Method of Target Recognition in High-resolution Remote Sensing Image Based on Visual Saliency Mechanism and ROI Region Extraction

Bing Liu[1], Tingwei Chen[1], Ping Fu[1], Yumei Zhen[1], Jeng-Shyang Pan[2,3,4]

[1] Department of Test and Control Engineering, Harbin Institute of Technology, China
[2] Department of College of Computer Science and Engineering, Shandong University of Science and Technology, China
[3] Fujian Provincial Key Laboratory of Big Data Mining and Applications, Fujian University of Technology, China
[4] Chaoyang University of Science and Technology, Taiwan

liubing66@hit.edu.cn, Ianisctw1212@outlook.com, fuping@hit.edu.cn, 52476571@qq.com, jengshyangpan@gmail.com

Abstract

Different from those in low-resolution tasks, targets in high-resolution remote sensing recognition tasks are closed. Usually targets with higher resolution, such as oil tanks and ships bear relatively simple features and gather densely in small areas. The recognition of such targets is more dependent on the fundamental features. Ships, tanks and aircraft targets with high resolution are closed targets, which contain obvious contour and have features that are quite different from the surrounding environment. Therefore, aiming at the targets in remote sensing images with high resolution of sub-meter that is below 1m, we proposed a recognition method based on region of interest (ROI) extraction using visual saliency mechanism. Combining the result of saliency detection, MeanShift algorithm is used to image segmentation and feature extraction. Then, machine learning algorithms are applied to perform the target recognition. We achieved a recognition accuracy of more than 85% on the dataset of aircraft, oil tank and ship targets in our work.

Keywords: High-resolution, Remote sensing, Visual saliency, ROI region, Target recognition

1 Introduction

For target detection in remote sensing images, there are several challenges. Usually targets in remote sensing images are quite small and dense. And the complexity of the circumstances also contributes to the difficulty of target detection in remote sensing images [1]. At present, the existing traditional remote sensing target recognition algorithms usually adopt the strategy of combining coarse and precise detection, extracting the ROI regions from the input images initially. After that, target recognition with higher accuracy is performed based on the extracted ROI regions to remove the negative regions and eventually locate the positive objects regions. Scholars have proposed many schemes for the process of extracting candidate regions in images. Among them, Zhu et al. [2] proposed a method of threshold segmentation, taking advantage of the gray information and edge information in images. Yang [3] and others proposed a local image binary model as the image target feature. The target recognition and detection techniques based on feature analysis and MeanShift algorithm was proposed in [4]. However, these algorithms are not meant for recognition of remote sensing targets. Hence, it is necessary to conduct a customized analysis of the feature of remote sensing targets. Chun et al. [5] proposed a coastal oil tank detection method based on recognition of T-shaped harbor. [6] presents an automatic recognition method for ship targets based of the local invariant feature extraction algorithm SIFT(Scale Invariant Feature Transform). Huang et al. [7] also adopts the SIFT algorithm, combing with nearest feature learning classification to propose a novel adaptive intelligent ship detection method-nearest feature learning classifier (NFLC). [8] presents an adaptive contrast threshold-SIFT (ACT-SIFT) procedure to gain better matching robustness. Wang et al. [9-11] present a remote target recognition approach, using joint feature dictionary for sparse representation based on different feature information for adaptive weighting. Hu et al. [12-14] create algorithms using CNN to achieve automatic object detection in remote sensing. Xia et al. [15-16] use CaffeNet network model to extract oil tank objects from remote sensing images of high-resolution.

Considering the specialty of remote sensing targets under high-resolution analysis, we proposed a new way to help detect these targets with high resolution of sub-meter that is below 1 meter. Our contribution will be listed below. First, we applied visual saliency ROI region extraction on high-resolution remote sensing

targets detection. Second, our scheme adopts MeanShift to segment the images and applies machine learning algorithms to detect the targets based on of visual saliency ROI region extraction. Finally, we have tested and analyzed the performance of our scheme on aircrafts, oil tanks and ships.

