Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 7, No. 3, 2023

Binocular Ranging Design Based on FPGA

Zhibo Chen, Shuai Guo, Guangshao Zhou, Xiaodong Zhang

Southwestern University for Nationalities, Chengdu, Sichuan, 610041, China

Abstract: In daily life, distance measurement is everywhere. In order to accurately measure the distance information of the target object, a system design for binocular ranging based on FPGA is proposed. The internal and external parameters of the binocular camera are determined through the Zhang Zhengyou calibration method (two-dimensional target), and then the basic model of the binocular camera is derived. The data model of the binocular camera is derived based on the binocular camera model. The calibrated binocular camera parameters are used to perform binocular correction and distortion correction on the image data collected by the binocular camera. The corrected left and right eye views are matched with pixel points through a stereo matching algorithm to form a disparity map. The principle of triangle similarity is used to estimate the target distance and obtain the coordinate system of the target. Finally, the data is measured by deploying it on the FPGA development board.

Keywords: Binocular ranging, Camera calibration, Binocular correction, FPGA.

1. Preface

Binocular vision technology is an important category of computer vision technology. Binocular vision technology simulates human eyes. Information about the depth (Z-axis) of the target is obtained by comparing the position difference (parallax) of the target in the left and right eyes. Then, the computer obtains the disparity of the image data at the required position through the stereo matching algorithm.

At present, most ranging system designs use computers as the development platform, which are large in size and inconvenient to move. It is not suitable for places with complex working environments such as underwater detection and portable robots. In this environment, it is of great research significance to deploy the binocular ranging system on embedded devices. But the commonly used embedded systems are those based on ARM architecture. Although it meets the volume and power consumption requirements, its computing power is poor. The data processing capabilities cannot meet the expected requirements.

Field Programmable Gate Array (FPGA) has the advantages of powerful computing power and fast data processing speed. This design uses the ZYNQ platform under ALINX Company, which is a system based on FPGA+ARM. FPGA can implement parallel computing and pipeline technology in algorithm design, which can meet the high speed and high bandwidth requirements of system design. The processing of images is implemented through the ARM core.

2. Theoretical Basis of Binocular Ranging

2.1. Camera model and four-clock reference coordinate system

In order to obtain the three-dimensional coordinate information of the distance target, changes between four reference coordinates need to be involved. The world coordinate system is a three-dimensional coordinate system used to describe the actual location of an object, usually using (X_W, Y_W, Z_W) to represent any point; the camera coordinate system is a coordinate system defined using the camera itself as the reference standard, usually using (X_C , Y_C, Z_C) to represent any point; the image coordinate system is a coordinate system with the center of the image as the origin, and (x, y) is usually used to represent any point; the pixel coordinate system is a coordinate system that uses a pixel as the basic unit System, usually (u, v) is used to represent any point of it. The four reference coordinate systems can be interchanged with each other, and the relationship between them is shown in Figure 1:

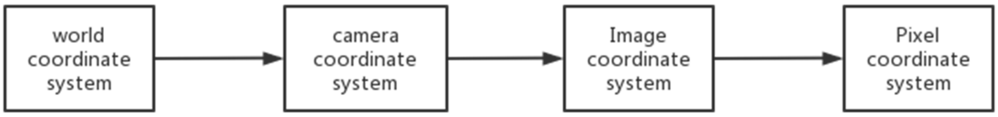

Figure 1. Conversion relationship diagram between four coordinate systems

In the above figure, the world coordinate system (X_W, Y_W, Z_W) is rotated and translated to obtain the camera coordinate system (X_C , Y_C, Z_C), the camera coordinate system (X_C , Y_C, Z_C) is obtained through perspective projection to obtain the image coordinate system (x, y), and the image coordinate system (x, y) After discretization and offset, the pixel coordinate system (u, v) is obtained. The conversion formula for any point from the world coordinate system to the pixel coordinate system is as shown in formula (1):

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R_c & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \quad (1)$$

Among them: f_x, f_y represents the scale factor of the coordinates, which is the ratio of the focal length f $(u_{0,} v_0)$ and d_x, d_y represents the optical center coordinates; R_C represents 3×3 the rotation matrix; T represents 3×1 the translation matrix.

