 doi:10.20944/preprints201710.0093.v1

Article

A New Visual Attention Model Designed for SAR Images Based on Texture Saliency

Wei XIONG[1], Yongli XU[1,*], Yafei LV[1], Libo YAO[1]

[1] Institute of Information Fusion, Naval Aeronautical and Astronautical University, Yantai 264001, China; xiongwei@csif.org.cn (W.X.); YFei_Lv@163.com (Y.L.); ylb_rs@126.com (L.Y.)

* Correspondence: xuylhjhy@126.com; Tel.: +86-178-5359-7576

Abstract: Object detection in synthetic aperture radar (SAR) images, which is a fundamental but challenging problem in the field of satellite image interpretation, plays an important role for a wide range of applications and is receiving significant attention in recent years. Recently, the ability of human visual system to detect targets with visual saliency is extraordinarily fast and reliable. However, visual computational modeling of SAR image scene still remains a challenge. This paper designs a visual attention model for SAR images. Firstly, we propose a novel approach for computing the local texture coarseness of the input image, then our model constructs the corresponding feature maps. Next a new mechanism of feature fusion is adopted to replace the linear additive mechanism of classical models to obtain the final saliency map. Moreover, the gray values of focus of attention (FOA) in all feature maps are taken into account, our model chooses the best saliency representation, the filter and threshold segmentation of saliency maps can be used to extract the salient regions accurately through the multi-scale competition strategy, thereby completing this operation for visual saliency detection in SAR images. Finally, the paper gives the framework based on classical ITTI model. In the paper, several types of satellite data, such as TerraSAR-X (TS-X) and Radarsat-2, are used to evaluate the performance of visual models. The results show that our model provides superior performance than classical models. By further contrasting with classical visual models, our model reduce the false alarm caused by speckle noise, its detection speed is greatly improved, and it is increased by 25% to 45%.

Keywords: SAR image; Visual attention model; Texture saliency; Feature map; Focus of attention

1. Introduction

Synthetic aperture radar (SAR), known as a kind of advanced active microwave sensors, with its all-weather, all-day, multi-polarization advantages, has been increasingly paid attention to by all countries seeking detection technology in remote sensing [1]. As the basis of its classification and identification, targets detection in SAR images is an important aspect of SAR application [2].

SAR images have the characteristics of low contrast, low signal-to-noise ratio and limited gray level, and so on. These characteristics cause targets in SAR images to be subject to noise interference, and the contrast between the target and the surrounding environment becomes low. This bring difficulties to target detection. What's more, with the successful launch of TerraSAR-X (TS-X), Radarsat-2 and other next-generation SAR sensors, SAR is gradually evolving towards higher resolution and larger width directions. The quality of SAR image is getting closer to the optical images, and the features of SAR images are complex. Traditional detection system cannot interpret and analyze complex features of high-resolution SAR images timely and effectively [3, 4]. In a word, since there is a contradiction between the large amount of information in SAR images and the limited

 doi:10.20944/preprints201710.0093.v1

computer processing power, the research for target detection technology in SAR images is currently a serious problem [5].

However, in the face of complex scenarios, the human visual system can quickly focus on several interesting targets, known as visual attention mechanism [6,7]. Some scholars have made great progress in human visual intelligence, and have adopted visual attention mechanism to select useful information from rich and complex information to complete target detection, which greatly improves the efficiency of processing. Consequently, scholars try to put forward the mathematical model to simulate human visual system.

At present, the study of visual attention models mainly includes two aspects: (1) The data-driven visual model, which can be divided into visual model based on image time domain and visual model based on image frequency domain. The visual model proposed by Itti simulated firstly the "attention" concept of the human eye in the mathematical level [8], and established the visual attention model based on the time domain; The visual model based on frequency domain transforms the processing level of the image from time domain to frequency domain. For example, the visual model based on the spectral residual presented by Hou et al. [9], Hou Model uses the fast Fourier transform of the amplitude spectrum as the significant representation of the image, and then returns to the time domain to obtain saliency map. The visual model proposed by Yu, which combined the discrete cosine transform [10], is also widely used. Although such models can obtain salient regions in a simple way, there are many false alarms in the saliency map. (2) The purpose-driven visual model, the most representative of which is the function model based on the psychological threshold proposed by Itti [11,12] and the visual model proposed by Oliva et al. based on Bayesian learning [13]. The shortcoming of such models is lack of self-adaptability.

Currently, commonly used visual model are ITTI model, AIM model and spectral residual model [14]. Recently, the ITTI attention model has been widely used in the field of computer vision, because the ITTI visual model has imitated the formation of the bottom-up saliency in the human visual system, so as to realize the saliency detection of the image.

Although the research of computer vision has made great progress in recent years, and a series of achievements have been obtained, the ability of human eye to process and to analyze information on the realistic scene is still more efficient. Recently, most of these existing visual attention models are designed for the nature scene image, and these models can obtain the saliency maps and extract the regions of interest. However, there are significant differences between SAR image and natural scene image. For instance, the characteristics of speckle noise in the background and targets are similar in SAR images [15]. The phenomenon makes the classical visual models difficult to get accurate results when the object regions are extracted from this type of SAR image.

The difficulties mainly include the following aspects: (1) Extraction of the underlying early visual features. The selected features in the classical ITTI model are local features such as brightness, color and orientation, but the global features of the target regions are not considered. Therefore, the model cannot accurately deal with the regions of interest whose local features are not obvious in the detected image. Among them, the texture, shape and other important features of targets in SAR image are not considered in ITTI model, which is also the reason for the poor performance of the model. (2) Strategy of feature fusion: In ITTI model, the fusion strategy of feature maps is linear combination. Usually, the model get the total saliency map by adding the feature maps linearly directly, ignoring the priority relationships of different features, which leads to the weakening of a dominant feature map in the merge process, so leads to a missed detection of the target areas. In conclusion, the ITTI model has no adaptability to the extraction of salient regions of SAR image.

Motivated by this, we propose an improved visual attention model for SAR images based on texture saliency. Firstly, we need to calculate and extract the texture and other features that can describe the SAR image better. In this step, we design a new calculation method for local texture coarseness. Then we construct the corresponding saliency maps of features. What's more, a new mechanism of feature fusion is adopted to replace the linear additive mechanism of classical models to obtain the overall saliency map, and a measurement method of the texture saliency is given. Finally, the gray-scale values of focus of attention in saliency maps of features are taken into account.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 October 2017 doi:10.20944/preprints201710.0093.v1

Our model choose the best significant representation. Through the multi-scale competition strategy, the filter and threshold segmentation of the saliency maps can be used to accurately select the salient regions, thereby completing this operation for the visual saliency detection in SAR images.

The paper is organized as follows. We provide a review of the existing computational models to visual attention with a brief description of their strengths and shortcomings in Section 1. And the motivation leading to the current improved work is described in the second part of Section 1. The proposed model is described by means of a pseudo-code and a graphical illustration in Section 2. In Section 3, we analyze and evaluate the performance of our calculation method for local texture coarseness through mathematical derivation and experiments; In Section 4, we describe the experiments carried out to validate the performance of the proposed method for the task of salient regions detection. Finally, we conclude this work by highlighting its current shortcomings with a brief discussion about the future directions of the current work in Section 5.

2. Design of Visual Model Based on Texture Saliency

2.1. Measurement of Texture Saliency

Under normal circumstances, local features are used to distinguish the target pixels and neighborhood pixels, and the global features can calculate the saliency of similar areas in the image from the perspective of global saliency, further highlighting the target areas. Considering these above, this paper takes the texture and shape features which is the dominant position in SAR image into the visual feature extraction category. Human beings have the perfect texture sensing mechanism, which can distinguish the fine texture difference. The features used by humans to distinguish textures include: coarseness, contrast, complexity, orientation, etc.

Texture feature is one of the important properties which are used to identify the target and region of interest [16]. The texture feature exists on the surface of every object, and contains the important information of the object's surface structure arrangement and their connection to the surrounding environment. Texture reflects the visual features of homogeneity, and is independent of the color or brightness in images [17]. Therefore, the visual attention model based on texture saliency is of great significance.