2 Algorithm Framework

In this paper, the mode of rough detection and precise detection is adopted. In this mode, the detection process can be divided into two parts. The first part is the quick acquisition of the candidate area while the second part is the description and recognition of the target. Taking the ship monitoring process as an example, the process of using the traditional machine learning algorithms to perform target recognition under high resolution is shown in Figure 1.
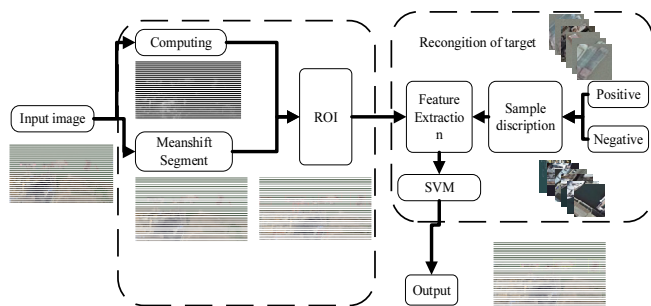

Figure 1. Flow chart of target recognition in remote sensing image with high resolution

The main steps are as follows: First, calculate the saliency map of the input image, and simultaneously use the "MeanShift" algorithm to segment the original image, and merge the fragmented area in the original image. By combining the saliency map with the results obtained by MeanShift segmentation and the basic shape features of the research target, the subsequent ROI regions can be filtered. The specific region extraction method will be detailed later. Thirdly, create a sample dataset by means of distributing the collected target sample files into positive samples and negative samples with a ratio of 1:3, and generate a sample description file. Subsequently, appropriate features are designed for the prepared sample description file to extract and describe the targets. After that, the feature descriptors are sent to the support vector machine (SVM) classifier that will output support vectors for subsequent identification.

Eventually, conducting the feature extraction and description on the selected ROI regions, input the images to the pre-trained SVM classifier and the recognition results will be output.

3 Algorithm Implementation and Result

3.1 The Extraction of the ROI Regions Based on Visual Saliency

It is highly necessary to locate the possible regions of the target while the background of the targets occupied most of the image area, eliminate redundant information. One of the commonly used research ideas is to extract regions by means of visual saliency. The ROI extraction algorithm based on the visual saliency mechanism draws on the human visual selection attention mechanism.

When the human visual system observes a certain scene, it will selectively extract the key information from the current scene, which is called the selective attention mechanism of human vision [17-18]. The current saliency detection approaches can be divided into two types. One approach is the bottom-up data-driven saliency detection, which usually utilize information such as local features, spectrum, and local contrast of the image to measure the saliency of the image. The second can be regarded as a top-down type visual saliency model. The generation of saliency maps is implemented by combining and regulating the scale, position, size, contour, and other features of the bottom-up detection result. The calculation of the saliency model is more complicated than the first approach.

In this paper, we adopt the saliency detection algorithm based upon color features of images to calculate the saliency map of the input image. In view of the visual saliency map, the implementation flow shown in Figure 2 is designed for the ROI region extraction in recognition of the high-resolution remote sensing targets.
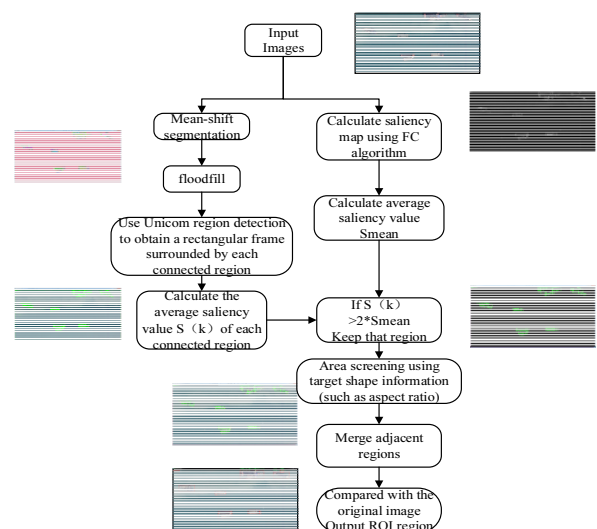

Figure 2. The extraction process of ROI regions

The main steps of the process of ROI region extraction are as follows:

(1) Calculate the saliency value $S(x,y)$ of every pixel through visual saliency detection algorithm and generate the saliency map of every input image.

(2) Calculate the mean saliency value S_{mean} of each image.

(3) For every input image, utilize MeanShift algorithm to segment the Gaussian filtered image, and then merge the small regions of the image and merge as much as possible the parts of high similarity in the background.

(4) Utilizing the joint area detection and removing part of the background area through geometric features of the target in image, the ROI was initially obtained and the mean of saliency $S(k)$ of each area k can be calculated.

(5) Compare $S(k)$ with S_{mean}, keep the region if $S(k) > 2 \times S_{mean}$.

(6) Select the regions by shape features of the targets and then merge the adjacent regions.

(7) Output the final ROI regions with help of the original image.

