, severe speckle noise and complex environments may lead to poor system performance when we extract salient regions of SAR images.

We here describe three cues that have been often used in SAR image analysis. First, the SAR shadow is a universal property when imaging. In SAR images, shadow regions provide us a strong cue to help target localization even if the target is badly corrupted. Combining shadow detection with a pre-screening phase may enhance the robustness and improve the system performance [33, 34]. Second, it is a fact that the radar cross section (RCS) of a man-made target is higher than that of the background, and that the pixels inside the compact object have stronger spatial autocorrelation than the pixels scattering in inner and outer areas. Such properties can catch the attention of the human visual system. Third, learning is one of the highly integrated skills of a biological central nervous system [35–37]. Target detection is generally considered as a binary classification problem [38], and it needs comprehensive background samples to train a classifier so that it is properly working. One-class SVM (OC-SVM) [39] could be a promising choice to cope with this problem.

Motivated by the three cues presented above, we here introduce a novel method for SAR target detection. Our method consists of a shadow proposal process, a brand new saliency detection process and a one-class classification process. More details are followed: 1) The shadow proposal stage: potential target chips are proposed after we have determined the spatial relationship between shadows and objects using a Mean Shift algorithm [40]. 2) The salient region extraction stage: firstly we define the saliency of SAR images, and then extract salient regions of the proposal chips using G statistics [41, 42] and significance tests (GSST). 3) The OC-SVM screening stage: the OC-SVM is introduced to screen the real targets, where only target chips are used as training samples.

During the training and testing stages, we use the visual features of the salient
 90 regions of the training and testing samples.

The remainder of this paper is structured as follows: section 2 presents the saliency detection method GSST. The proposed SAR target detection method is detailed in section 3. Section 4 reports the experimental results. Conclusion is given in the last section.

95 2. Saliency Detection by G Statistics and Significance Tests

Vehicle targets and other man-made objects have higher RCS than their background. A SAR image is of high intensity pixels aggregated together whilst compact objects are with higher brightness and sharper contours. There is a strong correlation between pixels within an object, leading to clear saliency
 100 that can be identified by the human visual system. Here, we define the saliency of SAR images as how dense high intensity pixels aggregate together. In this section, we attempt to develop a method to detect saliency. Global Moran's I statistics [43] were proposed to measure spatial autocorrelations, where there are three kinds of spatial distribution patterns: clustered, dispersed and random
 105 patterns, as shown in Figure 1. Moran's I is often used to analyze the distribution pattern of population, economy and resources in a specific geographical region. It is an overall evaluation of the cluster degree of the elements in the studied region, but it cannot determine the attributive characters of the clustered elements, and cannot determine the specific location of the clusters. Thus
 110 important local information is missing if we use Moran's I.

Getis et al. [42] proposed statistics G to measure the local spatial autocorrelation, which determines the hot or cold spots, and a specific area can be detected by significance tests. The idea of using statistics G to measure the degree of local spatial autocorrelations coincides with the saliency detection
 115 described above. Therefore, we here propose a new saliency detection method based on statistics G and significance tests (GSST). Statistics G is defined as

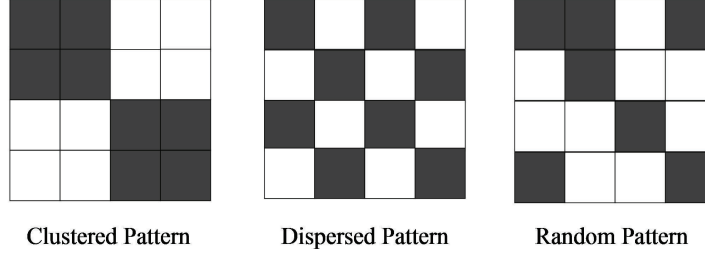


Figure 1: Three types of spatial distribution patterns.

follows:

$$G_i(d) = \frac{\sum_{j=1}^n w_{ij}(d)x_j}{\sum_{j=1}^n x_j}, j \neq i \quad (1)$$

Eq.(1) is used to calculate the correlation between the pixel i and the other pixels in a circular neighborhood of d , where d is the radius, and $w_{ij}(d)$ is the weight between pixels i and j . The shorter the distance is, the larger the correlation is, and the reciprocal of the distance is used here to compute the weights, namely $w_{ij} = \frac{1}{\|d_{ij}\|_2}$.

There are n pixels in the studied region. We fix the value for the center pixel i and take into account the set of $(n-1)!$ random permutations of the remaining values in space. Under the null hypothesis of spatial independence, these permutations are equally likely. Assuming X_j to be the random variable which describes the value assigned to pixel j , then we have $P(X_j = x_r) = \frac{1}{n-1}, r \neq i$. Hence the theoretical expectation and variance of $G_i(d)$ can be calculated as:

$$E(G_i(d)) = \frac{W_i}{n-1} \quad (2)$$

$$\text{Var}(G_i(d)) = \frac{W_i(n-1-W_i)}{(n-1)^2(n-2)} \left[\frac{s(i)}{\bar{x}(i)} \right]^2 \quad (3)$$

where $W_i = \sum_{j=1}^n w_{ij}(d), j \neq i$, $\bar{x}(i) = \frac{\sum_{j=1}^n x_j}{n-1}, j \neq i$ and $s^2(i) = \frac{\sum_{j=1}^n x_j^2}{n-1} - [\bar{x}(i)]^2$. x_j represents the intensity of pixel j . When d is large enough, the

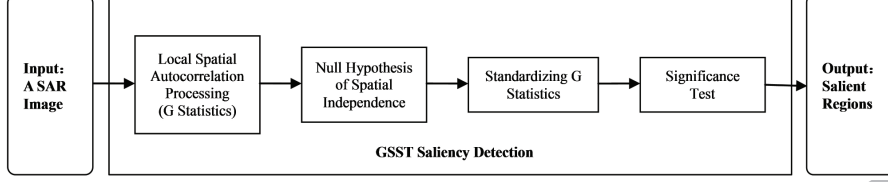


Figure 2: Flowchart of the proposed GSST saliency detection method.

permutation distribution of G_i under the null hypothesis approaches normality.

Thus we standardize statistics G as statistics Z , written as follows:

$$\begin{aligned}
 Z_i(d) &= \frac{G_i(d) - E(G_i(d))}{\sqrt{\text{Var}(G_i(d))}} \\
 &= \frac{\sum_{j=1}^n w_{ij}(d)x_j - W_i \bar{x}(i)}{s(i) \{[(n-1)S_{1i} - W_i^2]/(n-2)\}^{\frac{1}{2}}}, j \neq i
 \end{aligned} \tag{4}$$

where $S_{1i} = \sum_{j=1}^n w_{ij}^2, j \neq i$.

135 Then we apply hypothesis tests to the spatial independence. Let $Z_i(d)$ represent the test statistics. We assume the significance level to be α , and its corresponding quantile of the standard normal distribution is denoted as $z_{\alpha/2}$. The null hypothesis is refused when $|Z_i(d) \geq z_{\alpha/2}|$, that is, the spatial autocorrelation of pixel i exists, especially when:

$$Z_i(d) \geq z_{\alpha/2} \tag{5}$$

140 It means a large number of the neighboring pixels (values larger than the mean x_j) fall within d of pixel i . In other words, there is a cluster of pixels which have relatively high intensity. Applying the above analysis to each pixel, we can separate several salient regions of a SAR image. The main procedures of the GSST saliency detection are shown in Figure 2.