2.2. Camera calibration

In the binocular ranging system, the accuracy of the internal and external parameters of the camera and the distortion coefficient have a great impact on the accuracy of the measurement results. The purpose of camera calibration is to establish the relationship between three-dimensional world coordinates and two-dimensional pixel coordinates. Zhang Zhengyou's calibration method can quickly and accurately determine the internal and external parameters and distortion parameters of the camera. The basic principle of Zhang Zhengyou's calibration method is to derive the internal and external parameters of the camera by observing and measuring the projection of the camera onto a calibration object of known geometric shape. The specific steps are: 1. Collect a set of images of calibration objects with known geometric shapes, such as checkerboards; 2. Detect the corners of the calibration objects from these images, generally using a corner detection algorithm; 3. According to the camera model , establish the mapping relationship between the camera coordinate system and the pixel coordinate system. This relationship can be represented by the pinhole camera model, which includes the internal parameters of the camera (such as focal length, principal point, etc.) and external parameters (such as rotation and translation matrices); 4. Using the found corner points and camera model, it can be calculated Internal and external parameters of the camera. The internal parameters include the focal length, the coordinates of the main point in the pixel coordinate system, etc., and the external parameters include the rotation and translation matrices of the camera; 5. Based on the known geometric shape of the calibration object and the internal and external parameters of the camera, each corner point can be calculated Three-dimensional coordinates in the camera coordinate system; 6. Finally, the calibration results are optimized through an optimization algorithm to improve the accuracy of the calibration. This design uses a 7×9 checkerboard as the calibration object, the size of each checkerboard square is $30\text{mm}\times30\text{mm}$, and the checkerboard is fixed on a cardboard, as shown in Figure 2:

Figure 2. Image of checkerboard calibration plate

The Zhang Zhengyou checkerboard calibration method used above has the advantages of high accuracy and simple operation.

2.3. Binocular correction and epipolar geometry

The calibration of binocular vision is based on the calibration of monocular vision by two cameras respectively. On this basis, epipolar correction is added to complete the calibration. Its ultimate purpose is to achieve complete parallelism of the optical axes of the left and right cameras and to completely align the epipolar lines of the captured left and right eye images, laying a solid foundation for the next step of stereo matching and further improving the efficiency of the next step of stereo matching.

2.3.1. Epipolar geometry

Binocular cameras are essentially two monocular cameras that capture the same object from different positions. If there is overlap between the captured left-eye image and right-eye image, then there must be a certain correspondence between the two images. Generally, we use epipolar geometry to describe this relationship. Using the perspective of epipolar geometry, we can establish the mathematical relationship between the left eye image and the right eye image (Figure 3). This is an important method used to study binocular vision. method.

As shown in the figure, a polar plane jointly determined by the optical centers C1 and C2 of the two cameras and a point P on the world coordinate system intersects with the left and right destination perspective planes (i.e., projection planes) in a straight line. This straight line is called The epipolar line, point P, is imaged and projected on the left and right perspective planes as PL and PR respectively, and the imaging points of the optical centers of the left and right cameras are defined as poles el, er. According to the relevant definitions, we can draw a conclusion: given that the imaging point PL and the pole el of one camera are known, the projection point PR of point P on another camera must be the intersection with the polar plane on another perspective plane. superior. Therefore, we can deduce: given a feature on an image, its matching feature on another image must be on the corresponding epipolar line, which is the epipolar constraint.

Therefore, as long as we can determine the epipolar geometric relationship between the left and right cameras and establish epipolar constraints, the two-dimensional matching relationship between the left and right images can become a one-dimensional matching, which greatly reduces the matching time. required calculations. Moreover, because the matching range is narrowed, many possible matching errors are eliminated.

