Research on Binocular Ranging Method Based on Feature Point Extraction and Matching

Ganquan Su[*a], Lei Cheng[b]

[a]School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan 430081, China;

[b]School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan 430081, China

[*] Corresponding author: id_08241@qq.com

ABSTRACT

At present, artificial intelligence has become a hot topic, and the development of its related fields is also developing rapidly, among which visual ranging and image processing are particularly important for the development of artificial intelligence. Now there are many ranging methods are not fast enough, and measurement accuracy is not high, leading to the resulting estimates there are large deviation distance and the actual distance, such as unmanned vehicle, unmanned aircraft in operation process, cannot be accurately driving distance and obstacle avoidance, and existing deviations in robot grab, which cause personnel life safety is threatened, economic property damage. In order to solve the above related problems, this paper uses SIFT algorithm and ORB algorithm to extract feature points, and then through BFmatcher, FlannBasedMatcher, KnnMatch to match, finally get the corresponding distance, from the algorithm and accuracy of the two aspects of relevant research. It is concluded from the experimental measurement that FlannBasedMatcher takes into account both speed and accuracy after SIFT algorithm extraction, while ORB algorithm is faster than SIFT algorithm.

Keywords: binocular technology; camera calibration; ranging technology; SIFT algorithm; ORB algorithm; BFmatcher matching; FlannBasedMatcher matching; KnnMatch match

1. INTRODUCTION

Nowadays, artificial intelligence is a hot social topic. Research on computer vision has also increasingly been conducted. As a key research direction in the related field, binocular technology is progressively drawing people's attention. Binocular technology has a broad prospect in robot vision, driverless prediction and obstacle avoidance. Binocular camera is employed in the binocular technology to perceive the depth information of 3D scene and thus to provide an effective basis for robot vision and 3D modeling [1]. There are many ways to implement ranging. Traditional measurement can be carried out with ruler. However, modern technology facilitates the measurement and calculation by means of calculation and geometric model construction. Thus the ultimate measurement results are worked out. Binocular ranging also involves laser ranging, ultrasonic ranging, infrared ranging, etc. These rangings may consequently lead to high accuracy, but much more complex equipments and high implementation cost are also inevitably required.

In the advantageous indoor environment, binocular stereo vision ranging technology is employed in the object ranging research. Binocular vision technology refers to the computerized simulation of human eyes for the observation of surrounding environment. A fixed binocular camera is used to capture simultaneously a pair of images from different angles [2]. Similar to the function of human eyes, the object image information is captured and fed back to the computer for a series of operations, so as to get the corresponding results. The key of binocular ranging technology lies in the calibration and matching of the camera. The Camera calibration is achieved by first utilizing VC++ and OpenCV to obtain images, and then transmitting the acquired images into the Stereo Camera Calibrator toolbox of MATLAB to find out the internal and external parameters of the Camera. The distance measurement is based on the local feature of SIFT algorithm and ORB algorithm to extract feature points. Then combined with BFmatcher[3], FlannBasedMatcher[4], and KnnMatch[5], the matching is carried out. The parallax is therefore calculated in accordance with the feature matching. After obtaining the disparity, the distance of the target object can be obtained in line with the principle of triangular

similarity [6]. On the computer, the final ranging result is obtained by programming in VS2019 software with the help of OpenCV [7].

2. SYSTEM DESIGN

2.1 Ranging Principle and System Process of Binocular Camera

The human eye can perceive the distance of the object because there is a visual difference between the two eyes on the captured image, which is termed as parallax. The farther the target is, the smaller the parallax will be; The closer the target is, the greater the parallax will be. This principle is applied in binocular vision ranging. Therefore, the distance measurement of the obtained image is achieved by virtue of geometric methods and the related camera knowledge.

Assuming that there is a motion point P, it can move freely within the range of the camera. As point P moves, the position of its imaging point on the left and right cameras will also change. Based on the principle of binocular parallax triangulation, the depth information of the object is obtained, and the distance measurement of the binocular camera is realized. Figure 1 shows the ranging principle. In the figure, P is the moving point position; O_l is the optical center position of the left camera; O_r is the optical center position of the right camera; F is the focal length of the camera; B is the baseline length; P_l is the left imaging point; P_r is the right imaging point.

