An Integrated Camera Parameters Calibration Approach for Robotic Monocular Vision Guidance

Xufeng Liang
School of Automation
Wuhan University of Technology
Wuhan, China
265655@whut.edu.cn

Yuhao Du*
School of Automation
Wuhan University of Technology
Wuhan, China
265599@whut.edu.cn

Dongyu Wei
School of Automation
Wuhan University of Technology
Wuhan, China

***Abstract*—A core problem in robotics is to determine the position and orientation of an industrial robot in its working environment. The basic principle of vision-based robotic positioning technology relies on optical sensors that can provide tremendous amount of data. Monocular vision guidance of the robot requires firstly computing the intrinsic and extrinsic camera parameters from a set of physical features. In this paper, a camera calibration method combining Zhang Zhengyou's calibration method and PnP solution together is introduced to establish the three dimensional position and orientation of camera. Compared with other methods, this method does not require too much experimental data, it only requires the aid of a checkerboard and a white paper with several feature points. Moreover, the obtained camera parameters have high precision that can meet the requirements of a general robot monocular vision guiding system.**

***Keywords*—camera calibration; robot positioning method; PnP solution; robotic vision.**

I. Introduction

In recent years, the developing industrial requirements made robotic applications attract wide public concern. Extrinsic information coming from a variety of sensors helps the robot to perceive environments, of which the extraction of proper image features for positioning is a key calibration work. The robot vision system combined with the camera obtains a two-dimensional image of the surrounding environment. After the image are processed, the robot control unit can obtain the position and posture of the target, and then uses the feedback method to control the posture of robot, which is so called vision control [1]. The workflow of the robot vision guidance system consists of two basic modules: the first stage is to achieve the camera parameter calibration of the robot vision system to obtain the spatial mapping relationship between the two-dimensional pixel points and the three-dimensional object points. The second stage lies in the edge detection and recognition of the target, and to obtain accurate three-dimensional coordinate information of the target, thereby guiding the robot to perform predetermined actions such as grabbing or product assembling [2].

In the process of control and guidance, we need to convert the image information obtained by the camera into the pose information that can be performed by the robot system. Camera calibration is the process of solving a series of parameters such as camera intrinsic parameters and extrinsic parameters. Meanwhile, depending on the number of cameras, current vision guidance can be divided into monocular vision guidance, binocular vision guidance, and multi-eye vision guidance etc.

Many prior contributions dealing with the similar research topic are concerned in this work: Artur Borges Pio proposed an online camera calibration method based in the readings of an IMU and in the track of a single 3D point projected into the camera image frames [3]. Z. Zhang proposed a flexible technique which only requires the camera to observe a planar pattern shown at a few (at least two) different orientations [4]. Yang Guo proposed a solution for camera pose and camera calibration based on PnP problem [5]. Among them, Zhang's calibration method is used to solve the intrinsic parameters of the camera, besides the PnP solution in OpenCV is used to solve the extrinsic parameters of the camera.

This paper mainly focuses on a monocular camera calibration methodology which combined Zhang's calibration method and PnP solution together. Based on pin-hole camera with perspective projection model, the intrinsic and extrinsic camera parameters are sequentially calculated through the combined algorithm, and then the relationship between camera coordinates and the world coordinates is represented. At last, relative calibration precision is also discussed in this work.

II. MONOCULAR CAMERA IMAGING MODEL

The imaging model of the monocular camera can be simplified to a pinhole imaging model involving conversion of the camera coordinate, the image coordinate, and the pixel coordinate [6]. As can be seen from Fig.1, the camera coordinate is centered on the camera, while the image coordinate is a two-dimensional imaging plane that contains the detection array. Moreover, it is generally considered that the pixel coordinate and the image coordinate are in the same two-dimensional plane, despite to the difference between their coordinate origin and the coordinate axes. Unlike the camera coordinate and image coordinate, the pixel coordinate is modeled and analyzed in pixels.