145 3. Proposed Method

A direct operation on an original SAR image using the GSST is time consuming and may cause incorrect detection under the low signal-clutter-ratio (SCR)

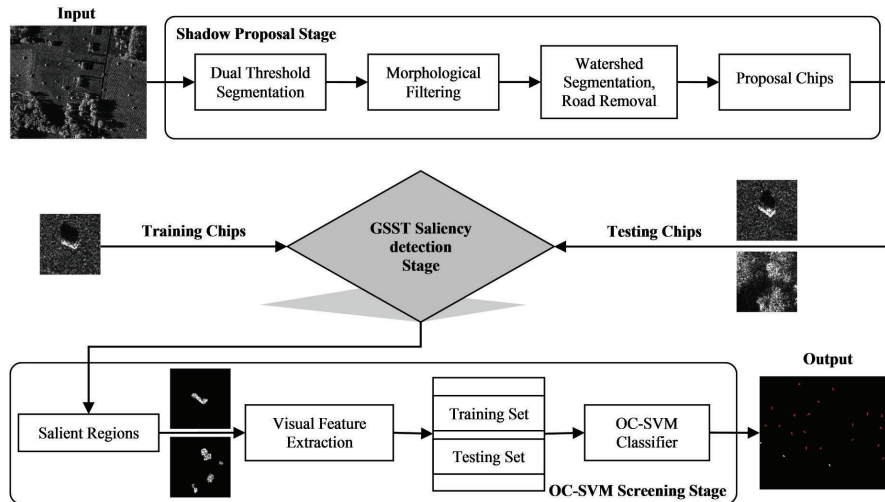


Figure 3: Flowchart of the proposed SAR target detection method.

conditions, and only relying on salient regions cannot guarantee to discriminate the real targets from clutters. So a sophisticate analysis is required. In this section, three stages for SAR target detection will be described in detail. Our algorithm is outlined as follows: the shadow proposal stage provides potential chips related to possible targets; the GSST saliency detection stage supplies salient regions of the potential chips; the OC-SVM screening stage presents real targets extracted from the detected salient regions. The flowchart of our method is shown in Figure 3.

3.1. Shadow proposal stage

There is remarkable difference between the imaging mechanisms of SAR and optical images, for example, shadow formation characteristics. SAR shadow can provide location cues for it is always located downrange from an object. The shadow proposal stage is to outline the approximate area of an image shadow, but not for precise detection. The proposal chips can be generated using the spatial relationship of objects and shadow regions.

3.1.1. Shadow detection

The shadow detection stage consists of three steps: thresholding, morpho-
 165 logical filtering, and watershed segmentation & road removal.

(i) *Thresholding.* A SAR image is composed of three parts: target, shadow, and background, so a single threshold for shadow segmentation will end up with poor outcomes. Thus we select the dual threshold Otsu method to segment the shadow regions for better accuracy. A SAR image is divided into $0 \sim m$ classes according to the intensities of the pixels. Assuming T_1 and T_2 are two thresholds, 170 the image is classified into three classes: $C_0 = \{0 \sim T_1\}$, $C_1 = \{T_1 + 1 \sim T_2\}$, $C_2 = \{T_2 + 1 \sim m\}$. T_1 and T_2 that maximise Eq.(6) are selected as the desired dual thresholds:

$$g(T_1, T_2) = \underset{0 < T_1 < T_2 < m}{\text{Arg max}} [\omega_0(\mu - \mu_0)^2 + \omega_1(\mu - \mu_1)^2 + \omega_2(\mu - \mu_2)^2] \quad (6)$$

where μ is the average intensity of the whole image. μ_0 , μ_1 , and μ_2 are the 175 average intensities of C_0 , C_1 , and C_2 , respectively. ω_0 , ω_1 , and ω_2 indicate the likelihoods of C_0 , C_1 , and C_2 .

(ii) *Morphological filtering.* After thresholding, we obtain the desired category C_0 . The open and close operations of morphological filtering are implemented in order to reduce noise and wrong shadows. Then the connected component 180 labeling is executed on the remaining regions, resulting in preliminary shadow regions.

(iii) *Watershed segmentation & road removal.* Because of the low RCS of roads or rivers, they are contained in the preliminary shadow regions as well. Moreover, the shadow of the target may be closely connected to roads or other shadow 185 regions. Considering these circumstances, we use a watershed scheme to segment the target shadow staying with other regions, and the Hough transform is performed to remove roads and rivers by detecting straight edges of the preliminary shadow regions. The final shadow regions will be determined after we have

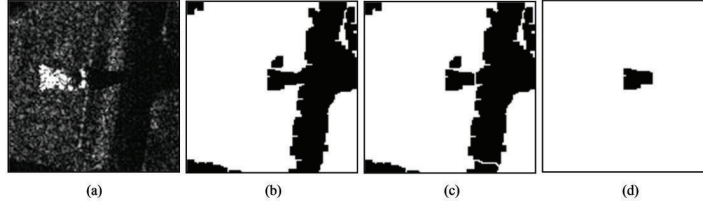


Figure 4: (a) Target embedded in the road; (b) Preliminary shadow regions; (c) Shadow regions after watershed segmentation; (d) Shadow regions after road removal.

removed the less possible regions according to spatial constraints. To illustrate the above process, Figure 4 (a)~(d) show the detection results step by step.

3.1.2. Relationship of spatial location

Objects with shadows such as trees, buildings and vehicles generally have high RCS, so the radar line of sight can be inferred according to the distribution of the intensities of the surrounding areas around the shadow regions. In other words, we can derive the spatial relationship of objects and shadow regions according to shadow characteristics. Therefore we need to study the intensity distribution around each shadow region. To make sure the whole or the most part of the object is included in the studied region, we outline a square area, whose side length R equals three times the diagonal length of the minimum rectangle enclosed in the shadow region.

The square area is shown in Figure 5 (a), and the center is represented as $P_0(x_0, y_0)$. Apart from the center pixel, we assume that there are L pixels inside the square area, and they are denoted as $P_i(x_i, y_i)$, $i = 1, \dots, L$, whose intensity is I_i , $i = 1, \dots, L$. Then a vector is formed and noted as $\mathbf{Z}_i = I_i \times \frac{\overrightarrow{P_0 P_i}}{\|P_0 P_i\|}$, where $\|\cdot\|_2$ denotes 2-norm. We define vector $\mathbf{M}(P_0)$ as:

$$\mathbf{M}(P_0) = \frac{1}{L} \sum_{i=1}^L \mathbf{Z}_i \quad (7)$$

where $\mathbf{M}(P_0)$ points to the location of an object. However, in practice, the calculation in this way is not straightforward, so a Mean Shift algorithm is undertaken. The idea of Mean Shift is to search for the location of the max-

imum probability density in a specific area. However, the pixels in the image
 210 are evenly distributed, thus the probability density function (PDF) cannot be
 defined according to the density. As expected, we understand that where the
 higher-intensity pixels gather, the greater of the probability density is. Actu-
 ally, the computation of the vector $\mathbf{M}(P_0)$ is a process of finding the location
 of the maximum probability density (mode) in the certain area. Two steps are
 215 involved indeed.

(i) *Calculation of the mode.* A circular area is extracted from Figure 5 (a) with
 the same center and with the radius of r ($r < R/2$). The purpose of this step is
 to estimate the PDF, and to find the mode in the circular area.

There are two kinds of information for each pixel: spatial ($[x_i, y_i]$) and in-
 220 tensity ($[I_i]$), both of which constitute the feature space, and each pixel can be
 denoted with a 3-dimensional vector $\mathbf{v}_i[x_i, y_i, I_i]$. Suppose there are N pixels in
 the area. The kernel function is introduced to estimate the PDF $\hat{f}(\mathbf{v})$:

$$\hat{f}(\mathbf{v}) = \frac{c_{k,s}}{Nh^s} \sum_{i=1}^N k \left(\left\| \frac{\mathbf{v} - \mathbf{v}_i}{h} \right\|^2 \right) \quad (8)$$

where $k(\cdot)$ is the profile of kernel function $K(\cdot)$. s is the dimension of the feature
 space, h is the bandwidth parameter, and $c_{k,s}$ is a constant ensuring the integral
 225 of $K(\cdot)$ to be 1. We define $g(\mathbf{v}) = -k'_N(\mathbf{v})$, and its corresponding kernel function
 is $G(\cdot)$. Thus the mode can be obtained by calculating the gradient of Eq.(8):

$$\begin{aligned} \hat{\nabla} f_{h,K}(\mathbf{v}) &= \frac{2c_{k,s}}{Nh^{s+2}} \left[\sum_{i=1}^N g \left(\left\| \frac{\mathbf{v} - \mathbf{v}_i}{h} \right\|^2 \right) \right] \\ &\times \left[\frac{\sum_{i=1}^N \mathbf{v}_i g \left(\left\| \frac{\mathbf{v} - \mathbf{v}_i}{h} \right\|^2 \right)}{\sum_{i=1}^N g \left(\left\| \frac{\mathbf{v} - \mathbf{v}_i}{h} \right\|^2 \right)} - \mathbf{v} \right] \end{aligned} \quad (9)$$

where,

$$\mathbf{m}_{h,G}(\mathbf{v}) = \frac{\sum_{i=1}^N \mathbf{v}_i g \left(\left\| \frac{\mathbf{v} - \mathbf{v}_i}{h} \right\|^2 \right)}{\sum_{i=1}^N g \left(\left\| \frac{\mathbf{v} - \mathbf{v}_i}{h} \right\|^2 \right)} - \mathbf{v} \quad (10)$$