The core of binocular ranging is to solve D, and D is the disparity:

$$d = x_l + x_r \tag{1}$$

According to the similarity triangle principle, the depth information z can be obtained:

$$\frac{z - f}{z} = \frac{b - d}{d} \tag{2}$$

Simplify to obtain:

$$z = \frac{fb}{d} \tag{3}$$

In the process of solving the binocular camera, the camera focal length F and the baseline length are fixed parameters, which are determined by the selected camera. It is the value of disparity D and depth information Z that need to be calculated by computer.
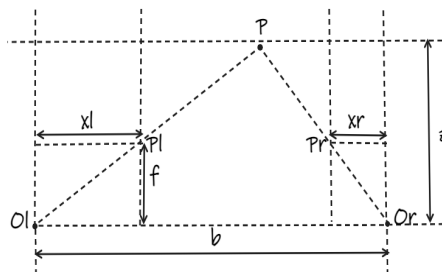

Figure 1. Schematic diagram of binocular camera ranging

The system process of binocular camera ranging is roughly divided into the following steps: image acquisition, camera calibration, stereo correction, feature point extraction and matching, and ranging. Image acquisition is realized by programming language with OpenCV. Camera calibration serves to obtain internal parameters such as focal length, image point, radial distortion, tangential distortion and external parameters involving translation parameter and rotation parameter. Stereo correction is to eliminate the distortion by calibrating the obtained parameters, and to achieve the goal of complete alignment for the need of ranging. Finally, the image pixels obtained from the left and right cameras are matched by feature point extraction and matching to find out the disparity and the calculated depth information. The ranging results are accordingly aquired. Figure 2 is the flow chart of the system.
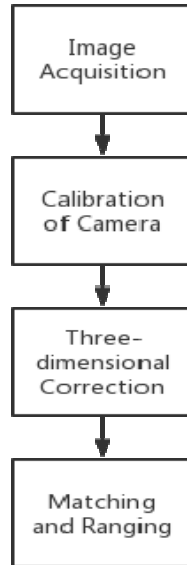
Figure 2. Flow chart of binocular camera ranging system

2.2 Image Acquisition and Camera Calibration

The acquisition of image means the application of OpenCV in controlling the binocular camera with programming language and then capturing the images from different angles.

The camera calibration is to obtain the internal and external parameters of the camera. These parameters are fixed and constant, only relating to the selected equipments. Once the relevant parameters are accurately obtained, they are directly taken in the future need. The internal and external parameters determine the correspondence between the image coordinate system and the world coordinate system. The inner parameter is a transformation from plane to pixel, which only depends on the physical characteristics of the camera itself. The outer parameter reflects the transformation between the camera coordinate system and the world coordinate system, and is determined by the inner parameter and the baseline length [8].

The commonly used checkerboard calibration method is employed in this experiment. The checkerboard template in this experiment is a rectangular black and white checkerboard with 10×7 corner points and 29mm×29mm side length. As shown in Figure 3.
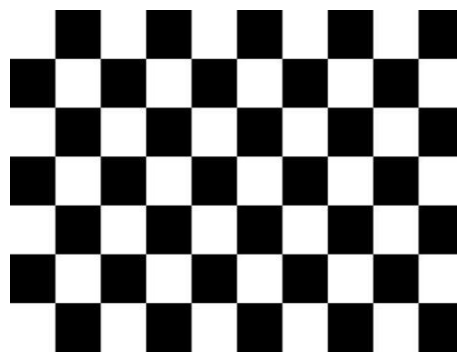

Figure 3. 10×7 black and white checkerboard

Three coordinate systems need to be used in the process of calibration, namely: image coordinate system, camera coordinate system and world coordinate system. Among them, the image coordinate system can be divided into pixel coordinate system and physical coordinate system. The pixel coordinate system is represented by U and V, which is the image captured by the camera and returned to the computer. After the computer processing, it is converted into a digital image, in which each element is called a pixel. The physical coordinate system is generally represented by X and Y, and the intersection between the optical axis of the lens and the front plane is set as the origin. Setting dx and dy as the size

of the pixel in the XY coordinate system, then the transformation relationship between pixel coordinates and physical coordinates is

$$\begin{bmatrix} u \\ v \\ l \end{bmatrix} = \begin{bmatrix} \dfrac{1}{dx} & 0 & u_0 \\ 0 & \dfrac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ l \end{bmatrix} \tag{4}$$

The camera coordinate system is represented by Xc, Yc and Zc. According to the camera coordinate system, the imaging position of the moving point P on the image is

$$\begin{cases} x = \dfrac{X_c * f}{Z_c} \\ y = \dfrac{Y_c * f}{Z_c} \end{cases} \tag{5}$$

