Target Detection in Colorful Imaging Sonar Based on HOG

XiaoWang[1], JingWang[2], FanYang[3], GuangyiZeng[4]
1,2. School of Information Science and Engineering
Yunnan University
Kunming, China
3. CSIC 750 Test Range
Kunming, China
4. CSG PGC
Guangzhou, China

Abstract*—TKIS-I helmet-mounted colorful imaging sonar is mounted on the helmet of divers to serve as their underwater eyes. Currently, there are more than two dozens of it serving the navy of China. However, under the complex underwater environment, divers usually take great risks while performing underwater operations. The work of this paper aims to achieve automatic underwater target detection of the imaging sonar in order for divers not to dive into water. The paper combined Histogram of Oriented Gradient(HOG) in computer vision for feature extraction and support vector machine (SVM) for classification to achieve quick underwater target detection. The results showed high detection rate. All these work are the foundation for automatic underwater target detection and recognition in the future.

***Keywords*— Colorful Imaging Sonar; HOG; SVM; Target Detection**

I.INTRODUCTION

With more exploration and development of rivers, lakes and shallow water areas near the shorelines in the country, higher requirements have been placed on underwater operations such as underwater search, rescue and salvation in which underwater target detection plays an important role. The Helmet-mounted colorful imaging sonar of high-resolution, TKIS-I, can easily acquire the shape and surface details of underwater targets, and then help divers to have a clear understanding of the underwater targets to make judgments. Machine learning is one of the target detection methods. The feature descriptors in machine learning include haar-like[1], HOG[2], LBP[3], edgelet[4] and so on. According to the features of sonar imaging, this paper applied HOG feature extraction, principal component analysis, support vector machine and other methods together for TKIS-I sonar target detection.

II. EXPERIMENTAL BACKGROUND

TKIS-I imaging sonar is a type of small, high-resolution, and colorful imaging sonar that a diver carries alone to work underwater. It is mainly composed of four parts: a waist control box, an onshore industrial control computer, video glasses, and a sonar probe which are showed in Fig.1 from left to right. Fig.2 is the South China Sea Salvage Site. When a diver performs underwater operations, he wears the control box at his waist and has the sonar probe and video glasses mounted on his helmet in Fig.3. The diver observes and detects targets through the video glasses while the sonar probe scans the underwater environment. The whole process is very challenging and time-consuming, and so dangerous for the diver. Therefore, the work of this paper aims to detect targets automatically while the sonar probe is mounted somewhere underwater, instead of the diver's helmet.

Fig.1.TKIS-I helmet-mounted colorful imaging sonar

Fig.2.TIKS-I is used in the South China Sea rescue site

Fig.3.Divers with TKIS-I helmet-mounted colorful imaging sonar

The target chosen for this experiment was the wood stakes in the school's wood stake pool. There are five stakes in the pool at the experimental site in Fig.4.Because one of the five stakes is located beyond the maximum range of the sonar scan, only four are imaged in Fig.5. The shape of

978-1-5386-7946-3/18/$31.00 ©2018 IEEE

these wood takes in the sonar is very obvious. Their that originally appeared in the sonar image will rotate when the direction of the sonar scan is changed. The size of these wood stakes that originally appeared in the sonar image will scale when the maximum range of the sonar scan is changed. These changes are shown in Fig.6 and Fig.7. This paper uses different training samples in the rotation direction to achieve the rotation invariance. The paper also scales the target image and unifies it into the same size detection window to achieve its scale invariance. The color of the target displayed by the sonar image can be adjusted manually. HOG has good invariability to color features, so it is stable to the change of the target color.

