Research on Stereo Matching Technology Based on Binocular Vision

Chang Su, Gongquan Tan, Yufeng Luo

School of Automation and Information Engineering, Sichuan University of Science & Engineering, Yibin, China
Email: 734867681@qq.com

Abstract

With the rapid development of machine vision, binocular stereo vision based on the principle of parallax has gradually become the core of scientific research. This paper briefly presents the background and research significance, elaborates the research status of binocular vision robot at home and abroad and studies the checkerboard calibration method, and uses Matlab to complete binocular camera calibration. Stereo matching technology is the core and most difficult part of binocular stereoscopic 3D reconstruction research. Firstly, the image acquired after calibration is enhanced by gray scale transformation to make the image clearness optimal, and then use NCC (normalization cross-compilation). The algorithm performs the matching of left and right image pairs in the Matlab environment to generate an optimal matching disparity map.

Subject Areas

Picture Processing

Keywords

Binocular Stereo Vision, Checkerboard Calibration, Binocular Camera Calibration, Stereo Matching, Gray Scale Transformation

1. Introduction

Vision has become the most widely integrated sensing system for industrial robots due to its features of contactless, enormous amount of information, easy to understand analysis, flexible detection and low installation cost. With the continuous development of social production and people's life, the system is becoming ever more intelligent and practical. Whether it can successfully identify the target object and accurately locate the target object becomes an important indi-

cator to measure the performance of the cerebral system. Machine vision can be divided into monocular, binocular and multi-vision depending on the number of visual sensors. Binocular and multi-vision vision technology has high complexity and real-time instability in robot integration, but because binocular vision simulates the way human eyes handle the surrounding three-dimensional environment, it is more in line with the biological structure and its stability relative to multi-eye vision [1]. More real-time, so more and more researchers at home and abroad have started to get involved in research in the field of binocular vision technology.

STROPPAL and CRISTALLICC of Loccioni have designed a binocular vision detection system for detecting the quality of plug-in on the production line. The final measurement results have shown that the measurement method based on binocular vision is excellent in resolution, detection speed and stability. For other non-contact measurement methods [2], LVRORAE of Valencia Polytechnic University reconstructed a three-dimensional model of the grape using binocular stereo vision with a reconstruction error of less than 1 mm [3]. In China, research on machine stereo vision started late. Zhang Wei, from Shanghai University of Technology, based on the principle of binaural measurement, used Matlab and Opencv to design a target positioning system based on binocular stereo vision [4].

This paper will use the checkerboard calibration method to calibrate the vision system, and enhance the gray scale transformation of the acquired image to improve the certain recognition degree. Finally, based on the NCC algorithm, the two images corresponding to the same scene are stereo-matched.

2. Camera Calibration

2.1. Calibration Principle

The three-dimensional pose information of the acquired object need to restore the three-dimensional information of the object in the three-dimensional space of the two-dimensional plane image captured by the camera. According to the camera imaging model, it can be seen that in this reduction process, there are coordinate transformation relationships, and the coordinate systems that need to be used are: world coordinate system, camera coordinate system, image coordinate system, and pixel coordinate system [5]. Camera calibration is the process of solving the projection matrix (composed of the camera's internal parameter matrix and external parameter matrix) based on the selected imaging model (this article selects the pinhole imaging model).

The calibration method of the camera is mainly divided into three types according to whether the calibration is needed: the traditional calibration method, the functional visual calibration method, and the self-calibration method. With the innovation of researchers' various thinking, the calibration methods are diverse. According to the advantages and disadvantages of most calibration methods, this section uses Zhang Zhengyou's pinhole camera perspective transformation matrix calibration method to calibrate the binocular camera [6].

In the pinhole imaging model, the four coordinate system conversion relationships in the process of converting the spatial coordinates of the object to the two-dimensional image coordinate are shown in **Figure 1**.