978-1-7281-3936-4/19/$31.00 ©2019 IEEE

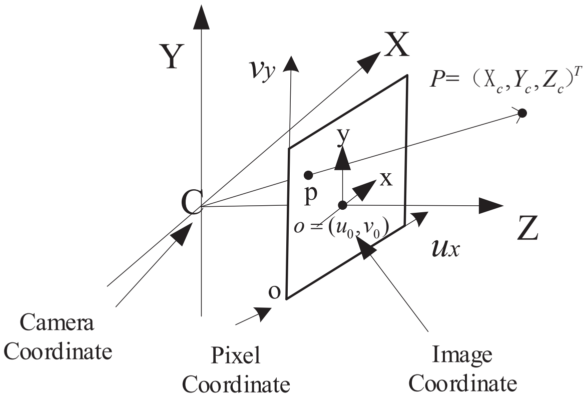


Fig. 1: Monocular camera imaging model.

As shown in Fig.1, point C is the origin of the camera coordinate, the vertical line from C to the image plane is called the main axis, and the intersection point between the main axis and the image plane is called the principal point o, that is, the origin of the image coordinate. And in the image plane, $Z = f$ (f is the focal length of the camera). Under the pinhole imaging model, the point $P_C=(X_C, Y_C, Z_C)^T$ in the camera coordinate system is mapped onto the image coordinate point p, which is also the intersection of the imaging plane and the line connecting the point P_C and the projection center C. According to the principle of similar triangles, the mathematical relationship between the point P_C in the camera coordinate and the point p mapped onto the imaging plane is expressed as follows:

$$\begin{cases} x = f\frac{X_C}{Z_C} \\ y = f\frac{Y_C}{Z_C} \end{cases} \tag{1}$$

Eq. 1 is perspective projection equation that can be employed to describe the relationship between a point in a three-dimensional space and an imaged point.

Since the image obtained by the camera is a discrete array of gray values, the pixel coordinates (u, v) can be used to describe the position of a pixel, where u, v represent the number of rows and columns of the pixel matrix, respectively. Based on it, the relationship between the imaging point (x, y) and the pixel point (u, v) are discussed.

The photosensitive cells in the camera sensor are typically 1-10 µm rectangles, the ratio of the number of horizontal and vertical pixels is called the aspect ratio, which is typically 4:3 or 16:9. The sensor is the same size as the display and measures about 1/3 inch (8 mm) in length according to the diagonal. The image plane information is amplified to obtain a digital image, which is, converted from an imaged point (x, y) to a pixel point (u, v). Assuming that the physical size of each pixel is $dx*dy$ mm, the origin coordinates of the image coordinate in the pixel coordinate are (u_0, v_0), then the relationship between image coordinates (x, y) and pixel coordinates (u, v) can be as shown as follows:

$$\begin{cases} u = \frac{x}{dx} + u_0 \\ v = \frac{y}{dy} + v_0 \end{cases} \tag{2}$$

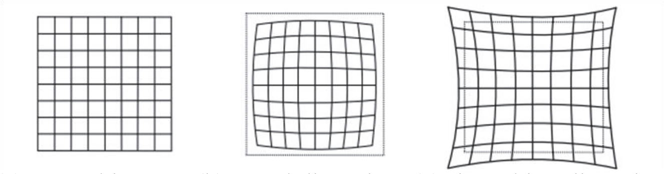
The combination of Eq. 1 and Eq. 2 can solve the problem of the transformation relationship from pixel coordinate to camera space coordinate, which is a critical equation for solving the intrinsic parameters of the camera.

III. CAMERA INTRINSIC PARAMETERS

Supposing the intrinsic parameters and the extrinsic parameters of camera need to be calibrated, the existing camera calibration methods is then divided into the following three categories: the traditional camera calibration method and the camera self-calibration method and the active vision-based calibration method.

The traditional camera calibration method is based on the experimental conditions and the reference objects with known shape and size (such as the checkerboard), and the mathematical parameters and calculations of the image are used to obtain the parameters of the camera (such as the Zhang's calibration method). The camera self-calibration method utilizes the constraints inherent in the intrinsic parameters, with high flexibility but low accuracy. The calibration method based on active vision is simple and has high robustness in algorithm, usually used to obtain linear solution, but its disadvantages are the high experimental cost and the expensive equipment of calibration. Zhang's calibration method overcomes the shortcoming of the high-precision calibration required for active vision-based calibration method and requires only one printed checkerboard. Meanwhile, this method also improves accuracy and operability compared to self-calibration method. Therefore, Zhang's calibration method is widely used in computer vision. Combining the advantages and disadvantages of the above three calibration methods, Zhang's calibration method is employed to obtain the intrinsic parameters of camera. The intrinsic parameters is obtained by the transformation relationship between pixel coordinate, image coordinate, and camera coordinate.

