

A binocular vision system for underwater target detection

Jie Shen, Hongye Sun, Huibin Wang, Zhe Chen

College of Computer and Information Engineering
Hohai University
Nanjing, China
Jieshen_hhu@163.com

Yi Wei

Institute of Communication Engineering
PLA University of Science and Technology
Nanjing, China

Abstract—For the underwater target detecting task, a binocular vision system specialized to the underwater optical environment is proposed. The hardware platform is comprised of a image acquiring unit, a image processing unit and a upper computer. Accordingly, the loaded software system is operated for the camera calibration, image preprocessing, feature point extraction, stereo matching and the three-dimensional restoration. The improved Harris operator is introduced for the three-dimensional reconstruction, considering the high scattering and strong attenuation by the underwater optical environment. The experiment results prove that the improved Harris operator is better adapt to the complex underwater optical environment and the whole system has the ability to obtain the three-dimensional coordinate of the underwater target more efficient and accurate.

Keywords—underwater target detection; binocular vision; feature matching; recovery of 3 d coordinate

I. INTRODUCTION

Binocular stereo vision [1] is an important issue in the computer vision, which utilize a couple of cameras to perceive three-dimensional (3D) world. First, a pair of binocular image of a single scene with different view angles are recorded from two viewpoints. Second, the space geometric shape and the position of object in scene are calculated by the parallax of the same scene in different image [2]. The binocular stereo vision method is economical, reliable, and out-performanced, which is very suitable for non-contact target detection. Recently, the binocular vision technology is increasingly generalized to the underwater target detection and underwater three-dimensional reconstruction [3] [4], for the demand in the marine resources exploration, marine ecological environment monitoring [5] as well as the marine military applications.

Differing from the atmosphere, the underwater optical imaging technology suffers from the uneven illumination, light absorption and scattering by the particles and the water medium. So the collected underwater is visually with low contrast and low resolution. It brings much complexity for the underwater binocular images processing. Hence, the design of the image processing is one of the important issue of the related system.

This paper is to design an underwater binocular vision detection system based on the underwater target detection, and emphasizes on exploring the image processing method.

II. SYSTEM DESIGN

A. The hardware structure of system

The binocular vision hardware system mainly includes three compartments: image acquisition subsystem, image processing subsystem and upper computer subsystem, as shown in Fig. 1

1) *The image acquisition subsystem*: The image acquisition subsystem includes two CCD cameras with water shield. The cameras are parallely loaded in the tank (as shown in fig. 1). The target images is collected by these two cameras simultaneously, The images are transmitted to the image processing subsystem through the USB interface.

2) *The image processing subsystem*: The image processing subsystem is the main part of the binocular vision system, which is operated for the image preprocessing, feature point extraction, stereo matching and the 3D coordinate restoration. Due to the advantages of the field programmable gate array (FPGA), such as the short design cycle, low cost, design flexibility and independent intellectual property rights, the image processing subsystem is based on the programmable chip system (SOPC). As the main chip we adopt high-performance Stratix II from ALTERA, which has Nios II embedded processor, the on-chip memory and the digital signal processing (DSP) block for higher performance. The design software on the chip includes four parts which are logic design by Quartus II, embedded design by SOPC Builder, Nios II Embedded Design Suite (EDS) and DSP design by DSP builder from ALTERA. Human-computer interaction and control function are completed by Nios II embedded processor. The accumulation and multiplication operator for image processing are completed by DSP block. The DSP Builder can load the advanced algorithm on the hardware by compiling the MATLAB and SIMULINK code into the HDL form. With the Ethernet interface, image processing subsystem receives internal and external parameters of the

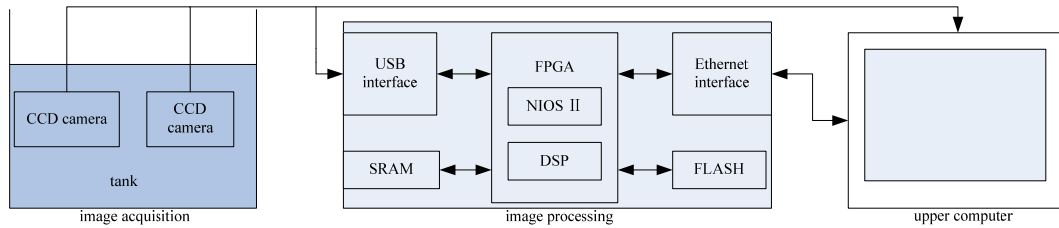


Figure 1. system structure chart.

binocular camera from the upper computer and feed the 3D coordinate back to them. Additionally, it receives the images data from the binocular camera by the USB interface.