As showing in **Figure 1**, the world coordinate system is O_W-$X_W Y_W Z_W$; the camera coordinate system is O-$X_c Y_c Z_c$; the image physical coordinate system is O_1-xy; and the pixel coordinate system is O_0-uv. Establish a world coordinate system to represent the camera's camera position, O_W-$X_W Y_W Z_W$. The world coordinate system can obtain the camera coordinate system through rotation transformation and translation transformation [7]. The homogeneous relation expression is as follows:

$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ 0^T & 1 \end{bmatrix} \cdot \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} = M_1 \cdot \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \tag{1}$$

where R is a 3×3 rotation matrix, T is a three-dimensional translation matrix, $0 = \begin{pmatrix} 0 & 0 & 0 \end{pmatrix}^T$, and M_1 is a 4×4 matrix.

In **Figure 1**, the camera optical center 0 is collinear with the target object point N and the perspective projection point Nu. The camera coordinate system can obtain the image physical coordinate system according to the similar triangle principle. The homogeneous relation expression is as follows:

$$Z_C \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} \tag{2}$$

In the formula, f is the focal length of the camera.

The unit of the image plane coordinate system is the physical unit, usually in mm, and the image pixel coordinate system is in pixels. The horizontal coordinate axis u of the image pixel coordinates is parallel to the x-axis of the image physical coordinate system, and the vertical coordinate axis v is parallel to the y-axis of the image physical coordinate system. Their homogeneous relational expressions are as follows:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \dfrac{1}{dx} & 0 & u_0 \\ 0 & \dfrac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{3}$$

where (u_0, v_0) is the coordinates of the principal point 01 in the image pixel coordinates system.

In the conversion process between the world coordinate system, the camera coordinate system to the image coordinate system and the pixel coordinate system, the projection relationship is easily obtained according to the perspective model of the camera:
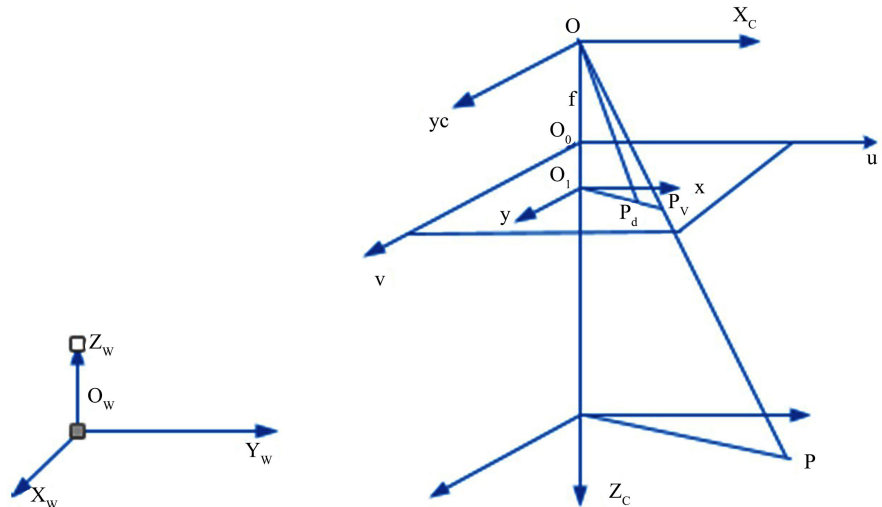
Figure 1. Coordinate transformation relationship.

$$
Z_C \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \dfrac{1}{dx} & 0 & u_0 \\ 0 & \dfrac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0^{\mathrm{T}} & 1 \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix}
$$

$$
= \begin{bmatrix} \alpha_x & 0 & u_0 & 0 \\ 0 & \alpha_y & y_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0^{\mathrm{T}} & 1 \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} = M_1 M_2 X_W = M X_W
$$

(4)

In the formula, M is a 3 × 4 matrix as a projection matrix. M_1 is the parameter matrix in the camera, and M_2 is the parameter matrix outside the camera.

