Industry Science and Engineering Vol. 1 No. 3, 2024

Academic Education Publishing House

Research on Underwater Distance Measurement of Binocular Vision by Semantic Separation Picture

Sijia Ren *

Leicester International Institute, Dalian University of Technology, Panjin, Liaoning, China

**Corresponding Author.*

Abstract: Aiming at the problem of large amount of calculation and slow speed of stereo matching in binocular vision ranging, this paper proposes an underwater distance measurement of binocular vision based on underwater target detection. By optimizing the SGBM algorithm, the disparity search range is limited in the underwater target recognition area, and an adaptive aggregation window is constructed to adapt to different texture changes. This method improves the matching accuracy and effectively reduces the computational burden of unrelated regions. At the same time, a texture similarity dynamic adjustment-matching window is established to flexibly respond to texture changes in different regions, thereby greatly improving the matching speed and accuracy. Experiments show that the ranging speed is increased by 26.8 %. This study can not only improve the efficiency and accuracy of underwater operations, but also lay a solid foundation for subsequent technical research and application promotion.

Keywords: Semantic Segmentation; Binocular Vision; Ranging; Underwater Imaging; Region Matching

1. Introduction

There are four main methods for underwater target detection and tracking, acoustic detection, optical detection, radiation detection and magnetic detector detection. Optical detection can visually observe the underwater situation, which is conducive to large-scale promotion [1,2]. However, the optical properties of water have certain limitations on the detection, resulting in short detection distance, low Picturequality, and obvious color difference and distortion. Xu et al. [3] developed an optical vision underwater pipeline automatic inspection system to achieve accurate detection of local edge changes. Chen and Nakajima [4] proposed an underwater pipeline detection method based on Hough transform. Zeng et al. [5] proposed an improved Hough transform method to improve the extraction effect of pipeline boundary. Li [6] proposed a line segment detection algorithm and improved the region growing algorithm. Zingaretti et al. [7] proposed a real-time underwater imaging system. Tang [8] constructed a shape-based dynamic window model based on the shape characteristics of underwater pipelines, which improved the accuracy and real-time performance of the system. Balasuriya et al. [9] proposed to use CCD cameras and acoustic sensors to obtain three-dimensional information from underwater pipelines.

At present, the binocular ranging technology for underwater targets is still blank. The target ranging of binocular vision simulates the visual system of human eyes, and infers the distance of the target object by calculating the parallax between the two cameras [10, 11]. The binocular ranging process mainly includes binocular camera calibration, binocular correction and stereo matching.

2. Binocular Vision Ranging Principle

The binocular camera consists of two cameras. Due to the baseline distance between the two cameras, when the same object appears in both picturesat the same time, the position of the object in the two pictureswill have a certain offset. The schematic diagram shown as Figure 1. P is the prediction target, O_l is the optical center of the left camera, O_r is the optical center of the right camera, f is the focal length of the camera, P_l is the coordinates of the Picturecoordinate system of P point on the left camera, and P_r is the coordinates of the Picturecoordinate system of P point on the right camera.

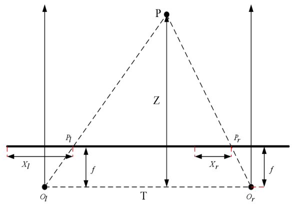

Figure 1. Principle of Binocular Visual Ranging

As shown in Figure 1, the formula is as follows:

$$\frac{T-(X_l-X_r)}{T}=\frac{Z-f}{Z} \quad (1)$$

The distance Z from the target to the camera can be obtained as follows:

$$Z=\frac{fT}{X_l-X_r}=\frac{fT}{d} \quad (2)$$

Among them, $d=X_l-X_r$, is the parallax of the two cameras, X_l and X_r are the abscissa of the pixels of the P point in the left and right picturesrespectively, f is the focal length, and T is the baseline.

