Visual Attention Model Based Vehicle Target Detection in Synthetic Aperture Radar Images: A Novel Approach

Fei Gao*, Ye Zhang, Jun Wang, Jinping Sun, Erfu Yang**, Amir Hussain**

School of Electronic and Information Engineering, Beihang University, Beijing, China

*Corresponding Author, e-mail: buaacyeah@126.com, Telephone: +86-18701319751, Fax: 8601082317240

**Cognitive Signal-Image and Control Processing Research (COSIPRA) Laboratory, School of Natural Sciences, University of Stirling, Stirling, UK

Abstract The human visual system (HVS) possesses a remarkable ability of real-time complex scene analysis despite the limited neuronal hardware available for such tasks. The HVS successfully overcomes the problem of information bottleneck by selecting potential regions of interest and reducing the amount of data transmitted to high-level visual processing. On the other hand, many man-made systems are also confronted with the same problem yet fail to achieve satisfactory performance. Among these, the synthetic aperture radar based automatic target recognition (SAR-ATR) system is a typical one, where the traditional detection algorithm employed is termed the constant false alarm rate (CFAR). It is known to exhibit a low probability of detection (PD) and consumes too much time. The visual attention model (VAM) is a computational model which aims to imitate the HVS in predicting where humans will look. The application of VAM to the SAR-ATR system could thus help solve the problem of effective real-time processing of complex large amounts of data. In this paper, we propose a new vehicle target detection algorithm for SAR images based on the VAM. The algorithm modifies the well-known Itti model according to the requirements of target detection in SAR images.

The modified Itti model locates salient regions in SAR images and following top-down processing reduces false alarms by using prior knowledge. Real SAR data are used to demonstrate the validity and effectiveness of the proposed algorithm, which is also benchmarked against the traditional CFAR algorithm. Simulation results show comparatively improved performance in terms of PD, number of false alarms and computing time.

Keywords Human Visual System (HVS), synthetic aperture radar automatic target recognition (SAR-ATR), visual attention model (VAM), vehicle target detection

1. Introduction

It is very easy for humans to instantaneously detect a target object in a complex visual scene. It seems they are born with the ability to detect various kinds of objects without any thought or effort, being able to "find" targets, without clearly knowing how. Visual attention is believed to play a key role in this process [1]. It is often understood as a built-in mechanism of the human visual system (HVS) that quickly selects regions in a visual scene, which are most likely to contain items of interest. Such a pre-selection mechanism focusing only on relevant data is of crucial importance in overcoming the problem of information bottleneck along the visual pathway. Mimicking visual attention could thus greatly help computers perform real-time object detection in complex scenes with the desired degree of accuracy.

Recent years have witnessed growing interest in developing computational models of visual attention. In general, visual attention models (VAM) can be divided into two categories, on the basis of specific tasks, targets or intentions: bottom-up models and top down models [2]. The so called feature integration theory [3] provides a solid foundation for bottom-up models, with the most significant work

being the saliency map model proposed by Itti et al. [4]. Based on the attention selection proposed by Koch et al. [1], this model decomposes an input image into three channels: intensity, color and orientation. In order to simulate the lateral surround suppression among cortical cells, it uses a center-surround operation to produce a set of feature maps, which are then normalized and combined across scales to create conspicuity maps of each channel. The normalized conspicuity maps are then linearly combined to form the overall saliency map. The Itti model, since its introduction in 1998, has become the most popular VAM and is often used as a yardstick to benchmark the performance of other models. Some of the other well-known bottom-up models include the STB model [5], AIM model [6], GBVS model [7], SR model [8], PFT model [9], PQFT model [10] and so on.

The above mentioned models make specific use of a given image to compute visual attention and saliency. However, it has often been pointed out that such models cannot completely explain entire visual attention systems, and a number of top-down concepts have thus been proposed. The so called model of guided search [11] introduces top-down knowledge on characteristics of target stimuli in visual search, and serves as a basis for recently reported top-down models. In general, top-down models fall into three classes based on how a model computes top-down saliency: weight modulation of bottom-up features, weighted combination of outputs from bottom-up and top-down models and joint learning of bottom-up and top-down features [2].

Though the current VAMs are far from a perfect imitation of HVS, they have already been shown to enhance a growing number of applications in computer vision and pattern recognition. Among these, the synthetic aperture radar automatic target recognition (SAR-ATR) is a promising and challenging application, where the SAR is exploited as a powerful tool for target detection due to its ability to work in all-weather conditions, day and night. However, a SAR-ATR system suffers from the same problem

of information bottleneck as the HVS [12], as it usually needs to process large amounts of SAR data in real-time with limited computing resource constraints. This problem poses significant challenges for real-time automatic target detection, warranting development of new fast and effective detection algorithms, which are indispensable for SAR-ATR.

It is important to note that since the detector is the first stage of SAR-ATR, the efficiency of the detector directly impacts the succeeding stages in the SAR-ATR processing chain. The detection algorithms for SAR images are generally categorized into 3 classes: single-feature-based, multifeature-based and expert-system-oriented[13]. The last one is the most sophisticated and utilizeds a multistage artificaial intelligence approach while multifeature based method uses two or more features extracted from the input image. The first one is the most common and widely used in literature and CFAR is the most most popular one among this class. This algorithm bases the search for regions of interest (ROIs) on radar cross sections (RCS) alone. It assumes that the background clutter can be roughly modeled by a certain probability distribution and CFAR detection is performed after estimating the model distribution parameters. The early one-parameter CFAR algorithm uses one parameter to characterize the distribution model. More realistic two-parameter CFAR use two-parameter distribution models to characterize clutter, such as Weibull distribution [14] and *K*-distribution [15]. It is assumed that target pixels obey a certain distribution and pixels in the reference window are used to estimate the parameters of the distribution model. The drawback of CFAR is obvious: as the size of the image and the reference window increases, the execution time increases dramatically. This severely restricts the key requirement of the SAR-ATR system, that its detector should be relatively computationally simple, in order to enable operation in real-time or near-real-time [13].

In order to achieve desired real-time detection in the SAR-ATR system, some researchers have

attempted to imitate the HVS. The HVS is often bombarded with large amounts of information and is still able to find a specific object in a complicated scene filled with various distractors. This success benefits from the visual attention mechanism, which selects part of the visual inputs and transmits them to high-level cortical processing [16, 17]. Exploiting this pre-selection mechanism would facilitate real-time target detection algorithm.