In Eq.(9), $\frac{2c_{k,s}}{Nh^{s+2}} \left[\sum_{i=1}^N g \left(\left\| \frac{\mathbf{v} - \mathbf{v}_i}{h} \right\|^2 \right) \right]$ is a constant. The gradient always points
 to the direction where the variation is the largest, so we learn that $\mathbf{m}_{h,G}(\mathbf{v})$

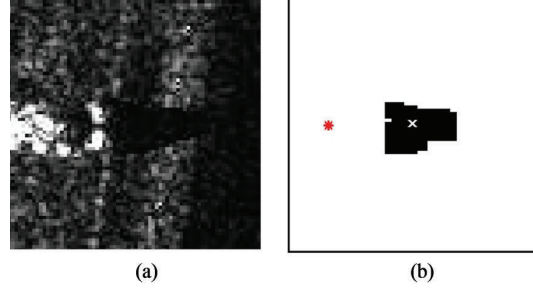


Figure 5: (a) The shadow region and its surrounding area; (b) The black area is the shadow region corresponding to (a), and the white \times represents the center of shadow region, besides, the red $*$ denotes the mode of (a).

230 points to the location of the mode. Therefore, the location of the mode in the circular area can be obtained:

$$\begin{aligned}
 x^{mode} &= \frac{\sum_{i=1}^N x_i g(\|\frac{I-I_i}{h}\|^2)}{\sum_{i=1}^N g(\|\frac{I-I_i}{h}\|^2)} \\
 y^{mode} &= \frac{\sum_{i=1}^N y_i g(\|\frac{I-I_i}{h}\|^2)}{\sum_{i=1}^N g(\|\frac{I-I_i}{h}\|^2)}
 \end{aligned} \tag{11}$$

(ii) *Judgement.* If $\|\mathbf{m}_{h,G}(\mathbf{v})\| > \varepsilon$ (ε represents the convergence threshold), we take the circle area with the center of (x^{mode}, y^{mode}) and the radius of r , and repeat the last step. Otherwise, (x^{mode}, y^{mode}) is the mode of the square area.

235 The mode for Figure 5 (a) can be obtained using the above two steps, and the result is shown in Figure 5 (b). Also, we can obtain the corresponding mode of each shadow region, whilst the spatial relationship between shadows and objects can be inferred according to the statistical distribution of the location of the modes. Furthermore, the square can be separated into four regions: up,
 240 down, left and right, as shown in Figure 6. We calculate the number of the modes in which region they are located, and the orientation of the object can be determined by voting. Based on this relationship, the proposal chips can be extracted from the original SAR image using prior information.

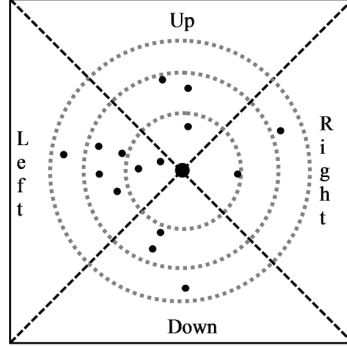


Figure 6: The spatial relationship of shadow areas and objects. The black spot in the center represents the shadow region, and other scattered black spots denote the location of the modes.

3.2. GSST saliency detection stage

245 Having obtained the proposal chips, we are to extract the areas that appear to be unique, where the GSST saliency detection method is employed. Let a potential chip be denoted by P , which contains $M \times N$ pixels. Saliency detection can be divided into the following three steps:

(i) *Local spatial autocorrelation process.* Eq.(1) is used to compute statistics G for each pixel of P , and we can obtain a statistics G matrix with the same size 250 as that of chip P :

$$G_P = \begin{bmatrix} G_{11} & G_{12} & \cdots & G_{1N} \\ G_{21} & G_{22} & \cdots & G_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ G_{M1} & G_{M2} & \cdots & G_{MN} \end{bmatrix} \quad (12)$$

where $G_{ij} (1 \leq i \leq M, 1 \leq j \leq N)$ represents the statistics G of the pixel at (i, j) in the image.

(ii) *Significance tests.* Eq.(4) is utilized to compute the statistics Z matrix of G_P , and we can obtain:

$$Z_P = \begin{bmatrix} Z_{11} & Z_{12} & \cdots & Z_{1N} \\ Z_{21} & Z_{22} & \cdots & Z_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ Z_{M1} & Z_{M2} & \cdots & Z_{MN} \end{bmatrix} \quad (13)$$

where $Z_{ij}(1 \leq i \leq M, 1 \leq j \leq N)$ represents the statistics Z of the pixel at (i, j) in the image.

Only the local neighborhood is considered when we calculate the spatial autocorrelation, so each element of Z_P is just a relevant value without considering its global impact. However, saliency is based on the whole chip, so in practice, the statistics matrix Z_P needs to be standardized globally:

$$Z'_P = \frac{Z_P - \bar{Z}_P}{s(Z_P)} = \begin{bmatrix} Z'_{11} & Z'_{12} & \cdots & Z'_{1N} \\ Z'_{21} & Z'_{22} & \cdots & Z'_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ Z'_{M1} & Z'_{M2} & \cdots & Z'_{MN} \end{bmatrix} \quad (14)$$

where \bar{Z} and $s(Z_P)$ represent the mean and standard deviation of Z_P respectively. $Z'_{ij}, 1 \leq i \leq M, 1 \leq j \leq N$ denotes the globally standardized statistics Z located at (i, j) in the image. We have made the null hypothesis of spatial independence above, so Eq.(5) is used to perform the significance tests for the globally standardized statistics matrix Z'_P :

$$Z'_{ij} \geq z_{\alpha/2} \quad (15)$$

(iii) *Salient regions extraction.* $z_{\alpha/2}$ is set as the threshold, and the regions $Z'_{ij} \geq z_{\alpha/2}$ are treated as the preliminary salient regions. Then some regions are discarded according to the thresholding. A sample of the salient region extraction process for a target and a background chip is illustrated in Figure 7.

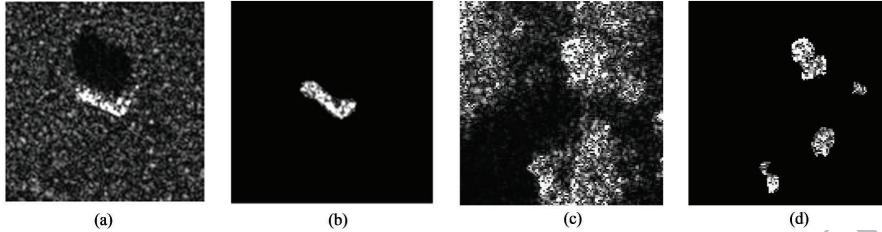


Figure 7: (a) and (b) are the target chip and its salient region; (c) and (d) represent the background chip and its salient regions.

3.3. One-Class SVM screening stage

It is observed in Figure 7 that the salient regions of the target chips are well matched with the areas of the real targets. Instead, the appearance of salient regions of the background chips are irregular. Conventional binary classification methods need both positive and negative training samples, but it is extremely difficult to guarantee that most of the negative states are included in the training. OC-SVM differs from traditional classifiers in that only one class training samples are required. Sarah et al. [44] handled high-dimensional and large-scale anomaly detection using a linear OC-SVM with deep learning. Yan et al. [45] applied OC-SVM to fault detection and obtained high accuracy rates. Victor et al. [46] concluded that OC-SVM is the most appropriate technique for anomaly detection in smart city wireless sensor networks. As we can see, OC-SVM have demonstrated good performance for anomaly and fault detection, and salient regions of background in this paper can be recognized as unexpected anomalies or faults as well [47].