2.3.2. Binocular Correction

In binocular correction, Fuseiello method and Bouguet method are two commonly used methods. Their main differences are as follows: 1. Correction model: Fuseiello method uses a model based on geometric relationships, while Bouguet's method uses a model based on parallax planes; 2. Parameter estimation: Fuseiello method needs to estimate the internal parameters and external parameters of the camera. parameters and distortion parameters, while the Bouguet method only needs to estimate the internal parameters and distortion parameters of the camera; 3. Pixel reprojection: The Fuseiello method will perform pixel reprojection on the corrected image to correct the image deformation caused by perspective projection. The Bouguet method does not perform pixel reprojection, but only performs geometric correction on the image; 4. Real-time performance: Since the Fuseiello method involves more parameter estimation and pixel reprojection operations, it may be slightly inferior to the Bouguet method in terms of real-time performance. To sum up, the Bouguet algorithm is used as the binocular correction algorithm in this design. Bouguet's method is to first rotate the coordinate system to be parallel, and then rotate the coordinate system to be coplanar.

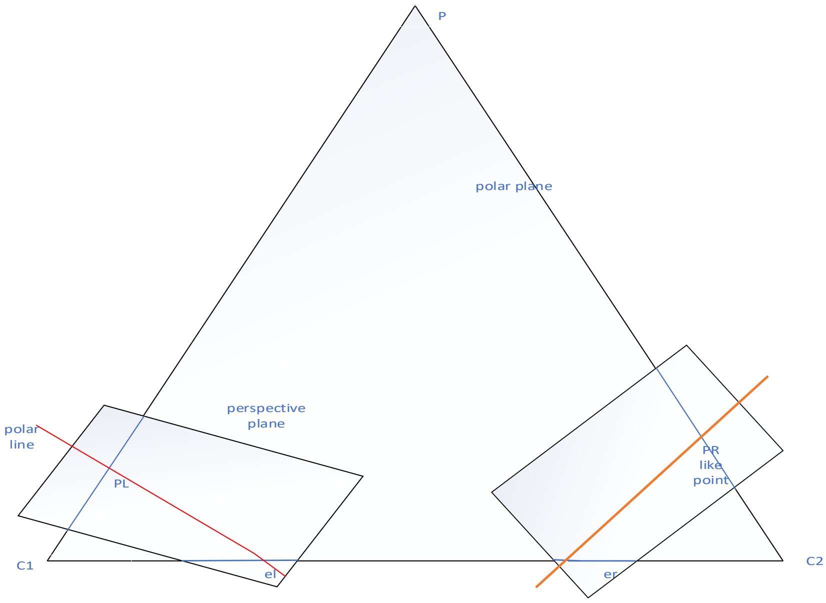

Figure 3. Mathematical model of epipolar geometry

Calibrate the intrinsic and extrinsic parameter matrices and distortion parameters obtained when the left and right cameras are calibrated respectively , and obtain the rotation matrix and translation vector T of the right camera relative to the left camera , as well as the camera intrinsic parameter matrix and after epipolar correction. Distortion vector. The purpose of the Bouguet algorithm is to minimize the number of reprojections of each of the left and right images while maximizing the observation area.

In order to minimize the number of reprojections for each of the left and right images, which is equivalent to minimizing the distortion of reprojections, the Bouguet algorithm divides the previously calculated rotation matrix R of the right camera plane relative to the left camera plane into R_C .The two parts R_L and R_R allow each camera to rotate half the angle to reduce reprojection distortion.

In order to obtain a matrix that can transform the left camera pole to infinity to obtain the epipolar horizontal alignment, R_{rect} here I use the translation vector T of the right camera relative to the left camera to achieve indirect implementation. The main steps are as follows:

Definition: The transformation between C1 and C2 coordinate systems is determined by R, T. As shown in Equation 2.

$$\begin{bmatrix} X_{\mathrm{cam\_right}} \\ Y_{\mathrm{cam\_right}} \\ Z_{\mathrm{cam\_right}} \end{bmatrix} = R_{3X3} \begin{bmatrix} X_{cam\_left} \\ Y_{cam\_left} \\ Z_{cam\_left} \end{bmatrix} + T_{3X1} \quad (2)$$

Among them, $R_{3\times3} = R^{\frac{1}{2}} R^{-\frac{1}{2}}$.