3) *The upper computer subsystem:* The upper computer subsystem is an assist part in a binocular vision system. It shows the visualized results and complete the manual calibration to the underwater binocular camera. We adopts the ordinary PC machine as the upper computer. It receives images from the underwater binocular camera through the USB interface and send internal and external camera parameters to the image processing subsystem through the Ethernet interface which is reused to receive the output results from image processing subsystem.

B. The function design of the system software

The work flow of the binocular vision system of underwater target detection is designed as below. Firstly, the image acquisition subsystem acquires and transmits images to the upper compute. Secondly, the upper calculates the internal and external parameters of the camera which is transmitted to the image processing subsystem. Then, the image processing subsystem achieves the image processing tasks, including the image preprocessing, feature point extraction, stereo matching and the 3D coordinates restoration; Finally, the result is feed back to the upper computer. The main phases in the image processing module are comprised of the camera calibrating, image preprocessing, feature point extracting, stereo matching and the 3D restoration.

1) *Camera calibrating:* Camera must be calibrated to eliminate the information distortion in the underwater imaging. First the parameters of the camera should be estimated by establishing the ideal linear model of the camera. Then these camera parameters that the radial distortion are calculated with the nonlinear method.

2) *Image preprocessing:* Because of the noise in the underwater image, the underwater image at first should be preprocessed . Both the image denoising and the image enhancement processors are taken in our system.

3) *Feature extracting:* Feature extraction is the foundation task for image stereo matching. The accuracy and quality of feature point extraction directly influence the result of stereo matching. For the 3D restoration our

interests are focused on the design of the corner detection algorithm.

4) *Stereo matching:* The core of the object matching is to find the corresponding points between the left ocular image and the right ocular image. However, the matching points are seriously impacted by the noise; algorithm optimization hence is very important in this phase.

5) *3D reconstruction:* With the 3D reconstruction method, we calculate the projection matrix of the camera and the 3D coordinate of the matching point.

III. THE METHODS IMPLEMENTATION

Based on the design mentioned above, the computing method includes camera calibrating, image preprocessing, corner feature extracting, feature matching, error eliminating and the 3D coordinate restoration, as shown in fig. 2.

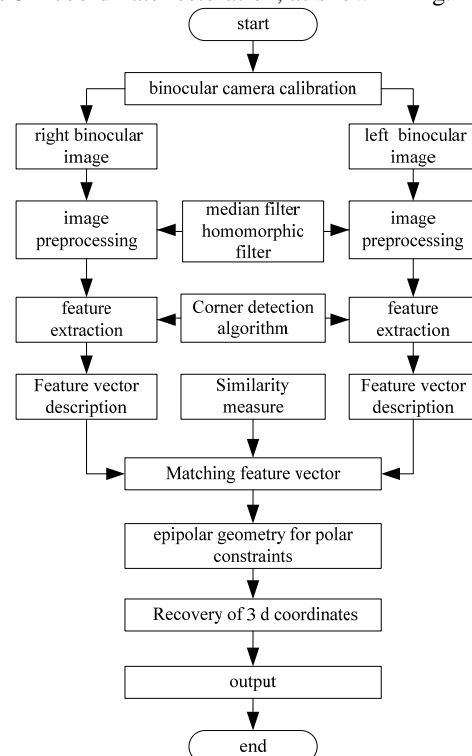


Figure 2. underwater target detection flow chart Based on binocular vision system.

A. Camera calibrating

Based on the plane calibration method [6], the chessboard templates is placed before the cameras with different view angle. As the result, many template images are obtained. Then established model is solved to obtain the internal and external camera parameters. When the calibration process complete, the mapping relationship between the calibration points M and the corresponding pixel m is indicated as

$$\lambda m' = HM' \quad (1)$$

Where, $m' = (u, v, 1)^T$ and $M' = (X, Y, Z, 1)^T$ are the homogeneous coordinates corresponding to m and M respectively.