Aiming at overcoming the uncertainty of background and improving the generalization ability of the proposed method, we treat the target detection as a one-class classification problem, where we use OC-SVM to detect target class and reject the others. Two steps are involved in the classification: feature extraction and OC-SVM classification.

(i) *Feature extraction.* Complex feature extractors composed of different components are generally applied in SAR target detection or recognition [48, 49].

However, GSST can extract accurate target areas, so visual features are sufficient to discriminate a target from clutters. In this paper, target chips are introduced as training samples while the background samples are neglected. Visual features such as area, perimeter, and aspect ratio of the salient regions of the training and testing samples are then extracted respectively, and the features of each region are concatenated and form a vector.

(ii) *OC-SVM classification.* The main idea of OC-SVM algorithm is to map the input samples to a high dimensional feature space with a kernel function, and search for a hyperplane to separate the mapping points from the origin (in this feature space). Given a training set $D = \{p_i\}, p_i \in R^K, 1 \leq i \leq n$, where p_i represents the i_{th} training sample, K is the number of visual features, and n denotes the total number of samples. It is assumed that there is a kernel function ϕ that maps from the original space R^K to the infinite dimensional space χ , which satisfies $\phi(p_i) \in \chi$. Thus the hyperplane is used to perform the classification:

$$\min \frac{1}{2} \|w\|^2 + \frac{1}{vn} \sum_{i=1}^n \xi_i - \rho \quad (16)$$

$$s.t. \quad w \cdot \phi(p_i) \geq \rho - \xi_i, \xi_i \geq 0$$

where w is the normal vector of the hyperplane, ρ represents the intervals between the hyperplane and the original point, ξ_i is the slack variable corresponding to the i_{th} sample to punish the points which deviate from the hyperplane, and $v \in [0, 1]$ indicates the compromise between a maximum interval and the penalty term. A Lagrange function is used to derive the hyperplane, and the decision equation is obtained:

$$f(p) = \text{sgn}((w \cdot \phi(p)) - \rho) \quad (17)$$

If $f(p) \geq 0$, then we assign p to the targets, otherwise to false alarms.

315 4. Experimental Analysis

The SAR image data used in this paper is Moving and Stationary Target Acquisition and Recognition (MSTAR) database which was collected using the Sandia National Laboratories Twin Otter SAR sensor payload operating at the X band with a high resolution of $0.3m$, spotlight mode, and HH single polarization [50].

4.1. Key parameter analysis

In the implementation of the proposed method, there are two free parameters needed to be set carefully. One is d in Eq.(1), and the other is $z_{\alpha/2}$ in Eq.(5). Both parameters have a direct impact on the salient regions extracted by GSST. To show how their values affect the performance of GSST, we perform the experiment on a list of vehicle chips and evaluate the F_{β} - measure based on manually labeled ground-truth by:

$$\begin{aligned}
 F_{\beta} &= \frac{(\beta^2 + 1) \times P \times R}{\beta^2 \times P + R} \\
 P &= \frac{TP}{TP + FP} \\
 R &= \frac{TP}{TP + FN}
 \end{aligned} \tag{18}$$

where P , R and F_{β} denote Precision, Recall and F_{β} - measure respectively. TP , FP , and FN represent the number of the true positive pixels, false positive pixels, and false negative pixels of the detected area respectively. F_{β} - measure is the weighted harmonic mean of precision and recall. Here we believe that the Precision is more important than Recall, so we take $\beta = 0.5$.

Figure 8 shows the F_{β} - measure curve over $z_{\alpha/2}$ and d . As can be seen F_{β} - measure is close to maximum at $z_{\alpha/2} = 2.5$ when $d = 25, 30, 35$. In consideration of the computational complexity, we select $d = 25$ and $z_{\alpha/2} = 2.5$ in our following experiments.

Another parameter needs attention is v in Eq.(16), which is empirically set to be 0.1 through our prior knowledge, and it is proven to work well empiri-

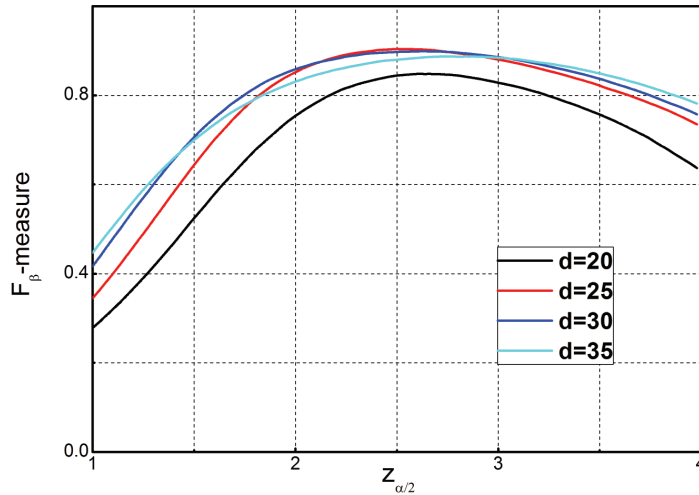


Figure 8: Influence of key parameters on the performance of the GSST.

cally. Additionally the sigmoid kernel function is chosen for the OC-SVM in our
 340 experiment.

4.2. Saliency comparison experiment

In this experiment, we aim to compare GSST on the SAR images with other
 four advanced methods, i.e. MC [25], GMR [28], RBD [29], and GBVS [17].
 MC is the saliency detection method which uses Markov chain, GMR realizes
 345 saliency detection via the graph-based manifold ranking, and RBD is derived
 from a robust background detection. All of these three methods show great
 performance reported in [24]. GBVS is a classical visual saliency detection
 method based on the graph theory.

There are seven types of vehicle targets: bmp2(sn_9563), bmp2 (sn_9566),
 350 bmp2 (sn_c21), btr70 (sn_c71), t72 (sn_132), t72 (sn_812) and t72 (sn_s7). In
 the experiment, 10 chips of each type are randomly selected, and typical target
 chip for each type of vehicle is shown in Figure 9 (a). According to the saliency

Table 1: F_β - measure of five methods.

Methods	F_β - measure
GSST	90.71%
MC	86.61%
GMR	88.59%
RBD	85.88%
GBVS	83.71%

theory, it is the target region that mainly raises visual attention in a chip, thus we manually intercept the target area as the ground-truth of the salient region, for example, Figure 9 (b) corresponding to Figure 9 (a). Precision, Recall and F_β - measure defined above are used to measure the performance of saliency detection.

Firstly, Eqs.(1)~(5), (12)~(14) are used to compute the saliency map of each target chip. Then the significance tests are carried out on the saliency map by changing α in Eq.(15), where we get $z_{\alpha/2} = 0, 0.05, 0.1, \dots, 4.45$, and thus 90 salient regions are extracted from each chip. Finally, Precision, Recall and F_β - measure are calculated using Eq.(18) for each pair of the detected salient region and its corresponding ground-truth. For all the 70 chips, we take the mean value of the data, and thus one group of Precision, Recall and F_β - measure data is obtained.

Also, MC, GMR, RBD and GBVS are handled in the same way as shown above to compute the saliency map, and a range of thresholds are used to extract salient regions from each chip, after which one group of Precision, Recall and F_β - measure data can be obtained for each method. Table 1 shows F_β - measure of the five methods when the salient region is accurately extracted. The Precision-Recall (PR) curves of five methods are plotted in Figure 10. Some detection samples of the five methods are shown in Figure 9 (c)~(g).

Typically, Precision and Recall are inversely related. To achieve a balance between them and compare the performance, the PR curves come in handy.