The calibration work requires photographing the calibration plate using a camera to obtain several (at least two) planar patterns displayed from different directions. Then, the homography matrix H between the image and the plane is obtained. Because of the existence of Gaussian noise, a closed-form solution and a nonlinear improvement method based on the maximum likelihood criterion are used in the Zhang's calibration method. In the picture processing process, the elimination of distortion is also necessary. The Zhang's calibration method takes into account the large radial distortion. The radial distortion can be divided into barrel distortion and pincushion distortion, as shown in the Fig.2.



(a)Normal image (b)Barrel distortion (c)Pincushion distortion

Fig. 2: Barrel distortion and pincushion distortion in radial distortion.

The mathematical expression of radial distortion is expressed as follows:

$$\begin{cases} x' - x = x + x[k_1(x^2 + y^2) + k_2(x^2 + y^2)^2] \\ y' - y = y + y[k_1(x^2 + y^2) + k_2(x^2 + y^2)^2] \end{cases} \quad (3)$$

$$\begin{cases} u' - u = (u - u_0)[k_1(x^2 + y^2) + k_2(x^2 + y^2)^2] \\ v' - v = (v - v_0)[k_1(x^2 + y^2) + k_2(x^2 + y^2)^2] \end{cases} \quad (4)$$

Where (x', y') is the actual distorted image coordinates, and (x, y) is the ideal undistorted image coordinates. Correspondingly, (u', v') and (u, v) are the actual diameter and the ideal pixel coordinates. And k_1 and k_2 are the first two orders of radial distortion parameters.

Referring to Zhang's calibration method and the relationship between the camera coordinate and the image coordinate in the monocular camera imaging model, the transformation relationship from the camera coordinate to the pixel coordinate is expressed as follows:

$$Z_C \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} 1/dx & 0 & u_0 \\ 0 & 1/dy & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} \quad (5)$$

After multiplying the matrix at the right of the equal sign of Eq. 5, it can be expressed as follows:

$$Z_C \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f/dx & 0 & u_0 & 0 \\ 0 & f/dy & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} \quad (6)$$

$$K = \begin{bmatrix} f/dx & 0 & u_0 & 0 \\ 0 & f/dy & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (7)$$

In Eq. 7, the matrix K is the intrinsic matrix of the camera that needs to be solved. Where f_x, f_y, u_0 and v_0 are the camera parameters to be sought. Furthermore, this paper will explain how to complete the acquisition of these parameters in the experiment.

During the process of calibration, a printed checkerboard calibration paper is used. First, a Basler grayscale camera is employed to take multiple photos of the checkerboard from different positions, different angles and different postures. Then, from the resulting series of images, 9 images are selected for calibration, as shown in Fig. 3.

The following work is to extract the base coordinates (based on the camera coordinate) and the pixel coordinates of the inner corner points (the corner points refers to the fixed portions of the squares that are black and white; the inner corner points are the internal corner points that are not in contact with the edge of the calibration checkerboard). The pixel coordinates of the inner corner points can be obtained by the library function in OpenCV. In order to ensure higher precision, the corresponding function can be used in the experiment to find the sub-pixel corner points and get their sub-pixel coordinates. The results obtained by this method are accurate enough and the operation is relatively simple.

Solving the intrinsic matrix is the last step. Due to its strong applicability, Zhang's calibration method has been packaged into some functions, tools and software for people to use, such as library functions in OpenCV and toolboxes in Matlab. The calibration method used in this paper relies on the library functions for camera calibration in OpenCV. In addition to the base coordinates and pixel coordinates of the inner corner points and camera distortion parameters obtained in the previous work, the camera intrinsic matrix can be obtained. The originally unknown parameters in the intrinsic matrix K are as follows:

$$\begin{cases} f_x = 1719.4898611106612 \\ f_y = 1718.8454248032479 \\ u_0 = 629.36488561709007 \\ v_0 = 503.91941749538387 \end{cases}$$