According to the application, only the first and second order of the radial distortion of the camera lens are considered. Where (u_0, v_0) is the center coordinate of the image plane. The center coordinate is a vertical projection of the camera lens optical center in the image plane and the center of the lens radial distortion. Here (u, v) is the image pixel coordinate without distortion and (u, v) is the actual image pixel coordinate. Then the camera's nonlinear model can be gain as

$$\begin{bmatrix} (u_u - u_0)r^2 & (u_u - u_0)r^4 \\ (v_u - v_0)r^2 & (v_u - v_0)r^4 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} u - u_u \\ v - v_u \end{bmatrix} \quad (2)$$

These equations can be obtained according to the $M \times N$ points of the given board template image when the internal and external parameters have been estimated by the ideal linear model. Next, the radial distortion parameters k_1 and k_2 of the camera is obtained.

B. Image preprocessing

The median filter and the homomorphic filter [7] are used for underwater image denoising and enhancement. The process of the median filter can be expressed as

$$g(x, y) = Med\{f(x-k, y-l), (k, l \in W)\} \quad (3)$$

Where, W is the selected window size, $f(x-k, y-l)$ is the pixel gray value of the window W . If the window length is odd, the median commonly equals to the gray value of the middle pixel. If the window length is even, the median equals to the mean value of the gray value of the two middle pixels. The whole process of homomorphic filter is shown in Fig. 3.

Where Ln is the logarithmic operator, FFT is the logarithmic discrete Fourier transform, $H(u, v)$ is the linear

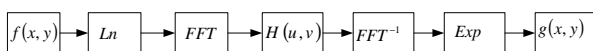


Figure 3. homomorphic filtering process of the underwater image.

high-pass filter, FFT^{-1} is the inverse Fourier transform, $f(x, y)$ is the original image while $g(x, y)$ is the image after the homomorphic filter.

In this paper we choose homomorphic filter function $H_L = 0.5$ and $H_H = 2.0$. The filtered image is with stronger contrast and clearer boundary compared to the original image. It provides a better quality of image for the following process.

C. Feature extraction

The improved Harris feature point detection algorithm [8] is employed for the feature extraction. First, we use the block method to divides the image into small area. Second, the Harris algorithm is used to detect the feature points within every underwater image block. Finally a certain number of feature points are extracted from the detected corner of each image block. When all image blocks are finished, the feature points with the larger T in image have been detected and reserved in each image area, The improved Harris feature point detection algorithm of the underwater image can be divided into six steps described as follow.

Step1 Calculating the gradient f_x and f_y of every image pixels in horizontal and vertical direction respectively; Computing the product of f_x and f_y and autocorrelation matrix M .

Step2 Calculating the new autocorrelation matrix M of the image by the Gaussian filter and the value of each pixel.

Step3 Calculation the length M and width N of the original underwater image and dividing the image into blocks while the length a and the width b respectively. So that original underwater image is divided into image rectangular blocks with

$$\frac{M}{a} \times \frac{N}{b} \quad (4)$$

Step4 Detecting the feature points of each image block, choosing the different initial threshold T . Since each image block is different, the number of the feature point of image block may vary considerably.

Step5 Setting the maximum number P of preserved feature point in the each image block, selecting the threshold T by the iteration dichotomy, achieving that the final selection threshold T which can ensure that the number of the feature points of the image block arrange within the range of $[0, P]$.

Step6 Calculating the row value and column value in the original image according to the initial value of each block.

D. Feature matching

In this paper, the right ocular image is used as the reference image, while the left ocular image is the matching image. The Nearest Neighbor (NN) algorithm is used to match the Harris corners, which, according to the Euclidean distance between two feature points, searches the nearest neighbor point and second-nearest neighbor point of the

feature point in another image. If the ratio is less than the threshold, it means that the feature point and the matching feature point of the image is the same point, the matching process is success. Note that the nearest feature point has the shortest Euclidean distance with the sample feature point and the second-nearest neighbor feature point has a longer distance.