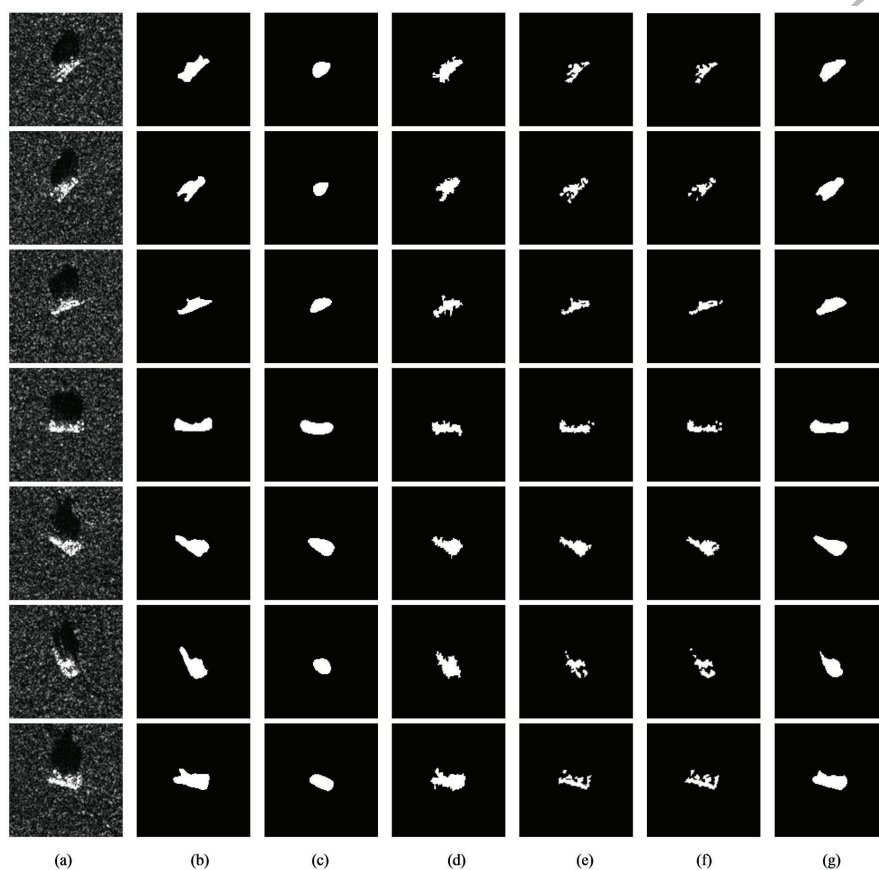


Figure 9: (a) Samples of seven types of vehicles. (b) Ground-truth of the salient area intercepted manually; (c) Detection results of GBVS; (d) Detection results of MC; (e) Detection results of GMR; (f) Detection results of RBD; (g) Detection results of GSST.

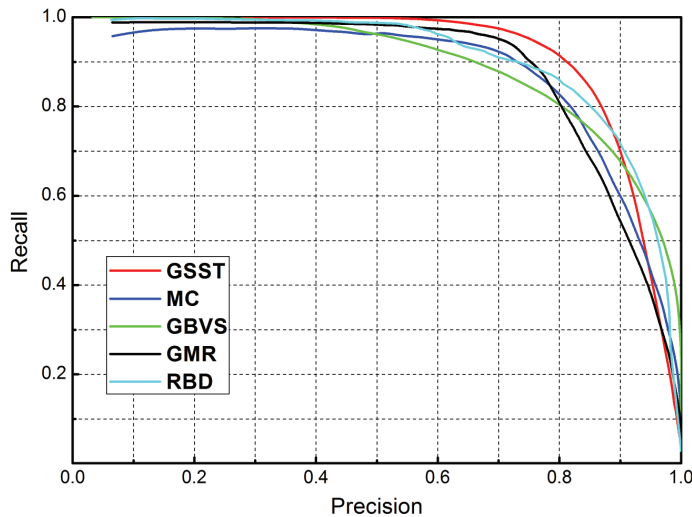


Figure 10: The PR curves of five kinds of methods on vehicle target chips.

375 Therefore the PR curve is capable of giving the informative plot of an algorithm's performance. In practice, a detection method is believed to be effective when the PR curve moves to the upper-right-hand corner [51] and F_β -measure possesses a large value. Thus it proves quantitatively that GSST has a better performance than the other four methods from Figure 10 and Table 1. Qualitatively, in Figure
 380 9 (c)~(e), we can see that the result of GSST is the best in agreement with the ground-truth, while MC, GMR, and RBD have some more burrs and GBVS misses some details. The reason for this outcome may be that MC, GMR, and RBD focus on the object interior while GBVS is able to pick up the boundary of the salient object. It is worth noting that the GMR has a close F_β -measure to the GSST, while it suffers from poor contours and the incomplete target region.
 385 Furthermore, the GMR and RBD depend on the background hypothesis that objects lie away from boundary of the image, which may cause certain failures if the targets are located nearby the boundary.

Computational complexity of GSST is related to the patches generated in

Table 2: Average computational time per image of five saliency detection methods.

Methods	GSST	MC	GBVS	GMR	RBD
Time(s)	0.087	0.713	1.227	0.780	3.08
code	Matlab	Matlab and C++	Matlab	Matlab	Matlab

390 the shadow proposal stage. Specifically, there are $20d^2N_P P$ addition and multiplication operations approximately, where d is the parameter in Eq.(1), P is the number of patches, and N_P is the total number of pixels in each patch. The average computational time per image of GSST, MC, GBVS, GMR and RBD are summarized in Table 2. The size of tested images is 128×128 , and all the
395 methods are experimented on Intel core 2 Duo CPU of 3.0 GHz with 8GB RAM. It can be seen that the GSST is the fastest among the five contrast methods.

4.3. Vehicle target detection experiment

To verify the adaptability of the proposed vehicle targets detection method in different environments, two scenes are selected for the evaluation. Again,
400 Precision, Recall and F_β -measure are used to measure the performance. What differs from section 4.1 is that we take regions as the basic unit rather than pixels. 10 vehicle targets for each type of BTR70 (sn.c71) and T72 (sn.132) are seen in two 1478×1784 full clutter MSTAR images, where both vehicle targets and the background are collected at 15° depression degree. The two scenes are
405 depicted in Figure 11 (a) and (b).

4.3.1. Scene 1

In this paper, the experiment can be divided into three stages. In the first stage, the structuring element of the morphological open and close operations are set to 5×5 and 3×3 respectively. The normal kernel function is selected for
410 Eq. (8). The spatial relationship between shadows and objects is obtained from Eq. (11). In Figure 11 (a), objects locate on the left of the shadow regions, based on which 146 proposal chips are generated by setting the size of the interception box to 128×128 .

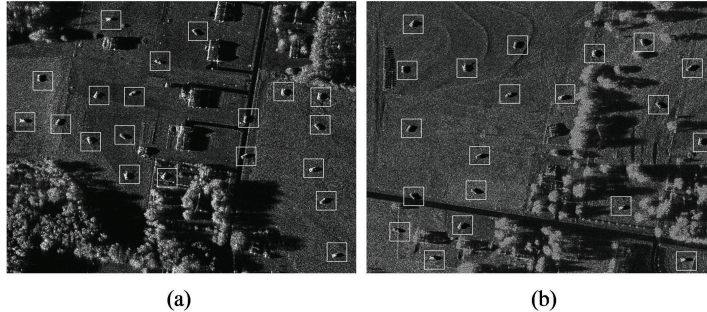


Figure 11: (a) Scene 1. 20 real targets are labeled by white rectangles. (b) Scene 2. A simpler background than scene 1. 20 real targets are labeled by white rectangles.

In the second stage, Eqs.(1)~(5), (12)~(15) are used to extract salient regions of the proposal chips. We here use Euclidean distance. When performing the significance tests, $z_{\alpha/2}$ is set to 2.5. Finally, 138 salient regions are obtained after some less possible regions are removed.

The OC-SVM is introduced in the third stage. BTR70 (sn.c71) and T72 (sn.132), both of which include 232 chips, are used as the training samples, whereas the background samples are not used. All the training samples are collected at 17° depression degree. The saliency detection process is applied to the training samples using Eqs.(1)~(5), (12)~(15), and 464 salient regions are extracted. Then, we compute visual features of salient regions of the training and testing samples. In this paper, the following five features are selected: area periphery ration, fractal dimension, weighted-rank fill ratio, maximum distance, and centrifugal rate of the ellipse which has the same second-order moments as the salient region. These five features have the form of a vector whose size is 1×5 . Afterwards, the OC-SVM model is trained using the feature vectors of the training samples. Finally, the feature vectors of the testing samples are cast into the trained model to detect vehicle targets. The final detection result is shown in Figure 12 (d), and we find that our method detects 18 targets and generates only 1 false alarm.