The matrix form is

$$Z_c * \begin{bmatrix} x \\ y \\ l \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} \tag{6}$$

The internal parameter matrix can be obtained by combining matrices (4) and (6)

$$Z_c \begin{bmatrix} u \\ v \\ l \end{bmatrix} = \begin{bmatrix} \dfrac{1}{dx} & 0 & u_0 \\ 0 & \dfrac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & f \end{bmatrix} * \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} \tag{7}$$

The world coordinate system is represented by Xw, Yw and Zw, which are all three-dimensional coordinate systems with the camera coordinate system. The transformation between them only requires rotation and translation, and the resulted matrix is the external parameter matrix

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \begin{bmatrix} R & T \\ 0^T & 1 \end{bmatrix} * \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{8}$$

After 30 pairs of images are input into the Stereo Camera Calibrator toolbox of MATLAB, 29 pairs of images satisfying the calibration conditions are acquired through relevant screening. Then, the relevant information images are secured through the operation of corner extraction, etc. Finally, the final 28 pairs of image information are garned after removing the one pair of images with large error. The calibration results are as follows:

$$M_l = \begin{bmatrix} 821.1276 & 0 & 332.7671 \\ 0 & 812.7568 & 219.3796 \\ 0 & 0 & 1 \end{bmatrix} \tag{9}$$

$$M_r = \begin{bmatrix} 817.6628 & 0 & 362.9843 \\ 0 & 807.7924 & 307.9630 \\ 0 & 0 & 1 \end{bmatrix} \qquad (10)$$

$$D_l = \begin{bmatrix} 0.0433 & -0.1262 & 0 & 0 & 0 \end{bmatrix} \qquad (11)$$

$$D_r = \begin{bmatrix} 0.0233 & -0.2614 & 0 & 0 & 0 \end{bmatrix} \qquad (12)$$

$$R = \begin{bmatrix} 0.9997 & 0.0084 & -0.0228 \\ -0.0089 & 0.9997 & -0.0241 \\ 0.0225 & 0.0243 & 0.9995 \end{bmatrix} \qquad (13)$$

$$T = \begin{bmatrix} 68.3663 & -1.0356 & -0.8737 \end{bmatrix} \qquad (14)$$

Where M_l and M_r are the internal parameters of the left and right cameras, D_l and D_r are the distortion parameters of the left and right cameras, R is the rotation parameter of the left and right cameras, and T is the translation parameter of the left and right cameras.

2.3 Stereo Correction

Stereo correction is to eliminate the distortion in the course of calibrating the image acquired by the left and right cameras. Then the strict correspondence of the already processed images is conducted, in which it is necessary to reprojecting the image plane of the two cameras.

This process paves the way for future acquision of the disparity through extraction and matching of the feature points. In this case, it is the simplest to calculate the stereo disparity, reducing the computational complexity of the matching process and improving the accuracy of feature point extraction and matching [9].

The Bouguet algorithm [10] uses the rotation parameter R and translation parameter T obtained by the binocular camera calibration above to minimize the number of reprojections and maximize the overlapping observation area of each left and right image. In order to minimize the image reprojection distortion, it is necessary to break down the matrix R that rotates the right camera image plane to the left camera image plane into two parts.

Firstly, the obtained calibration data are entered into the program and used as parameters. Then the correlation function in OpenCV is used to complete the calibration. Finally, the output of the corrected image is achieved.

Before completing the stereo correction, the corner points of the black and white checkerboard in the picture cannot match correctly (as shown in Figure 5), which could be clearly seen by pictures shot by the left and right cameras. This will bring great difficulty to the matching calculation later.
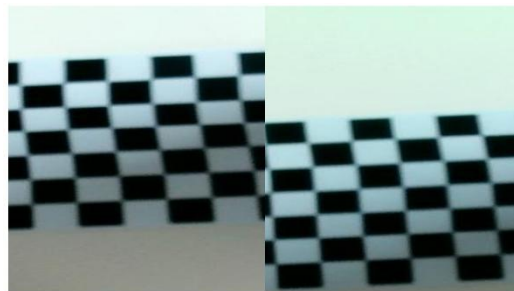