Since we only compare the Euclidean distance between neighboring area points, so the complexity of matching process reduces greatly. Meantime, since the matching metric value of the feature point is more sensitive on the change of position so the precision degree of match can improve greatly.

Although the coarse matching set of two images is gotten, the noise originated from the light scattering may lead to the matching error between a pair of underwater images. Hence, next we should eliminate the false matching points. This paper the polar constraints are introduced by the epipolar geometry method. If the image feature points are in line with the polar constraint conditions, the matching feature point in the left image must be in the corresponding external polar in the right image. However, because of the camera distortion and other external factors, matching point may deviate from polar position. A range (unit is pixel number) is set. When the distance between candidates point and polar is longer than the set range, the candidate points is the false match point and the matching point of this feature point will be deleted.

E. 3D coordinates recovering

In this system the Structure from Motion (SFM) algorithm [9] is used to reconstruct the 3D coordinate for underwater target. This algorithm calculates a projection matrix of any point in the space according to the camera parameters, gets the corresponding coordinate of its projection point in two camera image coordinate system, and finally establish the linear equation in the world coordinate system, restoring the 3D coordinates of the space discrete point on underwater targets.

We assume that P_1 and P_2 are the image point detected from binocular image of two camera C_1 and C_2 . According to the matching result, P_1 and P_2 are the homologous points. Because two cameras have been calibrated, the projection matrix of the homologous points P_1 and P_2 are M_1 and M_2 respectively, then the camera imaging model is as follows.

$$Z_{c1} \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} = M_1 \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11}^1 & m_{12}^1 & m_{13}^1 & m_{14}^1 \\ m_{21}^1 & m_{22}^1 & m_{23}^1 & m_{24}^1 \\ m_{31}^1 & m_{32}^1 & m_{33}^1 & m_{34}^1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (5)$$

$$Z_{c2} \begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} = M_2 \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11}^2 & m_{12}^2 & m_{13}^2 & m_{14}^2 \\ m_{21}^2 & m_{22}^2 & m_{23}^2 & m_{24}^2 \\ m_{31}^2 & m_{32}^2 & m_{33}^2 & m_{34}^2 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (6)$$

Equation (5) and (6) are solved as

$$\begin{bmatrix} M_{13u1i} - M_{11} \\ M_{13v1i} - M_{12} \\ M_{23u2i} - M_{21} \\ M_{23v2i} - M_{22} \end{bmatrix} X_i = 0. \quad (7)$$

Where M_{11} , M_{12} and M_{13} are three-row vector of the projection matrix M_1 , M_{21} , M_{22} and M_{23} are three row vector of projection matrix M_2 , $(u_{1i}, v_{1i}, 1)^T$ and $(u_{2i}, v_{2i}, 1)^T$ are the image homogeneous coordinates of the matching point and X_i is the space homogeneous coordinates corresponding to the match point.

The least square method can be used to solve X_w , Y_w and Z_w . We can define

$$N = \begin{bmatrix} M_{13u1i} - M_{11} \\ M_{13v1i} - M_{12} \\ M_{23u2i} - M_{21} \\ M_{23v2i} - M_{22} \end{bmatrix}. \quad (8)$$

Equation (7) can be rewritten as

$$NX_i = 0. \quad (9)$$

According to the least square method, solution of (9) is the feature vector of the matrix minimum eigenvalues. Equation (10) is the singular value decomposition available of N .

$$N = USV^T \quad (10)$$

The last column of the matrix V is the 3D coordinate (X_w, Y_w, Z_w) of a space point.

IV. EXPERIMENT AND ANALYSIS

In this paper, three image sets are taken in the experiments. They are coin images with high light and low noise, humanoid images with low light and low noise, box images with low light and low noise. The right ocular image is the reference image and the left ocular image is the corresponding matching image. The image preprocessing experiments and the results are below.

A. Image preprocessing

The parameters $H_L = 0.5$ and $H_H = 2.0$ for median filter and $c = 1.2$ for homomorphic filter are selected for the image preprocessing. Because the processed results of left and right binocular images are the same, we give the left ocular image after preprocessing shown in the fig. 4, 5, and 6.