The process of saliency detection is shown in Figure 3, which mainly utilizes the color and brightness information of the image in Lab space. In this way, an input image is first filtered by a Gaussian filter kernel to remove part of the noise. The saliency calculation function is as shown in (1):
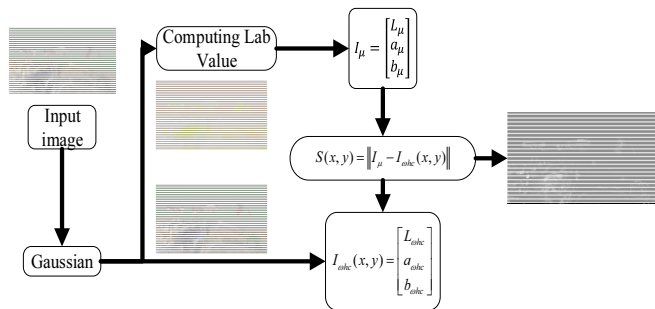
Figure 3. Saliency map acquisition

$$S(k) = \left\| I_\mu - I_{whc}(x,y) \right\| \tag{1}$$

Where $S(x,y)$ denotes the saliency value of the pixel at coordinates (x,y) in the image. And I_μ is the mean value of each channel mapped to the Lab space after Gaussian filtering whose expression is:

$$I_\mu = \begin{bmatrix} L_\mu \\ a_\mu \\ b_\mu \end{bmatrix} \tag{2}$$

where L_μ, a_μ and a_μ denote the mean values of the three channels L, a, b respectively in Lab space. $I_{whc}(x,y)$ denotes the vector that describes the mapping relation between RGB space and Lab space, shown in (3)

$$I_{whc}(x,y) = \begin{bmatrix} L_{whc} \\ a_{whc} \\ b_{whc} \end{bmatrix} \tag{3}$$

The comparison between the saliency map obtained by the saliency detection algorithm and the original image is shown in Figure 4. The MeanShift algorithm in Figure 4 is essentially a clustering algorithm. It is a parameterless clustering algorithm for feature space whose calculation is dependent on the estimate of probability density which has no requirement on either the number of the input of clusters and or the shape of the cluster. The feature space can be regarded as one of its posterior probability density functions. In the unknown probability density functions, modes correspond to the part with higher density, and the same cluster is composed of the data that bear the same mode. If there is a certain d-dimensional space R^d, given n sample points $x_i(i = 0,1,2,3,...,n)$, then for the point x, its corresponding MeanShift vector can be described as the form of (4):

$$M_h(x) = \frac{1}{n} \sum_{x_i \in S_h} (x_i - x) \tag{4}$$

S_h denotes a high-dimensional sphere with a radius h. The definition of S_h is:

$$S_h(x) = \left(y \,\middle|\, (y - x)(y - x)^T \le h^2 \right) \tag{5}$$
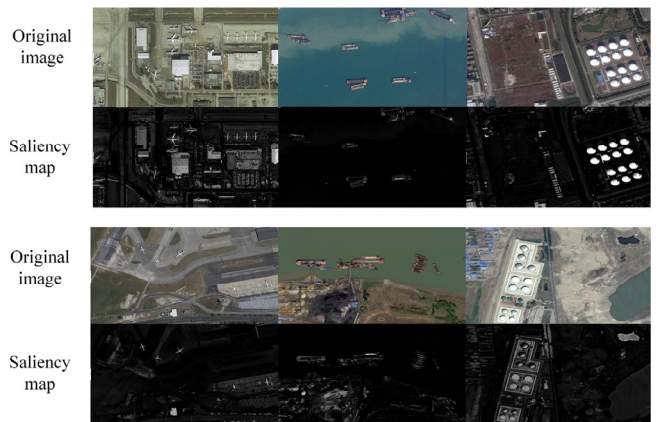
Figure 4. Saliency map of high-resolution images

In real circumstance, each point has a different contribution to point x in the region S_h. Therefore, in order to better describe the situation, the concept of

kernel function and sample weight is introduced into the MeanShift vector. The introduction of the kernel function causes the influence of the offset on the MeanShift vector to vary with distance as the sample distance is shifted by the offset point. After introducing the kernel function and the sample weight parameter, the MeanShift vector can be expressed as:

$$M_h(x) = \frac{\sum_{i=1}^{n} G_H(x_i - x)\omega(x_i)(x_i - x)}{\sum_{i=1}^{n} G_H(x_i - x)\omega(x_i)} \quad (6)$$

Where $G(x)$ and $\omega(x_i)$ denote the introduced kernel function and sample-weight parameter. H is a positive-definite bandwidth matrix. The form of H is as shown in (7):

$$A = \begin{bmatrix} h_1^2 & 0 & \cdots & 0 \\ 0 & h_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_d^2 \end{bmatrix}_{d \times d} \quad (7)$$

The MeanShift vector can be revised into the form of (8) after introducing the kernel function and sample-weight parameter.

$$M_h(x) = \frac{\sum_{i=1}^{n} G\left(\frac{x_i}{h_i}\right)\omega(x_i)(x_i - x)}{\sum_{i=1}^{n} G\left(\frac{x_i}{h_i}\right)\omega(x_i)} \quad (8)$$

Technically MeanShift vector can be interpreted as a kind of regularized probability density gradient, the direction of which is the weighted average of the direction vectors of the individual data points. Therefore, MeanShift algorithm is actually a method that uses the probability density gradient to obtain the local optimal solution in the sample points.