In the paper, four operational factors of features are designed, they include the local coarseness, standard deviation, orientation, and global contrast feature of SAR image.

2.1.1. Local texture coarseness

Tamura et al. proposed the expression of Tamura texture feature based on the psychological research on the human visual perception of texture, which has been widely used in image recognition and image retrieval in recent years [18,19]. The Tamura texture feature includes six properties that correspond to the texture features in the psychological point of view: coarseness, contrast, orientation, linearity, regularity and roughness. Among them, the coarseness is the most basic and important texture feature. From the narrow point of view, the texture is coarseness.

Coarseness feature is a quantity that reflects the granularity of texture, when the two texture patterns are only different in the dimension of element. The pattern with a larger dimension of element and fewer repetition units is more cruder [20]. The calculation of texture coarseness can be divided into the following steps:

(1) Calculate the average intensity of pixels in the activity window in the image. The size of the activity window is $2^k \times 2^k$. Assuming $I(i,j)$ is the input image; the average intensity value is:

$$A_k(x,y) = \sum_{i=x-2^{k-1}}^{x+2^{k-1}-1} \sum_{j=y-2^{k-1}}^{y+2^{k-1}-1} I(i,j) / 2^{2k} \tag{1}$$

Among that, k=0,1,2,…, L_{max}; L_{max} is the maximum window scale;

(2) For each pixel in the image, the average intensity difference between the non-overlapping windows in horizontal and vertical directions is calculated separately:

$$E_{k,h}(x,y) = \left| A_k(x+2^{k-1},y) - A_k(x-2^{k-1},y) \right| \tag{2a}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 October 2017 doi:10.20944/preprints201710.0093.v1

$$E_{k,v}(x,y)=\left|A_k(x,y+2^{k-1})-A_k(x,y-2^{k-1})\right| \quad (2b)$$

(3) The size of the maximum average intensity difference is set to optimum size at each pixel:

$$E_k = E_{\max} = \max(E_{k,h}, E_{k,v}) \quad (3)$$

$$Z_{opt} = 2^k \quad (4)$$

Where: Z_{opt} is the optimum size at current pixel. If there is $k>k_{max}$; $E_k>t\cdot E_{max}$, then, $k_{max}=k$; in the original, t takes the empirical value 0.9;

(4) Calculate the mean value of Z_{opt} at every pixel, that is $Z_{opt}(x,y)$. The coarseness of the input image (F_{crs}) is gotten:

$$F_{crs} = \frac{1}{M\times N}\sum_{x=1}^{M}\sum_{y=1}^{N} Z_{opt}(x,y) \quad (5)$$

It can be seen that, Tamura coarseness is a measurement of texture coarseness in a global perspective, and it can only extract coarseness from an entire image or a larger image block, but cannot accurately measure local texture coarseness. Due to the limitation of Tamura's algorithm, a new local texture coarseness calculation algorithm with more general noise robustness is proposed.

We put the principle of Tamura's algorithm shown in figure 1(a). A spike with a width of d is arranged in a spaced D cycles. The optimum output size of each pixel is shown in figure 1(b). As can be seen from the figure 1(b), the optimal size is the expression associated with d and D, and the final output result is $F_{crs}=(3d+D)/4$. The results are in line with the facts, when the value of d and D is larger. The element dimension is larger, the repeating unit is lesser, and the texture coarseness is greater.

Normally, complex texture feature is composed of some simple texture elements [21]. However, the texture element is still a vague concept. There is lack of a good mathematical model to describe it [22]. In this section, we construct mathematical models to analyze these problems. The general texture element in the image has a uniform gray image block, and we can think the image block is just an isolated pixel. The image of figure 1(a) can also be considered as two texture elements with two different dimensions and gray values, and their dimensions are d and D. If the image only includes a texture element, the optimal size is $Z_{opt}(x,y)$; The output is shown in figure 1(c). When $M=N=1$, there is $F_{crs}=Z_{opt}$. Therefore, the optimal size of $Z_{opt}(x,y)$ is used to calculate the local texture coarseness of the pixel point (x,y). The output of $Z_{opt}(x,y)$ should be figure 1(d).

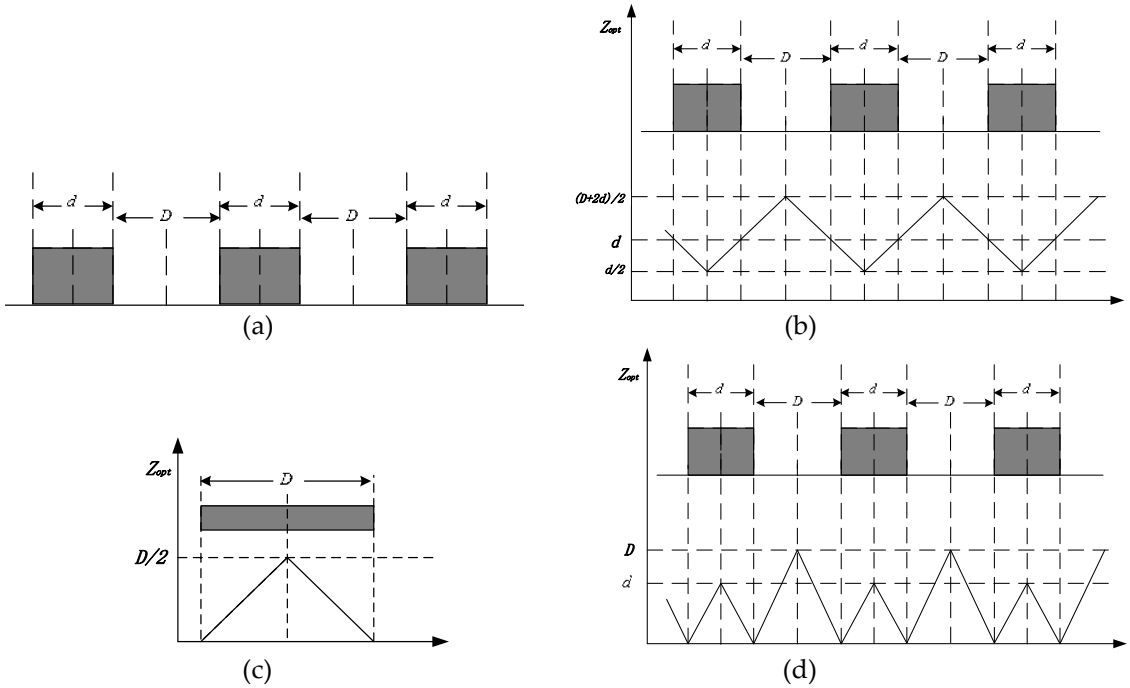


Fig.1 Analysis of local texture coarseness

In summary, the local coarseness is the largest at the center of texture element. The local texture coarseness are minimal at the boundary points of texture element; The pixels between the center and the boundary points have a middle-value local coarseness. the more far away from the center, the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 October 2017 doi:10.20944/preprints201710.0093.v1

smaller the local coarseness. For different texture elements, the greater the dimension, the greater the coarseness at the center of the texture element. Therefore local texture coarseness can be measured by the optimal size of pixels.

Now the pseudo code for the calculation is given.

Algorithm 1: Measurement of The Local Texture Coarseness.

