Author identifying information

Fei Gao[1]• Fei Ma[1]•Yaotian Zhang[1]*•Jun Wang[1]•Jinping Sun[1] •Erfu Yang[2]•Amir Hussain[3]

1. School of Electronic and Information Engineering, Beihang University, Beijing 100191, China

2. Space Mechatronic Systems Technology Laboratory, Department of Design, Manufacture and Engineering Management, University of Strathclyde, Glasgow G1 1XJ, UK

3. Cognitive Signal-Image and Control Processing Research Laboratory, School of Natural Sciences, University of Stirling, Stirling FK9 4LA, UK

*Corresponding Author, e-mail: zhangyaotian@buaa.edu.cn, Telephone: 010-82317240, Fax: 010-8231724

Biologically-Inspired Progressive Enhancement Target Detection from Heavy Cluttered SAR Images

Abstract High-resolution SAR (synthetic aperture radar) can provide a rich information source for target detection and greatly increase the types and number of target characteristics. How to and efficiently extract the target of interest from large amounts of SAR images is the main research issue. Inspired by the biological visual systems, researchers have put forward a variety of biologically-inspired visual models for target detection, such as classical saliency map and HMAX. But these methods only model the retina or visual cortex in the visual system, which limit their ability to extract and integrate targets characteristics, thus their detection accuracy and efficiency can be easily disturbed in complex environment. Based on the analysis of retina and visual cortex in biological visual systems, a progressive enhancement detection method for SAR targets is proposed in this paper. The detection process is divided into RET, PVC, and AVC three stages which simulate the information processing chain of retina, primary and advanced visual cortical, respectively. RET stage is responsible for eliminating the redundant information of input SAR image, enhancing inputs' features, and transforming them to excitation signals. PVC stage obtains primary features through the competition mechanism between the neurons and the combination of characteristics, and then completes the rough detection. In the AVC stage, the neurons with more receptive field compound more precise advanced features, completing the final fine detection. The experimental results obtained in this study show that the proposed approach has better detection results in comparison with the traditional methods in complex scenes.

Keyword: Cortex-Like Mechanisms, synthetic aperture radar (SAR), Hierarchical models, target detection

Introduction

With the increase of SAR systems and the improvement of their image quality, the targets in SAR images contain richer detail features. The common SAR target detection systems adopt the two-parameter Constant False Alarm Rate (CFAR) algorithm based on Gaussian distribution proposed by Lincoln Laboratory [1]. It takes advantage of differences between the statistical model [2] of background and the targets to achieve targets detection. Whereas, imaging scenes of SAR often contain a large number of artificial buildings, trees, grass， etc, which form heavy clutters [3] in images and reduce Signal to Clutter Ratio (SCR). The randomness of the clutter makes the estimation of background statistical models more difficult. Moreover, since this method does not make full use of outline, texture, shape and many other features of targets, its detection results are unsatisfied especially under low SCR condition.

Biological visual system can search the targets or areas of interests quickly and accurately through properly extracting and processing various kinds of target features, without knowing the model of clutters. Inspired by the unique processing mechanism of the biological vision systems we propose a new idea for the study of SAR target detection. The development of cognitive psychology, neurophysiology, and other related science has unceasingly deepened our understanding for biological vision system [4], which provides the possibility for this attempt.

The biological visual system is a complicated multi-level system [5], where low-level areas firstly receive inputs, and then processes these visual information and compound more complex features before transferring them to the higher level. Its main information processing units include retina, primary visual cortex (PVC) and advanced visual cortex (AVC).

Being the first part of biological visual system, retina realizes the first step, i.e., input processing. The retina includes the receptor cells, bipolar cells and ganglion cells from the outside to the inside. Specifically, with the aid of unique ON-center or OFF-center receptive fields, ganglion cells can remove some redundant information by ignoring those fields with no changes and transmit the interested information to the brain in the form of pulses. Rodieck [6] first proposed that the Difference of Gaussian (DOG) can be used to simulate these receptive fields, which has been applied in many saliency map models, e.g. Itti [7]. Whereas the DOG represents the initial stage of the visual processing, it is only sensitive to the limited simple feature and unable to manage the higher order feature. The saliency map [8] is easily influenced by the background clutters, which could only be applied in the relatively simple scenes.

The pulse signals are transferred to the primary visual cortex (PVC) through the lateral geniculate body. David Hubel and Toresten Wiesel [9] found that most of the neurons in primary visual cortex have directional selectivity, i.e., they are sensitive to some fine bright bar (or dark bar) in a certain direction. Jones [10] proposed a 2D Gabor filter with similar property which achieves optimal localization properties in both spatial domain and frequency domain.

Biological target search in the visual field is a gradual enhancement and clear process. The preliminary integration of information in the primary visual cortex could eliminate a large number of false alarms with obvious features in a quick look. Only a small amount of residual signal is fed into the advanced visual cortex for future processing. In the next level, the advanced Visual cortex (AVC) has neurons with larger receptive field, can respond to some subjective contour in a more sophisticated manner.

Based on the Hubel-Wiesel structure, Researchers made a number of heuristic target detection models. Amoon M [11] introduced Zernike moments and Particle Swarm Optimization (PSO) to features extraction and selection, improving target recognition efficiency. Tu et al. [12] proposed a moving target detection method based on independent component analysis (ICA) and principal component analysis (PCA). Ho-Phuoc et al. [13] established a statistical model of visual features in the focus prediction, available on compositing saliency maps. Poggio[14] put forward the HMAX model. HMAX model applies Gabor filtering and the local maximum pooling operation to simulate the mechanism of the primate visual cortex, which has excellent robustness to rotation, translation, and scaling. An airborne or space-borne SAR image may cover a very large area including hundreds of suspicious regions needed to identify, so the elimination of redundant information is necessary. However, the HMAX only simulates the visual cortex which greatly reduces the calculation efficiency because it has no screening role like the retina. In addition, this model adopts a same processing with respect to different suspicious targets regardless of feature types or amounts. A large number of unnecessary calculations limit the application of HMAX.

Because whether the saliency map or the HMAX can only model the partial function of biological visual systems, they cannot obtain the satisfactory results in both efficiency and effects at the same time. To solve this problem, by simulating the whole process of the biological vision system from the retina to the visual cortex, this paper presents a target detection method from SAR image to progressively enhance its performance. It takes advantage of the integration and eliminating the need of each unit to improve the efficiency of the information processing. Based on the collaborative mechanism of the retina, the primary and advanced visual cortex, the detection process is divided into the following three stages. The first stage simulates the retina, namely RET, which is responsible for extraction, distortion-rectifying and coding of the preliminary features of suspicious targets. The second stage PVC models the primary visual cortex to remove the false alarms for the objects having the obvious difference with the real targets by the “rough detection” based on the “primary feature” of each suspicious target. The third stage AVC plays the same role as the neurons in the advanced visual cortex. It further produces more accurate “advanced feature” for the remaining suspicious targets. With the help of the classifier, the final detection results are obtained through the “fine detection”. We conducted a series of experiments and made a comparison with the two-parameter CFAR to illustrate this proposed s method.