Therefore a successful combination of VAM coupled with the SAR-ATR could potentially result in a detector with better performance than CFAR. There are already some studies relating to VAM based target detection in SAR images [18-22]. Though these methods differ from each other, they all focus on the detection of ships. In this case, there is only one specific type of clutter, i.e. sea clutter, and the target number in these studies is less than 5 (apart from the work in [18]), which simplifies the detection task. And nearly all of them are merely a direct application of VAM followed by a simple threshold function. These studies can be viewed as attempts to apply VAM in SAR-ATR rather than complete algorithms. Our work, however, attempts to detect 20 targets in complicated background of grass and woods and will concentrate on this. Though it may not be directly applicable for the detection of ship or other targets, its idea could be referred to and inspiring.

This paper presents a novel complete vehicle target detection algorithm based on VAM for SAR images. The innovative aspects of our algorithm comprise the following: original application of VAM in vehicle target detection for SAR images; Adaptation of the powerful Itti model for SAR image application by eliminating the color channel and reducing pyramid scales; Adoption of top-down processing steps using prior knowledge to remove false alarms in the detection result. Surely it does not present an algorithm of breakthrough in SAR image target detection. But it is the first time VAM is applied to detect vehicle targets in complicated background of grass and woods while previous work

using VAM can only deal with simple background of sea. Our algorithm operates in two main stages: the first stage involves selecting salient regions with the modified Itti model; the second is top-down processing using prior knowledge. The first stage can be viewed as imitating the HVS in predicting where the human will look in a SAR image, whereas the second stage attempts to remove the clutter as the HVS ignores irrelevant distractors.

2. Proposed Method

2.1 Modified Itti Model

The pre-selection mechanism of the HVS plays a key role in processing visual scenes by locating potential ROIs and reducing the data transmitted to high level visual processing. Inspired by this cognitive mechanism, we apply VAM in our SAR-ATR system to locate potential targets and reduce SAR data transmitted to later processing modules. We choose the well-developed Itti model and modify it according to our application requirement. The input of the model is in the form of static intensity image $I(x, y)$. The details of the modified Itti model are described in the steps below.

Step 1: Use dyadic Gaussian pyramids [23] to create five spatial scales. The pyramid is built by low-pass filtering the original image and subsampling it progressively. The Gaussian low-pass filter is given by:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp(-\frac{x^2 + y^2}{2\sigma^2}) \tag{1}$$

where (x, y) is the coordinate of a pixel and $\sigma \in \{1, 2, \cdots, 5\}$ is the scale parameter.

Step 2: Use oriented Gabor pyramids $O(\sigma, \theta)$ to obtain local orientation information, where $\sigma \in \{1, 2, \cdots, 5\}$ represents the scale and $\theta \in \{0°, 45°, 90°, 135°\}$ is the preferred orientation [23]. The Gabor filter is:

$$H(x,y,\sigma,\theta)=\frac{1}{\sigma^2}\exp(-\pi\frac{x^2+y^2}{\sigma^2})\{\exp[i2\pi(x\cos\theta+y\sin\theta)]-\exp(-\frac{\pi^2}{2})\} \quad (2)$$

Step 3: Compute each feature map by performing a set of linear "center-surround" operations. Center-surround is implemented as the difference between fine and corresponding coarse scales. The center is a pixel at scale $c\in\{2,3\}$, and the surround is the pixel at scale $s=c+d$, with $d=3$. The intensity and orientation feature maps are respectively:

$$I(c,s)=\left|I(c)\Theta I(s)\right| \quad (3)$$

$$O(c,s,\theta)=\left|O(c,\theta)\Theta O(s,\theta)\right| \quad (4)$$

where "Θ" is computing the across-scale difference between two maps by interpolation to the finer scale and point-by-point subtraction. This "center-surround" operation is computational imitation of the typical visual neurons which are most sensitive in a small region of the center of the visual space [4]. Compared with the original Itti model, the center scale *c* and delta scale *d* is reduced because experiments show that reduced scale does not affect the detection result much and permit more time economy. In total, 10 feature maps are computed: 2 for intensity and 8 for orientation.

Step 4: Normalize the feature maps obtained in the previous step by applying operator $N[\cdot]$ to them. This operator consists of the following steps [4]: normalize the values in the map to a fixed range $[0,M]$; locate the map's global maximum *M* and compute the average *m* of all its other local maxima; multiply the map by $(M-\bar{m})^2$.

Step 5: Combine the normalized feature maps into two "conspicuity maps", by across-scale adding:

$$\bar{I}=\bigoplus_{c=2}^{3}\bigoplus_{s=c+3}N[I(c,s)] \quad (5)$$

$$\bar{O}=\sum_{\theta\in\{0°,45°,90°,135°\}}N[\bigoplus_{c=2}^{3}\bigoplus_{s=c+3}N[O(c,s,\theta)]] \quad (6)$$

where "$\oplus$" consists of reduction of each map to scale three and point-by-point addition.

Step 6: Normalize the two conspicuity maps and sum into the final saliency map *S*:

$$S = N[\bar{I}] + N[\bar{O}] \tag{7}$$

Saliency map is a prediction of the locations where the HVS will look. In our algorithm, saliency map will direct subsequent computation resource to process the regions which are most likely to contain a target. In contrast, traditional CFAR algorithm processes every pixel with the same computation. Equal amount of computation resource is allocated to every pixel in spite of their different possibility of containing a target. It is obvious that this computational scheme is the main reason of CFAR's deficiency.

So far we have provided every detail of the modified Itti model and we could now compare it with the original one. The modified one discards the colour channel and reduces the scales in which analysis is performed. The reason of the former modification is that SAR images are grey-scale image. A color channel will only generate meaningless result while wasting computation resourse. The reason of the latter one is that reduced scales permit time economy without severely affecting the result. The result of using the original Itti model for our detection purpose is far from satisfactory and consumes more time. So we do not demonstrate it in this paper.

An ideal VAM only presents the regions containing a target. Our current model, however, is not an ideal one. Therefore model parameters are set to let in more regions so that targets will not be missed. An inevitable consequence of this is that more regions of clutter can be seen in the saliency map. We will use other methods to deal with these in the next section.

2.2 Top-Down Processing

The modified Itti model provides us with locations where the HVS will allocate attention. Next we

need to scrutinize these areas to check whether targets really exist there. In a saliency map, a brighter area is known to indicate greater attention allocation. The areas located by our modified Itti model are obviously easier to process than the complete SAR image. Most areas of clutter from grass, roads and trees are already rejected in the visual attention process. In fact, when humans look at such a SAR image, they would quickly identify those areas of grass and trees by using the split version without the need to carefully examine them. Next we need to scrutinize the saliency map and check how many targets are contained in the regions suggested by the modified Itti model. The proposed steps are outlined below.