Input: (1) $I(i,j)$—The gray value of input pixel in (i,j) of SAR image
(2) $M\times N$—The size of the input SAR image
(3) $4k\times 4k$—The size of active windows
(4) L_{max}—The maximum window scale
(5) c_1, c_2—Two parameters for adjusting thresholds (T_2,T_3)

Output: F_{crs}—The local texture coarseness

Method

Step 1: Calculate the average intensity value of pixels in the activity window in the image
if k=0
The size of active windows is set to 3×3
else end
for k=1 to L_{max}

$$A_k(x,y)=\sum_{i=x-2^{k-1}}^{x+2^{k-1}-1}\sum_{j=y-2^{k-1}}^{y+2^{k-1}-1} I(i,j)/(4k)^2$$

end-k

Step 2: Calculate the deviation scale of the two windows (L_b)
if $L_{max}\geq 5$
α=3
else $\alpha=\min(2, L_{\max}-1)$
end-L_{max}
$L_b = L_{\max}-\alpha$

Step 3: Calculate the eccentricity of the two windows (ρ)
$k' = \max(k-L_b, 0)$
$\rho = 2k'+1$

Step 4: For each pixel in the image, the average intensity difference between the non-overlapping windows in horizontal and vertical directions is calculated separately
for x=1 to M
for y=1 to N
$E_{k,h}(x,y)=\left|A_{k'}(x+\rho,y)-A_k(x,y)\right|$
$E_{k,v}(x,y)=\left|A_{k'}(x,y+\rho)-A_k(x,y)\right|$
end-x
end-y

Step 5: Calculate the optimum size at each pixel
for x=1 to M
for y=1 to N
$Z_{opt}(x,y)=4\times k_{\max}$
$E_k(x,y)=\max(E_{k,h}(x,y), E_{k,v}(x,y))$
end-x
end-y
$E_{\max}=\max(E_k(x,y))$
$E_{\min}=\min(E_k(x,y))$

Step 6: There is a parameter k_{max}, which is determined by the following methods: it is divided into three situations: The pixel points in the boundary, the pixel points within the larger and smaller dimensional texture elements.
if k=0

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 October 2017 doi:10.20944/preprints201710.0093.v1

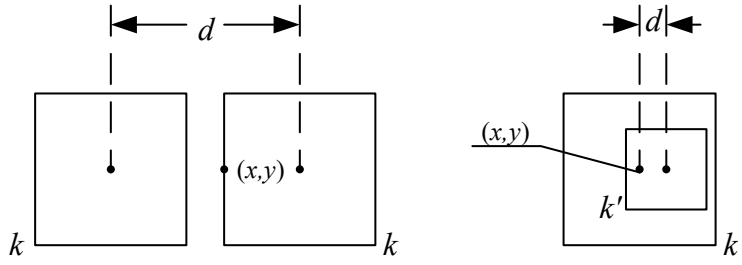
Get the mean of the local non-zero maxima of all the pixels of E_0,that is T_1,
if $E_k(x,y)>T_1$
k_{max}=0
else end
else end-k
In this situation, the current pixel is the point on the boundary of the texture element;
$\overline{E_{\min}} = mean(E_{\min})$
Experiment on many texture images, and found the values of T_2, T_3 are related to $\overline{E_{\min}}$
$T_2 = \overline{E_{\min}} / c_1$
$T_3 = \overline{E_{\min}} \cdot c_2$
$DE_k = |E_k - E_{k-1}|$
if $E_{max}<T_3$
if Numel($DE_k<T_2$)=L_{max}-1
$k_{max} = L_{max}$
else end
In this situation, the current pixel is the point within the larger dimensional element;
else $k_{max} = argmax(E_k)$
In this situation, the current pixel is the point within the smaller dimensional element;
end

Step 7: Calculate the local coarseness of the pixel point according to the optimal size of each pixel in the image

To increase the contrast, we put the power transformation to Z_{opt}
for x=1 to M
for y=1 to N

$$F_{crs}(x,y) = Z_{opt}(x,y)^{\gamma}$$

end-x
end-y



(a) (b)

Fig.2 The windows used to make a difference (a) is the window of Tamura's algorithm, which has the same size. In our model, The two windows that make the difference are the eccentric overlapping windows. There is a deviation in the size of windows, as shown in Fig.2 (b).

Moreover, our model have completed the measurement of local texture coarseness, it is:

$$F^n_{crs}(x,y) = Z_{opt}(x,y)^{\gamma} \tag{6}$$

After the normalization operation and significant treatment of the feature matrix, the feature map is as follows:

$$S_{crs} = N\left(F^n_{crs}(x,y)\right) \tag{7}$$

In the formula and later formulas, $N(\bullet)$is the normalization operation: the feature maps of the model will be normalized to the range [0,N], N is any gray value within the gray range of the input image, this process will reduce the saliency of background, the feature maps after the normalization operation is F'; And then we multiply F' and a coefficient;

$$S = (M - \bar{m})^2 \cdot F' \tag{8}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 October 2017 doi:10.20944/preprints201710.0093.v1

Where: M is the global maximum of F', and M is the average gray value of remaining pixels except the pixels with the gray value of M in F'. S is the initial saliency map of current feature.

2.1.2. Standard deviation

Normally, standard deviation can effectively reflect local features of images, such as the edge and shape features. The SAR images have rich edge information, and the target edge and contour information can be enhanced by extracting the standard deviation of the image, and realize detecting targets from the background [23]. Standard deviation can guarantee the difference between targets and background in a low computational complexity.

The standard deviation is calculated by sliding the filter in the detection image. The size of the filter sliding window used here is related to this balance, which is used to coordinate the relationship between the time consumed by the model and the effectiveness of the saliency inhibition of the background. Assuming $I(i,j)$ is the input image, the size of the filter is $N\times N$, the central coordinate of the filter is (m,n), the formulas for standard deviation are derived as follows:

Firstly, we need to calculate out the average value of the local areas in the image, that is $m_x(i,j)$, M is a parameter related to the size of the filter, $M=0.5\times(N-1)$.

$$m_x(i,j)=\frac{1}{(2M+1)^2}\sum_{i=n-M}^{n+M}\sum_{j=m-M}^{m+M}I(i,j) \tag{9}$$

Standard deviation (F_{std}) can be obtained:

$$\left(F_{std}(n,m)\right)^2=\frac{1}{(2M+1)^2}\left[\sum_{i=n-M}^{n+M}\sum_{j=m-M}^{m+M}\left(I(i,j)-m_x(i,j)\right)^2\right] \tag{10}$$

Pixels on the boundary of target areas is larger than pixels in their adjacent field, so the STD values calculated are larger, which are the highlights in the feature map. From the perspective of image comprehension, the corresponding F_{std} values of each pixels constitute the feature matrix of standard deviation. Moreover, after the normalization operation and significant treatment of the feature matrix, the feature map of standard deviation is obtained. The calculation is as follows:

$$S_{std}=N\left(F_{std}(n,m)\right) \tag{11}$$

2.1.3. Orientation

The orientation feature usually is used to represent the targets with different directions [24]. In our model, the orientation features of pixels in the input image are extracted by Gabor filter, and the filter is shown in the formula:

$$H(x,y,\theta_k,\lambda,\alpha,\beta)=\frac{1}{2\pi\alpha\beta}\exp\left\{-\pi\left[\left(\frac{x_{\theta_k}}{\alpha}\right)^2+\left(\frac{y_{\theta_k}}{\beta}\right)^2\right]\right\}\exp\frac{2\pi i x_{\theta_k}}{\lambda} \tag{12}$$

θ_k is the orientation of the sine wave; and λ is the wavelength of the sine wave; α and β refer to the standard deviation of the Gaussian function in the X-axis and the Y-axis, respectively.

The orientation the sine wave (θ_k) can be obtained by the formula:

$$\theta_k=\frac{\pi}{n}(k-1),k=1,2,\ldots,n \tag{13}$$

Normally, the orientations of the sine wave are periodic, our model selects four orientations (n=4), they are 0°, 45°, 90°, 135°; The calculation of the parameters (consist of x_{θ_k} and y_{θ_k}) is:

$$\begin{bmatrix}x_{\theta_k}\\ y_{\theta_k}\end{bmatrix}=\begin{bmatrix}\cos(\theta_k) & \sin(\theta_k)\\ -\sin(\theta_k) & \cos(\theta_k)\end{bmatrix}\cdot\begin{bmatrix}x\\ y\end{bmatrix} \tag{14}$$

Then our model substitutes the obtained parameters into the filter formula. We get the Gabor filters in four orientations, so we get the feature maps based on these four orientations. They are $F_{ori}(0)$, $F_{ori}(45)$, $F_{ori}(90)$, $F_{ori}(135)$.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 October 2017 doi:10.20944/preprints201710.0093.v1

Last but not least, since our model does not build the Gaussian pyramid structure, the final feature map of orientation (S_{ori}) is obtained by the linear addition of four feature maps. The calculation is as follows:

$$S_{ori} = \bigoplus_{\theta \in \{0,45,90,135\}} N\left(F_{ori}\left(\theta\right)\right) \tag{15}$$

2.1.4. Global contrast

Global contrast is a quantification of the difference between the current local region and the whole image [25]. Generally, the gray value of target areas in the SAR image is relatively high, and the saliency map of the global contrast feature can be used to enhance the difference of the target and the background, so as to highlight the significance of targets. The calculation processes of the global contrast feature based on pixel level are shown as follows.