The specific frame of this paper is as follows. The second session describes each stage of the proposed method in details. The experiment results of each step in complex scene and the comparison with the CFAR are given in the session 3. The fourth session is the conclusion.

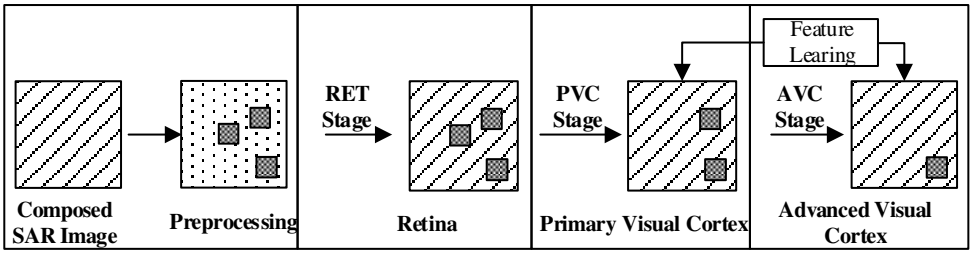


Fig1 Flowchart of the proposed method

Proposed Method

The proposed method is shown in

Fig1. A SAR image is firstly preprocessed to locate those suspicious regions where real targets may exist. In the RET stage, we use the DOG filter, the distortion rectifying and coding algorithm to handle the suspicious regions like retina in biological visual system. The PVC stage calculates the "primary feature" of each suspicious region through Gabor filtering and the maximum pooling operation. The AVC stage obtains more accurate "advanced features" to achieve the final results with the help of classifiers.

Preprocessing

The preprocessing includes the edge preserving, the gray level quantization and the connected component labeling. The result of the preprocessing for a composed SAR image is given in Fig. 2, where Fig. 2 (a) shows the composed SAR image; Fig. 2 (b) is the binary image after the preprocessing and the white areas indicate the suspicious areas. As shown in Fig. 2 (c), a total of N_T suspicious areas with the size of $W \times H$ extracted from (a) form a suspicious region set. Let $t_l \in T(l = 1, ..., N_T)$ represent the lth suspicious area.

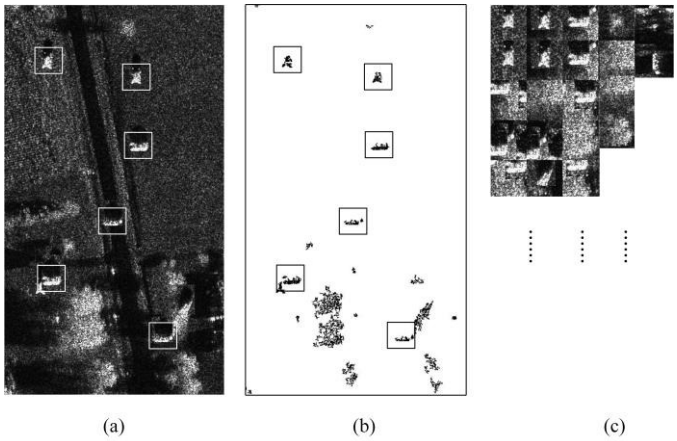


Fig. 2 Preprocessing (a) Composed SAR image, where the red rectangular labeled areas indicate real targets (b) The binary image after the preprocessing (c) A total of N_T suspicious regions extracted from the original SAR image with the size of $W \times H$.

RET Stage

We first use the DOG operator to filter each suspicious area to strengthen the important features like corner, edge and so on, which plays a similar role of the retinal ganglion cell in the vision information processing. As shown in equation (1), the DOG filter corresponds to the difference between the two images filtered by the Gaussian filers with different parameters.

$$f(x,u,\sigma_1,\sigma_2) = \frac{1}{\sigma_1\sqrt{2\pi}}\exp\left(-\frac{(x-u)^2}{2\sigma_1^2}\right) - \frac{1}{\sigma_2\sqrt{2\pi}}\exp\left(-\frac{(x-u)^2}{2\sigma_2^2}\right) \quad (1)$$

Where, f is a suspicious area after filtering; x is a pixel of f; u is the mean, which is usually taken to be 0 for the sake of simplicity; σ_1 and σ_2 are the standard deviations of the two Gaussian functions. Since the index computing in the DOG filtering process leads to a nonlinear distortion of the pixel values, a distortion rectifying is needed in the following formula:

$$f(x) = a\log_{10}\left[b \cdot f^2(x,u,\sigma_1,\sigma_2)\right] \quad (2)$$

Where a, b are the constant coefficients.

In biological vision system, there is a competitive relationship between the adjacent cells, called the "winner-take-all" where those relatively weak signals in a small area will be ignored while only the strongest will win the competition and become the excitation signal, transmitted to the visual cortex via the lateral geniculate nucleus (LGN). Based on this theory, we use a simple and effective mean filter to encode the suspicious region after the distortion rectifying

$$y(x)=\begin{cases} f(x) & \text{if } f(x)\geq \text{mean}(f(x)) \\ 0 & \text{else} \end{cases} \quad (3)$$

Through retaining those pixels greater than the mean, the coding can highlight the features like corner, edge, inflection point while ruling out the weaker background. It improves the efficiency of the subsequent treatments.

PVC Stage

Having received the signals from the retina, the primary visual cortex could get a "primary feature" for each suspicious region image through fast extracting and combining the target features. Then, based on the competition mechanism between the neurons, the "rough detection" is in charge of eliminating parts of false alarms and transmitting the remaining suspicious targets to the advanced visual cortex. The calculation flow diagram of the primary feature of a suspicious area is shown in Fig. 3. It is a feature on the basis of the correlation between the suspicious areas and the priori information, namely the patch-based features. Here we first introduce the patch-based feature learning and screening.

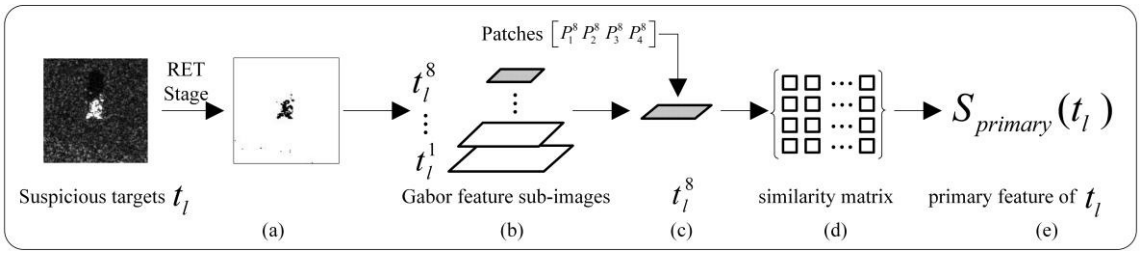


Fig. 3 Calculation flow diagram of the primary feature Results after the RET stage of processing (b) 8 Gabor feature sub-images of t_l (c) Selection of eighth sub-image to compute the correlation coefficients (d) Similarity matrix of the suspicious region t_l (e) The primary feature of t_l with the averaging method.