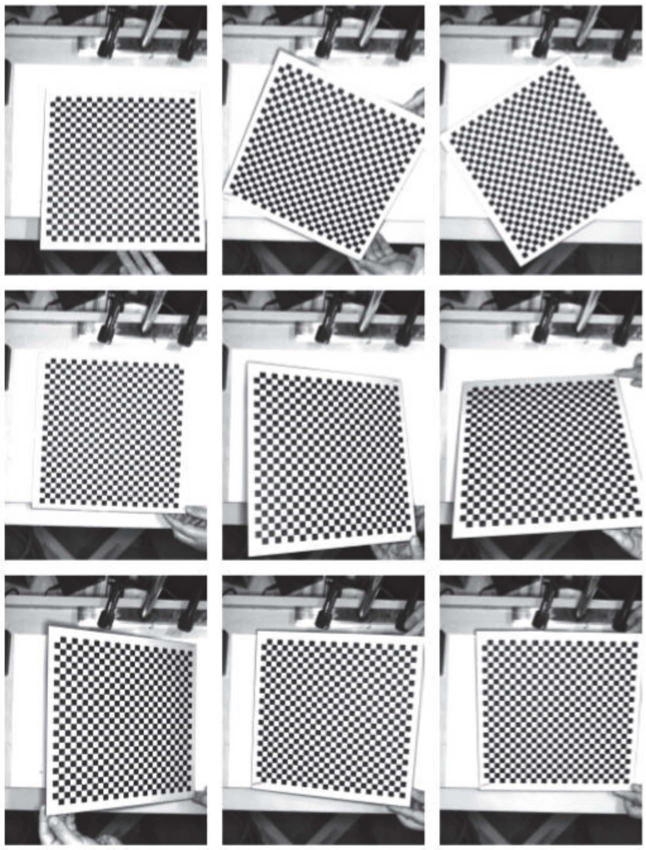


Fig. 3: Photographs of the checkerboard in the experiment.

IV. CAMERA EXTRINSIC PARAMETERS

Through the solution of the camera's intrinsic parameters, we can get the transformation relationship from the pixel coordinate of the image to the camera coordinate. However, applying vision guidance to the robot requires converting the resulting position of the camera coordinate to the position of the robot-based world coordinates. Therefore, the solution of the pose transformation relationship between the camera coordinate and the world coordinate is the problem need to be solved further.

Based on the previous vision guidance analysis, position-based vision control is selected in this work, as shown in Fig. 4. The world coordinate can be transformed into the pose of the camera coordinate by translation in space and rotation around three axes (X_W, Y_W, and Z_W).

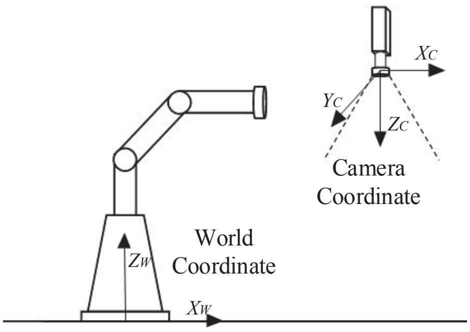


Fig. 4: Schematic diagram of the pose of camera and robot.

Therefore, the transformation from the world coordinate to the camera coordinate can be expressed as a mathematical relationship in Eq. 8.

$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} = \begin{bmatrix} R_{3\times3} & T_{3\times1} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \tag{8}$$

$$M = \begin{bmatrix} R_{3\times3} & T_{3\times1} \\ 0 & 1 \end{bmatrix} \tag{9}$$

In Eq. 9, M is the extrinsic matrix, and contains a 3*3 rotation matrix R and a 3*1 translation vector T, which represent the rotation and translation relations from the world coordinate to the camera coordinate.

In the case of known intrinsic parameters and distortion parameters of the camera, the determination of the camera pose can be solved by the world coordinates and pixel coordinates of several known feature points. So the key problem is a PnP (Perspective-n-Point) problem. For the camera pose solving problem, the world coordinates of n feature points and their corresponding n pixel coordinates are needed. In PnP problem, When $n \geq 3$, it is possible to obtain finite effective solutions.

The P3P problem is the most basic problem in the PnP problem. P3P problems are commonly solved by algebraic methods, that is, using the geometric relationship between the given three points [7, 8]. As shown in Fig. 5, there are three-dimensional coordinate points A, B, C and corresponding two-dimensional coordinate points a, b, c.