It is found that in fig4 there are no obvious differences after the process by the median and homomorphic filter while in fig5, the image contrast is enhanced by the homomorphic filter. In fig6, the image noise is reduced by the median filter and image contrast is enhanced by the

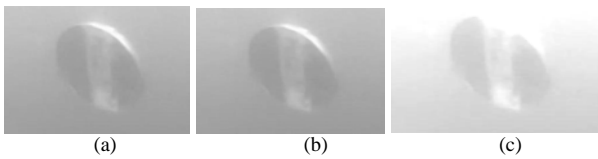


Figure 4. Coin images with high light and low noise condition (left ocular image): (a)original image, (b)image after median filter, (c)image after homomorphic filter.

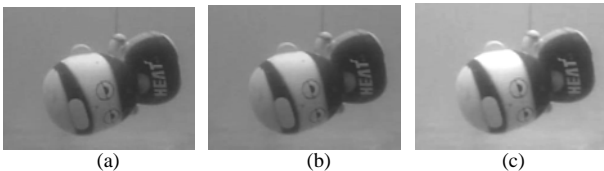


Figure 5. Humanoid images with low light and low noise (left ocular image): (a)original image, (b)image after median filter, (c)image after homomorphic filter.

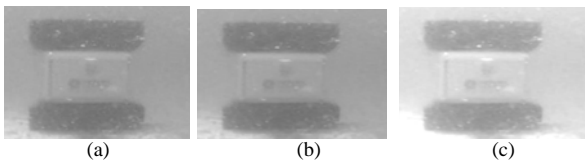


Figure 6. Box images with low light and low noise (left ocular image): (a)original image, (b)image after median filter, (c)image after homomorphic filter.

homomorphic filter. It proves that median filter has better inhibition effect on underwater image noise and the homomorphic filter has better enhancement effect on underwater image contrast.

B. Feature points extracting

This paper respectively adopts the traditional Harris corner detection algorithm and improved Harris corner detection algorithm for extracting the feature points in the underwater image. Fig 7 and 8 show the experimental result. The size of the images are 320×240 .

Fig 7 shows the corners with different threshold value in traditional Harris corner detection algorithm. There are 31 corners in (a) with the threshold $T = 10000$, 151 corners in (b) with the threshold $T = 4000$ and 386 corners in (c) with the threshold $T = 500$.

The main characteristic of the improved Harris corner detection algorithm is that every underwater image is divided into 32 small image block with $P = 6$.

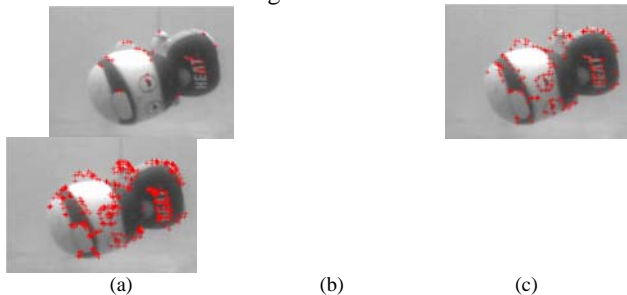


Figure 7. The detection results of different threshold value with traditional Harris corner detection algorithm: (a) $T = 10000$, (b) $T = 4000$, (c) $T = 500$.

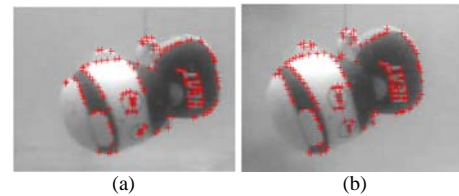


Figure 8. The detection results with improved Harris corner detection algorithm: (a) the left ocular image, (b) the right ocular image.

Fig 8 shows the left and right ocular images calculated by the improved Harris corner detection algorithm. There are 152 corners in every image. From the graph, we can see that the feature points is reduced, uniform distributed and have the ability to present the whole image characteristics.

Comparing these two kinds of algorithms, it can be seen that the improved Harris corner detection algorithm has several advantages compare to the traditional Harris corner detection algorithm in underwater environment. It have less feature points, which are easy to calculating. The feature points are uniformly distributed.

C. Feature point matching

1) *Corner matching results with the polar constraint:* In order to get higher matching accuracy, we select 200 corners,. The NN threshold value is set as 0.6. 191 matched points can be got with the epipolar geometry constraint. Fig. 9(a) shows the matching results. In the improved Harris corner detection algorithm, we select 152 extractive corners. 147 matched points can be got with the epipolar geometry constraint. Fig. 9(b) shows the matching results.