Step 1: Apply a threshold to the saliency map:

$$T_1 = \mu_S + c\sigma_S \tag{8}$$

$$R_s(x,y) = \begin{cases} 1 & S(x,y) > T_1 \\ 0 & S(x,y) \le T_1 \end{cases} \tag{9}$$

where μ_S and σ_S are the mean and standard deviation of the saliency map respectively and c is set to 1.1 based on experiment. We term $R_s(x,y)$ the region of saliency.

Step 2: Use the area of targets to remove those regions which are not likely to contain a target. Those regions with a very small area contain only some isolated strong reflection points; whereas those with a large area are woods. Some clutter regions can be removed in this step, using the following:

$$R_{s_i}(x,y) = 0 \quad \text{when } A[R_{s_i}(x,y)] \le 0.85 \times \min\{A_j\} \text{ or } A[R_{s_i}(x,y)] \ge 10 \times \max\{A_j\} \tag{10}$$

where $R_{s_i}(x,y)$ is the ith region of $R_s(x,y)$, A_j is the area of the jth target and $A[\cdot]$ is calculating the area. $R_s(x,y)$ processed by Eq. 10 regions is considered to be scrutinized. To be specific, the area of a target is computed by applying the same thresholding process to the target images provided by MSTAR. Of course, these values in Eq.(10) will probably be invalid for other SAR images. They are chosen not for their portability in all conditions, but to demonstrate the idea of using simple threshold

to achieve desired function of locating the possible areas containing target. To be noted, woods do not appear in every SAR images. For our experiments, we use those images with woods, which is clearly a more complicated case than those without woods.

Under low signal clutter ratio (SCR) conditions, the area of a region with a target is smaller than its normal number because the VAM will consider it to be less salient. Therefore we consider those regions whose area is less than $\min\{A_j\}$ but larger than $0.85\times\min\{A_j\}$ to be possible candidates. Under high SCR conditions, it is the opposite situation and we consider those regions with area less than $10\times\max\{A_j\}$ but larger than $\max\{A_j\}$ to be possible candidates. Clutter regions are further removed so that less data will be transmitted to the next processing unit. In this step, we use the prior knowledge of the targets, i.e. area, so it is viewed as a top-down processing step.

In terms of a physical interpretation of this step, the HVS may look at some regions due to their outstanding brightness, but will move to the next region quickly if this region is too small (treated as an isolated strong reflection point) or too large (for the case of woods). Of course, this behavior is determined by the application task at hand: finding vehicle targets in our case. If the task is aimed at finding woods, the HVS will behave differently. The equations above and below are used to mimic the HVS's possible behavior.

Step 3: Obtain the regions of the original image suggested by the result of the previous step and apply a threshold to it. Although the result of the previous step offers information about areas the HVS will scrutinize, it does not tell us what the HVS really sees in these regions. This can be obtained as follows:

$$I_S(x,y) = I(x,y) \otimes R_s(x,y) \tag{11}$$

$$I_s^{'}(x,y) = \begin{cases} 1 & I_s^{'}(x,y) > T_2 \\ 0 & I_s^{'}(x,y) \le T_2 \end{cases} \tag{12}$$

$$T_2 = \mu_I + d\sigma_I \tag{13}$$

where "$\otimes$" is point-by-point multiplication, μ_I and σ_I are the mean and standard deviation of the original SAR image respectively, and *d* is empirically set to 2 (based on experiments). After threshold processing, we perform an open operation with a 6×6 square structuring element followed by a close operation with a 4×4 square structuring element. This affiliated step aims at joining narrow breaks and removing small clutter. Here we only used fixed structuring elements because we try to focus on the development of the detection algorithm and future work could be done to adaptively choose the size of the structuring elements. The same reason also applies to the choice of several parameters in later process.

Step 4: Use prior information to further remove false alarms. If the area of a certain region in $I_s^{'}(x,y)$ is too small compared with the corresponding region in $R_s(x,y)$, remove this region as follows:

$$I_s^{'}(x,y) = 0 \quad \text{when } A[I_{s_i}^{'}(x,y)] \le 0.08 \times A[R_{s_i}(x,y)] \tag{14}$$

Eq. 14 above is attempting to remove those isolated strong reflection points not removed previously by Eq. 10. If there are several separated parts in one region to be scrutinized, those separated parts are removed as they are isolated strong reflection points. Note that the vehicle target is an area target and will not appear as several separated parts. Next, use area to remove clutter as follows:

$$D(x,y) = \begin{cases} 1 & A[I_{s_i}^{'}(x,y)] \ge 0.6 \times \min\{A_j\} \text{ or } A[I_{s_i}^{'}(x,y)] \le 1.1 \times \max\{A_j\} \\ 0 & \text{else} \end{cases} \tag{15}$$

The targets' area will fluctuate when the SCR changes. Especially when SCR is low, the area detected is much smaller than its ordinary number. Hoping not to miss targets with low SCR, we do not use $A[R_{s_i}(x,y)] \ge \min\{A_j\}$ or $A[R_{s_i}(x,y)] \le \max\{A_j\}$ as the possible range for targets' area in Eq. 15. Instead we use $A[R_{s_i}(x,y)] \ge 0.6 \times \min\{A_j\}$ or $A[R_{s_i}(x,y)] \le 1.1 \times \max\{A_j\}$. Still, some clutter has

similar area with targets and also survives previous elimination steps. As the first stage of a SAR-ATR system, a detection module should detect as many targets as possible. False alarms are acceptable as long as the number is not too large, as an appropriate follow-up module can deal with these. On the other hand, the detection module should not be too complicated, so we will not add any other processing steps in our algorithm.

This section has described all the steps of our proposed VAM based vehicle target detection algorithm. A flowchart of the complete algorithm is given in Fig. 1. From the above discussion, we could see that we only use area characteristic for our detection purpose. Adding other features or even models will probably increase the detection performance. However, as has been mentioned early in this paper, we attempts to develop an algorithm for detector, the first stage of a SAR-ATR system and should concentrate on detecting as many as target candidates without using too much time. Adding more characteristics will improve detection performance but also lead to significant time consumption. So we just keep it simple and efficient.

3. Simulation Results and Discussion

In this section, we will carry out a comparative evaluation of the proposed VAM based vehicle target detection algorithm using SAR images and the final detection results of the proposed algorithm compared with the CFAR. The simulation experiments are aimed at demonstrating the validity and effectiveness of the proposed algorithm.

This section is composed of three parts. The proposed algorithm is first used to process a SAR image and the intermediate results of each step are presented to show the function of each step. Then the proposed algorithm is compared with CFAR in detecting vehicle targets under different SCR

conditions. The curves of probability of detection (PD) and number of false alarms are computed to illustrate the comparative performance of both algorithms. Finally, the execution time of both algorithms is compared.