Three methods are used for comparison. The first one is MC+OC-SVM,

which substitutes the GSST stage using the MC algorithm, and other stages
 remain unchanged. The notation '+' means the composition relationship. The
 435 second one is CFAR+OC-SVM, in which the CFAR algorithm is used to roughly
 extract suspicious regions, and then the OC-SVM screening stage is applied. An-
 other one is currently popular Faster R-CNN [52], which consists of a Region
 Proposal Network stage and a Fast R-CNN stage. The VGG-16 model [53] is
 440 used as a base network to share features between two stages. As suggested
 in [52], a pragmatic 4-step training algorithm is adopted to learn the shared
 features via the alternating optimization. In the first step, we train the RPN
 which is initialized with an ImageNet-pre-trained model and finely tuned us-
 ing the MSTAR training set for the region proposal task. In the second step,
 445 the proposals generated by the step-1 RPN are used to train the Fast R-CNN
 network, which is initialized by the ImageNet-pre-trained model as well. Then
 the trained detection network is applied to initializing the RPN, where we keep
 the shared convolutional layers fixed and fine-tune the part unique to the RPN.
 In the last stage, we still fix the shared convolutional layers and fine-tune the
 450 unique layers belonging to the Fast R-CNN. We use the learning rate 0.001, the
 momentum 0.9, and the weight decay 0.0005 in all the four steps. Considering
 the network structure of the Faster R-CNN, we randomly embed vehicle targets
 into 600×600 scenes as training set, in which data augmentation strategies such
 as shift and rotation are adopted to extend training samples. The ground-truth
 455 of training chips is manually labelled. It is worth noting that the correspond-
 ing shadow regions and some extra background areas are included, which is
 extremely important. Since the Faster R-CNN only boxes the detected targets
 in the original image, we binarize the results with the standard Otsu method
 for convenience of comparison.

460 The final results of the four methods are shown in Figure 12 (a)~(d). The
 real targets that have been detected are labeled by red regions, and the false
 alarms are marked by white regions, while the white circle-labeled regions repre-
 sent the missing targets. We witness that MC+OC-SVM detects 18 targets and
 introduces 3 false alarms, and CFAR+OC-SVM detects 16 targets but gener-

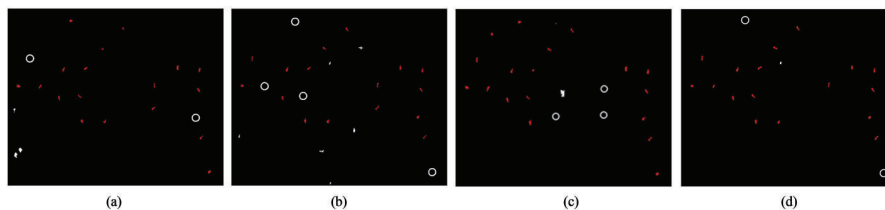


Figure 12: The detection results in scene 1. (a), (b), (c), and (d) show the detection results of MC+OC-SVM, CFAR+OC-SVM, Faster R-CNN, and the proposed method, respectively. The real targets that have been detected are labeled by red regions, and the false alarms are marked by white regions, while the white circle-labeled regions represent the missed.

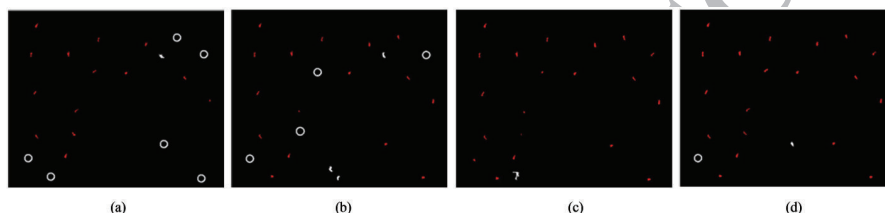


Figure 13: The detection results in scene 2. (a), (b), (c), and (d) show the detection results of MC+OC-SVM, CFAR+OC-SVM, Faster R-CNN, and the proposed method, respectively. The real targets that have been detected are labeled by red regions, and the false alarms are marked by white regions, while the white circle-labeled regions represent the missed.

465 ates 6 false alarms. The Faster R-CNN only generates 2 false alarms but misses three targets. The reason why the Faster R-CNN misses three targets is because the neighborhood information presented in the surrounding of target regions of training samples cannot cover all the real scenes. However, as mentioned above, our method is able to detect 18 targets with 1 false alarm only. Precision, Recall and F_β - *measure* are introduced to analyze the performance of the four
 470 methods where β is set to 1. The quantitative comparison is shown in Table 3.

4.3.2. Scene 2

Figure 11 (b) shows the second scene. There are less man-made objects and distracters in scene 2, which is significantly different from scene 1. We
 475 still adopt the above four methods to perform the experiment. The proposed method detects 19 of the 20 targets, and only one false alarm is generated. In

Table 3: The contrast results of four methods in scene 1.

Methods	Precision	Recall	$F_\beta - measure$
Ours	94.74%	90%	92.31%
MC+OC-SVM	85.71%	90%	87.80%
CFAR+OC-SVM	72.73%	80%	76.19%
Faster R-CNN	89.47%	85%	87.18%

Table 4: The contrast results of four methods in scene 2

Methods	Precision	Recall	$F_\beta - measure$
Ours	95%	95%	95 %
MC+OC-SVM	93.33%	70%	80%
CFAR+OC-SVM	84.21%	80%	82.05%
FasterR-CNN	95%	100%	97.44%

contrast, MC+OC-SVM detects only 14 targets and introduces one false alarm, and CFAR+OC-SVM has 4 targets undetected and generates 3 false alarms. The Faster R-CNN detects all the targets and generates only one false alarm. The detection results of the four methods are shown in Figure 13 (a)~(d), and the quantitative comparison is given in Table 4.

As it can be seen from Table 3, the Recall and $F_\beta - measure$ of our method are the highest among the four methods in scene 1. Comparably, MC+OC-SVM, CFAR+OC-SVM, and Faster R-CNN fall behind our method by 4.51%, 16.12%, and 5.13% using $F_\beta - measure$ metrics. In scene 2, our method has a little improvement in each index than those in scene 1, and is only 2.44% lower than the Faster R-CNN in $F_\beta - measure$. The CFAR+OC-SVM has obvious improvement in the three indexes, but still is much less successful than our proposed method. The $F_\beta - measure$ of the MC+OC-SVM drops 7.8% in the simpler environment. It is worth noting that the Faster R-CNN might be effective when there are enough training scenes and training chips, but our method is more adaptive in various scenes. Additionally, we analyse that why

Table 5: Average computational time per image of four vehicle target detection methods.

Methods	Ours	MC+OC-SVM	CFAR+OC-SVM	Faster R-CNN
Time(s)	28.723	105.715	154.828	9.912
code	Matlab	Matlab and C++	Matlab	Python

MC+OC-SVM and CFAR-SVM fall behind ours is that the detected target regions by the GSST agree with the ground-truth, which can be concluded from Figure 12 (a), (c), (d) and Figure 13 (a), (c), (d).

Also, we report average running time of these four methods in Table 5. We find that our method is much less time-consuming than CFAR+OC-SVM and MC+OC-SVM, and it is only behind the Faster R-CNN. The size of tested images is 1784×1478 . Faster R-CNN is tested on a 4-core Inter XEON E5506 CPU of 2.133MHz with 42GB RAM while other methods are experimented on Intel core 2 Duo CPU of 3.0 GHz with 8GB RAM.

4.3.3. Analysis of robustness

To further analyze the robustness of the above four methods, we perform the experiments under different SCR conditions, which is calculated as follows:

$$SCR = 20 \lg \frac{\bar{I}_{tar}}{\bar{I}_{bcg}} \quad (19)$$

where \bar{I}_{tar} and \bar{I}_{bcg} denote the mean intensities of the target and its surrounding background.