It is essential to translate the problems into a density estimation problem when applying MeanShift to solve them. We mainly focus on color and coordinates information when involving image processing. Segmentation issues concentrates on locating the class center of each pixel in the input image, while all the points with the same center can be regarded as one group of cluster. When using MeanShift to segment images, each point (x, y) of the input image can be expressed as a set of multidimensional data (x, y, r, g, b) consisting of coordinates and RGB color values, and the algorithm will locate the region with the highest density in the input multidimensional data using the window scan space. The spatial position of each point in an image, i.e., the range of variation of its corresponding coordinates in the image, differs vastly from the range of variation of RGB color values. Therefore, two different sizes of windows can be used to clustering in the algorithm in these two dimensions.

When the MeanShift window moves, all the points that converge to the same peak value after window transformation will form the same cluster corresponding to this peak. Such implementation on the image will lead to image segmentation.

For the high-resolution remote sensing images in this paper, the result of MeanShift segmentation is shown in Figure 5. The same color in the figure represents the same area. It is clear that for the dataset used in this paper, MeanShift can distinguish the targets to be detected from the background. Despite of the small fragment areas in the detection result, background with large areas such as waters will be eliminate efficiently. The regions obtained from the segmentation map obtained through MeanShift can be screened according to the shape features of the high-resolution image objects, despite the fact that vast small fragmented areas can be found. Comparing to the entire input image, target with high resolution such as ships or toil tanks in this paper are small agglomerate targets so that the small regions can be eliminated utilizing the shape information of the area of each communication regions, the aspect ratio, etc.. For the filtered regions, we calculate the average saliency value $S(k)$ with the assist of the results of the saliency test and contrast $S(k)$ against the S_{mean} of each image. Further filtering towards the segmentation result will be processing by means of this flexible threshold screening pair. If $S(k) > 2 \times S_{mean}$, the region is reserved, and larger saliency region constituted of the adjacent regions with less difference in saliency will be extracted as a candidate ROI, otherwise it is regarded as a background.
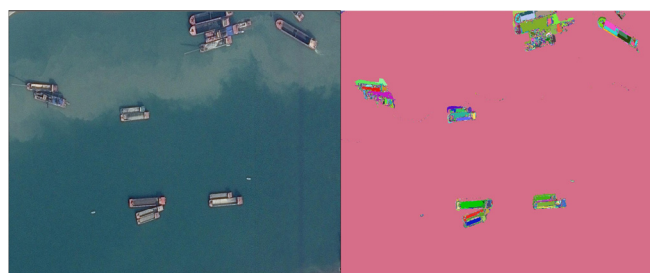

Figure 5. The result of MeanShift segmentation

Now that most of the backgrounds in the reserved ROI area have been eliminated, small regions around the target is the next to be addressed. In order to merge the target regions into one region, the adjacent regions are amalgamated due to the surrounding rectangular frame.

After the merge, the final extraction result of ROI region is output. Finishing the above implementation, the extraction of ROI region in the process of high-resolution remote sensing image target recognition is completed. The result of the target ROI region extraction of the final high-resolution remote sensing image is shown in Figure 6.
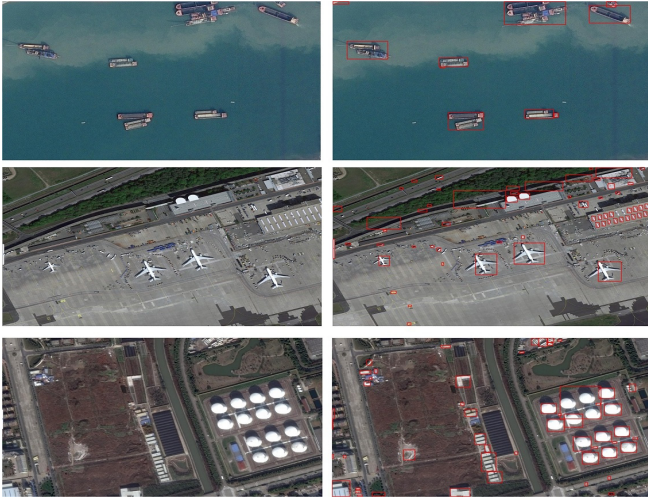
Figure 6. The schematic diagram of ROI region extraction results

3.2 The Description and Recognition of the Target

The major task after obtaining the ROI regions is to get a segmentation plane between the target and the negative sample using the classifier through selecting a proper feature descriptor for the ROI region and target. In this paper, high-resolution remote sensing targets are mainly small and multi-package targets such as ships, oil tanks, and airplanes. The difficulty in identification is that aircrafts, ships bear different types and scales, while the huge difference in imaging angle and target brightness of the same target in the dataset also contributes to the difficulty of recognition.