3. Underwater Picture Analysis

The process of the light reflected by the underwater object reaching the camera requires two refractions. One is the refraction of water and lens, and the other is the refraction of lens and air. It is assumed that the optical axis of the camera is perpendicular to the horizontal plane of refraction, and the refraction of lens and air is neglected. The simplified model of underwater imaging is shown in Figure 2, the incident angle as θ_1, the reflection angle as θ_2, target point as P, actual position of the cameraas O_1, and O_2 is the intersection of the incident light in the water and the extension line of the camera. n_1 is the refractive index of water, n_2 is the refractive index of air. According to the law of refraction:

$$\frac{\sin\theta_1}{\sin\theta_2}=\frac{n_2}{n_1} \quad (3)$$

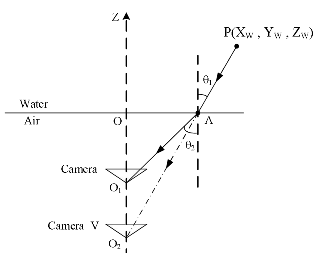

Figure 2. Underwater Imaging Model

X_1 is the imaging length of OA in the real time camera, X_2 is the imaging length of OA in the virtual camera, and $X_1=X_2$, the following results can be obtained:

$$\frac{X_1}{OA}=\frac{f_1}{OO_1} \quad (4)$$

$$\frac{X_2}{OA}=\frac{f_2}{OO_2} \quad (5)$$

$$\frac{f_1}{f_2}=\frac{OO_1}{OO_2}=\frac{\tan\theta_1}{\tan\theta_2}\approx\frac{\theta_1}{\theta_2}\approx\frac{\sin\theta_1}{\sin\theta_2}=\frac{n_2}{n_1} \quad (6)$$

From the above analysis, it can be seen that the camera is affected by refraction under the water, and its focal length has changed. The focal length of the underwater camera is about 1.33 times of the actual focal length.

4. Coordinate Transformation

Due to the error of the camera during installation, the main point is generally not at the center point, and it is necessary to introduce offsets c_x and c_y,

that is:

$$x_{\text{screen}}=f_x\left(\frac{X}{Z}\right)+c_x,\quad y_{\text{screen}}=f_y\left(\frac{Y}{Z}\right)+c_y \quad (7)$$

The matrix form is as follows:

$$Z_c\begin{bmatrix}u\\v\\1\end{bmatrix}=\begin{bmatrix}f_x&0&c_x&0\\0&f_y&c_y&0\\0&0&1&0\end{bmatrix}\begin{bmatrix}X_c\\Y_c\\Z_c\\1\end{bmatrix} \quad (8)$$

$$K=\begin{bmatrix}f_x&0&c_x\\0&f_y&c_y\\0&0&1\end{bmatrix} \quad (9)$$

Among them, the pixel coordinates of the principal point Pictureare (c_x, c_y), u and v are the pixel coordinates on the u axis (x axis) and the v axis (y axis), respectively. f is the actual physical focal length of the camera, f_x and f_y are the pixel focal length of the camera, and K is the internal parameter of the camera.

5. SGBM Optimization

The entire underwater target expressed as:

$$T_{\text{pile-leg}}=N\times t_m \quad (10)$$

The matching time expressed as:

$$T_{total}=W\times t_m \quad (11)$$

The ratio of semantic segmentation picture matching time to original Picture matching time

$$P\%=\frac{T_{pile-leg}}{T_{total}}=\frac{N\times t_m}{W\times t_m}=\frac{N}{W}\times 100\% \quad (12)$$

Therefore, using semantic segmentation picturesfor binocular vision ranging, the matching time can be reduced greatly.