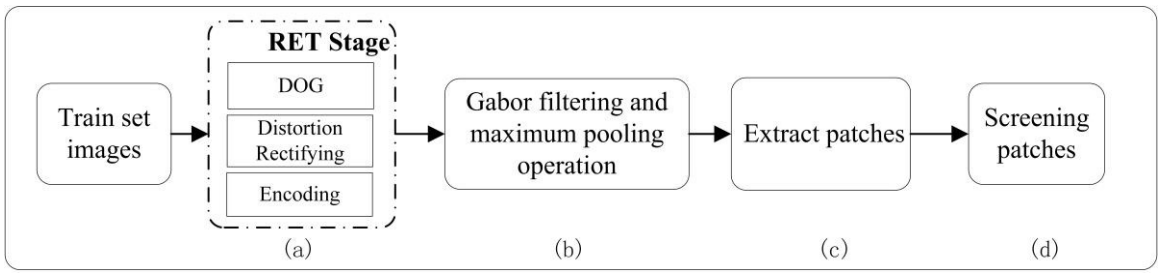


Fig.4 Flowchart of the patch-based feature learning and screening (a) Training samples first go through the RET stage to wipe out the clutters and enhance the features, including the DOG filtering, distortion rectifying and encoding (b) Simulating the simple and complex cells in the primary visual cortex. The Gabor filtering and maximum pooling operation are used to get the Gabor features of the training samples (c) Extracted patches from the Gabor features (d) Screening patches.

Patch-based Feature Learning and Screening

Fig.4 illustrates the flowchart of the patch-based feature learning and screening. The training samples first go through the RET stage described in last Section to enhance the image features. And then the Gabor filtering and the maximum pooling operation are used to get the Gabor feature sub-images of these training images, which simulate the simple and complex cells in the primary visual cortex with the bar-like receptive fields. After that, the patch-based features of the training samples are extracted randomly from these Gabor feature sub-images. In order to reduce the redundant patches, the last step is to screen patches. The detailed process is as follows.

After the processing in the RET stage, the training samples are handled according to the method described in S1 layer and C1 layer HMAX model, namely, the Gabor filter

$$h(x,y)=\frac{1}{2\pi\delta_u\delta_v}\exp\{-\frac{1}{2}(\frac{u^2}{\delta_u^2}+\frac{v^2}{\delta_v^2})\}\cos\frac{2\pi}{\lambda}u \quad (4)$$
$$u=x\cos\theta+y\sin\theta \qquad v=x\cos\theta-y\sin\theta$$

Where, $h(x,y)$ is the filtering result of a pixel located in (x,y) of the sample image; $\theta=[-45^0,0^0,45^0,90^0]$ is the angle of the filters; the axis u is parallel to θ and the axis v is perpendicular to θ; δ_u、δ_v are the standard deviations of the

Gaussian envelopes in the axis u and v, respectively; λ represents the wavelength which has 16 values: $7, 9, 11, \ldots, 33, 35, 37$. Each training image has 8 different Gabor sub-images after filtering [15] expressed by $G^k (k = 1, ..., 8)$ respectively. Assuming the size of the training sample is $W \times H$, the kth Gabor feature G^k is $W/R_k \times H/R_k \times 4$, where $R_k = [\ 8\ \ 10\ \ 12\ \ ...\ \ 22\ ]$ shows the down-sampling parameter and 4 corresponds to 4 filter angles.

After filtering, according to the filters of different wavelengths, the Gabor feature images of all the training samples can be divided into 8 groups. To ensure that the patches can cover the features in all scale space, we extract 1000 patches from each group, a total of 1000×8 patches. At the same time, these patches have 4 sizes $S_i \times S_i \times 4 (i = 1, 2, 3, 4)$, each having 250×8, which contributes the patches to have a robustness for the target scale transformation. The obtained patches can be represented by the following matrix

$$P = \begin{pmatrix} P_1^1 & \cdots & P_4^1 \\ \vdots & \ddots & \vdots \\ P_1^8 & \cdots & P_4^8 \end{pmatrix} \tag{5}$$

Where, the subscript of P_i^k represents the size of patches; the superscript k indicates that these patches are extracted randomly from the Gabor features G^k of all the training samples. In other words, P_i^k represents a total of 250 patches extracted from the kth Gabor feature G^k with the size of $S_i \times S_i \times 4$. And the jth patch in P_i^k is expressed as $p_{ij}^k \in P_i^k (i = 1, ..., 4, k = 1, ..., 8, j = 1, ..., 250)$. In the HMAX model of Poggio, each patch p_{ij}^k contains the features in 4 directions to ensure that the patches have the robustness for the rotation variation of the training samples, which increases the following calculation by 4 times. According to the "winner-takes-all" principle of the visual cortex during the signal processing, we used the method of [16] to compress the patches. It could reduce patches with four directions into one by retaining the maximum pixel at the same position but different directions and removing other three pixels, namely.

$$p_{ij}^k = \max_{o=1}^{o=4} p_{ij,o}^k \tag{6}$$

After the compression, the $S_i \times S_i \times 4$ patch p_{ij}^k becomes $S_i \times S_i$, and the total computation is reduced by three quarters on the base of retaining important features in every direction.

Because the above patches are selected randomly, the amount of features in each patch is also random. In this case, a large number of patches are needed to maintain a high recognition accuracy [15]. However, too many patches greatly increase the calculation time for detection. It is necessary to remove those patches with little characteristics and select the "better" patches. Here, the Support Vector Machine (SVM) is applied to screen the patches [16]. Firstly, we use the HMAX model to calculate the feature vectors of all the training samples and each element of the feature vector corresponds to a patch. Then the SVM is used to classify these feature vectors. In the process of classification, the classifiers would give each element a weight. Greater weight means this element plays a more important role in the classification and the corresponding patch would contain more useful features. By rearrange all the patches in P_i^k with the weights from big to small and reserve front $\bar{N}$, the patch screening is completed.

Acquisition of the Primary Feature

Similar to the processing method for the training samples, we first carry on the Gabor filtering and maximum pooling operation to handle each suspicious region having gone through the RET stage in processing. Likewise, according to the different wavelengths of filters, a suspicious region t_l has 8 Gabor feature sub-images expressed as t_l^k whose size is $W/R_k \times K/R_k \times 4 (k = 1, ..., 8)$. Then under the principle of the "winner-take-all", t_l^k can be compressed into $W/R_k \times K/R_k (k = 1, ..., 8)$ by retaining the maximum pixels in 4 directions.

The primary feature of a suspicious region is on the basis of image similarity, which can be described by the two correlation coefficients between the two images

$$s(m, p) = \frac{\sum_{g=1}^{N_g} \sum_{k=1}^{N_k} (m_{gk} - \bar{m})(p_{gk} - \bar{p})}{\sqrt{\sum_{g=1}^{N_g} \sum_{k=1}^{N_k} (p_{gk} - \bar{p})^2 \sum_{g=1}^{N_g} \sum_{k=1}^{N_k} (m_{gk} - \bar{m})^2}} \tag{7}$$

Wherein m_{gk} and p_{gk} are the pixel values on line

g column k of images m and p ,respectively; $\bar{m}$, $\bar{p}$ are the means of the two images, N_g , N_k represent their numbers of rows and columns.