However, the high-resolution targets are more closed than the surrounding background, and its shape features are prominent. Therefore, when designing the target feature-description vector, it is significant to highlight the contour and shape features of the target, meanwhile, the extracted features should have certain scale and rotation invariance.

In our work, target shape feature (aspect ratio, area size, proportion in the image) are used, Hu invariant moment [19] and hog gradient histogram are used in the recognition of targets.

Contour is a very important feature information in the recognition of high-resolution targets, since the geometry feature of those targets is obvious. While ships, airplanes and other targets bear different docking direction, color and size, the overall inner-class structure of each target is similar. Therefore, when designing features, the extracted shape features must have rotation invariance, translation invariance, and scale invariance. The moment of the image contains the features of the image in terms of size, position, direction and contour, which can be used to describe the global shape feature of an image.

In mathematics, if there is a discrete random variable X, a constant c, a positive integer k, then $E\left(\left|X-c\right|\right)^{k}$ will be defined as the k-order moment of X in relation of c. For the case that $c=0$, defining $E\left(\left|X\right|\right)^{k}$ as the k-order origin moment of the variable X. While $c=E\left(\left|X\right|\right)^{k}$, $E\left(\left|X-c\right|\right)^{k}$ is called the k-order center moment of the variable X. For a discrete grayscale image, it can be regarded as a plane with uneven thickness. The grayscale value of the point (x,y) in the image can represent the density value of the plane at this position, and the whole image can be interpreted as a two-dimensional probability density function. Therefore, the moment can be used to describe the features of the image.

For a discrete grayscale image sizing $M \times N$, set $f(x,y)$ as the pixel value at the point (x,y). Then for the grayscale image, its $(p+q)$-order origin moment can be expressed as the form of (9):

$$m_{pq} = \sum_{x=1}^{M}\sum_{y=1}^{N} x^{p} y^{q} f(x,y) \tag{9}$$

The central moment is defined as the form of (10):

$$\mu_{pq} = \sum_{x=1}^{M}\sum_{y=1}^{N} (x-x_{c})^{p}(y-y_{c})^{q} f(x,y) \tag{10}$$

When setting $p,q=0$ in (10), we can obtain the 0-order origin moment and 1-order origin moment of the image. (11) is the form of the 0-order origin moment while (12) represents the 1-order origin moment.

$$m_{00} = \sum_{x=1}^{M}\sum_{y=1}^{N} f(x,y) \tag{11}$$

$$\begin{cases} m_{01} = \displaystyle\sum_{x=1}^{M}\sum_{y=1}^{N} y f(x,y) \\ m_{10} = \displaystyle\sum_{x=1}^{M}\sum_{y=1}^{N} x f(x,y) \end{cases} \tag{12}$$

Then the 0-order origin moment can represent the sum of all grayscale values of the image. Combining the 1-order moment of the image with the 0-order moment, we can calculate the centroid of the image. The centroid coordinates of the image are:

$$\left(x_{c} = \frac{m_{10}}{m_{00}}, y_{c} = \frac{m_{01}}{m_{00}} \right) \tag{13}$$

Each image has three 2-order moments, m_{11}, m_{02} and m_{20}, which can describe the main axis of the target image. There are $2^n - 1$ n-order moments of the image, and it is more complicated to analyze the higher-order moments. The moment of the image can greatly describe the basic geometric features of the image, and to some extent resist the change of rotational translation and scaling, so it is also called the invariant moment of the image. Experiments has indicated that simply utilizing the origin moment or central moment of the image is insufficient to have rotation invariance, translation invariance and scale invariance. Hu proposed the Hu moment in 1962 and gave the definition of continuous moment and the basic properties of the moment. In addition, Hu proved the translation invariance, the rotation invariance and the proportional invariance of the moment. The Hu moment is mainly constructed based on the 2-order central moment and the 3-order central moment of the image. In order to eliminate the influence of rotation and scale changes on the image, define the normalized central moment as:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}} \qquad (14)$$

Where $\gamma = \frac{p+q}{2}$, $p+q = 2, 3, \cdots$. Using the 2-order and 3-order normalized moments, Hu proposes the following invariant moment groups of seven invariant moments in his paper, as shown in (15). They form the feature vectors of an input image, i.e. the Hu moment. The Hu moment can resist the change of the features brought by the rotation translation and scaling to a certain degree. However, since the Hu moment only adopts low-order moments, the detailed features of the images is relatively lacking. The Hu moment is suitable for describing the basic shape of an image. In addition to the Hu moment, our paper also applies the HOG feature to describe the targets, which will be described in detail later.