2.2. Camera Calibration

The calibration of the camera can be realized by various means. In this paper, the Matlab calibration toolbox based on Zhang Zhengyou calibration method is used to calibrate the camera to obtain the internal and external parameters of the camera. The implementation process is as follows [8]:

1) According to the camera shooting distance, shooting field of view and the size of the calibration plate is not less than one third of the field of view. The size of the checkerboard image is 5 × 9, and the size of each grid is 30 mm × 30 mm.

2) Keep the left or right camera position fixed, constantly change the spatial position of the calibration plate, and collect 17 images of the image at different positions by the camera.

3) After loading the image successfully, in the camera calibration main interface, select "Extract grid corners", and then extract all the corner points of each image in turn, as showing in **Figure 2**.

4) After all the corner points of the 17 images are extracted, select "Calibration"

in the toolbox to complete the internal parameter calibration of the single camera, and the initial result and the optimized result of the calibration of the parameters in the camera will be obtained.

5) The reproduction error of the internal parameters can be obtained by selecting "Reproject on images" in the toolbox, as showed in **Figure 3**. The re-projection error map has the same color cross as the pixel error of the same calibration plate image. The more the cross is concentrated near the origin of the coordinates (0, 0), the smaller the error, and the greater the error.

6) After the left and right cameras are respectively calibrated, the dual-target setting toolbox is enabled, the internal parameters of the left and right cameras are optimized, and the binocular camera is calibrated, and finally the rotation vector R and the translation vector T are obtained. The positional relationship between binocular cameras is shown in **Figure 4**.
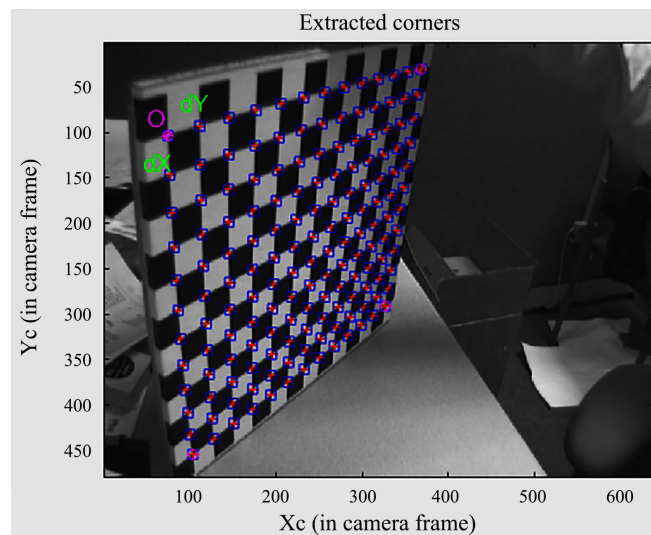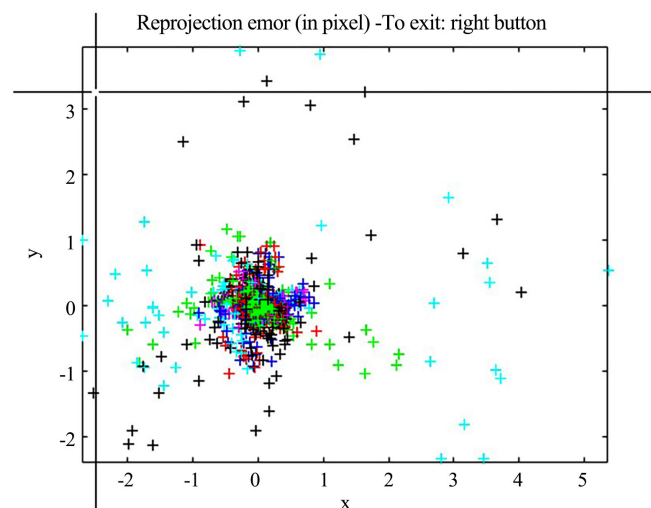

Figure 2. Extract corner points.