Assuming $I(i,j)$ is the input image, the size of the input image is $M \times N$. The feature value of global contrast ($val(i,j)$) is calculate by the formula:

$$val\left(i,j\right) = I\left(i,j\right) - \frac{1}{M \times N}\sum_{i \in N}\sum_{j \in M} I\left(i,j\right) \tag{16}$$

Next, the binarization result of the feature is obtained by a criterion:

$$g\left(i,j\right) = \begin{cases} val\left(i,j\right), val\left(i,j\right) \geq \gamma \\ 0, val\left(i,j\right) < \gamma \end{cases} \tag{17}$$

Among that, $g(i,j)$ refers to the final feature value of global contrast for the current pixel. γ is an experience parameter associated with grayscale. Last but not least, normalize it.

$$S_{global} = N\left(g\left(i,j\right)\right) \tag{18}$$

In the formula, S_{global} represents the feature map of global contrast.

2.2. Calculation of Saliency Map

In most cases, the visual models end up with a saliency map that is a synthesis of all the feature maps. The meaning of different feature maps corresponds to the different channels of "attention" [21]. The magnitude of the response of the feature maps corresponding to silent regions is quite different. Some feature maps correspond to more than one silent regions with strong responses. Some feature maps may include only one silent region with relatively weak. Therefore, the mechanism of feature fusion needs to be completed based on the significant levels of the features, rather than the simple linear addition.

According to the features extracted from SAR image in the last section, the feature maps are obtained. The fusion strategy of multiple features used in our model is: first of all, the feature maps of the local coarseness, standard deviation and orientation are adopted in a linear additive way and normalization operation, the saliency of the silent areas in the image has been enhanced by this step. And then we can eliminate the inconspicuous regions in the feature map of global contrast by the multiplication operation, at the same time to strengthen the saliency of the silent regions contained in local feature maps and the global feature map. We calculate the weighted sum of feature maps, after its normalization and we can get a coefficient, and then multiply it by the global feature map, then the total saliency map is calculated as follows:

$$S = N\left(\frac{S_{std} + S_{ori} + c \cdot S_{crs}}{2 + c}\right) \cdot S_{global} \tag{19}$$

Among that, the parameter can be understood as sample weight, c is an empirical adjustment parameter that ranges between 1.5 and 2.2.

2.3. Saliency Detection

Our model optimizes the competition strategy in the ITTI classic model and uses multi-scale segmentation in the saliency map to realize the detection and extraction of silent regions.

2.3.1. Determination of *FOA*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 October 2017 doi:10.20944/preprints201710.0093.v1

Focus of Attention (*FOA*) is the focus of visual attention. In general, *FOA* is the pixel that has the maximum value of the grayscale in the saliency map *S*. If there are more than one pixels with the maximum gray value at the same time, our model mimics the mechanism of the human visual system to deal with multiple regions of interest, and regards the region nearest to the center of the image as the most significant area of visual attention. In this situation, the calculation formulas of the focus of attention are as follows.

Firstly, we need to determine the distance from the center of the image to the current pixel;

$$dis\left[(x,y),(x_0,y_0)\right]=\sqrt{(x-x_0)^2+(y-y_0)^2} \tag{20}$$

Among that, the coordinates of the center of the input image are (x_0,y_0); the coordinates of *FOA* are (x,y); Finally, *FOA* is calculated as follows:

$$FOA=\min\left(dis\left[(x,y),(x_0,y_0)\right]\right) \tag{21}$$

2.3.2. Acquisition of binarization template

According to the method for determining *FOA*, our model calculates the *FOA* of the saliency map *S*, and then we can obtain four pixels corresponding to this *FOA* point in four feature maps, and take the feature map with the maximum gray value among the four pixels as the next saliency map (S') for detecting silent regions, and the binarization operation is carried out by using the *FOA* of the saliency map S' as the center. Last but not least, in the binarization operation, the global threshold segmentation is realized by the traditional Otsu method. The judgment criterion of binarization is:

$$B(x,y)=\begin{cases}1, & \dfrac{S'(x,y)}{sVal}\geq T_0\\ 0, & \dfrac{S'(x,y)}{sVal}<T_0\end{cases} \tag{22}$$

Among that, *sVal* is the gray value of the *FOA* pixel in S', and T_0 is the threshold for image segmentation; $B(x,y)$ is the result of binarization.

The model performs the following two steps on the binarization results obtained above. First of all, Gaussian filtering is performed in the binary image. The parameters of the filter need to be determined according to the prior knowledge of targets. The size of the filter in this article is set as the estimate value of the pixels of the actual target in the image. In the second phase, our model is used to judge the regions after Gaussian filtering.

Assuming that Num' is the number of pixels in the regions to be detected; and $N'\times M'$ is the size of the input image. The proportional parameter needed to be used in the judgment process is *ratio*; it is calculated by the formula.

$$ratio=\frac{Num'}{N'\times M'} \tag{23}$$

β is the ratio between the estimated value of the actual pixel and the number of all pixels in the image, and the criterion are as follows:

$$\begin{cases}ratio\geq\beta, & \text{The current region belongs to the target area}\\ ratio<\beta, & \text{The current region belongs to the background area}\end{cases} \tag{24}$$

Finally, the binary image judged to be the target area is saved.

2.3.3. The multi-scale acquisition of silent regions

In order to avoid the process of finding the next silent region entering into the cycle of death, an operation of "inhibition of return" (IOR) is done. The specific operation is that when the binarization template in a silent region is obtained, this silent region is set to zero, and then the operation of the next *FOA* is started until the retrieval of all silent regions is completed. The detection result of silent regions is obtained through the above operations.

The optimization and improvement of competitive strategy in our model is mainly reflected in the following aspects.

(1) Comparing and analyzing the gray value of *FOA* in each saliency map, the best saliency representation of every positions are selected;

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 October 2017 doi:10.20944/preprints201710.0093.v1

(2) Then the filtering and image segmentation of saliency map is accomplished from several aspects. Finally, our model realizes the accurate screening of the silent regions.

This paper analyzes the classic ITTI model and considers the features of targets in SAR images. In order to improve the accuracy of saliency detection and the edge sharpness of targets in the image, our model draws lessons from the framework of classical ITTI model. Firstly, our model extracts some new different underlying early primary features from the input image. Then we use a new mechanism of feature fusion to get the saliency map. Finally, a new multi-scale competition strategy is adopted to further complete the extraction of salient regions from the input image, as shown in figure 3.