The Histogram of Oriented Gradient (HOG) is a feature extraction means commonly applied in image processing. In computer vision, local features of an image can often be described by image gradients or directional density distributions of edges. The HOG feature is built on statistical information of gradients of the image, and the HOG feature reflects local feature information such as edges and gradients of the image.

$$
\begin{cases}
hu[0] = \eta_{20} + \eta_{02} \\
hu[1] = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\
hu[2] = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\
hu[3] = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \\
hu[4] = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12}) \times \\
\quad ((\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2) + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) \\
\quad \times (3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2) \\
hu[5] = (\eta_{20} - \eta_{02})((\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2) \\
\quad + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\
hu[6] = 3(\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) \\
\quad (3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2) \\
\quad - (3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2) \times (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})
\end{cases} \qquad (15)
$$

The main idea of extracting the HOG feature is to first divide the image into small connected blocks, called cell units, shown in Figure 7, and then calculate and merge the gradient histograms of the small blocks. The specific implementation manner lies in the following steps. First, collect the gradient value and direction of all the pixels in the input image, using the calculation method is of convolving the input image with $[-1, 0, 1]^T$ and $[1, 0, -1]^T$ after obtaining the gradient values I_x and I_y of the input image in the x, y directions. The magnitude of the gradient $M(x, y)$ and direction of the gradient $\theta(x, y)$ of the image at point (x, y) can be calculated from (16) and (17).

Figure 7. Cell units

$$M(x,y) = \sqrt[2]{I_x^2 + I_y^2} \qquad (16)$$

$$\theta(x,y) = \tan^{-1}\frac{I_y}{I_x} \in \left[0,360^o\right) or \left[0,180^o\right) \qquad (17)$$

The image is then referred to as several cell units in the manner presented in Figure 8. The gradient histogram of each cell unit is calculated. For each cell unit, a statistical result can be calculated to form a feature vector. After that, in order to guarantee these histogram features of better illumination invariance and geometric invariance, each four cell units is combined to form a larger block to make sure the contrast is normalized. For each block, the feature description vectors of all the cell units are stitched together to form the feature vector corresponding to it, and meanwhile the feature vectors of all blocks make up the gradient histogram description vector of the entire image, which is the final output of the HOG description vector for the original image. A schematic diagram of the HOG feature is shown in Figure 8. The description of the local features by the HOG feature combined with the description of the overall shape and contour features of the entire image by the Hu invariant moment conclude the entire feature description vector, expressed as $Vector_{H-HOG} = [Hu - in\,var\,iant\,moment,$ $HOG - description\,vector]$. The experiments show that this composite feature has excellent stability to remote sensing targets under high resolution, as well as considerable translation invariance and rotation invariance. In the training process of the high-resolution remote sensing targets in our work, the ratio of positive and negative samples is set to 1:3. After the description feature $Vector_{H-HOG}$ is obtained, SVM is utilized to be trained on the dataset to achieve the recognition of the ROI region and the target confirmation.
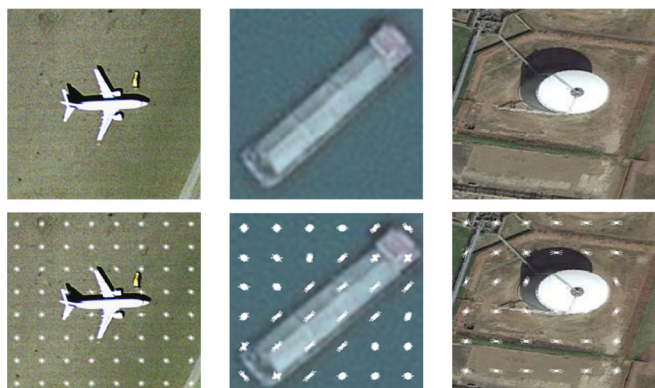
The training process is the process of finding the best segmentation hyper-plane for positive and negative samples. SVM performs excellent in the learning of small samples. Through introducing the kernel function and the relaxation factor, to some extent, the linear inseparability of the sample can be solved. The relaxation factor is a tolerance for the degree of relaxation of outlier sample points. In this paper, we use the linear kernel function to find the optimal segmentation hyper-plane in the original feature space of the sample. Set the relaxation factor to 0.01, and determine the iteration termination condition to 1000 or when the error is less than a fixed value. Then train the positive and negative samples to obtain a classification model. During the test process, extract the $Vector_{H-HOG}$ of each untested ROI region and send it to the trained SVM classifier for prediction on whether it belongs to the positive sample or not. If so, map back to the original image to calculate the corresponding position of the target on the original image. Otherwise, abandon the ROI region and output the result of the recognition of the original image.