Based on this formula, we can get the correlation coefficient between a Gabor feature sub-image t_l^k of the suspicious region t_l and a patch p_{ij}^k , namely $s\left(t_l^k, p_{ij}^k\right)$. In order to ensure the accuracy of correlation coefficient, t_l^k and p_{ij}^k must be in the same scale space which means that their corresponding Gabor filters have the same wavelengths or their superscripts k are the same. It also reflects the relative independence of the information flow between the various areas of visual cortex.

The eighth Gabor feature sub-images t_l^8 corresponds the longest wavelength of filters. Therefore they have the smallest size and retains the more overall features, which will greatly simplify the following calculation. As a result, we only use t_l^8 to calculate the primary feature. The correlation coefficients between t_l^8 of all the suspicious regions t_l and p_{ij}^8 compose the corresponding similarity matrix $S(t_l)$:

$$S(t_l)=\begin{pmatrix} s\left(t_l^8, p_{11}^8\right) & s\left(t_l^8, p_{12}^8\right) & \cdots & s\left(t_l^8, p_{1\bar{N}}^8\right) \\ s\left(t_l^8, p_{21}^8\right) & s\left(t_l^8, p_{22}^8\right) & \cdots & s\left(t_l^8, p_{2\bar{N}}^8\right) \\ s\left(t_l^8, p_{31}^8\right) & s\left(t_l^8, p_{32}^8\right) & \cdots & s\left(t_l^8, p_{3\bar{N}}^8\right) \\ s\left(t_l^8, p_{41}^8\right) & s\left(t_l^8, p_{42}^8\right) & \cdots & s\left(t_l^8, p_{4\bar{N}}^8\right) \end{pmatrix} \quad (8)$$

The primary feature $S_{primary}(t_l)$ of the suspicious region t_l is the mean of all elements in the similarity matrix

$$S_{primary}(t_l)=\frac{1}{4\bar{N}}\sum_{i=1}^{4}\sum_{j=1}^{\bar{N}}\left(s\left(t_l^8, p_{ij}^8\right)\right) \quad (9)$$

The primary feature $S_{primary}(t_l)$ represents the average correlation coefficient between this suspicious region t_l and the prior information. It reveals the probability of this suspicious region containing the real targets. The greater $S_{primary}(t_l)$ means a higher possibility.

Rough Detection

According to the competition mechanism of the visual cortex, the rough detection for the suspicious regions is completed using the obtained primary features. First, we need to set a threshold $S_{primary}(t_0)$

$$S_{primary}(t_0)=\mu_0+c\sigma_0 \quad (10)$$

Among them, $\mu_0=\frac{1}{N_T}\sum_{l=1}^{N_T}S_{primary}(t_l)$, $\sigma_0=\frac{1}{N_T}\sum_{l=1}^{N_T}\left(S_{primary}(t_l)-\mu_0\right)^2$ represent the means and standard deviation of all the primary features; N_T indicates the total number of the suspicious regions after the pretreatments; c is an adjustable parameter. The rough detection is finished through excluding the suspicious regions whose primary feature is less than the threshold.

AVC Stage

Rough detection has eliminated a large number of suspicious areas having poor correlation with the prior features, so finer features are needed to identify the targets and false alarms. In the AVC Stage, we simulate the neurons with the bigger receptive fields in the advanced visual cortex to further process the suspicious areas. More accurate "advanced feature" is extracted from each suspicious region and classifiers are used to separate the false alarms and real targets, namely "fine detection". The whole process is shown in

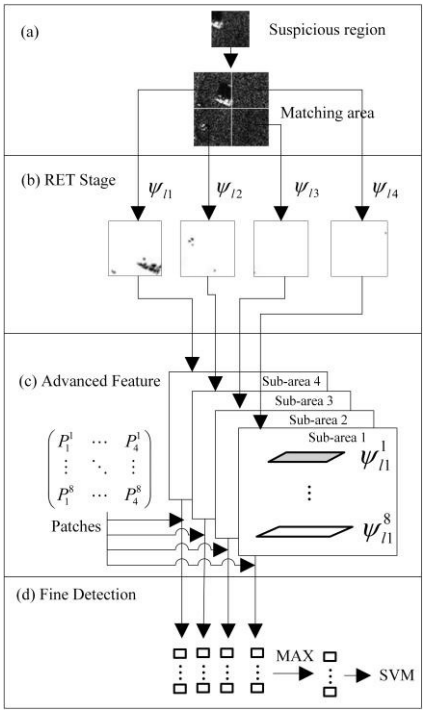


Fig 6. In the calculation of the advanced feature, the "matching area" [17] is introduced to compensate the position errors of the suspicious regions to improve the accuracy of advanced features.

Matching area

As the suspicious areas are obtained by the preprocessing,

their positions have relatively big errors. As shown in Fig. 5(a), only a portion of the real target is in the suspicious region, which reduces the quantity of the useful features in this suspicious region, influencing the accuracy of the detection. The schematic diagram of the matching area is shown in Fig. 5 (b). First, let us keep the center of the suspicious region t_1 ($W \times H$) unchanged and expand its area 4 times, which is the matching area of this suspicious region, expressed by Ψ_1. Then, the matching area is divided into the 4 sub-areas with the size $W \times H$, expressed by $\psi_{1\xi}(\xi=1,..,4)$. Now the main body of the real target will be located in one of these subareas. In the subsequent steps, we need to further deal with all the four sub-regions and calculate the advanced features on this basis.

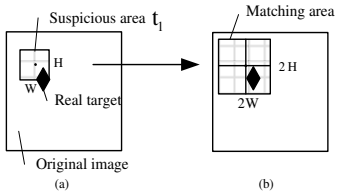


Fig. 5 Schematic diagram of the matching area (a) The relationship between a real target and a suspicious area (b) Location-relation of a real target and matching area

Advanced Feature and Fine Detection.

Similarly, the 4 sub-regions $\psi_{1\xi}$ of the suspicious region t_1 first go through the RET stage of the detection, the Gabor filtering, the maximum pooling operation and the compressing in different directions. Each $\psi_{1\xi}$ has 8 Gabor feature sub-images with the size of $W/R_k \times H/R_k$, $\psi_{1\xi}^k(k=1,...,8)$. According to the related calculation formula, we could get the correlation coefficient between $\psi_{1\xi}^k$ and its corresponding patches p_{ij}^k, namely $s(\psi_{1\xi}^k, p_{ij}^k)$. It also should be noted that $\psi_{1\xi}^k$ and p_{ij}^k must correspond to the Gabor filters of the same wavelength. Among the 4 correlation coefficients $s(\psi_{1\xi}^k, p_{ij}^k)(\xi=1,2,3,4)$, the larger correlation coefficient means that its corresponding sub-region $\psi_{1\xi}^k$ may contain more features of the target, hence we take the maximum as the correlation coefficient between t_1^k and p_{ij}^k, $s(t_1^k, p_{ij}^k)$:

$$s(t_1^k, p_{ij}^k) = \max_{\xi=1}^{4} s(\psi_{1\xi}^k, p_{ij}^k) \tag{11}$$

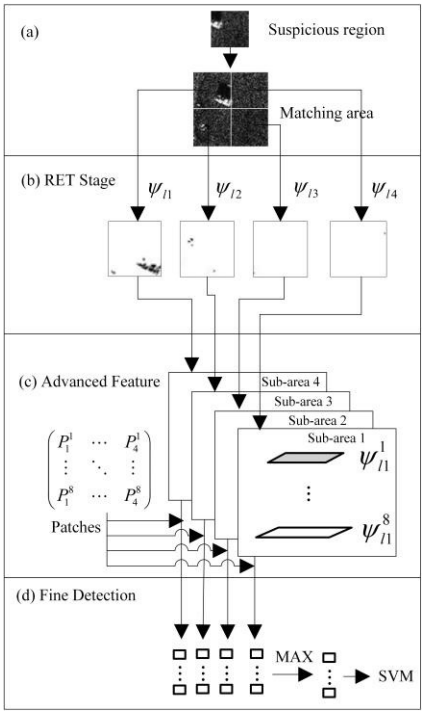


Fig 6 Processing flowchart of the AVC stage (a) Matching area of a suspicious area t_1 (b) RET stage process of 4 sub-images of a matching area $\psi_{1\xi}(\xi=1,2,3,4)$ (c) Calculate the advanced features on the base of Gabor features (d) Complete the fine detection by the means of classifiers

In this case, we get a new similar matrix about t_1 composing of the correlation coefficients of all Gabor sub-images of this suspicious area t_1 and the corresponding patches, namely $S_{advance}(t_1)$

$$S_{advance}(t_1) = \begin{pmatrix} s(t_1^1, P_1^1) & s(t_1^1, P_2^1) & s(t_1^1, P_3^1) & s(t_1^1, P_4^1) \\ s(t_1^2, P_1^2) & s(t_1^2, P_2^2) & s(t_1^2, P_3^2) & s(t_1^2, P_4^2) \\ \vdots & \vdots & \vdots & \vdots \\ s(t_1^8, P_1^8) & s(t_1^8, P_2^8) & s(t_1^8, P_3^8) & s(t_1^8, P_4^8) \end{pmatrix} \tag{12}$$

Where $s(t_1^k, P_i^k) = \left[s(t_1^k, p_{i1}^k) \quad s(t_1^k, p_{i2}^k) \quad \cdots \quad s(t_1^k, p_{i\bar{N}}^k) \right]$. Each of its elements indicates the correlation coefficient between t_1^k and a patch in P_i^k. As each P_i^k has $\bar{N}$ patches after the screening described in last section, $s(t_1^k, P_i^k)$ is a $1 \times \bar{N}$ vector. The similarity matrix $S_{advance}(t_1)$ is the obviously different matrix $S(t_1)$ in the PVC stage. $S(t_1)$ only consists of the correlation coefficients between t_1^8 and P_i^8 while $S_{advance}(t_1)$ contains all the Gabor feature sub-images, which preserves the characteristics of the suspicious targets in

various scale spaces and reflects the advanced visual cortex vast capable of synthesizing complex information.

After the above calculation, each remaining suspicious target t_1 has $4\times8\times\bar{N}$ correlation coefficients, composing a $4\times8\times\bar{N}$ vector, which is the named "advanced feature" of the suspicious target t_1. Compared with the primary feature of only one element, the advanced feature can comprehensively describe a target as a vector. Finally, we use the SVM to play the role of nerve center by separating the false alarms from real targets.

To summarize, the particular flow of the proposed method is shown in Fig. 7.

Experimental Results and Analysis

The original images used in our experiments are the real MSTAR amplitude images with a resolution of 0.3m and the 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 detection objects are 128×128 slice images of the various vehicles taken from different angles and they are divided into a training set and a detection set, respectively. The training set is used to extract the patch-based features while the detection set is used to compound the detecting SAR images.

In the example,

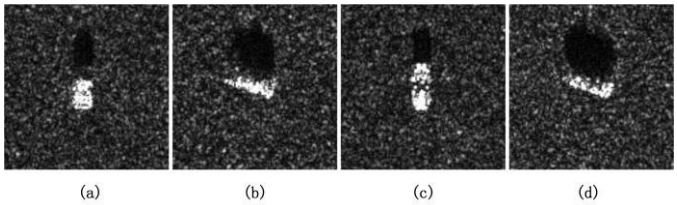


Fig. 8 shows the four slice images of two kind tanks bmp2_9563 and btr70 in different directions.

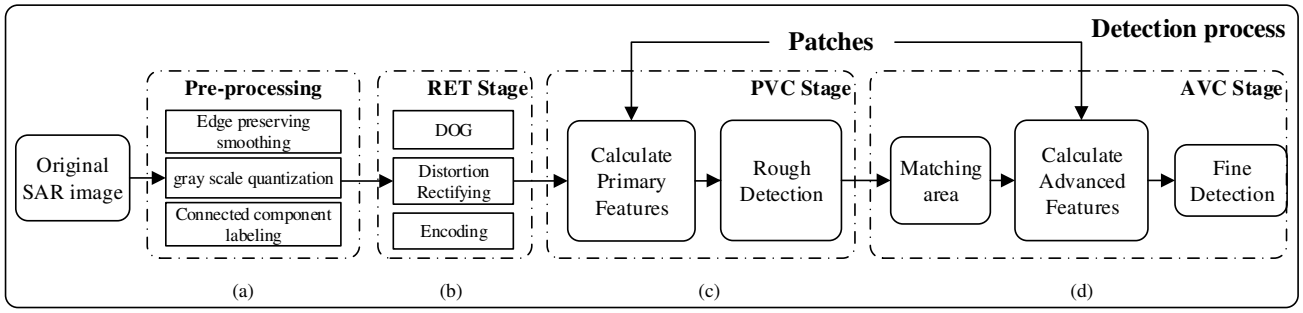


Fig. 7 The detailed flowchart of the proposed method (a) Find the suspicious regions by the pre-processing (b) RET stage processes to enhance the useful features and suppress the clutters, including the DOG filter, distortion rectifying, coding (c) PVC stage calculates the primary features of the suspicious regions on the base of the patch-based prior features to eliminate the parts of suspicious targets through the rough detection (d) AVC stage obtains more accurate advanced features to separate the real targets from the remaining suspicious regions in fine detection.

Validity Experiments

The original SAR images were selected as follows. 20 targets randomly selected from the detection set were added into the SAR images of the MSTAR public database with various SCR.

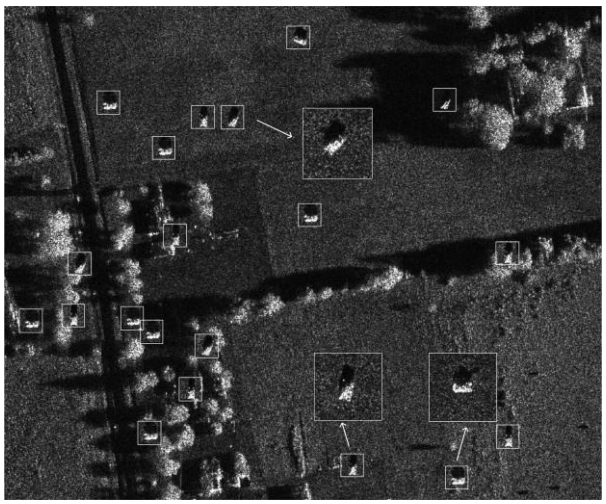


Fig. 9 shows a composed SAR image when $SCR = 2.0$ In order to observe their locations easily, the targets are marked

with the rectangle symbols. Some of the targets are obscured by forest and tussock adding the complexity of detection.