The images used in our experiments are real MSTAR amplitude images with a resolution of 0.3m and size 1748×1478. The well-known MSTAR public database was collected using the Sandia National Laboratories Twin Otter SAR sensor payload operating at X band with a high resolution of 0.3 m, spotlight mode and HH single polarization. The original image, examples of vehicle targets and an image with 20 targets added with $SCR = 2.0$ (In our work, SCR is the mean value of the target pixlel divided by the mean value of the background pixel in a 100×100 area) are shown in Figs. 2-4 respectively. Vehicle targets have been marked in Fig. 4.

3.1 Simulation Results Using the Proposed Algorithm

The saliency map of Fig. 4 (unmarked version) is shown in Fig. 5. As can be seen from Fig. 5, the modified Itti model successfully locates every target although many clutter regions from woods are also present. It should be noted that speckle noise does not cause the undesired influence in saliency map as it does in the CFAR algorithm. This is attributed to the modified Itti model which ignores the speckle noise, and is consistent with the behavior of the HVS in real scene analysis. As long as the noise is not too strong, the HVS will still be able to ignore the noise and detect targets. On the other hand, the modified Itti model also ignores some strong reflection points, which will be present in the CFAR detection result.

The region of saliency is obtained by applying threshold to Fig. 5 and is shown in Fig. 6. (To be noted, Fig. 6 and later images have been inverted in color for ink economy in printing. And in order to

make the boundry of each image clear, each of them has been added with a black frame.) There are several very small regions in Fig. 6 which is evidence that the modified Itti model is not a perfect duplication of the HVS. The HVS will ignore these small regions naturally. Despite this, the performance of our proposed VAM based algorithm is very satisfying. The very large regions in Fig. 6 are woods and should also be removed. Figure 7 presents the result of Fig. 6 after the removal of obvious non-target regions using Eq. 10, and what our VAM "sees" in these regions is given in Fig. 8. When the HVS looks at each region suggested by Fig. 7, it will see the corresponding shape in Fig. 8. The regions with a very small area or several broken parts are not targets and should therefore be removed. Figure 9 gives the result of Fig. 8 after removing broken parts and very small regions. Finally after applying Eq. 15 to Fig. 9, we obtain the final detection result using the proposed algorithm, illustrated in Fig. 10. The detected targets have been marked and there are 20 targets detected, indicating 100% of the targets added to the original SAR image are detected. The number of false alarms is 7, which is quite acceptable.

The detection result of the same SAR image using CFAR is shown in Fig. 11. The detected targets have been marked and there are 14 targets detected, indicating 70% of the targets added to the original SAR images are detected. The number of false alarm is 9. We can conclude from Fig. 10 and 11 that the proposed vehicle target detection algorithm gives better results for Fig. 2 both in terms of PD and number of false alarms.

To validate the generalization of the proposed algorithm, we test the proposed algorithm on images with different backgrounds. The simulation results also demonstrate its effectiveness and advantage over CFAR. One of the detection results is shown in Fig. 12-14. Fig. 12 is the original image with 20 vehicle targets added, Fig. 13 and 14 are respectively the detection result using the proposed

algorithm and CFAR. It can be seen from Fig. 13 and 14 that the proposed algorithm detects all the targets while CFAR only detects 9. As for the number of false alarm, both our method and CFAR produce 9. Another example is shown in Fig. 15-17. In this case, the proposed algorithm still manages to detect every target while CFAR detects only 12. Though CFAR produces less false alarm in this situation, the result of CFAR is not considered superior to that of the proposed method. This is because, as has been pointed out early in this paper, detector is the first stage of SAR-ATR and should focus on the detection of targets. From the above discussion and corresponding figures, we can conclude that the proposed algorithm is an effective one and can be applied in different situations.

3.2 Vehicle Target Detection Using the Proposed Algorithm and comparison with CFAR

In order to further validate the effectiveness of the proposed algorithm, we used both algorithms to detect SAR images with 20 vehicle targets randomly added under different SCR conditions. SCR ranged from 0.8 to 4.0. The position of the targets was randomly generated. The detection result under every SCR condition was computed as the average from 30 independent runs of experiments(10 runs for each background image shown in Fig.4, 12, 15. The location of the target of each run is different since it is generated randomly). The curves of PD and number of false alarms are shown in Figs. 18 and 19 respectively. The probability of detection is calculated using:

$$P_d = \frac{\text{Number of detected targets}}{\text{Total number of targets}} \tag{16}$$

We can see from Fig. 18 that the proposed algorithm can detect more targets under every SCR condition compared to CFAR. Specifically, when the SCR is 0.8, the proposed algorithm achieves a PD of 83.8%, and for the case of SCR larger than 1.0, the proposed algorithm can detect all targets. In contrast, the PD of CFAR never exceeds 80% and fluctuates frequently and sometimes even falls below

40%. The proposed algorithm, on average, detects 7.9 targets more than the CFAR approach.

Next we turn to the number of false alarms. The number of false alarms of the proposed algorithm is found to be below 9 under all SCR conditions. In contrast, the number of false alarms of CFAR usually fluctuate around 13 and even exceed 25 at times. The proposed algorithm, on average, generates 6.6 false alarms less than CFAR.

3.3 Execution Time of the Proposed Algorithm and CFAR

Finally, we compare the computational complexity of the two algorithms by using the CPU execution time of programs as estimation. MATLAB non-optimized routines of the proposed algorithm and CFAR were implemented on an Intel Dual 2.5-GHz CPU with 2GB RAM. The execution times were estimated at about 6.6s and 227s respectively. The time consumption of the proposed algorithm is thus only 3% of that of the CFAR which demonstrates the outstanding computational efficiency of the proposed algorithm.

4. Conclusion

In this paper, a novel visual attention model based vehicle target detection algorithm has been developed for SAR images. The proposed algorithm modifies the well-known Itti model to suit the requirement and specialty of SAR image applications and applies a top-down processing stage to remove clutter using prior knowledge. The performance of the proposed algorithm has been comparatively evaluated using real SAR images with 20 vehicle targets added under different SCR conditions and in 3 different SAR images. Experimental results have demonstrated that the proposed algorithm can achieve high performance in PD and minimization of number of false alarms

simultaneously, with very little time consumption. The proposed algorithm, on average, detects 7.9 more targets and 6.6 less false alarms than the conventional CFAR algorithm, and its execution time is only 3% that of CFAR. This shows that our attempt to imitate the human visual system in detecting target by applying the VAM and top-down processing is successful and could possibly inspire other researchers in their study of SAR image target detection. We believe this work makes it possible to detect vehicle targets using SAR in real-time applications and serves as a promising example of how the HVS can effectively contribute to object detection and recognition. Our current and future work is focused on further evaluating the new algorithm, for real-time implementation, using additional real SAR data benchmarked against other state-of-the-art approaches, in addition to exploring new visual attention models.