Based on the above process, the schematic diagram of the recognition result of the high-resolution remote sensing images in this paper is shown from Figure 9 to Figure 14. For the targets of oil tanks, airplanes, ships, etc., in our work, take 20% of the source dataset obtained from the cooperative as a testset and conduct test on it, the target recognition accuracy of the test results meets the project indicators that it exceeded 85%.

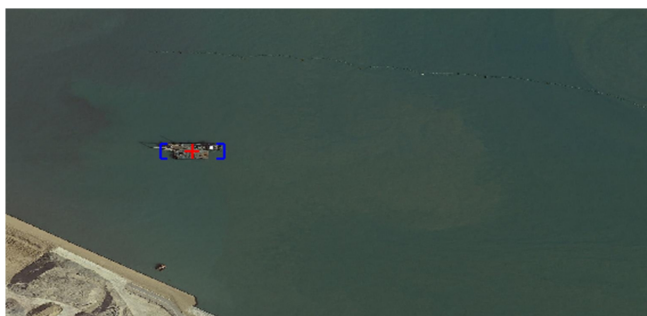**Figure 9.** The recognition result of ship target-1

Figure 8. Schematic diagram of HOG features of high-resolution samples

Figure 10. The recognition result of ship target-2

Figure 11. The recognition result of oil tank target-1

Figure 12. The recognition result of oil tank target-2

Figure 13. The recognition result of plane target-1

Figure 14. The recognition result of plane target-2

4 Conclusion

Aiming at the issue of remote sensing target recognition under high resolution, we proposed a method based on visual saliency mechanism and ROI region extraction, machine learning algorithms are used to perform segmentation and classifying. We

achieved a high recognition accuracy of over 85% on our dataset consisting of the targets of planes, oil tanks and ships. More importantly, our method can be used for recognition of other similar remote sensing targets of high resolution.

Acknowledgment

References

[1] H. Long, Y. Chung, Z. Liu, S. Bu, Object Detection in Aerial Images Using Feature Fusion Deep Networks, *IEEE Access*, Vol. 7, No. 6, pp. 30980-30990, March, 2019.

[2] C. Zhu, H. Zhou, R. Wang, J. Guo, A Novel Hierarchical Method of Ship Detection from Spaceborne Optical Image Based on Shape and Texture Features, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 48, No. 9, pp. 3446-3456, September, 2010.

[3] F. Yang, Q. Xu, F. Gao, L. Hu, Ship Detection from Optical Satellite Images Based on Visual Search Mechanism, *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Milan, Italy, 2015, pp. 3679-3682.

[4] D. Comaniciu, P. Meer, Mean Shift: A Robust Approach toward Feature Space Analysis, *IEEE Transactions on Pattern Analysis & Machine Intelligence*, Vol. 24, No. 5, pp. 603-619, May, 2002.

[5] L. Chun, X. Chunhua, Y. Jian, X. Yingying, B. Junliang, A Method for Coastal Oil Tank Detection in Polarimetrie Sar Images Based on Recognition of t-shaped Harbor, *Journal of Systems Engineering and Electronics*, Vol. 29, No. 3, pp. 499-509, June, 2018

[6] T. Shuai, K. Sun, B. Shi, J. Chen, A Ship Target Automatic Recognition Method for Sub-meter Remote Sensing Images, *2016 4th International Workshop on Earth Observation and Remote Sensing Applications (EORSA)*, Guangzhou, China, 2016, pp. 153-156.

[7] B. Huang, T. Xu, Y. Luo, S. Chen, B. Liu, B. Yuan, A Novel Nearest Feature Learning Classifier for Ship Target Detection in Optical Remote Sensing Images, *International Conference in Communications, Signal Processing, and Systems*, Harbin, China, 2017, pp. 600-606.

[8] M. Safdari, P. Moallem, M. Satari, Sift Detector Boosted by Adaptive Contrast Threshold to Improve Matching Robustness of Remote Sensing Panchromatic Images, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 12, No. 2, pp. 675-684, February, 2019.