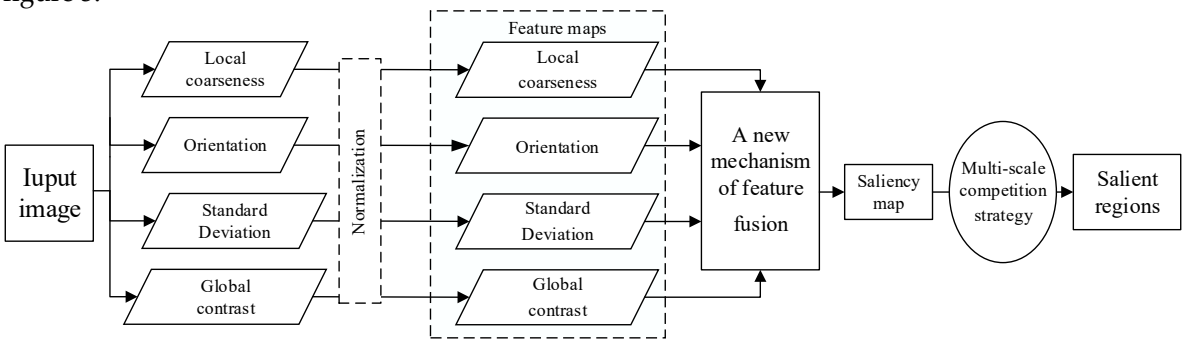


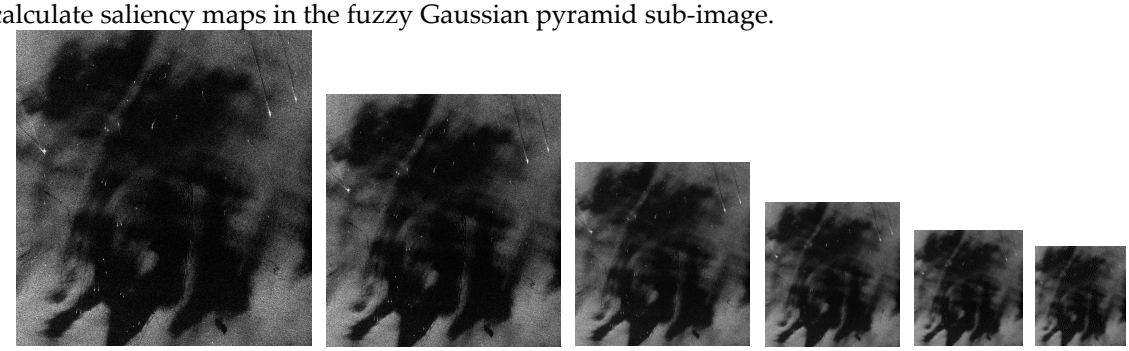
Fig.3 The flow of visual attention model algorithm

3. Evaluation of Texture Saliency Extraction Algorithm

A new method is proposed to measure the texture saliency. In this section, the extraction algorithm is discussed from the following two aspects.

3.1. Complexity Analysis of Algorithm

In terms of the operation process, this algorithm does not need to do the construction of Gaussian pyramid structure. The ITTI model uses the nine-story pyramid structure to simulate the human visual attention system, so as to realize multi-scale representation of images. However, the areas of ships and other targets in SAR image that occupy the detected images are relatively small. Excessive down sampling makes the target information in the detected image missing. It makes no sense to calculate saliency maps in the fuzzy Gaussian pyramid sub-image.



(a) (b) (c) (d) (e) (f)

Fig.4 Results of Gaussian pyramid structure

The figure 4 (a) is the original SAR image. (b) to (f) are the sub-image obtained by Gauss pyramid. In figure (c) to (f), we can clearly see that these sub-image become blurred, and the target information is missing, and feature extraction on them to build saliency map have no practical significance, it also can cause the waste of time.

In terms of the time complexity, our algorithm deals with pixels in two categories. The pixels in the image can be divided into points in texture boundary and points inside the texture elements.

For the interior point of the texture element, the current window size (k) is less than the texture dimension, and it satisfies $E_k \approx 0$; When k is larger than the dimension of the texture element, it is clear that $E_k \gg 0$, and the maximum value is E_{max}, then $k_{max}=k$; When the size of the texture element is very

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 October 2017 doi:10.20944/preprints201710.0093.v1

large, the values of all E_k are small and they are similar. At this point, $k_{max}=L_{max}$. Use the constraint conditions: Numel($DE_k<T_2$)=L_{max}-1 and $E_{max}<T_3$ to judge.

For the boundary points, E_k is larger and $E_k \gg 0$. At this point, set $k_{max}=0$. Because E_0 contains the information of original texture boundary, so we use the condition: $E_0>T_3$ to judge the boundary points. The value of T_3 is set as the average of all local non-zero maximums in pixels of E_0. we can get: $k=0$, $E_0<T_3$.

In a word, it is fast and effective to distinguish the internal points of the affected texture elements from the boundary points.

3.2. Noise Robustness

3.2.1. The mathematical model

The image is disturbed by noise in the process of acquisition and propagation. The extraction algorithm proposed in this paper is applied to SAR images, while the SAR image has higher noise than the general optical image. Therefore, the noise robustness of algorithms must be considered.

Considering the additive noise $n(i,j)$, the intensity value of pixel (i,j) in SAR image ($I(i,j)$) is changed into $f(i,j)$: $f(i,j)=I(i,j)+n(i,j)$ .Then, there are:

$$\begin{aligned} E_{k,h} &= \left| \frac{1}{N_{k'}} \sum_{(i,j)\in A_{k'}} f(i,j) - \frac{1}{N_k} \sum_{(i,j)\in A_k} f(i,j) \right| \\ &= \left| \frac{1}{N_{k'}} \sum_{(i,j)\in A_{k'}} I(i,j) - \frac{1}{N_k} \sum_{(i,j)\in A_k} I(i,j) + \frac{1}{N_{k'}} \sum_{(i,j)\in A_{k'}} n(i,j) - \frac{1}{N_k} \sum_{(i,j)\in A_k} n(i,j) \right| \end{aligned} \tag{25}$$

N_k is the total number of pixels in the window area (A_k). When the area A_k and $A_{k'}$ are in the same texture element, we can get:

$$E_{k,h} = \left| \frac{1}{N_{k'}} \sum_{(i,j)\in A_{k'}} n(i,j) - \frac{1}{N_k} \sum_{(i,j)\in A_k} n(i,j) \right| \tag{26}$$

There is a condition: when the radius of the probability distribution of $n(i,j)$ is small, that is r, the values of N_k and $N_{k'}$ are greater than the value of r, considering Wiener-khinchin law of large numbers, we can get:

$$\lim_{N\to\infty} P\left\{ \left| \frac{1}{N} \sum_{i=1}^{N} a_i - \mu \right| < \varepsilon \right\} = 1 \tag{27}$$

Then we can derive the following formulas:

$$\frac{1}{N_{k'}} \sum_{(i,j)\in A_{k'}} n(i,j) \approx \mu_n \tag{28a}$$

$$\frac{1}{N_k} \sum_{(i,j)\in A_k} n(i,j) \approx \mu_n \tag{28b}$$

Among that, μ_n is the mean of the noise $n(i,j)$. Then we can get:

$$E_{k,h} \approx 0 \; ; E_{k,v} \approx 0 \tag{29a}$$

$$E_k \approx 0 \tag{29b}$$

Considering the values of N_k and $N_{k'}$ are greater than the value of r, we can get:

$$E_{k,h} = \left| \frac{1}{N_{k'}} \sum_{(i,j)\in A_{k'}} I(i,j) - \frac{1}{N_k} \sum_{(i,j)\in A_k} I(i,j) \right| \tag{30a}$$