Figure 3. Reprojection error map of internal parameters.
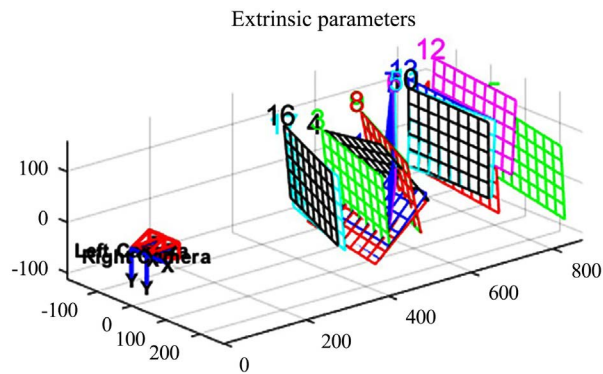
Extrinsic parameters

Figure 4. Position diagram between binocular cameras.

3. Image Enhancement

In the process of image acquisition, due to various factors such as noise interference, unstable light source, sensitivity of the sensor and data transmission medium, the captured image cannot achieve satisfactory results and will affect the image acquisition process. Therefore, image preprocessing is usually required first. The most important technology in image preprocessing is image enhancement. Its main purpose is to eliminate irrelevant information in the image and enhance the recovery of useful and relevant information. Thereby improving feature extraction, image matching, three-dimensional reconstruction accuracy and reliability.

Common image enhancement processing techniques include gray scale transformation, image sharpening, histogram equalization and normalization, radiation calibration. Fourier transform, and color enhancement. Although image enhancement processing is various, image enhancement is closely linked to people's visual habits and ways of observing objects of interest. Therefore, it is highly targeted, and no image enhancement algorithm can be applied on all occasions. In this section, the gray scale image processing technology is used to directly adjust the gray value of the acquired image to increase the contrast of the image, thereby achieving the purpose of image enhancement.

The gray scale enhancement steps are as follows:

1) by graying out the collected color pictures in Matlab and obtaining the main concentrated area of the gray value: AB (where: $A \geq 0, B \leq 255$), if the gray value is not evenly distributed between 0 - 255. If the image does not reach the sharpest state, the gray scale image can be made sharper by adjusting the gray scale value.

2) Set the gray value smaller than A to 0, and the gray value larger than B to 255. Set the gray value of the original image of *x*, and the gray value of the enhanced image of *y*, then:

$$\frac{x-A}{B-x} = \frac{y-0}{255-y} \tag{5}$$

Simplified to:

$$y = \frac{255(x-A)}{B-A} \tag{6}$$

3) Finally, through Matlab, the gray value can be effectively adjusted, and the image enhancement effect is obviously achieved.

4. Stereo Matching

In binocular stereo vision, the principle of stereo matching is to perform matching search on the corrected standard outer pole line along the left and right stereo images according to the selected matching primitives, and then calculate the matching obtained according to the corresponding relationship. The parallax of the image matching point is obtained, and the disparity map is obtained, thereby completing the work of stereoscopic image of three-dimensional reconstruction.

In binocular stereo vision, the stereo matching algorithm is crucial for the complete stereoscopic system to accurately and efficiently complete 3D reconstruction. Common stereo matching algorithms mainly include feature-based matching algorithms and region-based matching algorithms.

In this paper, the NCC (normalization cross-compilation) matching algorithm is used to complete the matching of left and right image pairs. It is a block matching algorithm that compares the original pixels, compensates the average gray value and its covariance in the sub-window, and has a short calculation time and can quickly find the optimal match. The implementation steps are as follows: after extracting two frames of images to be matched, correct them so that their optical centers are on the same horizontal line (polar line level), which can reduce the consumption of more computing resources. A given point on an image is the pixel position to be matched, and a 3 × 3 neighborhood matching window is constructed at this position, and the target pixel position is searched according to a certain similarity criterion along the polar line where the point is located. Find the subgraph that is most similar to the subwindow image. The similarity S of the two images is:

$$S(p,d)$$
$$= \frac{\sum\limits_{(x,y)\in W_p} \left[I_1(x,y) - \overline{I_1}(p_x,p_y) \right] \cdot \left[I_2(x+d,y) - \overline{I_2}(p_x+d,p_y) \right]}{\sqrt{\sum\limits_{(x,y)\in W_P} \left[I_1(x,y) - \overline{I_1}(p_x,p_y) \right]^2 \cdot \sum\limits_{(x,y)\in W_p} \left[I_2(x+d,y) - \overline{I_2}(p_x+d,p_y) \right]^2}} \quad (7)$$

wherein, W_p represents a 3 × 3 matching window centered on the coordinates of the pixel to be matched; I_1 represents a pixel value of a pixel position in the matching window; $\overline{I_1}$ represents the mean value of all pixels of the matching window; I_2 is the same; $S(p,d) \in [-1,1]$.

The high value of the similarity S indicates that the two images have high similarity, so the correlation between the two images is high when $S = 1$; the two images are not correlated when $S = -1$. After the end of the matching, the parallax d, that is, the difference $d = x' - x$ between the horizontal direction x of the pixel to be measured and the horizontal direction of the matching pixel is recorded, and finally a disparity map D having the same size as the original image is obtained.

5. Experimental Result

The calibration experiment was carried out using 17 chessboard pictures taken by the left and right cameras. The internal and external parameters of the camera are shown in Table 1.

The image is acquired by the calibrated binocular camera on the same scene, and the left and right stereo image pairs are obtained, as showed in Figure 5 and Figure 6.

In this paper, under the experimental environment of Matlab, the gray scale image value and gray scale enhancement program are written, and the image enhancement experiment is carried out in Figure 7. Depending on the second section of this paper, the value of A is 10 and the value of B is 230. The gray value is mostly concentrated in the area of 10 - 230. The image enhancement can be carried out in Matlab by formula (6) (Figure 8). After completing the enhancement on the left, do the same thing for the image on the right.

After the image enhancement is completed, the experiment is based on Matlab software, and the binocular visual stereo matching program is programmed, and the acquired image is stereo-matched. The image is obtained by the calibrated binocular camera on the same scene, and the left and right stereo image pairs are obtained, and then the stereo matching experiment is performed. The result is shown in Figure 9.