2) *the comparison of number and accuracy of matching:* Fig. 10 is the result of matching precision comparison between the improved Harris operator, traditional Harris operator and SIFT. It is clear that, the image matching effect is better under the condition of high light and low noise. But the lighting condition has little influence on the image matching. In the same conditions, with the increasing of the noise, image matching becomes more and more difficult, the number of successful matching points reducing. Also in the same condition, improved Harris operator can extract more feature points than the Harris operator and SIFT operator.

D. 3D reconstruction result of underwater target

An excellent style manual for science writers is [7].

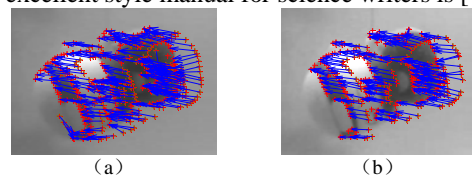


Figure 9. The result of Harris corner matching: (a) The traditional Harris corner matching, (b)The improved Harris corner matching.

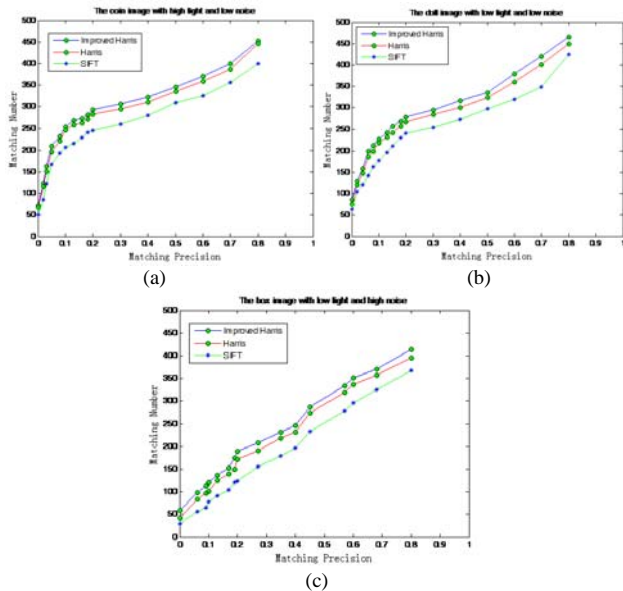


Figure 10. The experimental results of three kinds of algorithm on three images: (a)coin images with high light and low noise, (b)humanoid images with low light and low noise, (c)box images with low light and low noise.

In order to get the spatial position of underwater target, the SFM algorithm[9] is used for 3D reconstruction. In order to test the detection accuracy of underwater target, 20 matching points are selected on the target surface, as shown in column 1, 2 of table 1.

Combining the coordinate of the above matching points in the image coordinate system with the internal and external parameters of the camera, 3D coordinate of the corresponding space point and its space position can be got. Column 3, 4 of table 1 show the 3D coordinate calculated by the reconstructed coordinate and the actual coordinates.

From column 3, 4 of table 1, we can see that the design of reconstruction system in this paper gains higher reconstruction precision, while the process is more simple and effective.

V. CONCLUSION

This paper presents a novel binocular vision system for underwater target detection. Experiments prove the matching accuracy and computational complexity of the improved Harris operator. At the same time, the 3D coordinates of space discrete points have been got through 3D reconstruction. By experiments result we also found that the image preprocessing quality and the scale sensitivity of algorithm seriously influence the precision of image matching under the special underwater optical environment. Therefore, the issues that how to further improve the effect of image preprocessing and reduce the scale sensitivity of the algorithm are need to be studied in the future work.