Acknowledgments

This research is supported by the National Natural Science Foundation of China, No. 61071139 and No. 61171122, the Foundation of ATR Key Lab, the Fundamental Research Funds for the Central Universities and "New Star in Blue Sky" Program Foundation of Beihang University. This research is also funded in part by the Royal Society of Edinburgh (RSE) and The National Natural Science Foundation of China (NNSFC) under the RSE-NNSFC joint project (2012-2014) [grant number 61211130309] with the University of Stirling, Scotland, UK.

References

1. Koch C, Ullman S. Shifts in selective visual attention: towards the underlying neural circuitry. Human neurobiology. 1985 1985;4(4):219-27. PubMed PMID: MEDLINE:3836989.

2. Kimura A, Yonetani R, Hirayama T. Computational Models of Human Visual Attention and Their Implementations: A Survey. Ieice Transactions on Information and Systems. 2013 Mar;E96D(3):562-78. PubMed PMID: WOS:000320214300024.

3. Treisman AM, Gelade G. A feature-integration theory of attention. Cognitive psychology. 1980;12(1):97-136.

4. Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 1998;20(11):1254-9.

5. Walther D, Koch C. Modeling attention to salient proto-objects. Neural Networks. 2006 Nov;19(9):1395-407. PubMed PMID: WOS:000242857400006.

6. Bruce N, Tsotsos J, editors. Saliency based on information maximization. Advances in neural information processing systems; 2005.

7. Harel J, Koch C, Perona P, editors. Graph-based visual saliency. Advances in neural information processing systems; 2006.

8. Hou X, Zhang L, Ieee. Saliency detection: A spectral residual approach. 2007 Ieee Conference on Computer Vision and Pattern Recognition, Vols 1-8. Proceedings - Ieee Computer Society Conference on Computer Vision and Pattern Recognition2007. p. 2280-7.

9. Guo C, Ma Q, Zhang L, Ieee. Spatio-temporal saliency detection using phase spectrum of Quaternion Fourier Transform. 2008 Ieee Conference on Computer Vision and Pattern Recognition, Vols 1-12. Proceedings - Ieee Computer Society Conference on Computer Vision and Pattern Recognition2008. p. 2908-15.

10. Chenlei G, Liming Z. A Novel Multiresolution Spatiotemporal Saliency Detection Model and Its Applications in Image and Video Compression. Image Processing, IEEE Transactions on.

2010;19(1):185-98.

11. Wolfe JM, Cave KR, Franzel SL. Guided search: an alternative to the feature integration model for visual search. Journal of Experimental Psychology: Human perception and performance. 1989;15(3):419.

12. Itti L, Koch C. Computational modelling of visual attention. Nature reviews neuroscience. 2001;2(3):194-203.

13. El-Darymli K, McGuire P, Power D, Moloney C. Target detection in synthetic aperture radar imagery: a state-of-the-art survey. Journal of Applied Remote Sensing. 2013;7(1):071598-.

14. Di Bisceglie M, Galdi C. CFAR detection of extended objects in high-resolution SAR images. Geoscience and Remote Sensing, IEEE Transactions on. 2005;43(4):833-43.

15. Kuttikkad S, Chellappa R, editors. Non-Gaussian CFAR techniques for target detection in high resolution SAR images. Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference; 1994: IEEE.

16. Crick F, Koch C. Constraints on cortical and thalamic projections: the no-strong-loops hypothesis. Nature. 1998;391(6664):245-50.

17. Desimone R, Duncan J. Neural mechanisms of selective visual attention. Annual review of neuroscience. 1995;18(1):193-222.

18. Biao H, Wei Y, Shuang W, Xiaojin H, editors. SAR image ship detection based on visual attention model. Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International; 2013 21-26 July 2013.

19. BO H, GU H, SUN Q. Visual Attention Based SAR Image Small Target Detection. Journal of Computational Information Systems. 2013;9(1):179-86.

20. Gao L, Bi F, Yang J. Visual attention based model for target detection in large-field images. Systems Engineering and Electronics, Journal of. 2011;22(1):150-6.

21. Yu Y, Wang B, Zhang L. Hebbian-based neural networks for bottom-up visual attention and its applications to ship detection in SAR images. Neurocomputing. 2011 May;74(11):2008-17. PubMed PMID: WOS:000291837800025.

22. Amoon M, Bozorgi A, Rezai-rad G-a. New method for ship detection in synthetic aperture radar imagery based on the human visual attention system. Journal of Applied Remote Sensing. 2013 Feb 19;7. PubMed PMID: WOS:000315269900001.

23. Greenspan H, Belongie S, Goodman R, Perona P, Rakshit S, Anderson CH, editors. Overcomplete steerable pyramid filters and rotation invariance. Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on; 1994: IEEE.