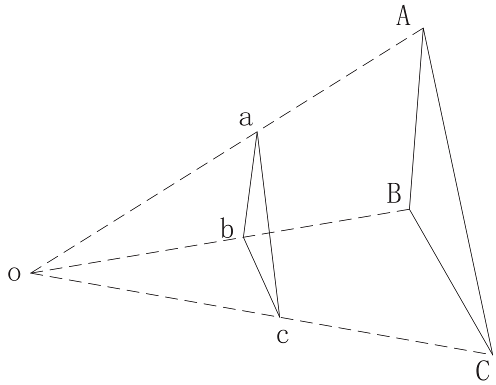


Fig. 5: P3P problem geometry model.

According to the triangle similarity principle, the following relationship can be obtained:

$$\Delta Oab \cong \Delta OAB \Leftrightarrow \Delta Obc \cong \Delta oBC \Leftrightarrow \Delta Oac \cong \Delta OAC$$

According to the cosine theorem, there are the following relations in the triangle:

$$\begin{cases} OA^2 + OB^2 - 2OA * OB\, cos(a,b) = AB^2 \\ OB^2 + OC^2 - 2OB * OC\, cos(b,c) = BC^2 \\ OA^2 + OC^2 - 2OA * OC\, cos(a,c) = AC^2 \end{cases} \tag{10}$$

Divide both sides of the (10) by OC^2 and let $x=OA/OC$ and $y=OB/OC$:

$$\begin{cases} x^2 + y^2 - 2xy\, cos(a,b) = AB^2/OC^2 \\ y^2 + 1^2 - 2y\, cos(b,c) = BC^2/OC^2 \\ x^2 + 1^2 - 2x\, cos(a,c) = AC^2/OC^2 \end{cases} \tag{11}$$

Let $v=AB^2/OC^2$, $uv=BC^2/OC^2$, $wv=OC^2/OC^2$, then:

$$\begin{cases} x^2 + y^2 - 2xy\, cos(a,b) - v = 0 \\ y^2 + 1^2 - 2y\, cos(b,c) - uv = 0 \\ x^2 + 1^2 - 2x\, cos(a,c) - wv = 0 \end{cases} \tag{12}$$

After transformed, it can be obtained as follows:

$$\begin{cases} (1-u)y^2 - ux^2 - y\cos(b,c) + 2uxy\cos(a,b) + 1 = 0 \\ (1-w)x^2 - wy^2 - x\cos(a,c) + 2wxy\cos(a,b) + 1 = 0 \end{cases} \tag{13}$$

The coordinate positions in the two-dimensional image are known, so the three cosine values are also known. Meanwhile, since u, v, and w are calculated by the world coordinates of A, B, and C, the values of u, v, and w do not change after changing to the camera coordinate system. In Eq.13, the values of x and y are to be solved, Eq. 13 is a polynomial equation for x and y.

Based on the solution of the P3P problem, this paper completed the calibration of the extrinsic parameters. As shown in Fig. 6, 5 points (P1, P2, P3, P4 and P5) are selected on a white paper. Among them, P1~P4 are used as four feature points to obtain the pose solution in the PnP problem, and P5 is used to calculate the reprojection error and verify the correctness of the obtained pose solution.

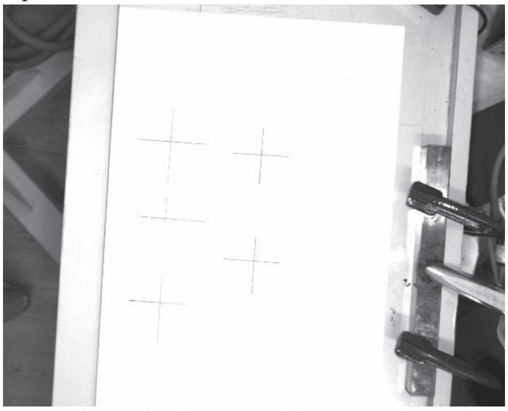


Fig. 6: Schematic diagram of feature points selection.

