

Application of Stereo Vision 3D Target Recognition Using Camera Calibration Algorithm

Guo-Shing Huang

Department of Electronic Engineering
National Chin-Yi University of Technology
Taichung, Taiwan
e-mail: hgs@ncut.edu.tw

Yu-Yong Tseng

Department of Electronic Engineering
National Chin-Yi University of Technology
Taichung, Taiwan
e-mail: happy62196227@yahoo.com.tw

Abstract—This paper proposes a method of 3D(three-dimensional) object recognition using binocular parallax. The calibration object is a 2D(two-dimensional) object and a calibration label is placed on its plane. The internal, external, and distortion parameters of the camera are collaborated by taking pictures of the calibration object in different angles. This research allows image distortions caused by the camera to be calibrated, allowing the image coordinate of the characteristic point to better match with its actual coordinate. After calibration, the depth, localization, and identification of stereo vision of the image developed by the left and right cameras are optimized; the coordinate, distance, and the angle of rotation can be precisely identified.

Keywords—pinhole imaging; internal parameter; external parameter; stereo vision.

I. INTRODUCTION

Computer vision can be divided into monocular and binocular visions. Monocular vision can only recognize a single plane of an object. Since monocular vision can only determine the 2D information of an object that is commonly used for image recognition and image tracking [1]; [2] discussed how to take an image of an object using multilocular approach; [3] proposed a vision algorithm theory that can be applied to binocular appliances using the visual distortions of the two images to produce a 3D image with depth; [4] combine both binocular vision system and controller for image servo that can be installed onto motile robots for environmental detections. Many robotic research has mainly used infrared [5] radiation, ultrasound, and laser detectors for distance detection. These methods, however, encompass some deficiencies, such as small angles of detection, high expenses, being vulnerable to distractions. Therefore, many research now utilizes visual detections [6][7][9]. Stereo vision system incorporates the coordinate of the characteristic point from both the images developed from the left and right camera. However, in the process of recognition, refraction and construction errors of the lens contribute to the distortion of 2D image. Therefore, camera calibration algorithm is necessary for calibration and increase the precision of the 2D image and 3D image recognition.

This paper is divided into six sections. The second section discusses the specification of webcam, stereo vision

and system construction. The third section describes camera models, the theory of pinhole imaging, and the internal and external parameters collected through calibration. The fourth section utilizes MATLAB (MATrix LABoratory) to calibrate camera, and the fifth section discusses the data collected following camera calibration. The last section concludes the results of this research.

II. SYSTEM ARCHITECTURE

The recording apparatus utilized in this research is the webcam produced by Logitech. The main specification is shown in Table I. Full HD supports high quality display, its resolution is 1920×1080; the resolution of the picture is maximized when the snapshot function is used. H.264 AVC (Advanced Video Coding), is an image compressing function. Carl Zeiss is a German company that produces optical system, industrial inspection, and medical instruments. Carl Zeiss optical lens transforms the four components design of the Tesser lens into a five components design. The lens focuses the light entering the sensor to prevent image distortion.

As shown in Fig. 1, every robot's platform servo differs slightly. The stereo vision platform here is compiled by CAD (Computer Aided Design) software and manufactured through laser excision technique. Acrylic sheet is the material used. As shown in Fig. 2, human-computer interface is compiled by Visual C#, while the visual system is processed by EmguCV. The system structure described in this paper is shown in Fig. 3. Images were retrieved from webcam and the resolution of the retrieved images is 640x480. Fig. 4 shows the flowchart for image retrieval.

TABLE I. CAMERA INTERNAL PARAMETERS

Resolution	Snapshots pixel	H.264 AVC compression	Photographic lens design
Full HD 1080p	15-Megapixel	Yes	Carl Zeiss optics



Figure 1. Stereo vision diagram.



Figure 2. Human-Computer Interface diagram.

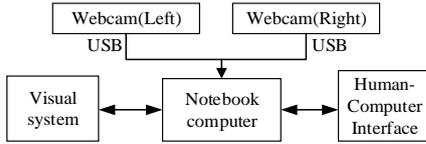


Figure 3. Stereo vision system architecture diagram.

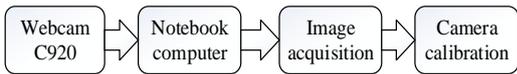


Figure 4. Image acquisition flow chart.