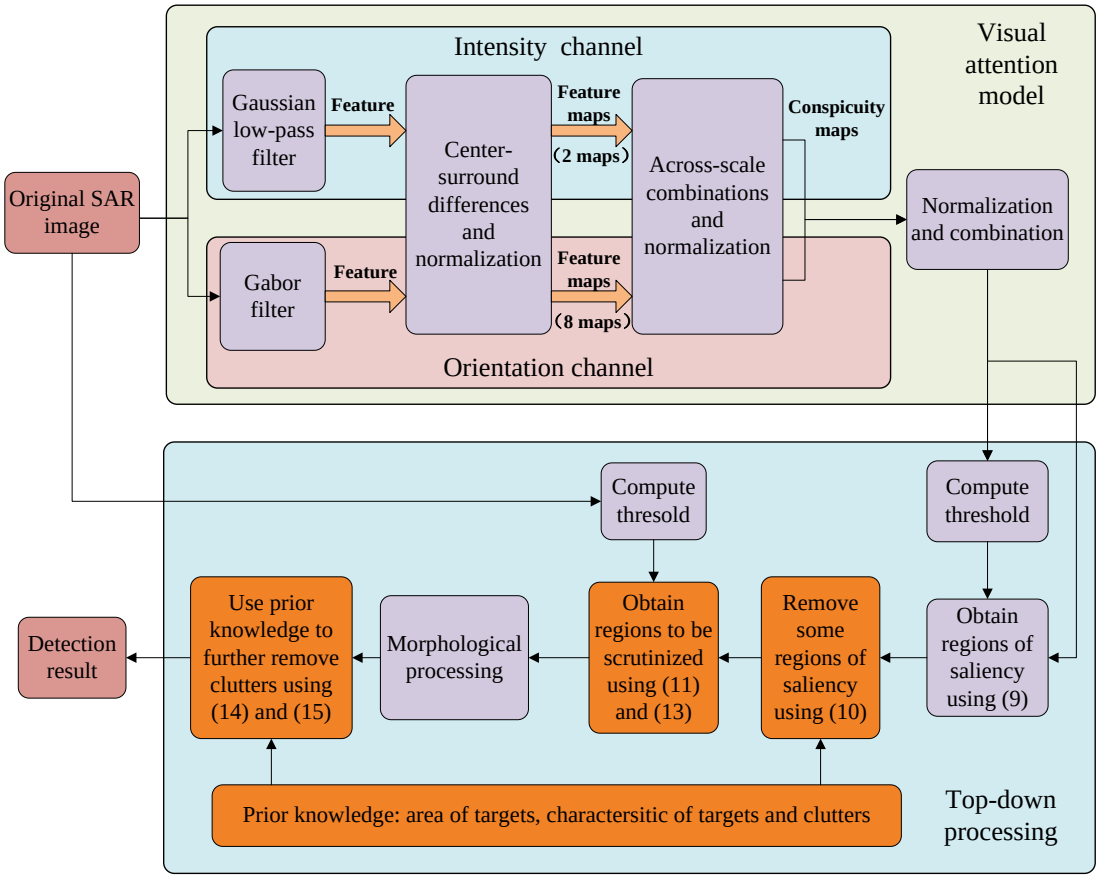

Fig. 1 Flowchart of the proposed visual attention model based vehicle target detection algorithm for SAR images. This algorithm is composed of two parts: the modified Itti model and top-down processing.

Fig. 2 Original SAR image.

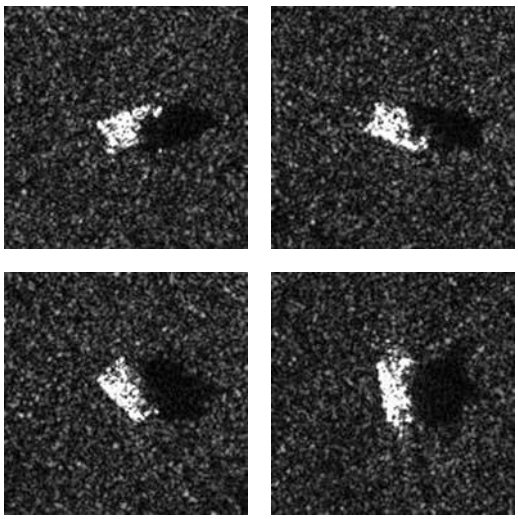

Fig. 3 Examples of vehicle targets to be added into the original SAR image

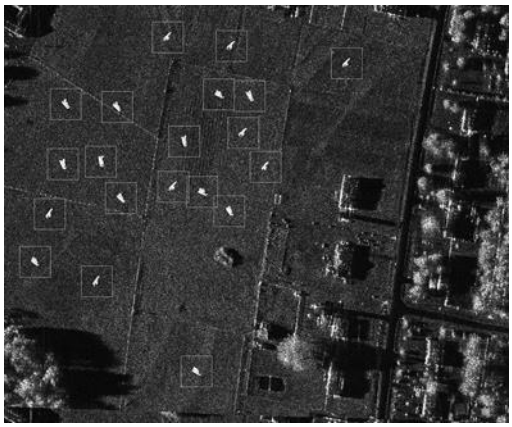

Fig. 4 SAR image with 20 targets added. It is generated by adding vehicle targets into Fig. 2 with SCR=2.0. The targets have been marked.

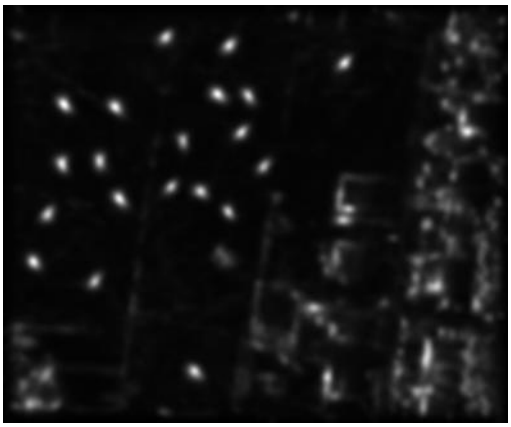

Fig. 5 Saliency map of Fig. 4 - resulting from application of the modified Itti model to Fig. 4 (unmarked version).

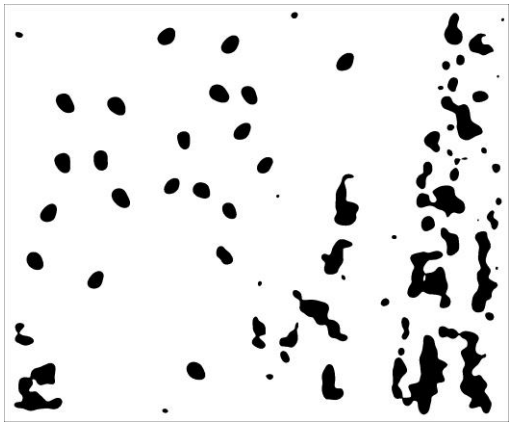

Fig. 6 Region of saliency - resulting from application of Eq. 9 to Fig 5 - showing the salient regions of Fig. 4.

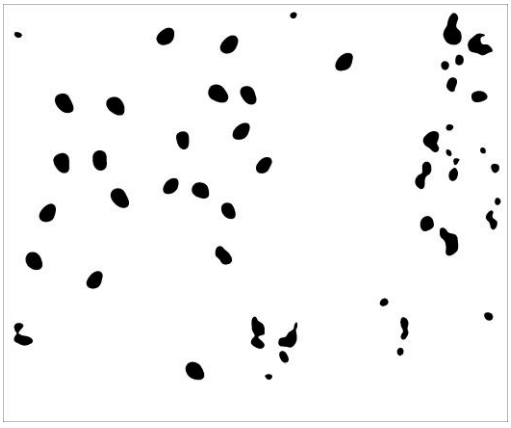

Fig. 7 Regions to be scrutinized - result of removing very small and large regions in Fig. 6.