Construct a unit vector e_1 oriented from the right camera to the left camera:

$$e_1 = \frac{T}{\|T\|} \quad (3)$$

Construct a unit vector e_2 that is orthogonal to the direction of the main optical axis and e_1 perpendicular to the direction. It can be seen that e_2 it can be obtained by the vector product of e_1 the unit vector in the direction of the main optical axis e_Z . And because e_1 and e_Z are not orthogonal, e_2 normalization is required to restore them to the identity matrix:

$$e_z = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$$
$$e_2 = \frac{T \times e_z^T}{\|T \times e_z^T\|} \quad (4)$$

Construct the unit vector e_3 . Since it is orthogonal to e_1 and e_2, we get:

$$e_3 = e_1 \times e_2 \quad (5)$$

In this way, we can get the matrix that transforms the extremes of the left eye camera to infinity R_{rect} :

$$R_{\mathrm{rect}} = \begin{bmatrix} e_1^T & e_2^T & e_3^T \end{bmatrix}^T \quad (6)$$

By R_{rect} transforming the matrix, you can obtain a new rotation matrix sum for the left and right R_L cameras R_R .

$$R_L = R_{rect} R^{\frac{1}{2}}$$
$$R_L = R_{rect} R^{-\frac{1}{2}} \tag{7}$$

Use R_L these R_R two matrices to rotate the camera coordinate system, calculate the difference, and then derive the correction result. Left eye camera correction result formula:

$$\begin{bmatrix} X_{cam\_rect\_left} \\ Y_{cam\_rect\_left} \\ Z_{cam\_rect\_left} \end{bmatrix} = R_L \begin{bmatrix} X_{cam\_left} \\ Y_{cam\_left} \\ Z_{cam\_left} \end{bmatrix} \tag{8}$$

Right eye camera correction result formula:

$$\begin{bmatrix} X_{cam\_rect\_\mathrm{right}} \\ Y_{cam\_rect\_\mathrm{right}} \\ Z_{cam\_rect\_\mathrm{right}} \end{bmatrix} = R_R \begin{bmatrix} X_{cam\_\mathrm{right}} \\ Y_{cam\_\mathrm{right}} \\ Z_{cam\_\mathrm{right}} \end{bmatrix} \tag{9}$$

2.3.3. 1.3.3 Calibration results

This design adopts Zhang Zhengyou's chessboard calibration method and uses MATLAB's calibration toolbox to obtain the required internal and external parameters. The final calibration parameter results obtained are as follows:

$$\mathrm{A}_1 = \begin{bmatrix} 470.20975 & 0 & 303.44027 \\ 0 & 470.74927 & 264.36546 \\ 0 & 0 & 1 \end{bmatrix}$$

$$DS_1 = [0.04820 \quad -0.02706 \quad 0.00927 \quad 0.00211 \quad 0.00000]; \%\mathrm{k}_1 k_2 p_1 p_2 k_3$$

$$A_2 = \begin{bmatrix} 468.48570 & 0 & 343.93897 \\ 0 & 469.48889 & 235.11195 \\ 0 & 0 & 1 \end{bmatrix}$$

$$DS_2 = [0.05700 \quad -0.06316 \quad 0.00366 \quad 0.00520 \quad 0.00000]; \%\mathrm{k}_1 k_2 p_1 p_2 k_3$$

$$T = [58.61932 \quad 0.14159 \quad -0.01300]$$

$$Rvec = [-0.00531 \quad -0.02928 \quad 0.00280]$$

Among them, Rvec is the rotation matrix, T is the translation matrix, A_1 and A_2 are the internal parameter matrices of the left and right cameras, DS_1 and DS_2 sum is the external parameter matrix of the left and right cameras.

The comparison of the image before camera calibration and the image after camera calibration is shown in Figure 4 and Figure 5.

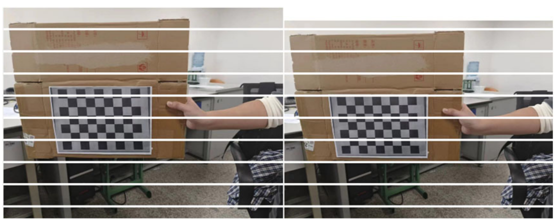

Figure 4. Before double target setting

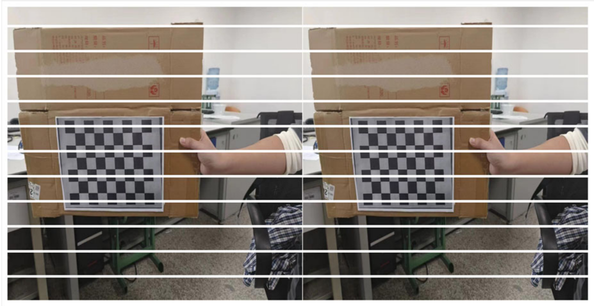

Figure 5. After camera calibration

Based on the basic principles of binocular vision and Zhang Zhengyou's calibration method, MATLAB was selected as the calibration tool. Camera calibration completed.