III. CAMERA CALIBRATION

Camera utilizes pinhole imaging theory for 2D image projection. The light that passes through the lens can reduce the exposure time, allowing the images to emerge in a shorter period of time. However, the lens can cause image distortion. As a result, calibration process is required to calibrate the image. The calibration process provides the internal and external parameters of the camera. The internal parameter includes matrix and phase distortion coefficient; the external parameter includes rotational matrix and translation matrix[8][14].

A. Internal Parameter

Matrix parameters include focal length, focal point, and phase distortion matrix. (1) can be determined by the model shown in Fig. 5 using similar triangles. From the model, it can be seen that a point in a space $Q(X, Y, Z)$ displayed in

a 2D coordinate, $q(x, y, f)$. Theoretically, the main point is the center of the imaging platform. However, due to the actual construction error, the center of the imaging sensor would not be located at the center. Therefore, adding C_x , C_y parameters can calibrate the center of the image. Moreover, the imaging platform utilizes pixels as its unit and pixels are in squares. The dimension pixels of the imaging sensor, however, cannot be 1:1, and thus, S_x , S_y parameters have to be added in for calibration as shown in (2). f and S_x , S_y all have the unit of mm, which can be derived into (3). Lastly, homogeneous coordinate can be used to transform the Q in 2D space into the q in 3D space, giving (4). In here, the w can be set as 1. If w is used as a distance parameter, then it has to be set as the reciprocal of the distance, if distance is not a parameter, then w is set as 0[15].

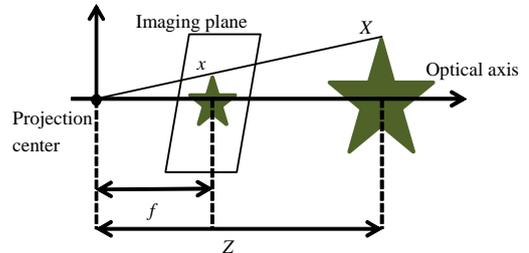


Figure 5. Pinhole imaging diagram.

$$x = f \frac{X}{Z}, y = f \frac{Y}{Z} \quad (1)$$

Where x and X express respectively image coordinate and world coordinate's axis of abscissa, the y and Y express respectively image coordinate and world coordinate's axis of ordinate, f represents x and the distance to the center of projection and Z represents X and the distance to the center of projection.

$$x = f \left[\frac{X}{Z} \right] \cdot S_x + C_x, y = f \left[\frac{Y}{Z} \right] \cdot S_y + C_y \quad (2)$$

Where S_x represents the scale factor of the axis of abscissa's on the image coordinate, S_y represents the scale factor of the axis of ordinate's on the image coordinate, C_x represents the deviation of the abscissa center on the image coordinate and C_y represents the deviation of the ordinate center on the image coordinate.

$$x = f_x \left[\frac{X}{Z} \right] + C_x, y = f_y \left[\frac{Y}{Z} \right] + C_y \quad (3)$$

Where f_x represents f and S_x are multiplied and f_y represents f and S_y are multiplied.

$$q = M \cdot Q. \quad (4)$$

$$q = \begin{bmatrix} x \\ y \\ w \end{bmatrix}, M = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, Q = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}.$$

Where q represents image coordinate displayed as a homogeneous coordinate, w represents distance parameter on the coordinate, M represents internal parameter expressed by homogeneous coordinate system and Q represents world coordinate displayed as a homogeneous coordinate.

Phase distortion coefficient is constructed by radial distortion and shear distortion, radial distortion is shaped by the lens shape, affecting the light's angle of refraction. The closer the light is to the center of the lens, the smaller the angle of refraction is. Shear distortion is caused by the error of camera construction. The main reason for this is the misalignment of the lens and sensor.

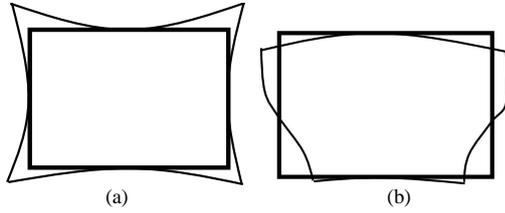


Figure 6. Image distortion. (a) radial distortion. (b) shear distortion.