Fig. 8 What our VAM "sees" in regions to be scrutinized (obtained using Eq. 11-13)

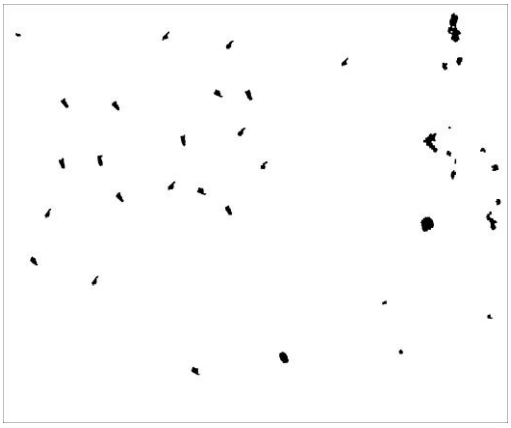

Fig. 9 Result of Fig. 8 after removing separated parts in one region (to be scrutinized) and very small regions using Eq. 14.

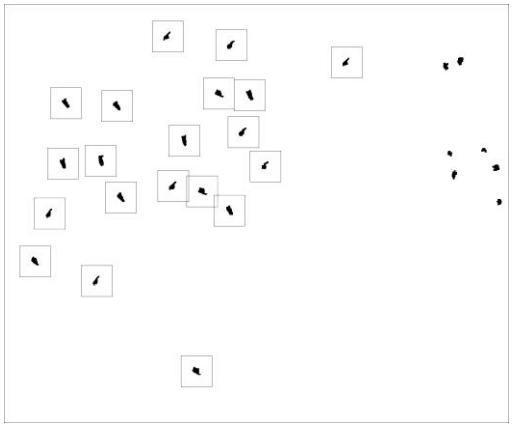

Fig. 10 Detection results for Fig. 4 using the proposed algorithm

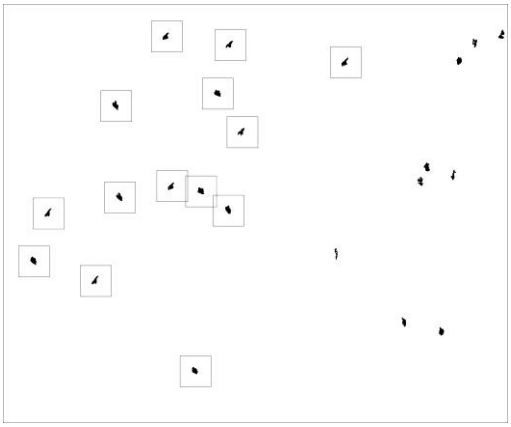

Fig. 11 Detection results for Fig. 4 using CFAR

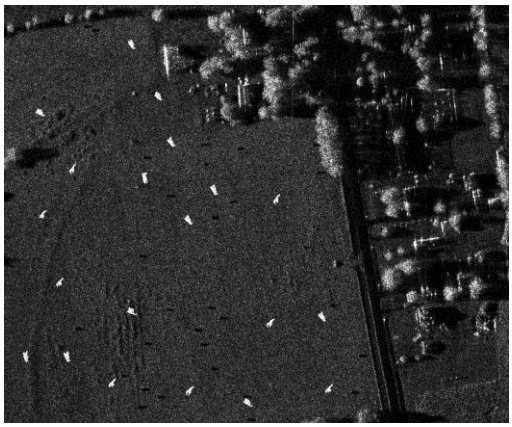

Fig. 12 Example 2: SAR image with 20 vehicle targets added

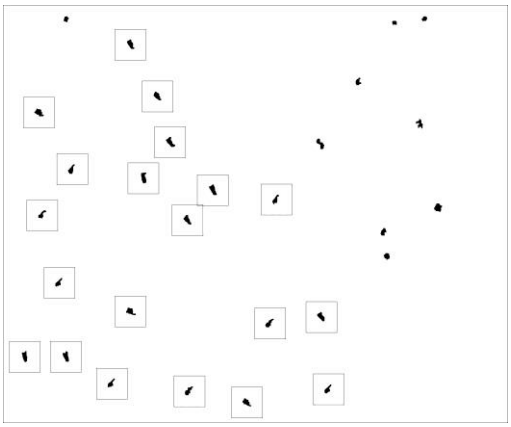

Fig. 13 Detection result of Fig. 12 using the proposed algorithm

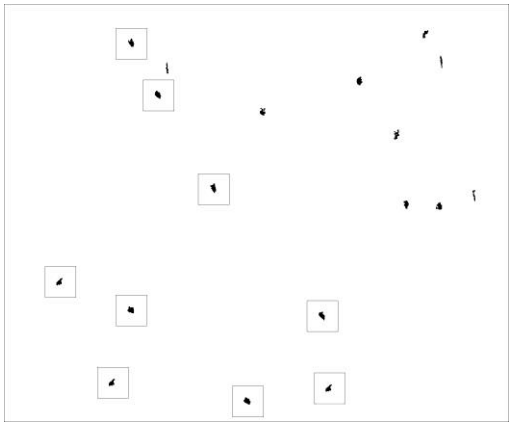