Figure 4. Images taken by left and right cameras without correction

After correction, the images (as shown in Figure 6) obtained by the left and right camera could clearly prove that the corresponding positions of the obtained black and white checkerboard are all on the same row.
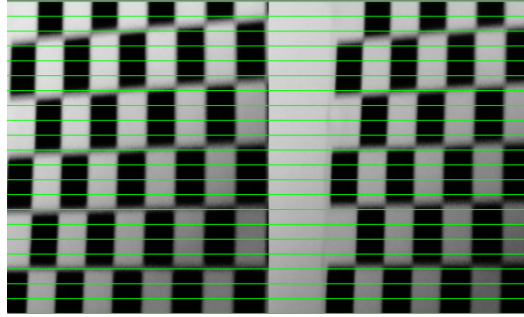
Figure 5. Image pair by left and right cameras after correction

2.4 Feature Point Extraction, Matching and Ranging

A pixel can be regarded as a feature of the image, which is accordingly entitled as a feature point. Feature point is characterized by repeatability, distinguishability, efficiency, locality, rotation invariance and scale invariance. The difference can be used for detection, and the repeatability is employed for matching. Feature points include key elements and descriptors. SIFT algorithm is a widely-used scale-invariant feature detection method. Each feature point in the picture is described by 128-dimensional vector. SIFT algorithm seeks extreme points in the scale space to extract location, scale and selection invariants [11]. The ORB algorithm is a combination of FAST detector [12] and Brief descriptor [13], and the ORB algorithm has scale invariance and rotation invariance [14]. After extracting feature points according to the two algorithms, the extracted feature points are matched by corresponding methods, and then the disparity is obtained according to the matching method. The purpose of ranging is eventually achieved in accordance with the disparity obtained and the depth calculation by the triangular similarity principle.

3. EXPERIMENT RESULTS AND ANALYSIS

3.1 Experiment Process and Results

The experiment was divided into five groups, and each group was divided into six different distance ranging experiments of the same target. The first group carried out SIFT+BFmatcher algorithm for image ranging; The second group implemented SIFT+FlannBasedMatcher algorithm for ranging; The third group conducted SIFT+KnnMatch algorithm for ranging. The fourth group used the ORB+BFmatcher algorithm for ranging. The fifth group of experiments employed the ORB+KnnMatch algorithm for ranging. The contents of each group are the same. The same target picture (as shown in Figure 6) is divided into 350mm, 400mm, 450mm, 500mm, 550mm and 600mm, and then ranging is carried out to check the accuracy. The left and right shots are shown in Figure 6.

Figure 6. Image of 350mm

Figure 7. Image of 400mm

Figure 8. Image of 450mm

Figure 9. Image of 500mm

Figure 10. Image of 550mm

Figure 11. Image of 600mm

First, fix the camera in the ready position and adjust the machine position so that the cameras on both sides are kept on the same horizontal line and perpendicular to the plane. Then use a ruler to measure the corresponding distance between the camera and the level. The picture is pasted and vertical to the plane and placed in the front end of the range corresponding position. Then the image acquisition and ranging are carried out.The distance measurement of SIFT+BFmatcher algorithm is shown in Table 1, SIFT+FlannBasedMatcher algorithm in Table 2, SIFT+KnnMatch algorithm in Table 3, ORB+BFmatcher algorithm in Table 4, and ORB+KnnMatch algorithm in Table 5. Taking 600mm ranging as an example, the time used to calculate each algorithm is indicated in Table 6.

Table1.   Experiment results of SIFT+BFmatcher binocular ranging

Serial Number	Experiment Data			
	Soft ruler ranging /mm	Binocular ranging /mm	Absolute error /mm	Relative error /%
1	350	352.6164	2.6164	0.7475
2	400	389.7640	10.2360	2.5590
3	450	433.5267	16.4733	3.6607
4	500	478.3204	21.6796	4.3359
5	550	520.8069	29.1931	5.3078
6	600	567.2150	32.7850	5.4642

Table2.   Experiment results of SIFT+FlannBasedMatcher binocular ranging

Serial Number	Experiment Data			
	Soft ruler ranging /mm	Binocular ranging /mm	Absolute error /mm	Relative error /%
1	350	352.6164	2.6164	0.7475
2	400	393.1052	6.8948	1.7237
3	450	433.5267	16.4733	3.6607

4	500	502.9664	2.9664	0.5933
5	550	569.5153	19.5153	3.5482
6	600	591.2716	8.7284	1.4547

Table3.   Experiment results of SIFT+KnnMatch binocular ranging

Serial Number	Experiment Data			
	Soft ruler ranging /mm	Binocular ranging /mm	Absolute error /mm	Relative error /%
1	350	352.6164	2.6164	0.7475
2	400	393.1052	6.8948	1.7237
3	450	433.5267	16.4733	3.6607
4	500	502.9664	2.9664	0.5933
5	550	569.5153	19.5153	3.5482
6	600	591.2716	8.7284	1.4547