In the example,

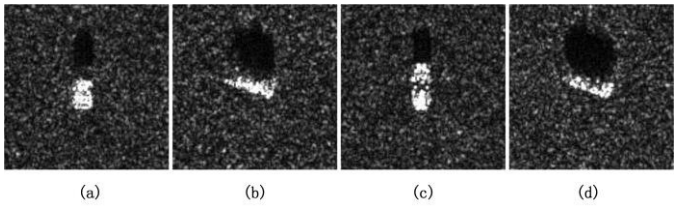


Fig. 8 slice images of some targets (a), (b) represent bmp2_9563 tanks in different perspectives. (c), (d) are btr70 tanks in different perspectives.

The result of the preprocessing for the above SAR image is shown in Fig. 10. As a large numbers of trees, construction and other clutters are treated as the suspicious regions, the results contain a total of 146 false alarms and 20 real targets.

Fig. 11 indicates the distribution characteristics of all the primary features of the suspicious areas calculated in the PVC stage, where the horizontal axis represents the values of the primary features and the vertical axis represents the numbers of the targets in different ranges. The false alarms and real targets are respectively represented by the blue, red bars. The vertical dotted line is the threshold for filtering the false alarms. As can be seen, compared with the false alarms, since the features of true targets are more similar to the transcendental characteristics, the correlation coefficient between the real targets and the patches are greater, so they have larger primary features. The rough detection eliminates 102 false alarms while remaining 44, misjudging 1 real target, as shown in **Error! Reference source not found.**.

Fig. 13 shows the fine detection results. In addition to a missed target in the rough detection, the remaining 19 targets are all detected in this stage and the number of false alarms is reduced from 44 to 3. Hence, the advanced feature could greatly reduce the false alarms and effectively improve the detection performance with the aid of more accurate advanced features.

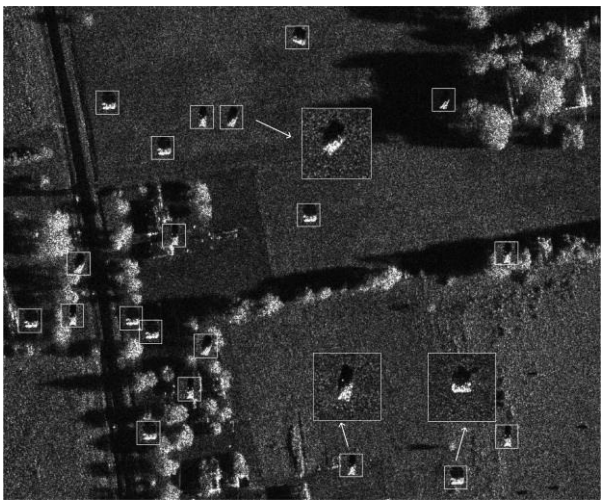


Fig. 9 Composed SAR image with SCR=2, which includes 20 real targets labeled by red rectangle

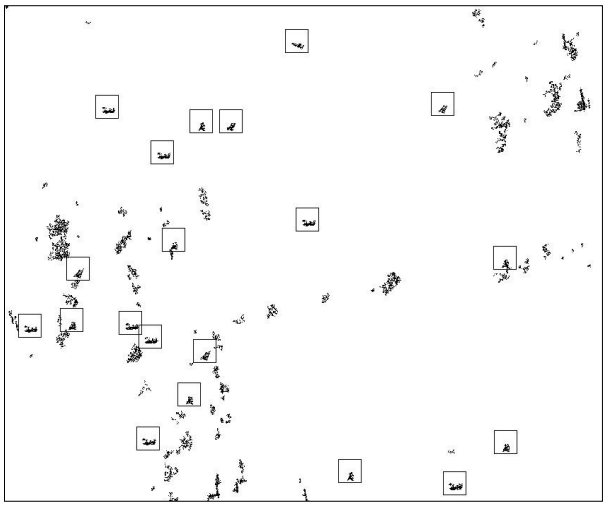


Fig. 10 The binary image after preprocessing where black areas indicate the suspicious targets and the rectangle labeled area are real targets

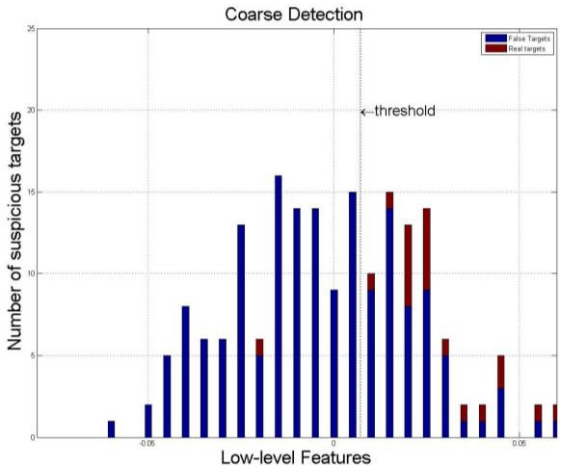


Fig. 11 Numbers of the suspicious targets in different range of the primary features

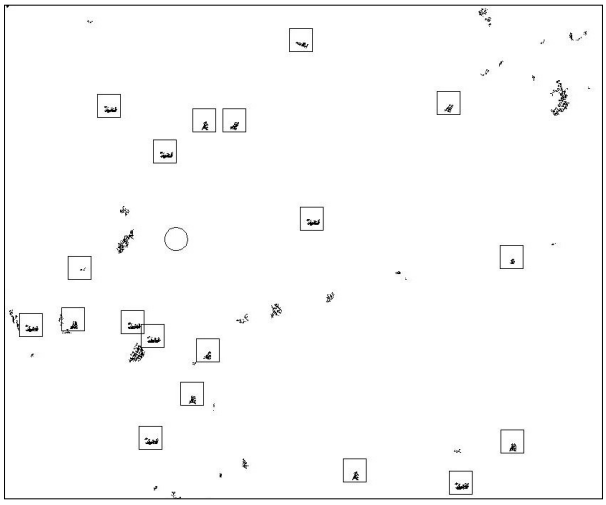


Fig. 12 Results of the rough detection on the base of primary features, a large number of false alarms have been eliminated. Rectangle labeled areas are detected for real targets and circle labeled region is the missed.