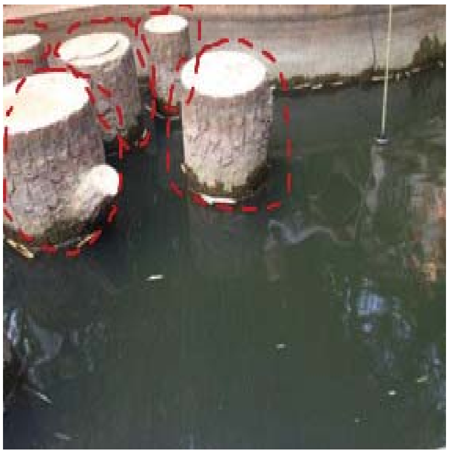

Fig.4. Experiment site of wood stakes

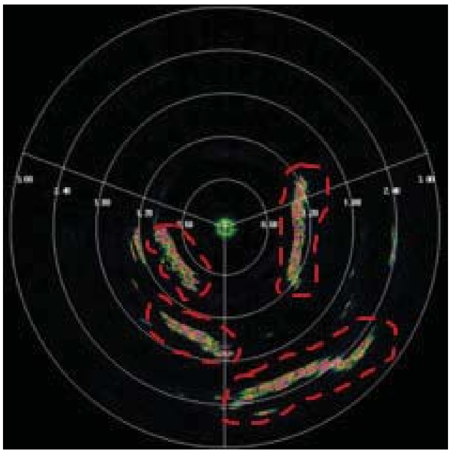

Fig.5.Original image of TKIS-I colorful imaging sonar for wood stakes

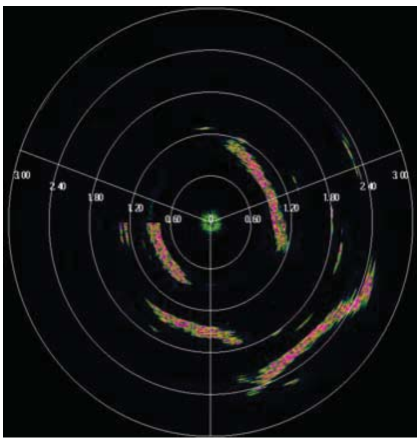

Fig.6.Wood stakes sonar imaging with varying rotation

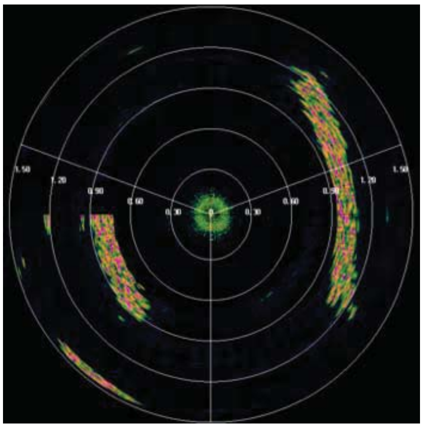

Fig.7.Wood stakes sonar imaging with scale extension

III. HOG FEATURE EXTRACTION

Put forward by Dalal and Triggs in an article published in 2005 by CVPR[1], HOG is mainly used as a pedestrian detection algorithm in computer vision. It is calculated on grid-sized uniform cell units. To improve performance, HOG applies overlapping local contrast normalizations. The specific steps for extracting HOG features are as follows:

(1) Standardize Gamma space and color space. Square root Gamma standardization can well eliminate the effects of overall image illumination and contrast.

(2) Calculate the pixel gradient. In the experiment, the [1,0,-1] template works well. First, the original image is convolved with the [-1,0,1] gradient operator to obtain the horizontal gradient component (right in the positive direction), and then the [1,0,-1] gradient operator performs convolution on the original image to obtain the vertical gradient component (upward in the positive direction). $G_x(x,y)$, $G_y(x,y)$ respectively represents the pixel horizontal gradient and vertical gradient of the input image at the pixel point. $G(x,y)$, $\alpha(x,y)$ respectively representsthe pixel gradient gradient and gradient direction.

$$G_x(x,y) = I(x+1,y)\text{-}I(x\text{-}1,y) \tag{1}$$

$$G_y(x,y) = I(x,y+1)\text{-}I(x,y\text{-}1) \tag{2}$$

$$G(x,y) = \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \tag{3}$$

$$\alpha(x,y) = tan^{\text{-}1}\left(\frac{G_y(x,y)}{G_x(x,y)}\right) \tag{4}$$

(3) Statistics the direction gradient histogram in the cell. Divide the image evenly into several square cells. Each cell contains8×8 pixels. Each cell also divides the gradient direction from 0 to 180 degrees (or 0 to 360 degrees, taking into account the positive and negative values) into 9 direction block, each pixel in the cell is weighted and projected in the histogram with a gradient direction, and is mapped into a corresponding angle range block, then a gradient direction histogram of the cell is obtained, which is

the cell corresponding to the 9-dimensional feature vector.

(4) Normalize the histogram in the block. 2 × 2 cells form a block, such a block forms a 36-dimensional feature vector, and then the entire block is normalized using L2-Hys to obtain the final feature vector.

(5) Reduce the dimension of the feature vector. Because of the high dimension of the HOG feature vector, there is a large amount of redundant information which will reduce the accuracy of detection and the classification speed. Therefore, we need to reduce the dimension of the feature vector. In this paper, the PCA (principal component analysis) is used to reduce the dimension of the feature vector. By transforming the original sample features, the original coordinates are projected to a new, low-dimensional and mutually orthogonal space.