Table 4.   Experiment results of ORB+BFmatcher binocular ranging

Serial Number	Experiment Data			
	Soft ruler ranging /mm	Binocular ranging /mm	Absolute error /mm	Relative error /%
1	350	348.9470	1.0530	0.3009
2	400	394.4507	5.5493	1.3873
3	450	435.1743	14.8257	3.2946
4	500	488.1520	11.8480	2.3696
5	550	550.3675	0.3675	0.0668
6	600	569.9466	30.0534	5.0089

Table 5.   Experiment results of ORB+KnnMatch binocular ranging

Serial Number	Experiment Data			
	Soft ruler ranging /mm	Binocular ranging /mm	Absolute error /mm	Relative error /%
1	350	343.9796	6.0204	1.7201
2	400	389.8436	10.1564	2.5391
3	450	435.1743	14.8257	3.2946
4	500	488.1520	11.8480	2.3696
5	550	535.1565	14.8435	2.6988
6	600	569.1147	30.8853	5.1476

Table 6.   Taking ranging 600mm as an example to compare the final time of each algorithm

	Experiment Data				
	SIFT+BFmatcher	SIFT+FlannBasedMatcher	SIFT+KnnMatch	ORB+BFmatcher	ORB+KnnMatch
Time(s)	1.2959	0.0586	0.0673	0.2475	0.1796

3.2 Analysis of the Experiment Results

According to the above five groups of experiment results, it is found that the experiment results obtained by SIFT+FlannBasedMatcher algorithm and SIFT+KnnMatch algorithm are the same, the feature points acquired by matching and screening are exactly the same, and therefore the ranging results are exactly the same. However, compared with SIFT+KnnMatch algorithm, SIFT+FlannBasedMatcher algorithm takes less time and is more efficient, which is the best matching scheme to extract through SIFT algorithm. However, ORB algorithm is faster for extracting feature points compared with SIFT algorithm. In terms of accuracy of ranging, SIFT+FlannBasedMatcher algorithm and SIFT+KnnMatch algorithm are relatively more precise in ranging. And ORB+BFmatcher algorithm has a higher accuracy in the experimental close-range measurement.In general, SIFT+FlannBasedMatcher algorithm is the most efficient one. In the five groups of experiments, this algorithm is both accurate and fast, which can better meet the needs of binocular camera ranging and other related academic research.
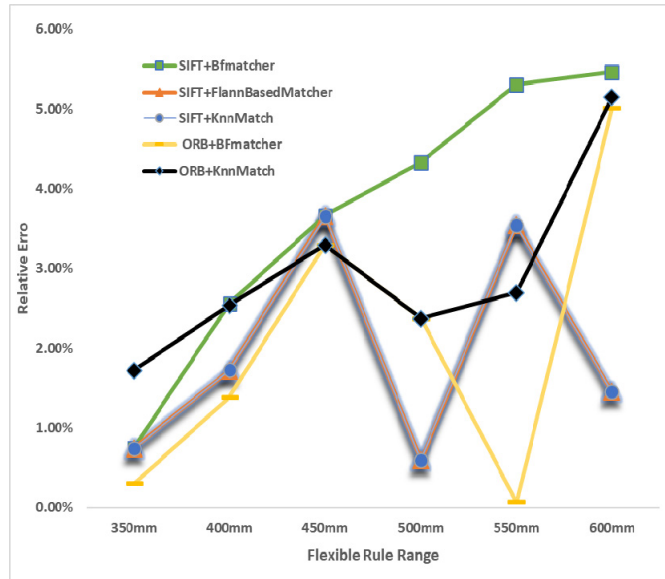
Figure 12. Relationship between the measurement and relative error of the tape ruler
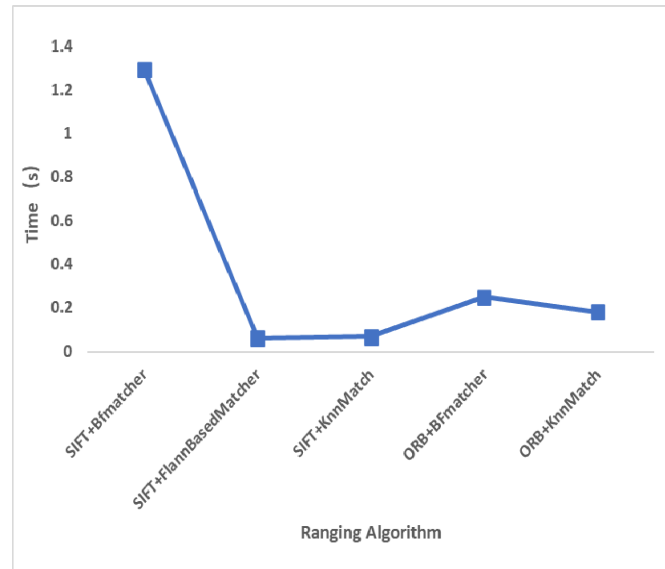