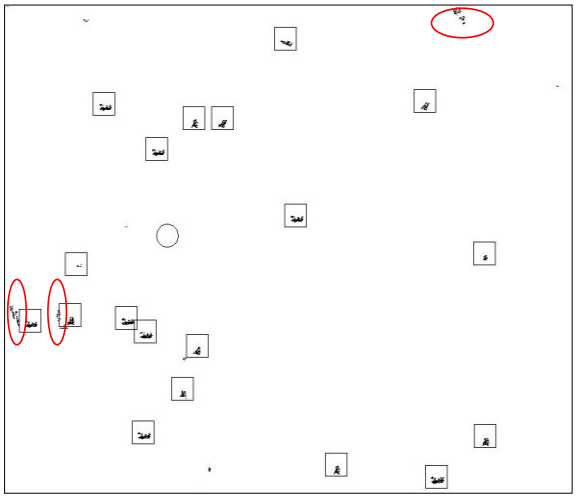


Fig. 13 Final results from the fine detection using the advanced features. There are only three false alarms labeled by ellipses. The rectangles labeled areas are detected for real targets and the circle labeled region is the missed.

Experiments about Environment Adaptability

To verify the adaptability of the proposed method for various clutter environments, we carried out the following two experiments. The background clutters in Fig.14 (a) mainly includes trees and grass, which may obscure the targets and form many highlights similar to the goals in the image. The clutters in Fig.14(b) are the artificial buildings randomly distributed across the background. Fig.14(c) and (d) are the detection results of the CFAR for two composed SAR images, which leave a large number of false alarms. Fig.14 (e) shows the final detection results of the proposed method for image (a). There are 4 false alarms and one missed target. Fig.14 (h) is the detection result of the image(b), where one target is missed and two false alarms are treated as targets. As can be seen, this algorithm has better performance than the commonly used CFAR method.

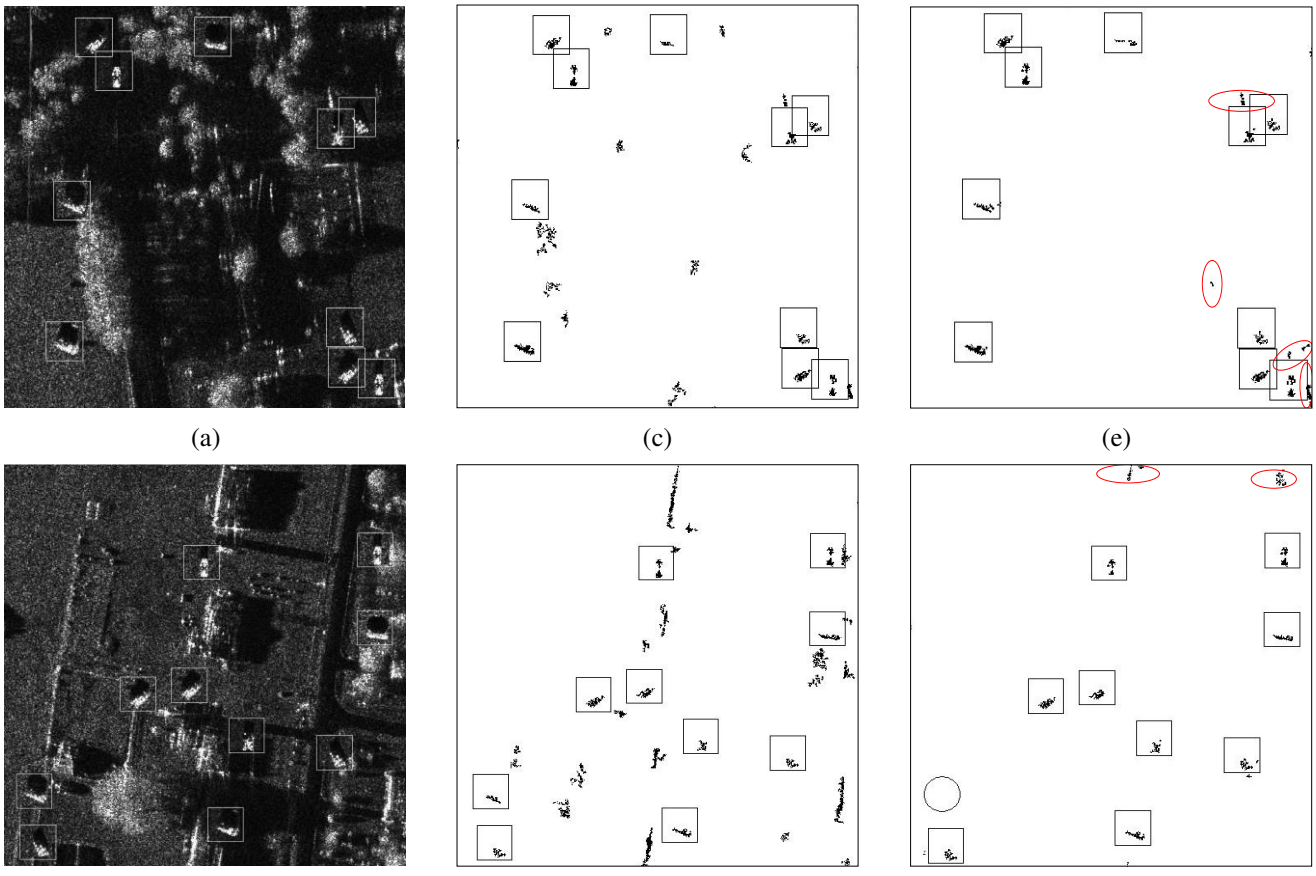
Matching Area Effect

In order to observe the effect of the matching area on the detection rate and false alarms, Fig. 15 compares the detection results of the 5 composed SAR images with and without the matching area. Fig. 15 (a) shows the detection rates in two cases while Fig. 15 (b) indicates their difference in false alarms. Results show that the introduction of the matching area increased the average detection rate by 8.2% and reduced the false alarms by 1.6%.

Performance analysis

Fig. 16 gives the detection performance of different stages of the proposed approach, CFAR and HMAX. As HMAX does not have the function of target searching, a same preprocessing stage is also applied to obtain image slices of all suspicious areas before feature extraction by HMAX and recognition by means of SVM. Similarly, patches used in HMAX are also selected randomly from target samples. In Fig. 16(a), the curve PVC shows the number of false alarms after the processing at the PVC stage without the RET stage under different SCR, the curve RET+PVC represents the false alarms on the condition that the SAR image first goes through the RET stage and then enters the PVC phase. Compared with the PVC curve, the average false alarm of the RET+PVC decreases by 26%. Obviously, the RET stage has played a vital role in the detection similar to the retina in a biological visual system, which could preliminarily screen the visual information and effectively enhance the target features so as to improve the detection results. The curve RET+PVC+AVC represents the false alarms through all the 3 stages of the proposed method. Its average false alarm falls by 85% compared with the RET+PVC curve and can maintain a low false alarm under the low SCR. Instead, false alarm of CFAR is far more than the proposed method and HMAX since CFAR is easily affected by SCR. As can be seen, from the PVC stage and AVC stage, the continued refinement for features suspicious of regions gradually narrows down the scope of detection and decreases false alarm. It also reflects the rules of transmitting and managing information in biological visual systems. Namely, the information flows are processed and transmitted from the retina to the primary visual

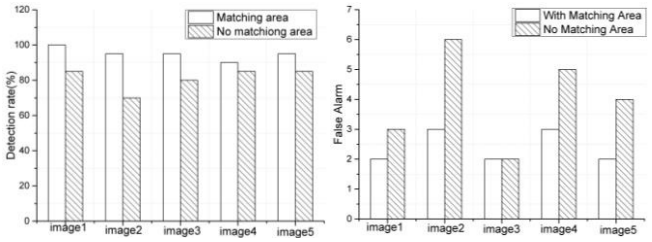
cortex and the advanced visual cells. Fig. 16. (b) compares the detection rates of above three methods. Compared with CFAR, the proposed method and HMAX have relatively higher detection rates. Meanwhile, performance of the proposed method is better than HMAX especially when SCR is less than 4, proving that, DOG filtering, distortion rectifying and encoding enhance this model's robustness by wiping out the clutters, and enhance the features.