In this paper, the PnP problem of camera pose can be solved by the library function solvePnP() in OpenCV. The final parameter flags of this function have three options in OpenCV: CV_P3P, CV_ITERATIVE, and CV_EPNP. For the case of inputting four feature points in this experiment, all three methods are applicable. After solving them separately by three methods, the reprojection verification is performed with P5 point, and finally the result obtained by the CV_P3P method with the smallest error is selected.

The resulting extrinsic matrix M is:

$$\begin{bmatrix} -0.9979 & 0.0083 & -0.0646 & -663.283290570084 \\ 0.0496 & 0.7390 & -0.6719 & -98.9599884828214 \\ 0.0422 & -0.6736 & -0.7379 & 1020.03798386183 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

According to the result of program, the calculated rotation vectors can be converted into a rotation matrix and Euler angles. Firstly, let the camera translate according to the calculated translation vector from the position of the world coordinate origin, and then let the camera rotate according to the calculated rotation Euler angles around the three axes of the camera coordinate, and finally the camera can be transformed to the actual pose.

At this point, the camera calibration work is completed. The results obtained from the calibration work can be applied to the vision guidance of the robot. By combining the intrinsic parameters and the extrinsic parameters, the conversion relationship between the worlds coordinates of the robot and the pixel coordinates of the image can be obtained as follows:

$$\mathrm{s}\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K * M * \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \tag{14}$$

In Eq. 14, s is the zoom factor between the physical resolution and the logical resolution of the camera imaging.

In the vision guidance of the robot, the focus is about how to obtain the position of the center point of the target object or the position of the entire target object bounding box. After the camera is taken, the pixel coordinates of the desired image point can be obtained by the program. The process of converting pixel coordinates into world coordinates can be expressed as follows:

$$\begin{bmatrix} X_W \\ Y_W \\ Z_W \end{bmatrix} = R_{3\times3}^{-1}\left(K^{-1}\mathrm{s}\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} - T_{3\times1}\right) \tag{15}$$

V. CONCLUSION

This article describes the process of Zhang's calibration method and PnP solution to complete the monocular camera calibration. Method of the intrinsic parameters and the extrinsic parameters of the camera in the form of mathematical formulas is deduced. Using the checkerboard and feature points, camera parameters with relatively high precision can be obtained. However, only the influence of the radial distortion in the Zhang's calibration method is discussed in this work, and the tangential distortion due to the non-parallel of the lens and the imaging plane will be considered in the future work. The following work is to work on the sensing schemes, the representations of environment and positioning technology.

ACKNOWLEDGMENT

The work is supported by National Natural Science Foundation of China (No. 61603281), the Fundamental Research Funds for the Central Universities (WUT: 2017IVB038), the Open Funds of Key Lab of Digital Signal and Image Processing of Guangdong Province (No. 2017GDDSIPL-04) and the WUT Start-Up Grant (40120173).

REFERENCES

[1] K. S. Fu, R. C. Gonzalez, C. S. G. Lee, "Robotics," New York: McGraw-Hill, 1987.

[2] Z. Chen, H. Zou, Y. Wang, B. Liang, and L. Yu, "A vision-based robotic grasping system using deep learning for garbage sorting," in Control Conference, 2017.

[3] A. B. Pio and G. A. Borges, "Online Inertial-Aided Monocular Camera Self-Calibration," in 2018 Latin American Robotic Symposium, 2018 Brazilian Symposium on Robotics (SBR) and 2018 Workshop on Robotics in Education (WRE), 2018, pp. 100-105.

[4] Zhang Z. 2000. A flexible new technique for camera calibration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22; 22(11; 11): 1330-1334.

[5] Y. Guo, "A Novel Solution to the P4P Problem for an Uncalibrated Camera," Journal of Mathematical Imaging and Vision, vol. 45, pp. 186-198, 2013-01-01 2013.

[6] J. Tian, Molecular imaging. Hangzhou : New York: Zhejiang University Press, Springer, 2012.

[7] G. Xiao-Shan, H. Xiao-Rong, T. Jianliang, and C. Hang-Fei, "Complete solution classification for the perspective-three-point problem," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, pp. 930-943, 2003-01-01 2003.

[8] T. Wang, "Determining model pose based on body of revolution," Chinese Journal of Computers, vol. 31, pp. 493-501, 2006-01-01 2006.