[9] W. Wang, J. Chen, J. Li, X. Wang, Remote Targets Recognition Based on Adaptive Weighting Feature Dictionaries and Joint Sparse Representations, *Journal of the Indian Society of Remote Sensing*, Vol. 46, No. 11, pp. 1863-1870, November, 2018.

[10] J.-S. Pan, L. Kong, T.-W. Sung, P.-W. Tsai, V. Snasel, alpha-Fraction First Strategy for Hierarchical Wireless Sensor

Networks, *Journal of Internet Technology*, Vol. 19, No. 6, pp. 1717-1726, June, 2018

[11] J.-S. Pan, C.-Y. Lee, A. Sghaier, M. Zeghid, J. Xie, Novel Systolization of Subquadratic Space Complexity Multipliers Based on Toeplitz Matrix-Vector Product Approach, *IEEE Transactions on Very Large Scale Integration Systems*, Vol. 27, No. 7, pp. 1614-1622, March, 2019.

[12] G. Hu, Z. Yang, J. Han, L. Huang, J. Gong, N. Xiong, Aircraft Detection in Remote Sensing Images Based on Saliency and convolution neural network, *EURASIP Journal on Wireless Communications and Networking*, Vol. 26, pp. 1-16, February, 2018.

[13] T. Nguyen, J.-S. Pan, T.-K. Dao, An Improved Flower Pollination Algorithm for Optimizing Layouts of Nodes in Wireless Sensor Network, *IEEE Access*, Vol. 7, No. 120, pp. 75985-75998, June, 2019.

[14] Z. Meng, J.-S. Pan, K.-K. Tseng, PaDE: An Enhanced Differential Evolution Algorithm with Novel Control Parameter Adaptstion Schemes for Numerical Optimization, *Knowledge-Based Systems*, Vol. 168, No. 9, pp. 80-99, March, 2019

[15] X. Xia, H. Liang, Y. RongFeng, Y. Kun, Oil Tank Extraction in High-resolution Remote Sensing Images Based on Deep Learning, *2018 26th International Conference on Geoinformatics*, Kunming, China, 2018, pp. 1-6.

[16] J.-S. Pan, L. Kong, T-W. Sung, P.-W. Tsai, V. Snášel, A Clustering Scheme for Wireless Sensor Networks Based on Genetic Algorithm and Dominating Set, *Journal of Internet Technology*, Volume 19, No. 4, pp. 1111-1118, June, 2018.

[17] B. Liu, M. Xu, P. Fu, Graph-based Clustering with Spatiotemporal Contour Energy for Video Salient Object detection, *Information Hiding and Multimedia Signal Processing*, Vol. 10, No. 2, pp. 368-383, March, 2019.

[18] M. Xu, B. Liu, P. Fu, J. Li, Y. H. Hu, Video Saliency Detection via Graph Clustering with Motion Energy and Spatiotemporal Objectness, *IEEE Transactions on Multimedia (Early Access)*, pp. 1-1, May, 2019.

[19] M. K. Hu, Visual Pattern Recognition by Moment Invariants, *IRE Transactions on Information Theory*, Vol. 8, No. 2, pp. 179-187, February, 1962.

Biographies

Bing Liu received his B.S, M.S and Ph.D. degree in electrical engineering from Harbin Institute of Technology, Harbin, China. He is currently with Department of Test and Control Engineering, Harbin Institute of Technology, China. His current research interests include computer vision and high-performance computing.
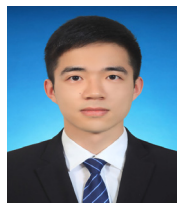
Tingwei Chen is currently pursuing his M.S degree at Harbin Institute of Technology, Harbin, China. His research area lies in Computer vision.

Ping Fu receieved the B.S. at University of Science and Technology of China, Hefei, China, the M.S and Ph.D. at Harbin Institute of Technology, Harbin, China. He is currently with Department of Test and Control Engineering, Harbin Institute of Technology, China. His research interest spans Digital Image Processing, Computer Vision.

Yumei Zhen is currently pursuing her M.S degree at Harbin Institute of Technology, Harbin, China. Her research area lies in Computer Vision.

Jeng-Shyang Pan received the B.S. degree in electronic engineering from the National Taiwan University of Science and Technology, Taipei, Taiwan, the M.S. degree in communication engineering from the National Chiao Tung University, Hsinchu, Taiwan, and the Ph.D. degree in electrical engineering from the University of Edinburgh, Edinburgh, U.K.. He is currently with the Department of College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao, China; with the Fujian Provincial Key Laboratory of Big Data Mining and Applications, Fujian University of Technology, Fuzhou, China; and with Chaoyang University of Science and Technology, Taichung, Taiwan. His current research interests include cryptographic computing, robot vision, and cloud computing.