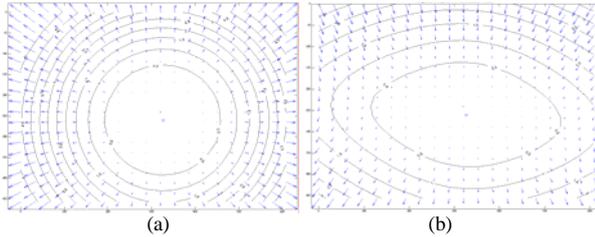


Figure 7. (a) The variable diagram of radial distortion of the lens. (b) The variable diagram of shear distortion of the lens.

Fig. 6 shows the two major distortions. Fig. 7(a) is the variable diagram of radial distortion of the lens. The directions of the arrow indicates the direction of the distortion. The length of the arrow indicates the magnitude of the distortion. Therefore, from the graph, it can be seen that the magnitude of the distortion is greater on the edges. Fig. 7(b) shows the variable diagram of shear distortion of the lens. The direction and the length of the arrow also indicates direction and magnitude of the distortion, respectively.

$$\begin{aligned} x' &= x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \\ y' &= y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \end{aligned} \quad (5)$$

Where x' represents radial distortion on platform: phase distortion coefficient of the axis of abscissa, y' represents radial distortion on platform: phase distortion coefficient of the axis of ordinate, k_1 , k_2 and k_3 represents radial

distortion coefficient and r represents the distance between the image coordinate and the world coordinate.

$$\begin{aligned} x'' &= x + [2p_1 y + p_2 (r^2 + 2x^2)] \\ y'' &= y + [p_1 (r^2 + 2y^2) + 2p_2 x] \end{aligned} \quad (6)$$

Where x'' represents shear distortion on platform: phase distortion coefficient of the axis of abscissa, y'' represents shear distortion on platform: phase distortion coefficient of the axis of ordinate, the p_1 and p_2 represents shear distortion coefficient.

Equation (5) is an equation for radial distortion, (6) is an equation for shear distortion. The k_1 , k_2 , k_3 , p_1 , and p_2 in the two equations represent phase distortion coefficients.

B. External Parameter

The external parameter is used to describe the relationship between the world and camera coordinates. The relationship is illustrated in Fig. 8. In order to produce meaningful coordinates from the image projection process, we used homogeneous coordinate system to describe the relationship. It can be represented by both the rotational and translation matrices[10][11].

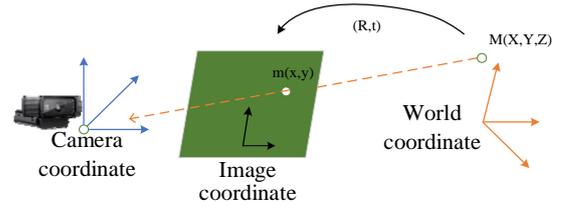


Figure 8. The world coordinate and the camera coordinate projection

Using the rotation matrix to align the rotational axis of the reference coordinate and the camera coordinate; the translation matrix is then utilized to shift the origin of the reference coordinate to the origin of the image coordinate. This can be represented in homogeneous coordinate as shown in (7).

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R_{3 \times 3} & T_{3 \times 1} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}. \quad (7)$$

Where X_c , Y_c and Z_c represents new coordinates on the world coordinates, $R_{3 \times 3}$ represents rotation matrix, $T_{3 \times 1}$ represents translation matrix, the X_w , Y_w and Z_w represents original coordinates on the world coordinates.

The location of the image platform can be derived from the external and internal parameters of the 3D coordinate in the space, as shown in (8).

$$S \begin{bmatrix} x_c \\ y_c \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}. \quad (8)$$

Where \tilde{p} represents homogeneous coordinate of the image plane, \tilde{P} represents homogeneous coordinate of the reference space, S represents scale factor, M_{int} represents the internal parameter's matrix of the camera and M_{ext} represents the external parameter's matrix of the camera.

The camera's internal and external parameter matrix can be defined as the project parameter shown in (9). (8) can also be organized into (10). After the left and right camera obtains the parameters, the depth of the image can be observed through stereo vision[12][13].

$$M_p = M_{int} M_{ext}. \quad (9)$$

$$S\tilde{p} = M_p \tilde{P}. \quad (10)$$

IV. SOFTWARE SIMULATION

Camera calibration MATLAB toolbox is a perfect camera calibration tool. It can effectively monitor the stereo vision's internal and external parameters proposed in this research. The 10×7 calibration plate is used in this research.

As shown in Fig. 9, the image of the calibration board was taken by the left and right camera. This method is then repeated for 20 times from different angles and distances. As shown in Fig. 10(a), the four corners of the calibration board were manually obtained. The positions of the corners of the squares were calculated using tools as shown in Fig. 10(b).