IV. ACTUAL SIGNAL PROCESSING

4.1 Create Training and Testing Data Sets

From the above this paper selects the wooden pile in the pond as the object of the test. The training sample used in this paper is the data collected by the laboratory TKIS-I helmet-mounted colorful imaging sonar. The sonar equipment is used to obtain the original picture of the wooden stakes in the sonar image, and then the self-written mark software is used for labeling. These pictures are positive samples. And those pictures that do not contain wooden stakes are used as negative samples. There are 700 positive sample pictures and 8000 negative sample pictures in this data. There are 300 positive sample images and 600 negative sample images in the test set. The sample data set is relatively clear, and the direction and size of the stakes are different. Some positive and negative samples are shown in the figures:

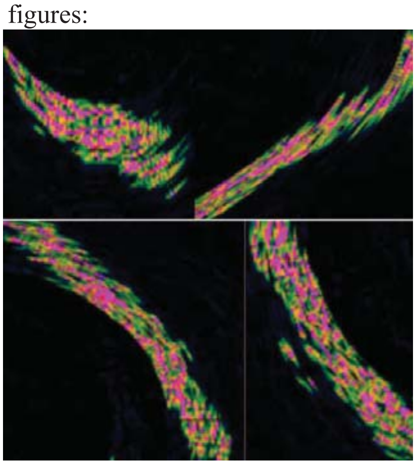

Fig.8. Some positive samples

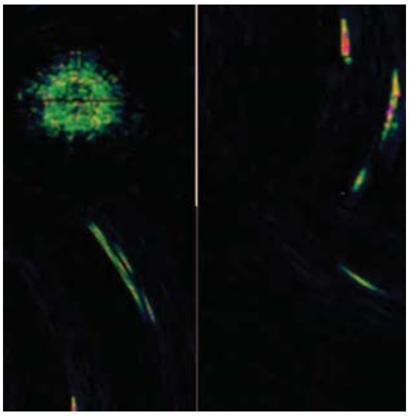

Fig.9. Some negative samples

4.2 Implement HOG Algorithm

The compiled language used in this article is Python. The development environment used is Spyder. Compared with other Python development environments, its greatest advantage is to imitate MATLAB's "working space." The function can easily observe and modify the value of the array.

The experiment first transforms our pending color image into a grayscale image because HOG feature extraction is based on grayscale images. Then, the target images marked in the sonar image are extracted and unified to a size of 128×128. In the experiment, 16×16 is segmented as the size of the block, and each block is normalized. Each block contains 2×2 cells, so each cell occupies 8×8 pixels, each block has a slide increment of 8×8, so each image contains 225 blocks. Each block contains 4 cells, each with a dimension of 9, so we have a HOG feature vector dimension of 8100 for each unified 128×128 window image. According to the PCA dimension reduction method mentioned above, we have tested that the PCA has the highest detection rate in the 1100dimension.

4.3 Optimize SVM for Feature Classification

This paper uses the SVM classifier to train the above features and chooses the radial basis function (RBF) as the kernel function of the SVM. Because the RBF kernel function can map the input data into the high-dimensional space to select the optimal classification surface. This achieves a non-linear mapping and the classification works well. Then we will carry out the initial SVM model trained Hard Negative Mining. The idea of Hard Negative Mining is to train the classifier first with positive and negative samples. Then use the trained classifier to classify the samples, put the hard negative samples into the negative sample set, and continue training the classifier. Repeatedly until the stop condition is reached (the classifier performance is no longer improved), so the method above the choice of positive and negative samples, the number of negative samples is much larger than the number of positive samples. So this paper selected 8000 negative 128×128 images. According to the above results, the optimized classification model was used as our final target detection model. Finally, the target detection model was used to test in the test set.

4.4 Result Analysis

According to the previous discussion of this paper, it can be seen that the experiment uses the PCA dimensionality reduction method. Then how much is the dimension selection of HOG-PCA appropriate? Therefore, this paper uses the sonar image data set labeled by the laboratory to compare multiple principal component values to determine the optimal detection rate. The following results were obtained:

TABLE I. HOG-PCA DIMENSION

HOG-PCA dimension	Detection rate (%)
200	87.75
300	88.67
400	89.23
500	90.62
700	91.12
900	91.68
1100	92.47
1300	91.88
1500	91.79
1700	91.28
1900	91.26