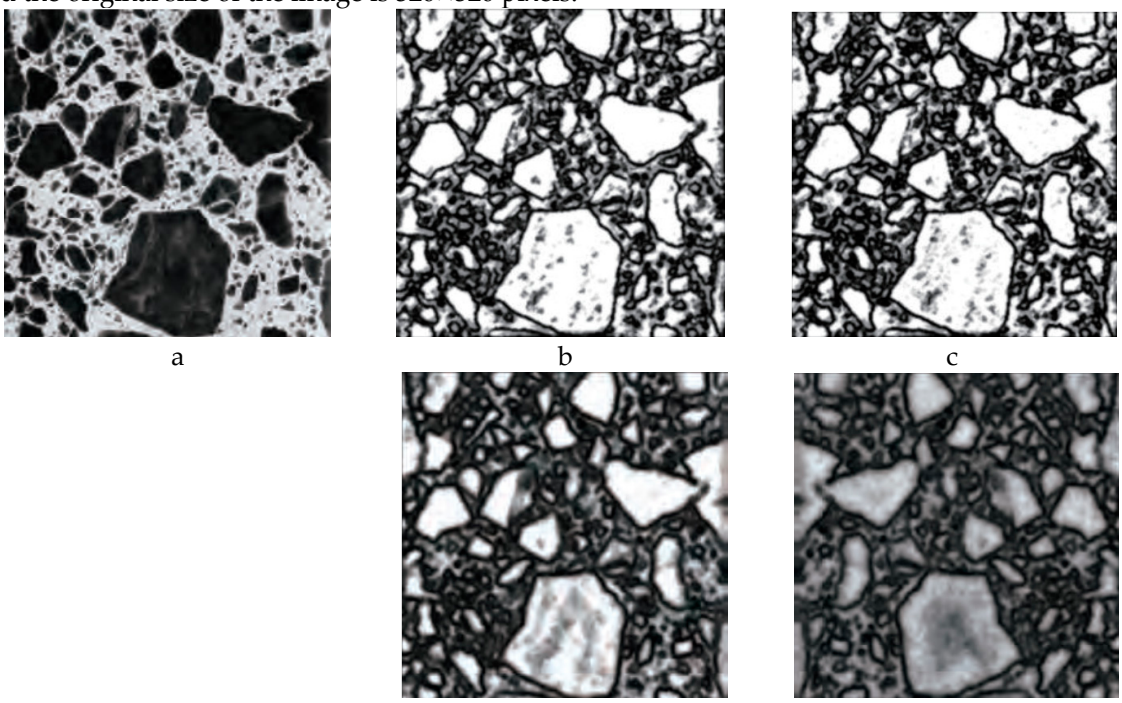
$$E_{k,v} = \left| \frac{1}{N_{k'}} \sum_{(i,j)\in A_{k'}} I(i,j) - \frac{1}{N_k} \sum_{(i,j)\in A_k} I(i,j) \right| \tag{30b}$$

Obviously, the bigger the value of N_k is, the better the condition is met, the noisy suppression effect. In fact, after the calculation of average intensity difference in the first few steps of the algorithm, $E_{k,h}$ and $E_{k,v}$ is the intensity difference after the mean filtering of the original image, and theoretically the algorithm also should have good anti-noise ability.

 doi:10.20944/preprints201710.0093.v1

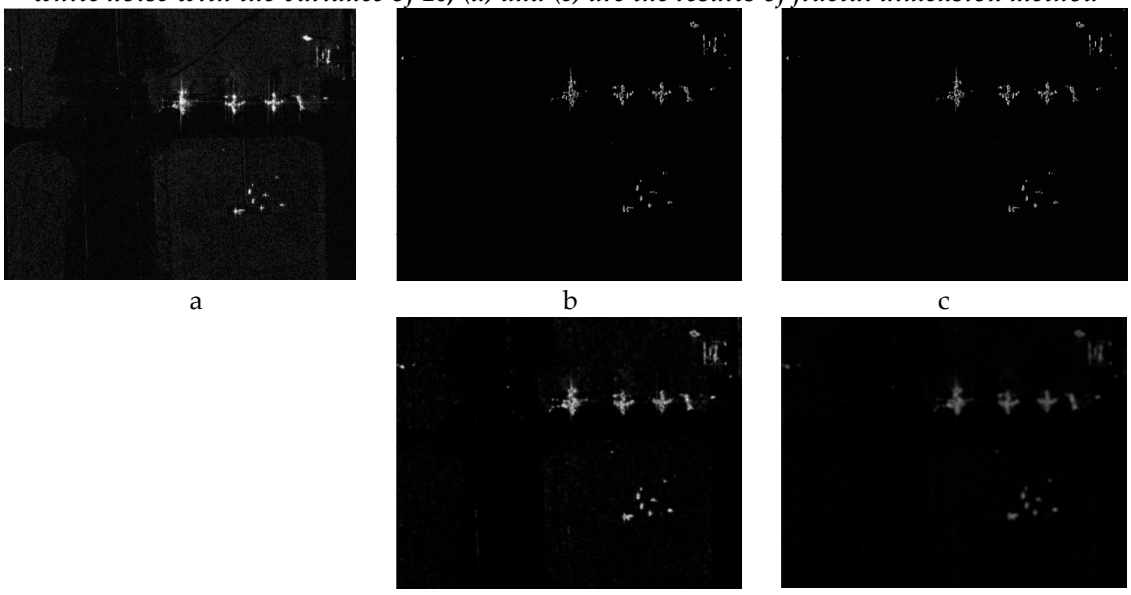
3.2.2. Analysis of simulation experiments

In order to prove the effectiveness of the extraction algorithm, the algorithm has carried on the experimental analysis to some images. Experimental images include images in Brodatz's texture library and some natural scene images, and the results are compared with Novianto's algorithm based on local fractal dimension [22].In order to show the consistency of the coarseness feature map obtained by the two methods, the coarseness feature map of fractal dimension method is inverted. Figure 5 is the processing result of the natural scene texture image from the Brodatz's texture library, and the original size of the image is 320×320 pixels.



a b c d e

Fig.5 Texture extraction of natural scene image (a) is the original image; (b) is the feature map of texture saliency in the original image; (c) is the feature map of the image after adding the Gaussian white noise with the variance of 10; (d) and (e) are the results of fractal dimension method



a b c d e

Fig.6 Texture extraction of SAR image The annotation of each image is the same as Fig. 5.

From the coarseness feature map extracted by the proposed algorithm we can see that the value of each pixel point corresponds to the value of the local coarseness of the image, and then the texture coarseness distribution of the original image is given accurately. Compared with the fractal

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 October 2017 doi:10.20944/preprints201710.0093.v1

dimension algorithm, the proposed algorithm is as effective as fractal dimension method. When the Gaussian white noise is added to the original image whose variance is 10, we can get two coarseness feature maps by these methods. It is easy to see that the noise has little impact on the result of the algorithm proposed in the paper, but has a bad impact on the result of fractal dimension method. In the experiments, we found that, even if we use the 5×5 window instead of the 3×3 window in fractal dimension algorithm, the feature map is still influenced by noise greatly, and we can't use simple post-processing such as median filtering to filter noise.

In a word, our algorithm have a good noise robustness.

4. Evaluation of Salient Regions Extraction

This section carries out the saliency detection experiments on the basis of visual saliency and the design theory of the proposed model. The results of the model are given for SAR images. Considering TS-X image is a typical high-resolution SAR image, several representative TS-X images are selected. Compared with classical models, the advantages and disadvantages of our model are analyzed.

4.1. Evaluation Index of Detection Results

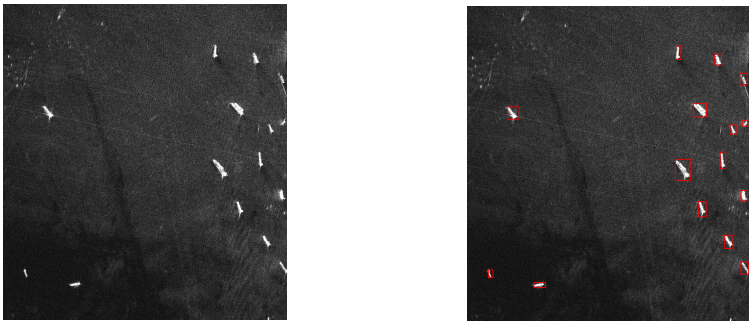
In order to verify whether the model is valid to SAR images, a TS-X image is selected as the experimental data, whose pixels are 4015×3616. The imaging area is Strait of Gibraltar, the sampling rate of the image is 1.25 meters, and the polarization mode is HH polarization. The distribution of sea in this area is complex and there are a lot of ships and strong speckle noise, and the area contains a certain region of non-uniform ocean background. In this area, combining with GPS and AIS data analysis, we collect and sort out the geographical information of the Strait of Gibraltar, and determine the number and location of ships in SAR image data, and improve the objectivity of evaluation in experimental results. For better evaluation of detection algorithms, define the detection rate (P_t) and the Figures of Merit (*FoM)* [26]:

$$P_t = \frac{N_{tt}}{N_{gt}} \times 100\% \tag{31}$$

$$FoM = \frac{N_{tt}}{N_{fa} + N_{gt}} \tag{32}$$

Among that, N_{tt} is the number of targets detected by current algorithm; N_{fa} is the number of false alarm; N_{gt} is the number of actual targets.