We still adopt the F_β - *measure* to execute the comprehensive evaluation. After a number of experiments, we obtain a curve of F_β - *measure* ($\beta = 1$) over SCR, as is shown in Figure 14. We find that the Faster R-CNN has the most robust detection results. Although our method is not good as the Faster R-CNN under the very low SCR conditions, it performs better when $SCR > 4\text{dB}$, which attributes to the robust proposals of shadow detection and accurate salient region extraction of the GSST. On the contrary, the F_β - *measure* of MC+OC-SVM drops rapidly when $SCR < 8\text{dB}$ for MC cannot extract accurate salient

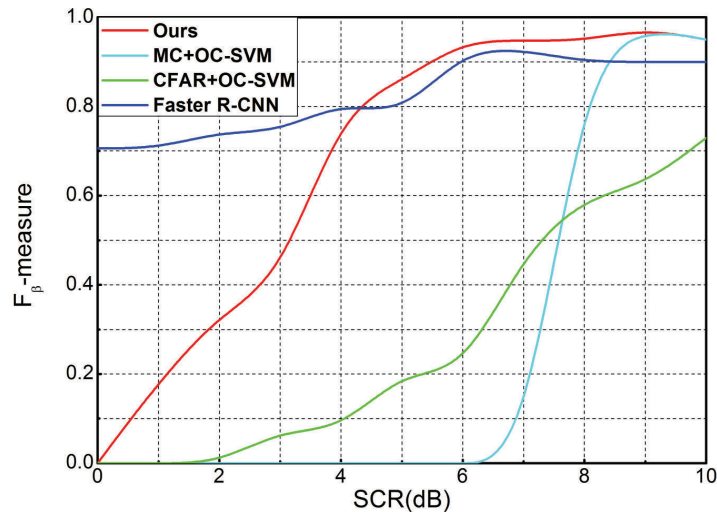


Figure 14: The F_{β} - measure of four methods under different SCR.

515 regions in complex environments. The CFAR+OC-SVM performs a little bit better than the MC+OC-SVM, but it is still less robust than ours. In summary, our method possesses relatively robust performance under the different SCR conditions, especially when $SCR > 4\text{dB}$.

5. Conclusion

520 By exploring the unique features of SAR image, we have presented a novel target detection scheme which consists of shadow proposal, saliency analysis and OC-SVM screening phases. The main novelty lies in the following three aspects. First, we used shadow detection to generate proposal chips, which makes the subsequent processing easier and enhances the robustness of our method. Second, we proposed a novel saliency analysis method, exploiting
525 the local spatial autocorrelation statistics and significance tests. It allows the extraction of salient regions in SAR images. Last but not least, OC-SVM was

employed to identify the real targets from salient regions, where only the visual features of salient regions of positive samples are needed to train the model.

530 The results of the comparison experiments quantitatively and qualitatively demonstrate that GSST extracts more accurate salient regions than the MC, GBVS, GMR and RBD models when applied to SAR images. Moreover, the results from the SAR vehicle target detection experiments show that our proposed method is adapted to both scenes, while maintaining the F_β -measure over 90%.
 535 On robustness and computational complexity, our method only lags behind Faster R-CNN. On the contrary, the results of MC+OC-SVM and CFAR+OC-SVM either lacks robustness or has low accuracy. Therefore, it suggests that the proposed method is effective and robust for SAR target detection.

Acknowledgements

540 This work was supported by the National Natural Science Foundation of China (61071139; 61471019; 61171122; 61501011; 61671035), the Aeronautical Science Foundation of China (20142051022), and the Pre-research Project(9140A07040515HK01009). Dr E. Yang is supported in part under the RSE-NNSFC Joint Project (2017-2019) (6161101383) with China University of Petroleum (Huadong). Dr H. Zhou
 545 is supported by UK EPSRC under Grants EP/N508664/1 and EP/N011074/1, and Royal Society-Newton Advanced Fellowship under Grant NA160342.

References

References

- 550 [1] L. M. Novak, M. C. Burl, W. W. Irving, Optimal polarimetric processing for enhanced target detection, IEEE Transactions on Aerospace & Electronic Systems 29 (1) (1993) 234–244.
- [2] L. M. Novak, S. D. Halversen, G. Owirka, M. Hiatt, Effects of polarization and resolution on sar atr, IEEE Transactions on Aerospace & Electronic Systems 33 (1) (1997) 102–116.

- 555 [3] H. Rohling, Radar cfar thresholding in clutter and multiple target situations, *IEEE Transactions on Aerospace and Electronic Systems* AES-19 (4) (1983) 608–621.
- [4] M. E. Smith, P. K. Varshney, Vi-cfar: a novel cfar algorithm based on data variability, in: *Proceedings of the 1997 IEEE National Radar Conference*, 1997, pp. 263–268.
- 560 [5] B. Alexe, T. Deselaers, V. Ferrari, What is an object?, in: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 73–80.
- [6] J. Hosang, R. Benenson, P. Dollr, B. Schiele, What makes for effective detection proposals?, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38 (4) (2016) 814–830.
- 565 [7] K. E. A. v. d. Sande, J. R. R. Uijlings, T. Gevers, A. W. M. Smeulders, Segmentation as selective search for object recognition, in: *2011 International Conference on Computer Vision*, 2011, pp. 1879–1886.
- [8] S. Manen, M. Guillaumin, L. V. Gool, Prime object proposals with randomized prim’s algorithm, in: *IEEE International Conference on Computer Vision*, 2013, pp. 2536–2543.
- 570 [9] P. Rantalankila, J. Kannala, E. Rahtu, Generating object segmentation proposals using global and local search, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2417 – 2424.
- 575 [10] A. Humayun, F. Li, J. M. Rehg, Rigor: Reusing inference in graph cuts for generating object regions, in: *Computer Vision and Pattern Recognition*, 2014, pp. 336–343.
- [11] P. Arbez, J. Pont-Tuset, J. Barron, F. Marques, J. Malik, Multiscale combinatorial grouping, in: *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 328–335.
- 580

- [12] M. M. Cheng, Z. Zhang, W. Y. Lin, P. Torr, Bing: Binarized normed gradients for objectness estimation at 300fps, in: IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 3286–3293.
- 585 [13] C. L. Zitnick, P. Dollr, Edge Boxes: Locating Object Proposals from Edges, Springer International Publishing, 2014.
- [14] Z. Zhang, J. Warrell, P. H. S. Torr, Proposal generation for object detection using cascaded ranking svms, in: CVPR 2011, 2011, pp. 1497–1504.
- 595 [15] F. Jie, Y. Wei, L. Tao, Z. Chao, J. Sun, Salient object detection by composition, in: 2011 International Conference on Computer Vision, 2011, pp. 1028–1035.
- [16] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, IEEE Transactions on Pattern Analysis & Machine Intelligence 20 (11) (1998) 1254–1259.
- 595 [17] B. Schlkopf, J. Platt, T. Hofmann, Graph-Based Visual Saliency, MIT Press, 2007, pp. 545–552.
- [18] L. Zhang, M. H. Tong, T. K. Marks, H. Shan, G. W. Cottrell, Sun: A Bayesian framework for saliency using natural statistics, Journal of Vision 8 (7) (2008) 1–20.
- 600 [19] N. Bruce, J. Tsotsos, Attention based on information maximization, Journal of Vision 7 (9) (2010) 950–950.
- [20] X. Hou, L. Zhang, Saliency detection: A spectral residual approach, in: 2007 IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.
- 605 [21] C. Guo, Q. Ma, L. Zhang, Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform, in: IEEE Computer Society Conference on Computer Vision & Pattern Recognition IEEE Computer Society Conference on Cvpr, 2008, pp. 1–8.