5.1 Underwater Target Region Matching (ROI)

The energy function is shown in Eq. (13) to minimize this energy function. That is, the optimal disparity is obtained.

$$E(D) = E_{\text{data}}(D) + \lambda E_{\text{smooth}}(D) \quad (13)$$

The energy function has the following expression:

$$\begin{aligned} E_{ROI}(D) &= \sum_{P \in ROI} C\left(P, D_p\right) \\ &+ \sum_{q \in N_p \cap ROI} P_1 I\left[\left|D_p - D_q\right| = 1\right] \\ &+ \sum_{q \in N_p \cap ROI} P_2 I\left[\left|D_p - D_q\right| > 1\right] \end{aligned} \quad (14)$$

Among them, D is the disparity map, $E_{data}(D)$ is the sum of the matching cost, $E_{smooth}(D)$ is the smoothing constraint term as the energy function, p and q represent the pixels, P_1 and P_2 are penalty coefficient, C is the data constraint term, λ is a weight parameter, which is used to balance the importance of the data term and the smoothing term. ROI represents the set of pixels in the underwater target area, $N_p \cap ROI$ represents the set of pixels in the neighborhood of Picturepoint p that are also located in the ROI area.

5.2 Adaptive Aggregation Window

Suppose that there is a central pixel i, and the set of surrounding pixels is denoted by N_i, which includes the central pixel and its surrounding pixels. For each pixel point $j \in N_i$, the texture similarity between the center pixel point i and the pixel point j can be calculated, which is recorded as S_{ij}.

$$S_{ij} = \sum_{p \in w} (I_i(p) - I_j(p))^2 \quad (15)$$

Among them, $I_i(p)$ and $I_j(p)$ represent the gray values of pixel i and pixel j at position p, respectively, and represents the matching window.

According to the texture similarity S_{ij}, the adaptive window size W_i of each pixel can be calculated. The following formula is usually used:

$$W_i = \alpha . \sqrt{\frac{1}{|N_i|} \sum_{j \in N_i} S_{ij}} \quad (16)$$

Here, α is a parameter that controls the degree of adjustment of the window size.

6. Binocular Ranging Experiment

6.1 Experimental Environment and Process

The experiment uses a binocular camera underwater shown as Figure 3. The software environment is Ubuntu, Matlab 2021a, OpenCV, python3.8.

The process: Firstly, calibrated the left and right cameras by Matlab toolbox to obtain the internal and external parameters of the binocular camera. Then, the underwater target picture captured by the binocular camera is semantically segmented to obtain the generated picture without the background. Finally, the semantic segmentation generated picture is binocularly corrected to measure the distance of the selected point of the underwater target.

Figure 3. Binocular Camera

6.2 Calibrate and Correct to Binocular Camera

By constantly moving the calibration plate, 20 groups of checkerboard pictures in different angles are taken form the air and water respectively, as shown in Figure 4.

The calibration parameters as shown in Table 1 and Table 2.

(a) Calibration in Air

(b) Underwater Calibration

Figure 4. Calibration Image

Table 1. Calibration Parameters in Air

parameter	left camera			right camera		
Intrinsic-matrix	634.1184	1.1761	401.3214	635.3728	1.0653	388.3534
	0.0	631.9297	305.6609	0.0	633.1938	299.6241
	0.0	0.0	1.0	0.0	0.0	1.0
Radial-distortion	-0.0734	0.4423	-0.6823	-0.0686	0.3454	-0.4197
Tangential-distortion	0.0003	0.0002		-0.0007	-0.0007	
Rotation-matrix	1.0		0.0015		-0.0036	
	-0.0015		0.9998		0.0111	
	0.0037		-0.0111		0.W 9989	
Translation-vector	634.1084		1.1661		401.3114	

Table 2. Calibration Parameters in Water

parameter	left camera			right camera		
Intrinsic-matrix	853.8806	1.1722	401.5821	864.1113	1.0976	388.9734
	0.0	851.2006	304.2702	0	867.1203	299.8742
	0.0	0.0	1.0	0.0	0.0	1.0
Radial-distortion	-0.1331	0.6660	-0.3658	-0.1372	0.6731	-0.3592
Tangential-distortion	0.0003	0.0015		0.007	-0.0012	
Rotation-matrix	0.9998		2.6812		-0.0074	
	-2.8947		0.9998		3.2319	
	0.0059		-0.0113		0.9998	
Translation-vector	-58.6332		-1.1854		1.6938	

6.3 Experimental Results and Analysis

The above pictures are semantically segmented to generate a hybrid picture of underwater target and background, and a generated picture after underwater target recognition. So, the distance data is obtained by using the binocular ranging formula. The actual distance, measured value, absolute error and relative error of the binocular ranging distance of 0.27 m and 0.59 m based on the semantic segmentation Pictureare converted into broken lines, as shown in Figure 5.