(a) (c) (e)

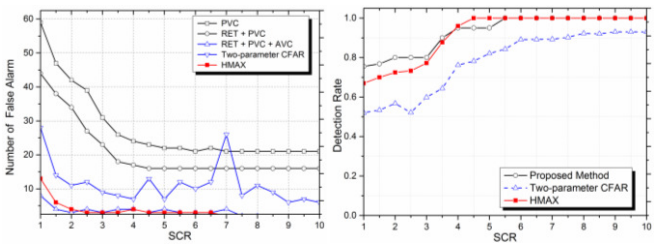
(b) (d) (f)

Fig.14 Detection results of the proposed method in two kinds of background clutters. (a) A composite SAR image by adding 10 vehicles in the wood and grass background with SCR=2. (b) A composite SAR image with manmade building and 10 vehicles with SCR=2. (c) The result of CFAR for image (a). (d) The result of CFAR for image (b). (e) The result of the proposed method on the image (a). (g) The result of the proposed method on image (b).



(a) (b)

Fig. 15 Comparative effects of with and without matching area (a) detection rate. (b) false alarms



(a) (b)

Fig. 16 The performance analysis of the proposed method, CFAR and HMAX (a) Comparison of false alarms in different SCR, where the curve PVC means the direct PVC detection without RET; curve RET+PVC represents the false alarms of rough detection; RET+PVC+AVC indicates the fine detection. (b) Comparison of the detection rate of the proposed method, CFAR, and HAMX.

Conclusion

In order to realize the rapid detection for the targets of interest, inspired by the cooperation mechanisms between the retina, primary and advanced visual cortex in the visual information processing, a progressive enhancement SAR

targets detection approach is proposed. Using the improved feature learning method, this method progressively extracts and processes the target features in three stages to obtain the primary features and the advanced features. On this basis, the rough detection and the fine detection complete the target search by gradually reducing the suspicious areas.

Our experimental results have shown that, whether for the isolated highlights or the various continuous clutters in SAR images, the proposed method can make the full use of the unique feature extraction and the processing mechanism of biological vision systems to accurately detect the targets with few false alarms. The introduction of the RET stage and the improvement of feature learning methods make the new method have the better detection rate and the lower false alarms even with the low SNR. At the same time, the matching area further improves the detection results by compensating the position errors of the suspicious regions. Finally, a series of comparative experiments with the traditional method have been carried out to demonstrated the superiority of the proposed approach.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (No.61071139, No.61471019, No.61171122), the Aeronautical Science Foundation of China (No. 20142051022), the Foundation of ATR Key Lab. . the Royal Society of Edinburgh (RSE) and The National Natural Science Foundation of China (NNSFC) under the RSE-NNSFC joint projects (2012-2015) [grant number 61211130309 and 61211130210] with Beihang University and Anhui University, China, respectively. It was also supported in part by the "Sino-UK Higher Education Research Partnership for PhD Studies" joint-project (2013-2015) funded by the British Council China and The China Scholarship Council (CSC).

Compliance with Ethical Standards

Funding: This study was funded by the National Natural Science Foundation of China (No.61071139, No.61471019, No.61171122), the Aeronautical Science Foundation of China (No. 20142051022), the Royal Society of Edinburgh (RSE) and The National Natural Science Foundation of China (NNSFC) under the RSE-NNSFC joint projects (2012-2015) [grant number 61211130309 and 61211130210] with Beihang University and Anhui University, China, respectively.

Conflict of Interest: Author Fei Gao et al. declare that he/she has no conflict of interest.

Ethical approval: This article does not contain any studies with human participants performed by any of the authors.

References

1. Novak L M, Halversen S D, Owirka G J, et al. Effects of polarization and resolution on SAR ATR. Aerospace and Electronic Systems. 1997, 33(1): 102-116.
2. Di Bisceglie M, Galdi C. CFAR detection of extended objects in high-resolution SAR images. Geoscience and Remote Sensing. 2005;43(4):833-43.
3. Bai X, Zhou F, Jin T. Enhancement of dim small target through modified top-hat transformation under the condition of heavy clutter [J]. Signal Processing, 2010, 90(5): 1643-1654.
4. DENG X, MA Y. PCNN Model Analysis and Its Automatic Parameters Determination in Image Segmentation and Edge Detection. Chinese Journal of Electronics. 2014, 23(1).
5. Francis Crick. Astonishing hypothesis: The scientific search for the soul. Simon and Schuster. 1995.
6. Rodieck R W. Quantitative analysis of cat retinal ganglion cell response to visual stimuli. Vision research. 1965, 5(12): 583-601.
7. Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on pattern analysis and machine intelligence. 1998, 20(11): 1254-1259.
8. Zhao J, Sun S, Liu X, et al. A Novel Biologically Inspired Visual Saliency Model. Cognitive Computation. 2014, 6(4): 841-848.
9. Hubel D H, Wiesel T N. Ferrier lecture: Functional architecture of macaque monkey visual cortex. Proceedings of the Royal Society of London. Series B, Biological Sciences, 1977: 1-59.
10. Jones J P, Palmer L A. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. Journal of neurophysiology. 1987, 58(6): 1233-1258.
11. Amoon M, Rezai-rad G, Daliri M R. PSO-Based Optimal Selection of Zernike Moments for Target Discrimination in High-Resolution SAR Imagery[J]. Journal of the Indian Society of Remote Sensing, 2014, 42(3): 483-493.
12. Tu Z, Zheng A, Yang E, et al. A biologically inspired vision-based approach for detecting multiple moving objects in complex outdoor scenes[J]. Cognitive Computation, 2015: 1-13.

13. Ho-Phuoc T, Guyader N, Guérin-Dugué A. A functional and statistical bottom-up saliency model to reveal the relative contributions of low-level visual guiding factors[J]. Cognitive Computation, 2010, 2(4): 344-359.
14. M. R. a. T. Poggio. Hierarchical models of object recognition in cortex. nature neuroscience. 1999.vol. 2, no. 11, pp. 7.
15. Serre T, Wolf L, Bileschi S, et al. Robust object recognition with cortex-like mechanisms. Pattern Analysis and Machine Intelligence. 2007, 29(3): 411-426.
16. Mutch J, Lowe D G. Multiclass object recognition with sparse, localized features Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. IEEE. 2006, 1: 11-18.
17. Zhang L, Tjondronegoro D. Facial expression recognition using facial movement features. Affective Computing. 2011, 2(4): 219-229.