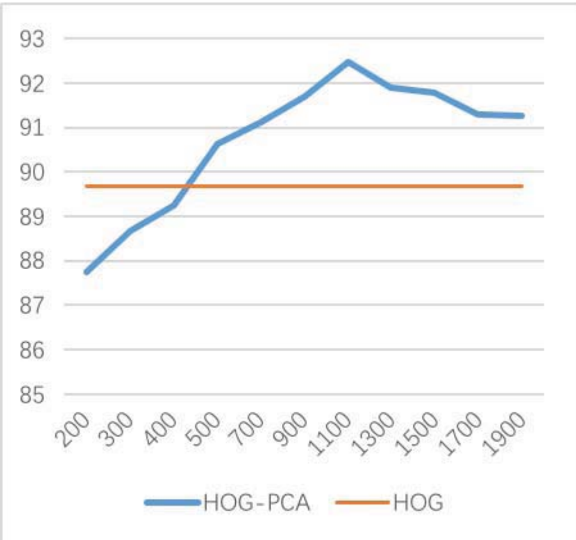

Fig.10. Relationship between detection rate in each dimension of HOG-PCA

Fig.10 shows that when the dimension of the principal component is1100, the detection rate is the highest. With the increase of the later dimension, the detection rate basically does not change much. In addition, the HOG feature after PCA dimensionality reduction is higher in the appropriate dimension than the HOG without dimensionality reduction. Therefore, the dimension selected by HOG-PCA is1100. Fig.11 shows the picture of the target detection in the sonar. The position of the wood stakes in the imaging sonar can be basically detected.

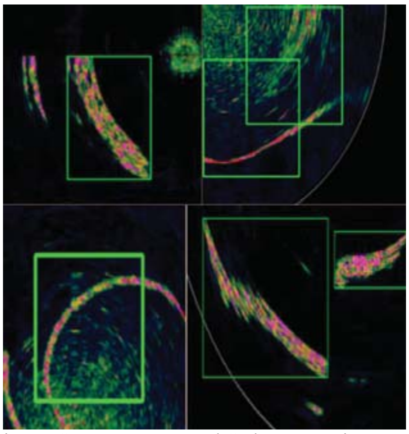

Fig.11. Sonar Image Wood Stakes Detection Image

V. CONCLUSIONS AND FUTURE DIRECTIONS

This paper demonstrated that HOG feature extraction techniques significantly improved the target detection rate of underwater sonar images. The target detection lays a foundation for the automatic target detection of underwater robots in the future. However, the accuracy of the detection will be affected when there are many disturbances or noises around the target, making the shape of the target unclear. This is also the direction we need to improve and explore.

ACKNOWLEDGMENT

1.XiaoWang: M.s, mainly engaged in machine learning, computer vision, sonar signal processing, E-mail: 1391458077@qq.com.

2.JingWang: Ph.D , Corresponding author, mainly engaged in underwater acoustic signal processing, imaging sonar development, E-mail: wj8538@163.com.

3.FanYang: B.s, mainly engaged in underwater robot , high frequency multibeam sonar, E-mail: 104623068@qq.com.

4.GuangyiZeng: B.s, mainly engaged in electrical automation and high voltage engineering E-mail: zenggy@pgc.csg.cn.

REFERENCES

[1] Viola P, Jones M. Robust Real-time Face Detection[J]. International Journal of Computer Vision, 2004, 57(2):137-154.

[2] Dalal, Navneet, Triggs, et al. Histograms of Oriented Gradients for Human Detection[C]// Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. IEEE, 2005:886-893.

[3] Ojala T, Pietikainen M, Harwood D. A comparative study of texture measures with classification based on featured distributions[J]. Pattern recognition, 1996, 29(1):51-59.

[4] Wu B, Nevatia R. Detection of Multiple, Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectors[C]// Tenth IEEE International Conference on Computer Vision. IEEE Computer Society, 2005:90-97.

[5] Wu B, Nevatia R. Detection of Multiple, Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectors[C]// Tenth IEEE International Conference on Computer Vision. IEEE Computer Society, 2005:90-97.

[6] P.Blondel, The Handbook of Sidescan Sonar. Springer, 2009.

[7] Bengio Y, Lecun Y. Scaling learning algorithms towards AI[C]// Large-Scale Kernel Machines. 2007:321-359.

[8] Murphy K P. Machine Learning: A Probabilistic Perspective[M]. MIT Press, 2012.

[9] Bengio Y. Learning Deep Architectures for AI[J]. Foundations & Trends® in Machine Learning, 2009, 2(1):1-127.

[10] Y. Anzai, Pattern recognition and machine learning. Elsevier, 2012.