Figure 9. Stereo vision obtained black chessboard distortion.

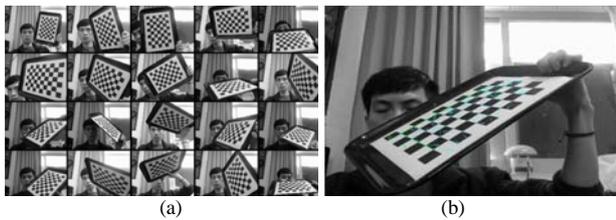


Figure 10. (a) Calibration image of camera. (b) Corner mark distortion.

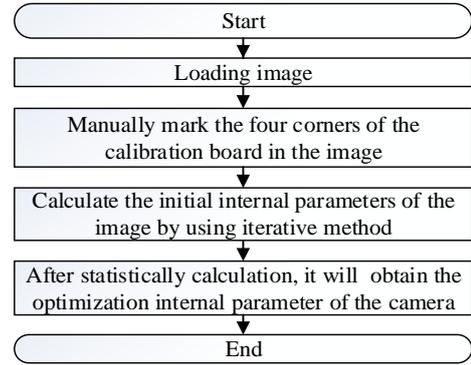


Figure 11. Calibration flow chart.

The calibration process is shown in Fig. 11. By fixing the four corners of the calibration board in the image and using the iterative method, the internal parameters of the left and right cameras can be calculated and the results of the individual camera calibration and the combined calibration can also be simulated in 3D space. The red squares in Fig. 12 are the 2D calibration plate. The green pyramid indicates the location and direction of the camera.

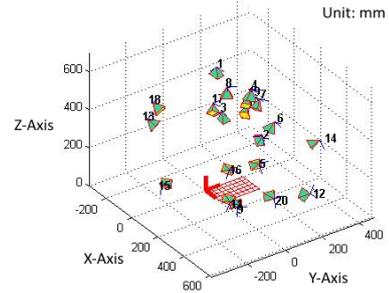


Figure 12. Diagram between camera and calibration image.

V. EXPERIMENTAL RESULTS

The camera's internal parameter like Table II. It can be observed that the focal length of the left and right camera are both approximately 612mm. The center coordinate in the 640×480 pixel image is theoretically (320, 240). However, the construction of the left and right cameras creates errors that can be considered into the calculation that is used to determine the relative positions of the characteristic point and the image's central point, increasing the precision of the left and right cameras identification. Among all phase distortion coefficients, k_1 , k_2 , and k_3 are radial distortion, p_1 and p_2 are shear distortion. The purpose of this research is to improve the camera's image distortion problem. The value of the distortion is less than 0.2mm. The pixel error is a critical factor that is considered into the recognition of the characteristic point. The error of the x and y axis are both within ± 0.18 pixel. The external parameter listed in Table III is based on the left camera. The radial matrix and the

translation matrix obtained from the right camera can be used to determine the 3D coordinate of the characteristic point in the world coordinate system after adjustment. r_1 , r_2 , and r_3 are the x, y and z axes' rotational value respectively, while t_1 , t_2 , and t_3 are the x, y and z axes' translational value respectively. Following calibration, the 3D coordinate of the characteristic point can be further confirmed by the left and right images.

TABLE II. THE LEFT AND RIGHT CAMERA INTERNAL PARAMETERS

Parameter	Left camera	Right camera
Focal length ^a	(612.624, 613.767)	(611.801, 613.167)
Center coordinates ^a	(325.858, 259.447)	(320.661, 248.231)
Phase distortion coefficient matrix ^b	[0.084 -0.122 0.007 -0.007 0.000]	[0.1 -0.234 0.009 -0.003 0.000]
Pixel error ^a	(0.173, 0.175)	(0.176, 0.171)

a. The x axis and y axis. b. The k1, k2, p1, p2 and k3.

TABLE III. STEREO VISION EXTERNAL PARAMETER

Parameter	Position of right camera with respect to left camera
Rotation matrix ^c	[-0.001 -0.0002 0.004]
Translation matrix ^c	[-93.212 -5.847 -4.568]

c. The x axis, y axis and z axis.