Figure 13. Relationship between ranging algorithm and operation speed

4. CONCLUSION

Research is carried out in this paper, regarding the differences in speed and accuracy between five ranging algorithms based on binocular technology. Starting from the ranging principle of binocular vision technology, a series of processes such as image acquisition, camera calibration, stereo calibration, feature point extraction and matching, and ranging are studied through experiments, and the ranging information required by the experiment is calculated. According to the experiments of five ranging methods, SIFT+FlannBasedMatcher algorithm leads to the best result. And the reasons for this conclusion are analyzed. The future research will be improved with the continuous progress of computer technology and upgrade of the related algorithms. Therefore, a more rapid and accurate ranging method will be worked out.

REFERENCES

[1] Xu Jie, Chen Yimin, Shi Zhilong. [1] Binocular Vision Zoom Ranging Technology [J]. Journal of Shanghai University (Natural Science Edition), 2009, 15(2): 169-174.

[2] Shasha Yu, Hao Huang, Yangjie Liu. A Low-complexity Autonomous 3D Localization Method for Unmanned Aerial Vehicles by Binocular Stereovision Technology [A]. In: IEEE. 2018 10th International Conference on Intelligent Human-Machine Systems and Cybernetics, 2018, Page(s): 344-347.

[3] Amila Jakubović, Jasmin Velagić. Image Feature Matching and Object Detection Using Brute-Force Matchers[C]. International Symposium ELMAR.2018

[4] Vineetha Vijayan, Pushpalatha Kp. FLANN Based Matching with SIFT Descriptors for Drowsy Features Extraction[C]. International Conference on Image Information Processing.2019

[5] Fengquan Zhang, Yahui Gao, Liuqing Xu. An adaptive image feature matching method using mixed rich-KD tree[J].Multimedia Tools and Applications, 2020, 23 (a24) : 16421-16439

[6] Wang Hao, Xu Zhiwen, Xie Kun. [6] Binocular Ranging System Based on OpenCV [J]. Journal of Jilin University (Information Science Edition), 2014,32(2):188-194.

[7] LI Dejun, MA Xiaohui. The Binocular Measuring System Research Based on OpenCV[C]//International Conference on Material   and Manufacturing technology. 2012

[8] Hu Jinbo, Zhang Feixiong, Wan Zekun, Huang Hao. [8] Research on Indoor Three-dimensional Measurement Algorithm Based on Binocular Technology [J]. Computer measurement and control, 2019,27(9):66-67.

[9] Zelin Meng, Xiangbo Kong, Lin Meng, Hiroyuki Tomiyama. Distance Measurement and Camera Calibration based on Binocular Vision Technology[C]. International Conference on Advanced Mechatronic Systems.2018

[10] Gunen Mehmet Akif, Besdok Erkan, Civicioglu Pinar, Atasever Umit Haluk. Camera Calibration by Using Weight Differential Evolution Algorithm: A Comparative Study with ABC, PSO, COBIDE, DE, CS, GWO, TLBO, MVMO, FOA, LSHADE, ZHANG and BOUGUET[J]. Neural computing & applications, 2020, 32 (23)

[11] David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints [J]. International Journal of Computer Vision, 60, pages91-110 (2004).

[12] Edward Rosten, Tom Drummond. Machine Learning for high-speed Corner Detection[C]//European Conference on Computer Vision(ECCV 2006) pt.1; 20060507-13; Graz(AT).2006

[13] Michael Calonder,Vincent Lepetit,Christoph Strecha,Pascal Fua. BRIEF: Binary Robust Independent Elementary Features[C]//European Conference on Computer Vision.2010

[14] Ethan Rublee;Vincent Rabaud;Kurt Konolige;ORB: An Efficient Alternative to SIFT or SURF [A].in: IEEE. 2011 International Conference on Computer Vision, 2011, Page(S): 2564-2571.