The experiment data used in the paper is shown in Figure 7 (a), which contains a large number of non-uniform sea background and different sizes of ship targets. Figure 7 (b) is the image after marking targets, and the actual number of ships in the experimental image is 15.



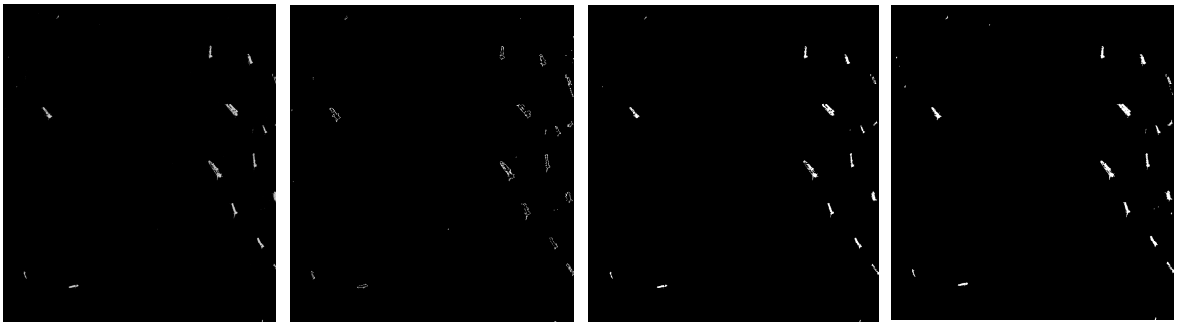
a. The original image b. The image after marking targets

Fig.7 Experimental simulation image

4.2. Analysis of Experimental Results

According to the model proposed in this paper, the image is processed and the feature maps are obtained. The parameters settings in the process of calculating feature maps include in the input image: the size of sliding window in the standard deviation feature is set as 5×5; The parameter γ in global contrast feature is 200.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 October 2017 doi:10.20944/preprints201710.0093.v1

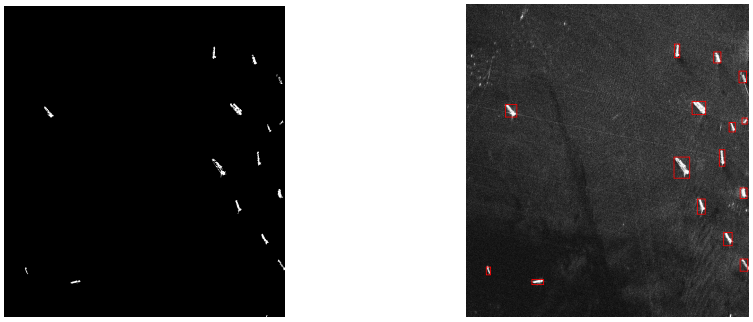


a. Local coarseness b. Standard deviation c. Orientation d. Global contrast

Fig.8 The feature maps of our model

As can be seen in Figure 8, the feature maps shown in this paper are of better clarity.

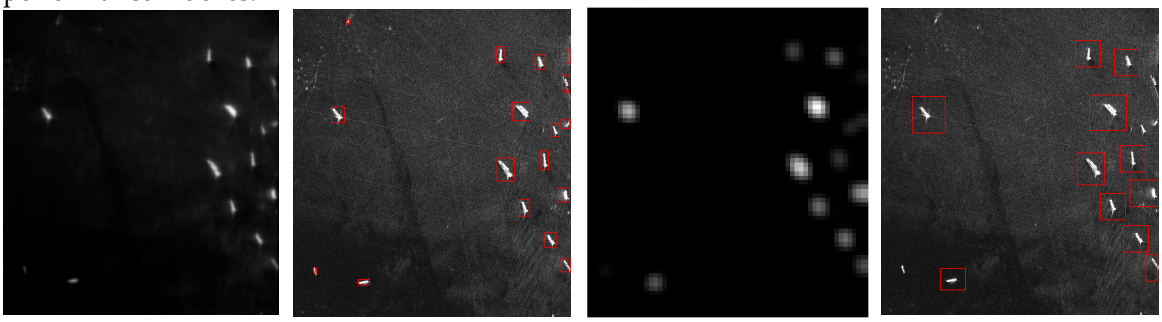
First of all, as can be seen from the feature map of texture coarseness, the method of measuring texture coarseness presented in this paper has a relatively accurate measurement output. The feature map can even provide us with all the accurate target information. And the standard deviation can correspond to the edge and texture features of targets in SAR image. The feature map of orientation characterizes and distinguishes the targets with different orientations. In the feature map of global contrast, the little stronger target information in the complex background is enhanced from a global perspective.



a. The saliency map b. The detection result

Fig.9 The results of silent regions in our model

Figure 9 shows the simulation results of the experimental data. (a) is the final saliency map of the input image; (b) is the detection result of silent regions. Figure 10 shows the results of two classical visual models. In this paper, three models are analyzed based on the quantitative analysis of performance indexes.



a b c d

Fig.10 The detection results of two classical models (a) is the saliency map of the Hou's model, (b) is the detection result of the model; (c) and (d) are the classical ITTI model

Table.1 Comparison of three detection results of the visual models in the TS-X image

Model	N_{tt}	N_{fa}	Detection time	P_t	*FoM*	Detail analysis
ITTI model	11	0	45.63s	73.3%	0.733	The silent regions are fuzzy and contain part of the background areas
HOU model	15	3	70.13s	100%	0.833	Some target areas are missing
Our model	15	1	30.24s	100%	0.938	The edge details of the result are relatively good

Table 1 shows the comparison of the performance of our model and two other traditional visual attention models. Among them, the evaluation indexes include the number of targets detected, the

 doi:10.20944/preprints201710.0093.v1

number of false alarms and the number of missing targets and detection time. Compared with classic ITTI model and HOU model, the target detection algorithm in SAR images proposed in this paper has a low leakage number and a low false alarm number, which guarantees the accuracy of target detection. That is because the algorithm takes texture features of image into account. Combining the location with the contrast information of target can effectively enhance the target saliency and inhibit the saliency of the background region at the same time. It provides the guarantee for accurate detection target.

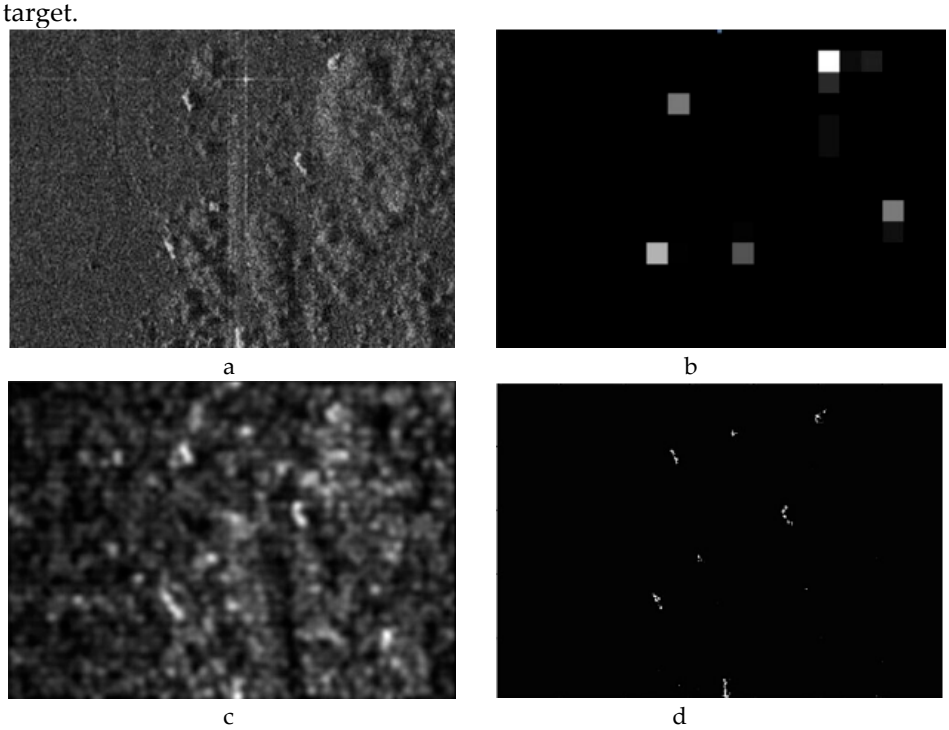


Fig.11 The detection results complex scene SAR image (a) is a complex scene SAR image, (b) is the saliency detection result of the ITTI model; (c) is the saliency map of the Hou's model; and (d) is the saliency detection result of our model.