Fig. 14 Detection result of Fig. 12 using CFAR

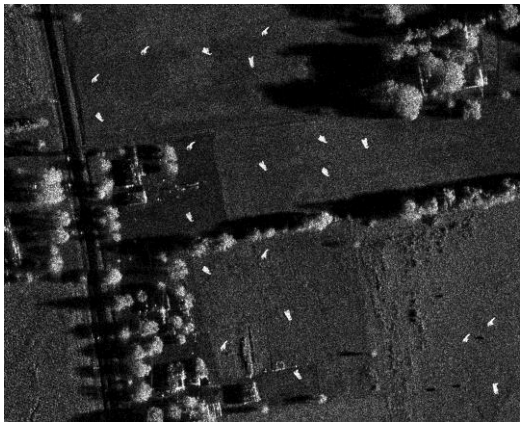

Fig. 15 Example 3: SAR image with 20 vehicle targets added

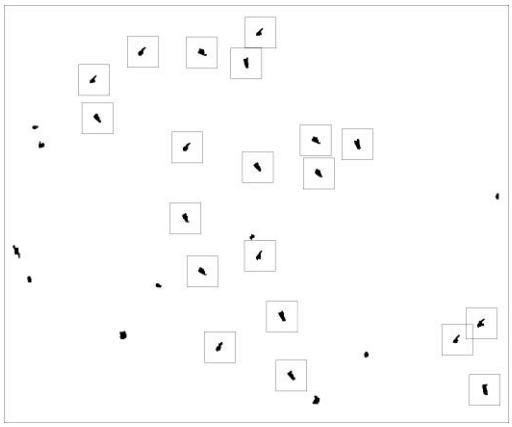

Fig. 16 Detection result of Fig. 15 using the proposed algorithm

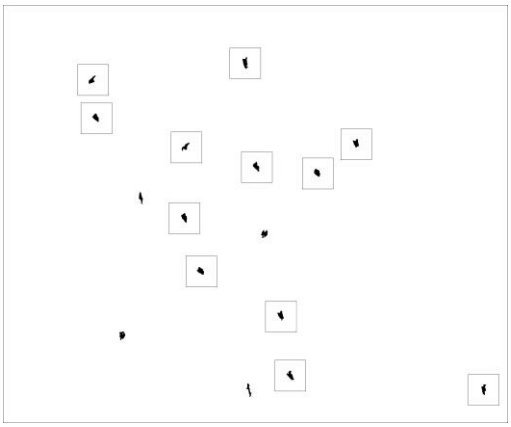

Fig. 17 Detection result of Fig. 15 using CFAR

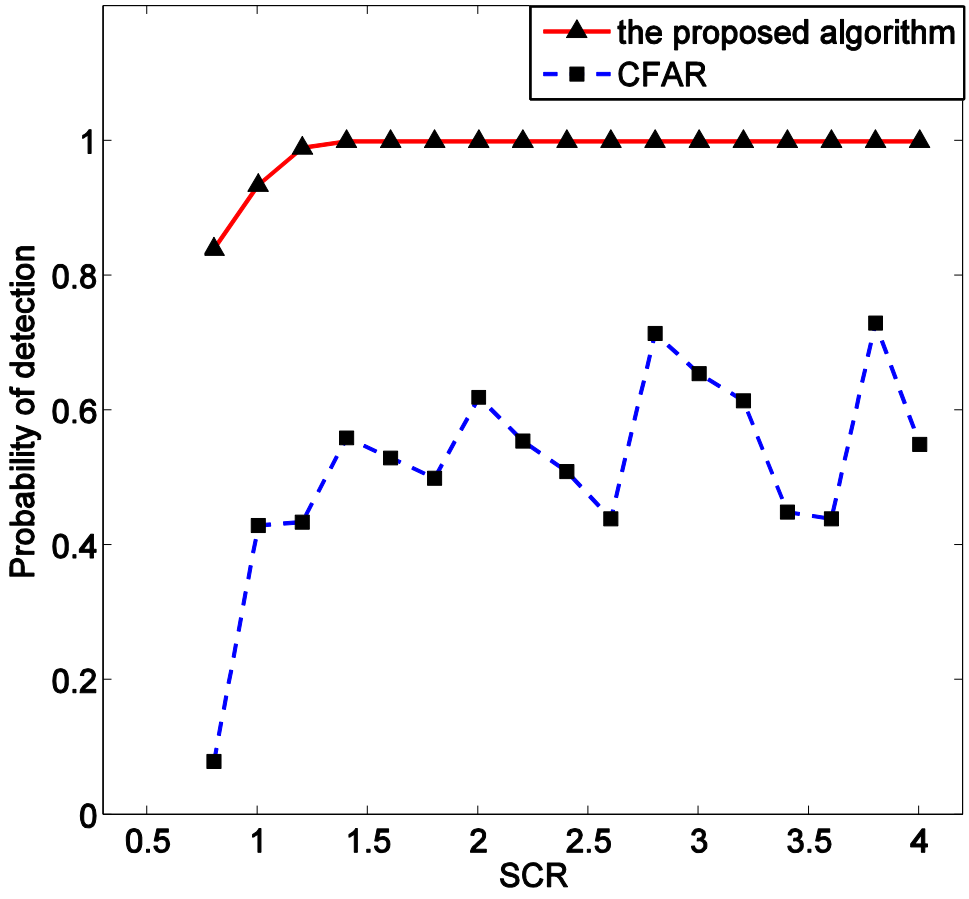

Fig. 18 The curves of probability of detection (PD) for the proposed algorithm and CFAR. The range of SCR is from 0.8 to 4.0.

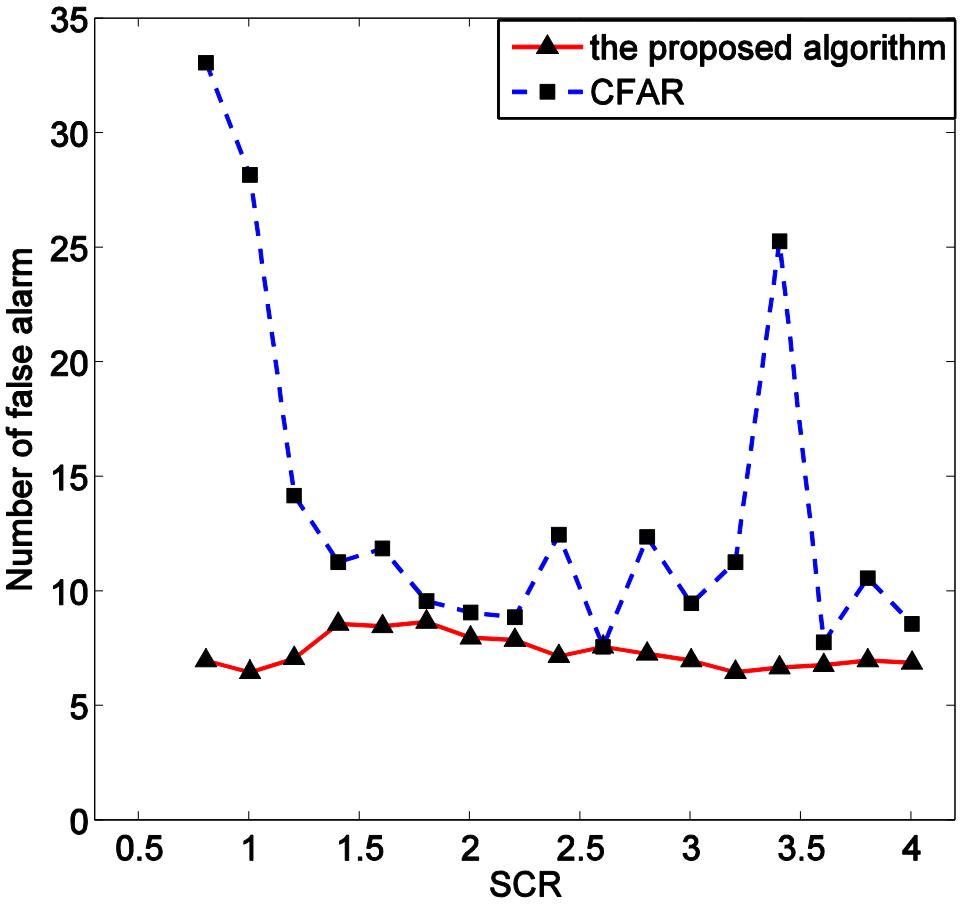

Fig. 19 The curves of number of false alarms for the proposed algorithm and CFAR. The range of SCR is from 0.8 to 4.0.