VI. CONCLUSIONS

This research determines the errors of the focal length, phase distortion coefficients, and pixel error. The error of focal length is less than 2mm, indicating that this stereo vision excels in identifying the distance of the characteristic point. The error of phase distortion coefficients are less than 0.2mm, showing that the camera itself encompass errors, that following calibration, can produce less distorted 2D images. The left and the right cameras both encompass pixel errors less than 0.18 pixel. This allows the camera to precisely measure the coordinate of the characteristic point and at simultaneous increase the precision of the stereo vision recognition. The image coordinates can be used in coordination with 3D reconstruction principle to obtain the 3D coordinates of the two access points on the left and right of the object, such that the pair of mechanical arms of order taking robot can more precisely handle trays. This paper aims to discuss 2D image calibration, which differs than 3D objects, 1D(one-dimensional) object and self-calibrated objects in the fact that 2D image calibration only requires the printout of the calibration object that is placed onto a non-distorted platform without knowing the rotation and transition paths. The calibration process can minimize the image identification error and promote the precision of the object's platform coordinate, distance, and angle of rotation, achieving the robustness of system, increasing the stability and the reliability of the results.

ACKNOWLEDGMENT

This work was supported by the National Science Council of Taiwan, R.O.C. under grant NSC 101-2221-E-167-020-MY2.

REFERENCES

- [1] W. S. Fife, and J. K. Archibald, "Improved Census Transforms for Resource-Optimized Stereo Vision," IEEE Circuits and Systems for Video Technology, 2013, pp. 60–73.
- [2] X. Fan, X. Wang, and Y. Xiao, "A shape-based stereo matching algorithm for binocular vision," IEEE Security, Pattern Analysis, and Cybernetics, 2014, pp. 70–74.
- [3] L. Zou, and Y. Li, "A method of stereo vision matching based on OpenCV," IEEE Audio Language and Image Processing, 2010, pp. 185–190.
- [4] P. Ben-Tzvi, and X. X. An, "Embedded feature-based stereo vision system for autonomous mobile robots," IEEE Robotic and Sensors Environments, 2010, pp. 1–6.
- [5] J. W. Starr, and B. Y. Lattimer, "Application of thermal infrared stereo vision in fire environments," IEEE Advanced Intelligent Mechatronics, 2013, pp. 1675–1680.
- [6] L. W. Zheng, Y. H. Chang, and Z. Z. LiA, "Study of 3D feature tracking and localization using a stereo vision system." IEEE Computer Symposium, 2010, pp. 402–407.
- [7] D. Gardeazabal, V. Ponomaryov, and I. Chairez, "Fuzzy control for obstacle avoiding in mobile robots using stereo vision algorithms," IEEE Electrical Engineering Computing Science and Automatic Control, 2011, pp. 1–6.
- [8] D. Sutton, and R. Green, "Evaluation of real time stereo vision system using web cameras," IEEE Image and Vision Computing New Zealand, 2010, pp. 1–10.
- [9] M. Samadi, M. F. Othman, and S. H. M. Amin, "Stereo vision based robots: Fast and robust obstacle detection method," IEEE Control Conference, 2013, pp. 1–5.
- [10] A. Woodward, D. Berry, and J. Dunning, "Real-time stereo vision on the Vision Server Framework for robot guidance," IEEE Image and Vision Computing New Zealand, 2010, pp. 1–8.
- [11] K. Al-Mutib, M. Emaduddin, M. Alsulaiman, H. Ramdane, and E. Mattar, "Motion periodicity based pedestrian detection and particle filter based pedestrian tracking using stereo vision camera," IEEE Mechatronics and Machine Vision in Practice, 2012, pp. 32–37.
- [12] R. A. Hamzah, K. A. A. Aziz, and A. S. M. Shokri, "A pixel to pixel correspondence and region of interest in stereo vision application," IEEE Computers & Informatics, 2012, pp. 193–197.
- [13] M. Chen, Z. Cai, and Y. Wang, "A method for mobile robot obstacle avoidance based on stereo vision," IEEE Industrial Informatics, 2012, pp. 94–98.
- [14] Y. Xu, Y. Zhao, F. Wu, and K. Yang, "Error analysis of calibration parameters estimation for binocular stereo vision system," IEEE Imaging Systems and Techniques, 2013, pp. 317–320.
- [15] R. A. Hamzah, A. M. A. Hamid, and S. I. M. Salim, "The solution of stereo correspondence problem using block matching algorithm in stereo vision mobile robot," IEEE Computer Research and Development, 2010, pp. 733–737.