Fig.11 is the result of a simulation experiment of a complex scene SAR image. The regions of targets and background in the SAR image are relatively similar, and the background is very complex. A saliency map of the image using the ITTI model is shown in the figure 11 (b), and it includes a higher missed rate and false alarm rate; (c) is a saliency map obtained by using Hou model, which cannot distinguish the target region and background clutter region, so that the obtained saliency map is meaningless, and the model fails. According to the saliency map obtained by the model and proposed in the paper, that is (d), our model has a better detection effect, and can inhibit the saliency of background clutter, filter the features of background, highlight the target contour shape, and enhance the saliency of targets.

5. Conclusions

Considering the characteristics of SAR image, this paper analyzed the basic theory of classical visual models. We focus on the problem of their poor performance when the classical visual models are applied to SAR images with complex background. A new visual model for detecting targets in SAR images is presented in the paper. Firstly, our model extracts several special features which can describe the SAR image better. After a series of calculations, the feature maps are obtained; Secondly, the model combines the feature maps to obtain the final saliency map by a new mechanism of feature fusion; Finally, the extraction of silent regions is achieved through a multi-scale competition strategy, so as to realize the saliency detection of SAR image.

In the end, the performance of our visual model and classical visual models are simulated in the uniform clutter environment. A number of experiments were performed in TS-X images with a complex background. The results show that our model has a better performance: the lower false

alarm rate and the better contour shape. Our model has a great advantage in the saliency detection of targets in SAR image.

The research of next phase is: (1) When we complete the extraction of visual features, how do we combine the attributes of targets to select a feature? how to select and extract features which can represent the targets accurately with a small amount of calculation as a prerequisite? (2) How to establish a new mechanism of feature fusion, which can adaptively adjust the proportion of each feature? (3) When carrying out the calculation of feature maps, it is necessary to further improve the parameters setting of these filters.

Acknowledgments: Received funds for covering the costs to publish in open access: National Natural Science Foundation of China: Basic Theory and Key Technology of Spatial Information Network for Continuous Observation of Sea Target. (Number: 42511133N)

Author Contributions: Yongli XU and Wei XIONG conceived and designed the experiments. Yongli XU and Libo YAO performed the experiments. Yongli XU, Libo YAO and Wei XIONG analyzed the data. Yongli XU and Yafei LV wrote the manuscript, Yongli XU and Wei XIONG reviewed the manuscript.

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

References

1. Moreira A. A tutorial on synthetic aperture radar[J]. *IEEE Geoscience and Remote Sensing Magazine*, **2013** 1(1): 6-43.
2. Yun-kai DENG, Feng-jun ZHAO, Yu WANG. Development trend and application of spaceborne SAR Technology[J]. *Journal of Radar*, **2012** 1(1): 1-10.
3. Jinsong CHONG, Yue OUYANG, Minhui ZHU. Detection of ocean target in synthetic aperture radar imagery[M]. *Ocean Publishing Firm*, **2006**.
4. Xiangwei XING, Kefeng JI. Review of ship surveillance technologies based on high-resoluion wide-swath synthetic aperture radar imaging[J]. *Journal of Radar*, **2015** 4(1): 107-121.
5. You HE, Jian GUAN, Yingning PENG. Automatic radar detection and constant false alarm rate processing[M]. *Tsinghua University Press*, **1999.**
6. L. Gagnon, H. Oppenheim and P. Valin. R & D activities in airborne SAR image processing/analysis at Lockheed Martin Canada[C]. *Proceeding of SPIE*, 998-1003.
7. Kazuo Ouchi, Shinsuke Tamaki, Hidenobu Yaguchi and Masato Iehara. Ship detection based on coherence images derived from cross correlation of multilook SAR images[J]. *IEEE Geoscience and Remote Sensing Letters*, **2004** 1(3): 184-187.
8. Hou X and Zhang L. Saliency detection: a spectral residual approach[C]. *Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition*, **2007**: 1-8.
9. Itti L, Koch C, Braun J. Revisiting spatial vision: toward a unifying model[J]. *Journal of the Optical Society of America*, **2000** 17(11): 1899-1917.
10. Xiangwei XING, Kefeng JI, Huanxin ZOU. Feature selection and weighted SVM classifier based ship detection in PolSAR imagery[J]. *International Journal of Remote Sensing*, **2013** 34(22): 7925-7944.
11. Gui GAO. A parzen-window-kernel-based CFAR algorithm for ship detection in SAR images. *IEEE Geoscience and Remote Sensing Letters*, **2011** 8(3): 557-561.
12. L. Itti. Models of bottom-up and top-down visual attention[D]. *California Institute of Technology*, **2000** 13-19, 32-34.
13. Licheng JIAO, Xiangrong ZHANG, Biao HOU. Intelligent SAR image processing and interpretation [M]. *Science Press*, **2007**.
14. Gao D, Han S, Vasconcelos N. Discriminant saliency, the detection of suspicious coincidences, and applications to visual recognition[J]. *Pattern Analysis and Machine Intelligence*, **2009** 31(6): 989-1005.
15. Harm G. Developments in detection algorithms at JRC[C]. *The Third Meeting of the DECLIMS Project*, **2004**: 1-7.
16. Achanta R, Estrada F, Wils P. Salient region detection and segmentation[C]. *Proceedings of the 6th International Conference on Computer Vision Systems*, **2008**: 66-75.
17. M S Livingstone, D H Hubel. Psychophysical evidence for separate channels for the perception of form, color, movement, and depth[J]. *Journal of Neuronscience*, **1987** 7: 3416-3468.

 doi:10.20944/preprints201710.0093.v1

18. P J Burt, E H Adelson.The laplacian pyramid as a compact image code[J]. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **2002** 31(4): 532-540.
19. Dirk Walthera, Christof Kochb. Modeling attention to salient proto-objects[J]. *Neural Networks*, **2006** (19): 1395-1407.
20. H. Kong, Y. Nie, A. Vetro. Coding artifact reduction using edge map guided adaptive and fuzzy filter[C]. *in Proc. IEEE Int. Conf. on Multimedia and Expo*, 2004, 1135-1138.
21. M. M. Cheng, G. X. Zhang, N. J. Mitra. Global contrast based salient region detection[C]. *Proc. IEEE CVPR,* **2011:** 409-416.
22. E V Abdelkaw, Mc Gaughy D, Wavelet-based image target detection methods[C]. *International Conference on Sensors and Control Techniques*, **2003**: 337-347.
23. Odelson, B.J., Rajamani, M.R.; Rawlings, J.B. A new autocovariance least-squares method for estimating noise covariances[C]. *Automatica 2006*, 42, 303–308.
24. Lan. H, Liang. Y, Yang. F, Wang. Z, Pan. Q. Joint estimation and identification for stochastic systems with unknown inputs[J]. *IET Control Theory Appl*. 2013 7: 1377-1386.
25. W. Wang, Y. Wang, Q. Huang, W. Gao, Measuring visual saliency by site entropy rate[C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, **2010**: 2368-2375.
26. M.H. Wilder, M.C. Mozer, C.D. Wickens. An integrative experience-based theory of attentional control [J]. *Journal of Vision,* **2011** 11(2): 1-30.