- [22] W. Wang, J. Shen, L. Shao, F. Porikli, Correspondence driven saliency transfer, *IEEE Transactions on Image Processing* 25 (11) (2016) 5025–5034.
610
- [23] W. Wang, J. Shen, L. Shao, Consistent video saliency using local gradient flow optimization and global refinement, *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society* 24 (11) (2015) 4185–96.
- [24] A. Borji, M. M. Cheng, H. Jiang, J. Li, Salient object detection: A benchmark, *IEEE Transactions on Image Processing* 24 (12) (2015) 5706–5722.
615
- [25] B. Jiang, L. Zhang, H. Lu, C. Yang, M. H. Yang, Saliency detection via absorbing Markov chain, in: *2013 IEEE International Conference on Computer Vision*, 2013, pp. 1665–1672.
- [26] J. Han, S. He, X. Qian, D. Wang, L. Guo, T. Liu, An object-oriented visual saliency detection framework based on sparse coding representations, *IEEE Transactions on Circuits and Systems for Video Technology* 23 (12) (2013) 2009–2021.
620
- [27] X. Qian, J. Han, G. Cheng, L. Guo, Optimal contrast based saliency detection, *Pattern Recognition Letters* 34 (11) (2013) 1270–1278.
625
- [28] C. Yang, L. Zhang, H. Lu, X. Ruan, M. H. Yang, Saliency detection via graph-based manifold ranking, in: *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3166–3173.
- [29] W. Zhu, S. Liang, Y. Wei, J. Sun, Saliency optimization from robust background detection, in: *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2814–2821.
630
- [30] F. Zhang, B. Du, L. Zhang, Saliency-guided unsupervised feature learning for scene classification, *IEEE Transactions on Geoscience & Remote Sensing* 53 (4) (2015) 2175–2184.

- 635 [31] J. Han, P. Zhou, D. Zhang, G. Cheng, L. Guo, Z. Liu, S. Bu, J. Wu, Efficient, simultaneous detection of multi-class geospatial targets based on visual saliency modeling and discriminative learning of sparse coding, *Isprs Journal of Photogrammetry & Remote Sensing* 89 (1) (2014) 37–48.
- [32] J. Han, D. Zhang, G. Cheng, L. Guo, Object detection in optical remote
640 sensing images based on weakly supervised learning and high-level feature learning, *Geoscience & Remote Sensing IEEE Transactions on* 53 (6) (2015) 3325–3337.
- [33] S. Papson, R. M. Narayanan, Classification via the shadow region in sar
645 imagery, *IEEE Transactions on Aerospace and Electronic Systems* 48 (2) (2012) 969–980.
- [34] P. Lombardo, M. Sciotti, L. M. Kaplan, Sar prescreening using both target and shadow information, in: *Proceedings of the 2001 IEEE Radar Conference (Cat. No.01CH37200)*, 2001, pp. 147–152.
- [35] G. Cheng, P. Zhou, J. Han, Learning rotation-invariant convolutional neural
650 networks for object detection in vhr optical remote sensing images *PP (99)* (2016) 1–11.
- [36] Y. Yuan, H. Lv, X. Lu, Semi-supervised change detection method for multi-temporal hyperspectral images, *Neurocomputing* 148 (2015) 363375.
- [37] X. Lu, Y. Yuan, X. Zheng, Joint dictionary learning for multispectral
655 change detection, *IEEE Transactions on Cybernetics* 47 (4) (2017) 884–897.
- [38] G. Cheng, J. Han, A survey on object detection in optical remote sensing images, *Isprs Journal of Photogrammetry & Remote Sensing* 117 (2016) 11–28.
- 660 [39] B. S. Olkopf, R. Williamson, A. Smola, J. Shawe-Taylor, J. Platt, Support vector method for novelty detection, *Advances in Neural Information Processing Systems* 12 (2000) 582–588.

- [40] D. Comaniciu, P. Meer, Mean shift: a robust approach toward feature space analysis, *IEEE Transactions on Pattern Analysis & Machine Intelligence* 24 (5) (2002) 603–619.
- [41] A. Getis, J. K. Ord, The analysis of spatial association by use of distance statistics, *Geographical Analysis* 24 (3) (1992) 189–206.
- [42] J. K. Ord, A. Getis, Local spatial autocorrelation statistics: Distributional issues and an application, *Geographical Analysis* 27 (4) (1995) 286–306.
- [43] P. A. P. Moran, Notes on continuous stochastic phenomena, *Biometrika* 37 (1-2) (1950) 17–23.
- [44] S. M. Erfani, S. Rajasegarar, S. Karunasekera, C. Leckie, High-dimensional and large-scale anomaly detection using a linear one-class svm with deep learning, *Pattern Recognition* 58 (2016) 121–134.
- [45] K. Yan, Z. Ji, W. Shen, Online fault detection methods for chillers combining extended kalman filter and recursive one-class svm, *Neurocomputing* 228 (2017) 205–212.
- [46] V. Garcia-Font, C. Garrigues, H. Rif-Pous, A comparative study of anomaly detection techniques for smart city wireless sensor networks, *Sensors* 16 (6) (2016) 868.
- [47] B. Du, L. Zhang, A discriminative metric learning based anomaly detection method, *IEEE Transactions on Geoscience & Remote Sensing* 52 (11) (2014) 6844–6857.
- [48] Z. Zhao, L. Jiao, B. Hou, S. Wang, J. Zhao, P. Chen, Locality-constraint discriminant feature learning for high-resolution sar image classification, *Neurocomputing* 207 (2016) 772–784.
- [49] S. Wang, L. Jiao, S. Yang, H. Liu, Sar image target recognition via complementary spatial pyramid coding, *Neurocomputing* 196 (C) (2016) 125–132.

- [50] E. R. Keydel, Mstar extended operating conditions: a tutorial, Proc Spie
690 2757 (1996) 228–242.
- [51] J. Davis, M. Goadrich, The relationship between precision-recall and roc
curves, in: International Conference on Machine Learning, 2006, pp. 233–
240.
- [52] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object
695 detection with region proposal networks, IEEE Transactions on Pattern
Analysis and Machine Intelligence PP (99) (2016) 1–1.
- [53] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-
scale image recognition, Computer Science 2014.

Fei Gao received the B.S. and M.S. degree from the Xi'an Petroleum Institute, Xi'an, China, in 1996 and 1999, respectively, and the Ph.D. degrees from Beijing University of Aeronautics and Astronautics (BUAA), Beijing, China, in 2005. He is currently an Associate Professor with the School of Electronic and Information Engineering, BUAA. He is interested in radar signal processing, moving target detection and image processing.

Jialing You received the B.S. degree in electronic and communication engineering from Beijing University of Aeronautics and Astronautics (BUAA), Beijing, China, in 2015. He is currently pursuing the M.S. degree in electronic and communication engineering at BUAA. His current research activities include computer vision and synthetic aperture radar image processing.

Jun Wang received the B.S. degree from the North western Polytechnical University, Xi'an, China, in 1995 and the M.S. and Ph.D. degrees from the Beijing University of Aeronautics and Astronautics (BUAA), Beijing, China, in 1998 and 2001, respectively. He is currently a Professor with the School of Electronic and Information Engineering, BUAA. He is interested in signal processing, DSP/FPGA real-time architecture, target recognition and tracking, and so on. His research has resulted in 38 papers in journals, books, and conference proceedings.

Jinping Sun received the M.Sc. and Ph.D. degrees from the Beijing University of Aeronautics and Astronautics (BUAA), Beijing, China, in 1998 and 2001, respectively. He is currently a Professor with the School of Electronic and Information Engineering, BUAA. His research interests include high-resolution radar signal processing, image understanding, and robust beam forming.

Erfu Yang is a Lecturer in the Department of Design, Manufacture and Engineering Management (DMEM) at the University of Strathclyde, Glasgow, UK. His main research interests include robotics, autonomous systems, mechatronics, manufacturing automation, computer vision, image/signal processing, nonlinear control, process modelling and simulation, condition monitoring, fault diagnosis, multi-objective optimizations, and applications of machine learning and artificial intelligence including multi-agent reinforcement learning, fuzzy logic, neural networks, bio-inspired algorithms, and cognitive computation, etc. He has over 60 publications in these areas, including more than 30 journal papers and 5 book chapters.

Huiyu Zhou received the B.Eng. degree in radio technology from Huazhong University of

Science and Technology, China, in 1990, the M.Sc. degree in biomedical engineering from the University of Dundee, Dundee, U.K., and the Ph.D. degree in computer vision from Heriot-Watt University, Edinburgh, U.K.

He is currently at the Institute of Electronics, Communications, and Information Technology, School of Electronics, Electrical Engineering, and Computer Science, Queen's University Belfast, Belfast, U.K. He has authored or coauthored widely in international journals and conferences. His research interests include computer vision, human motion analysis, intelligent systems, and human-computer interface.

ACCEPTED MANUSCRIPT



Fei Gao



Jialing You



Jun Wang

ACCEPTED MANUSCRIPT