2.4. Stereo matching

Stereo matching is the most critical step in the binocular ranging system. The basic principle is to find a pair of corresponding points in the same scene in the left and right views. The main constraints in stereo matching are: 1. Epipolar constraints: a pair of matching points must be located on the corresponding epipolar lines in the left and right views. 2. Uniqueness constraint: There is only one corresponding point in the left and right views. 3. Parallax continuity constraint: Except for occlusion areas and discontinuous areas, the change of parallax should be smooth. 4. Sequence constraint: A group of four treasures on the epipolar line of a certain view have the same order on the epipolar line of the corresponding view.

There are many types of algorithms for calculating binocular disparity, which can be mainly divided into local stereo matching algorithms and global stereo matching algorithms. Stereo matching algorithm flow:

a) Cost calculation (local correlation)
b) cost aggregation
c) Parallax calculation
d) Parallax results optimization

The differences between stereo matching algorithms are mainly reflected in the methods of cost calculation and cost aggregation. The implementation of these two process algorithms is related to the efficiency of the disparity map obtained by the stereo matching algorithm and the matching error rate. In the design of this system, the SGM algorithm, a semi-global matching algorithm with relatively good speed and accuracy, is used. According to the disparity gradient information between adjacent pixels, different penalties are applied to the initial matching cost to solve the disparity.

3. FPGA Design Process

3.1. Overall design of hardware system

Vivado is an integrated design environment released by FPGA manufacturer Xilinx. It provides a complete design process, including design, simulation, synthesis, implementation and verification functions. The FPGA development board used in this design is the Zynq-7020 series developed and designed by Xilinx. The overall hardware development process of this system design is shown in 6, which mainly includes image acquisition module, binocular correction module, stereo matching module and VGA display module. Data transfer between the four modules is achieved by relying on the memory interconnection protocol and the stream data transfer protocol.

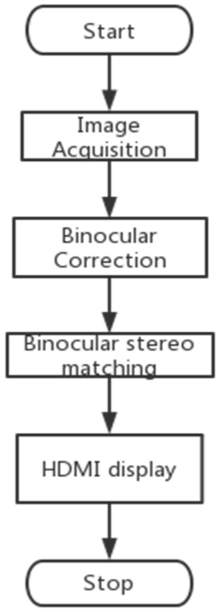

Figure 6. Overall hardware design

Transmits the original binocular image data collected by the binocular camera AN5642 to the hardware system , and stores it in the memory, waiting for subsequent processing.

The binocular correction module mainly converts the original binocular image into a binocular image after epipolar correction through the remapping table , so that the binocular stereo matching changes from two-dimensional pixel matching to one- dimensional pixel matching.

The binocular stereo matching module performs stereo matching on the corrected binocular images , calculates the binocular disparity, and outputs a disparity map.

The HDMI display module converts the RGB image in the disparity map into HDMI output format, and transmits the data to the display through the HDMI data transmission line .

3.2. Image acquisition module design

The binocular camera AN5642 transmits the original binocular image data collected to the hardware system and stores it in the memory, waiting for the next step of processing. The process of collecting image data is shown in Figure 7.

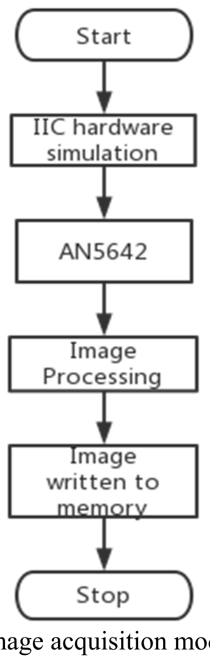

Figure 7. Image acquisition module process

The register configuration in the AN5642 binocular camera is implemented through Verilog code written in Vivado. Then, the RGB565 format data output by the AN5642 binocular camera is converted into the RGB88 format commonly used in image processing. Finally, image storage is implemented by calling the VDMA IP core in Vivado, and the image data is stored in memory.