REFERENCES

- [1] Morales, R.; Pereira, D.; , "Prototype of Stereoscopic Vision System to Improve the Image Quality in Turbid Waters during Underwater Inspections," OCEANS 2006 , Sept. 2006, pp.1-4, 18-21, doi: 10.1109/OCEANS.2006.306786.
- [2] T. Hanning, A. Lasaruk, T. Tatschke. "Calibration and Low-Level Data Fusion Algorithms for a Parallel 2D/3D-Camera," Information Fusion, Vol. 12, Jan. 2011, pp. 37-47.
- [3] H. Aliakbarpour, J. Dias, "Three-dimensional reconstruction based on multiple virtual planes by using fusion-based camera network," Computer Vision, IET, vol.6, July. 2012, no.4, pp. 355-369, doi: 10.1049/iet-cvi.2011.0078.
- [4] C. Beall, B.J. Lawrence, V. Ila, F. Dellaert, "3D reconstruction of underwater structures," Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on, Oct. 2010, pp.4418-4423, 18-22, doi: 10.1109/IROS.2010.5649213.
- [5] Camilli, R.; Bowen, A.; Farr, N.; , "Bright blue: Advanced technologies for marine environmental monitoring and offshore energy," OCEANS 2010 IEEE - Sydney , May. 2010, pp.1-7, 24-27, doi: 10.1109/OCEANSSYD.2010.5
- [6] Z. Zhang, "A flexible new technique for camera calibration," Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.22, Nov. 2000, pp. 1330- 1334, doi: 10.1109/34.888718.
- [7] Y.J. Zhang, Image Engineering(1) Image Processing, 2nd ed., tsinghua: Beijing, 2006, pp.151-153.
- [8] W.J. Zhao, S.R. Gong, C.P. Liu, X.J. Shen, "Adaptive Harris Corner Detection Algorith," Computer Engineering, Vol. 34, May. 2008, pp.212-217, doi:1000-3428(2008)10-0212-03.
- [9] G.H. Wang, Q.M.J. Wu, H. Wei, "Structure and motion factorization under quasi-perspective projection with missing data in tracking matrix," Pattern Recognition, 2008. ICPR 2008. 19th International Conference on , Dec. 2008, pp.1-4, 8-11, doi: 10.1109/ICPR.2008.4761345.

TABLE I. THE COORDINATE VALUE OF THE MATCHING POINT IN THE IMAGE COORDINATE SYSTEM

Left ocular image coordinate	right ocular image coordinate	actual coordinate	reconstruct coordinate
(87,103)	(57,99)	(2.74,2.43,1.45)	(2.8326,2.6841,1.4469)
(75,114)	(44,107)	(2.30,2.68,1.39)	(2.3187,2.7003,1.3820)
(102,133)	(70,127)	(2.92,3.30,1.05)	(3.0067,3.4256,1.1233)
(100,204)	(76,200)	(2.79,4.61,1.20)	(2.8245,4.6832,1.3057)
(133,79)	(100,74)	(3.40,2.18,1.30)	(3.4804,2.2040,1.3864)
(126,123)	(98,112)	(3.31,3.00,1.11)	(3.3052,2.9411,1.1409)
(143,157)	(133,150)	(3.61,3.80,1.02)	(3.6248,3.7922,1.0368)
(139,213)	(110,207)	(3.59,4.82,1.09)	(3.6289,4.9003,1.0865)
(193,51)	(159,43)	(4.62,1.60,1.50)	(4.6565,1.6228,1.5490)
(198,90)	(163,81)	(4.78,2.40,1.36)	(4.8201,2.3792,1.3588)
(179,156)	(145,150)	(4.40,3.70,1.06)	(4.4698,3.8004,1.0591)
(190,198)	(156,189)	(4.60,4.59,1.08)	(4.5790,4.5688,1.0827)
(232,44)	(199,38)	(5.41,1.50,1.48)	(5.4267,1.4902,1.4902)
(238,195)	(201,187)	(5.59,2.45,1.07)	(5.6108,2.4603,1.0816)
(254,148)	(222,143)	(5.89,3.59,1.06)	(5.9204,3.6034,1.0681)
(224,178)	(196,171)	(5.24,4.18,1.11)	(5.2656,4.1864,1.1208)
(277,56)	(245,50)	(6.34,1.71,1.32)	(6.3502,1.7062,1.3308)
(292,86)	(257,80)	(6.62,2.32,1.28)	(6.6407,2.3314,1.2867)
(298,120)	(263,114)	(6.76,3.00,1.19)	(6.7724,3.0156,1.2033)
(281,157)	(249,150)	(6.42,3.74,1.16)	(6.4355,3.7519,1.1702)