Figure 5. Left view.
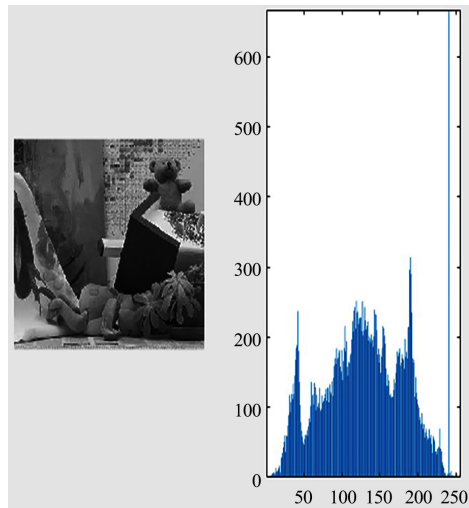

Figure 6. Right view.
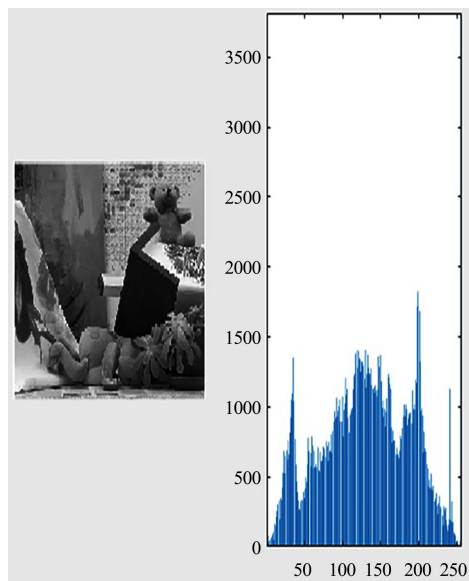
Figure 7. Gray valued area.
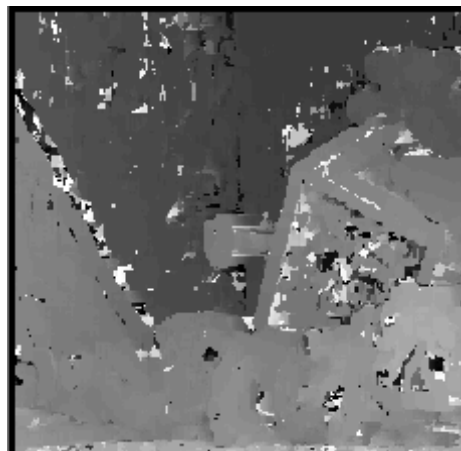

Figure 8. Enhanced gray value concentration area.

Figure 9. Disparity diagram D.

Table 1. Camera internal and external parameters.

Parameter	Left camera	Right camera
Focal length	[920.68679 927.82792] ± [8.61095 8.80952]	[919.82145 928.83736] ± [8.31756 8.71341]
Main point	[300.16341 237.88936] ± [16.37088 11.21697]	[300.81203 241.95676] ± [11.44344 8.79339]
Distortion	[−0.38425 0.61066 −0.00118 −0.00816 0.00000] ± [0.04731 0.30424 0.00196 0.00278 0.00000]	[−0.48467 1.24990 −0.00240 −0.00006 0.00000] ± [0.04898 0.40751 0.00192 0.00180 0.00000]
Balance vector	[0.00025 0.00777 0.00019] ± [0.01131 0.01630 0.00072]	
Rotation vector	[−44.38603 −0.54727 0.59312] ± [0.31080 0.09119 1.74577]	

6. Conclusion

Founded on the principle of binocular stereo vision, this paper studies the stereo calibration technology, image enhancement technology and stereo matching technology. The Zhang Zhengyou method is used to calibrate the vision system to obtain the internal parameters and the relative pose of the two cameras. The image is enhanced by the gradation transformation enhancement. By analyzing the main concentrated distribution area of the image gray value and adjusting it, the experimental results show that the image has an obvious enhancement effect. Finally, based on the NCC algorithm, this paper performs stereo matching on the two images corresponding to the same scene, and obtains the disparity map.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Chen, X.H. and Yuan, W. (2017) Target Positioning Based on Binocular Stereo Vision. *Automation Technology and Application*, **36**, 102-105.

[2] Zhang, W. and Hua, Y.S. (2018) Research on Target Location Based on Binocular Stereo Vision. *Software Guide*, **17**, 198-201.

[3] Gao, R.X. and Wang, J.M. (2014) Target Recognition and Location Based on Binocular Vision. *Journal of Henan University of Technology* (*Natural Science Edition*), **33**, 495-500.

[4] Wang, P.Q. (2016) Target Positioning and Grabbing of Robotic Arm Based on Binocular Vision. Polytechnic University, Tianjin.

[5] Jin, S., Ma, Y. and Han, Y. (2017) Camera Calibration and Its Application of Binocular Stereo Vision Based on Artificial Neural Network. *International Congress on Image & Signal Processing*, Shanghai, 14-16 October 2017, 66-72.

[6] Zhao, M. and Han, C.L. (2017) Tong Y. Research on the Calibration Method of "Bi-Binocular" Stereo Vision System. *Optical Technique*, **43**, 385-393.

[7] Yu, S., *et al.* (2017) Encoded Light Image Active Feature Matching Approach in Binocular Stereo Vision. 2016 *International Forum on Strategic Technology*, Ulsan, 31 May-2 June 2017, 45-65.

[8] Xu, P. (2017) Preliminary Study on the Mechanical Arm Demagnetizing Device with Binocular Vision Positioning. South China University of Technology, Guangzhou.