3.3. 2.3 Design of binocular correction module

Mainly through the remapping table, the collected original binocular image data is converted into epipolar-corrected binocular image data, so that the binocular stereo matching changes from two-dimensional pixel matching to one-dimensional pixel matching. The workflow of the binocular correction module is shown in Figure 8.

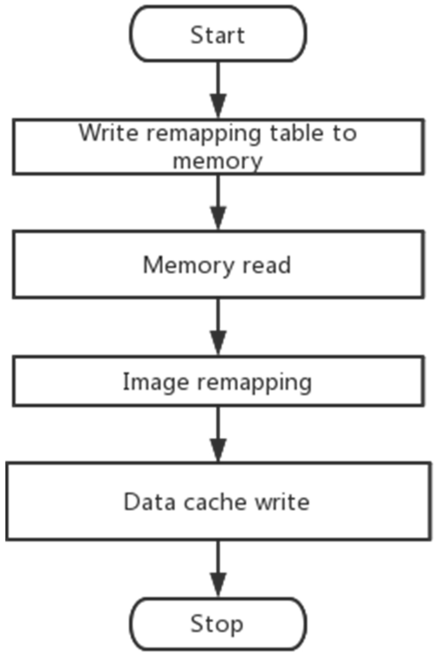

Figure 8. Binocular correction module process

The remapping table is written into the memory in the form of a one-dimensional array. Since the remapping table is fixed, the program only needs to continuously read from the written address during operation. Then, the collected binocular image data and remapping table data are simultaneously taken out in AX14-Stream format and transmitted to the image remapping module. Finally, the image remapping module corrects the collected original binocular images according to the remapping table, and outputs the epipolar corrected image.

3.4. Binocular stereo matching module design

The binocular stereo matching module performs stereo matching based on the binocular image pairs after epipolar correction, calculates the binocular disparity of the corrected images based on the calculated internal and external camera parameters, and outputs the calculated results. The design process of the binocular stereo matching module is shown in Figure 9.

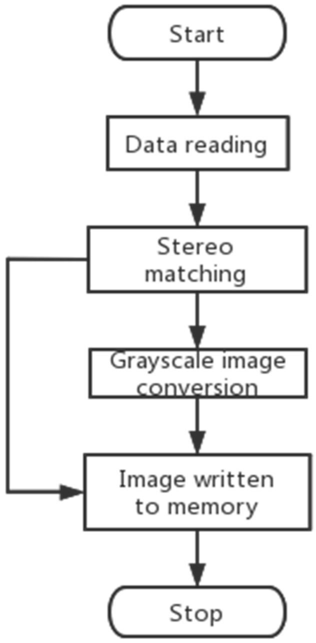

Figure 9. Binocular stereo matching module design process

The stereo matching module outputs a 16-bit single-channel grayscale disparity image by calculating the parallax of the corrected left and right eye images, and then converts the 16-bit grayscale disparity image obtained in the previous module into an 8-bit three-channel pseudo-color image. , and output the corrected image of the left eye at the same time as a control image.

3.5. HDMI display module design

HDMI is a fully digital video and sound transmission interface that can send uncompressed audio and video signals. In this system design, PGB format data is converted into HDMI format data for output. Display the output data on an external monitor.

4. Results Display

The test work is completed in the laboratory through the binocular ranging system designed in this article. The internal and external parameters of camera calibration are obtained through binocular correction and epipolar geometry, and the correction is obtained to obtain an image pair. Then use the stereo matching algorithm to obtain the corresponding disparity map. Finally, the distance information of the target is obtained based on the principle of triangle similarity.