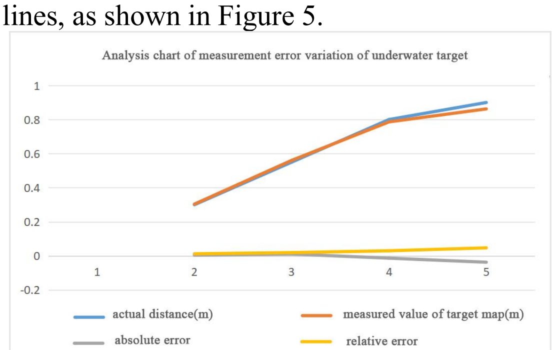

Figure 5. Analysis Diagram of Underwater Target Mapping Error Change

7. Conclusion

In this paper, the DeepLabV3 + semantic segmentation model is used as the core algorithm of underwater target recognition, and the model is optimized and improved. The method of binocular vision ranging by semantic segmentation pictureis is proposed, which realizes the accurate measurement of underwater target distance and improves the measurement speed. This method effectively improves the matching accuracy of the stereo matching algorithm in the area with large texture changes and reduces the amount of calculation by limiting the disparity search range in the underwater target area and dynamically adjusting the matching window method with texture similarity. This application not only expands the application scope of semantic segmentation technology, but also provides a new solution to the ranging problem in the field of marine engineering.

References

[1] Ortiz A Simo M Oliver G. A vision system for an underwater cable tracker. Machine Vision and Applications, 2002 13(3):129-140.
[2] Li Zhen. Design and research of automatic tracking ROV for underwater pipeline. Dalian University of Technology, 2018, 4.
[3] Xu Peng-fei, Hu Zhen, Cui Wei-cheng, et al. Application of Vision-based System in Underwater Pipeline Active Inspection. ShipBuilding of China, 2010, 51(03):142-151.
[4] Chen C, Nakajima M. A Study on

Underwater Cable Automatic Recognition Using Hough Transformation. Proceedings of IAPR Workshop on Machine Vision Applications, Kawasaki, Japan.1994.
[5] Zeng Wen-jing, Xu Yu-ru, Wan Lei, et al. g. Robotics Vision-Based System of Autonomous Underwater Vehicle for an Underwater Pipeline Tracker. Journal of S hanghai Jiaotong University, 2012, 46(02):178-183.
[6] Li Shuangshuang. Research on AUV underwater pipeline detection method based on gradient information. Harbin Engineering University, 2016, 3.
[7] Zingaretti P, Tascini G, Puliti P, et al. Imaging approach to real-time tracking of submarine pipeline. Proceedings of SPIE - The International Society for Optical Engineering, 1996, 2661:129-137.
[8] Tang Xudong. Research on underwater pipeline detection and tracking technology of intelligent underwater vehicle. Harbin Engineering University, 2010, 12.
[9] Balasuriya B, Takai M, Lam W, et al. Vision based autonomous underwater vehicle navigation: underwater cable tracking. OCEANS'97 MTS/IEEE Conference Proceedings, Halifax Cannada: Marine Technology Society 1997:1418-1424.
[10] Cai Sijing, Wang Yanyu. Improved DDeepLabV3+ semantic segmentation network. Journal of Fujian University of Technology. 2024, 22 (01):95-102.
[11] Cai Yan-xiang, Xie Hai-cheng, Yu Tian-qi, et al. Ranging model based on convergent binocular stereo vision. Laser & Infrared, 2023, 53 (08):1272-1278.