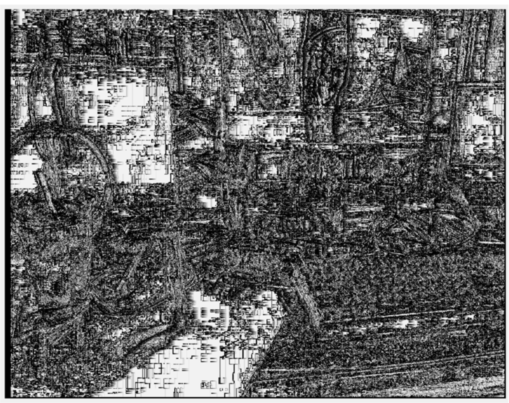

Figure 10. Disparity map

Place the ointment at monitoring distances of 200mm, 250mm, 300mm, 350mm, 400mm, 450mm, 500mm, 550mm, 600mm, 650mm, 700mm, 750mm, and 800mm from the camera. Measure each location 5 times, and take the average of the 5 measurement results as the final measurement result. The measurement chart is as follows.

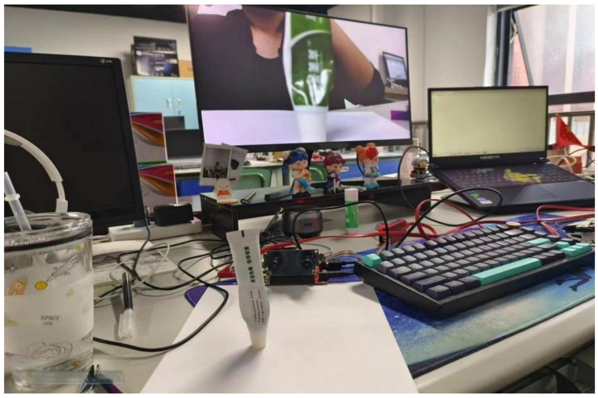

Figure 11. Measurement results

In order to further carry out error analysis, the actual value, the distance measured by binocular vision, the error value and the accuracy rate were put into a table for analysis, as shown in Table 1.

Table 1. Test results

serial number	Actual value (mm)	Measurement value (mm)	Error value (mm)
1	200	201.628	1.628
2	250	251.855	1.855
3	300	302.457	2.457
4	350	352.687	2.687
5	400	403.892	3.892
6	450	455.621	5.621
7	500	506.211	6.211
8	550	557.517	7.517
9	600	606.546	6.546
10	650	655.963	5.963
11	700	708.451	8.451
12	750	757.687	7.687
13	800	808.709	8.709

According to the analysis in Table 1, the measurement results of this system at close range are relatively accurate. It shows that the binocular ranging system has high accuracy in close measurement. Generally speaking, the ranging system based on binocular vision can more accurately measure the distance between the target and the binocular camera, and has certain feasibility.

5. Conclusion

Based on binocular vision technology, the ranging technology for targets was studied at different distances. Relying on the reliable performance of FPGA, a binocular ranging system based on FPGA was proposed. Through binocular cameras, video images and other information in real scenes are collected, and Zhang Zhengyou's plane calibration, stereo correction, stereo matching and triangular similarity principles are used to achieve real-time distance calculation of the target. However, the system is easily affected by human errors and systematic errors, and the calculation results are unstable and need to be further improved and applied to specific engineering projects.

References

[1] Wang Mingji, Chen Qiumeng, Ren Fushen. Target ranging system based on binocular vision [J]. Automation and Instrumentation, 2022.07.05.

[2] Han Qian. Design and implementation of camera ranging system based on binocular vision [D]. Hebei: Hebei University of Science and Technology, 2019.

[3] Zhao Fuyang. Research on binocular stereo vision system based on Zynq [D]. Tianjin: Tianjin University of Science and Technology, 2017.

[4] Wang Qiang. Research on binocular vision system based on FPGA [D]. Beijing: Tsinghua University, 2015.

[5] Yue Xiaofeng, Qi Huan. Binocular stereo vision system calibration based on Zhang Zhengyou's plane template method [J]. Mechanical Engineer, 2014, (02): 1-3.

[6] Pan Lei, Ge Zhongqin, Zhuang Jianjun, Zhang Zhijian. System design of HDMI video stream image processing based on FPGA [J]. Laboratory Research and Exploration, 2015, 34 (10): 76-80.

[7] Wu Zhen. Research and FPGA implementation of stereo matching algorithm based on binocular vision [D]. Harbin: Harbin University of Science and Technology, 2016.