Check for updates

3D Pose Estimation of Manipulator Based on Multi View

Xuechun Geng and Cheng Cai(✉)

Shanghai DianJi University, Shanghai 200240, China
caic@sdju.edu.cn

Abstract. 3D pose estimation plays an important role in human-computer cooperation and intelligent control. At present, the manipulator with sensor can realize spatial three-dimensional pose estimation, but the cost is too high. In this paper, the three-dimensional spatial positioning of the manipulator is carried out through the visual label. The two-dimensional pixel coordinates of the visual label on the manipulator are obtained by pasting the visual label on the manipulator and placing three cameras in different directions. By solving the three-dimensional coordinates of the manipulator through the three-dimensional vision fusion proposed in this paper, combined with the least square method to optimize the measurement results, multiple constraint equations can be established through multi-objective vision to improve the measurement accuracy, expand the motion range of the measured object, and have high robustness and economy. In the binocular vision pose detection, 89% of the measurement error within 2 m is less than 10 mm. The detection speed of the robot pose measurement system based on multi vision reaches 16 fps, and more than 96% of the measurement error within 2 m does not exceed 3 mm. Considering the needs of real-time monitoring of industrial detection in the production environment, it needs to have high robustness, and the multi vision pose detection scheme is recommended.

Keywords: Multi-vision · Mechanical arm · Visual label · 3D pose estimation

1 Introduction

At present, the multi-objective vision pose estimation based on visual label [1] has been applied in many fields. Most of the mechanical arms on the market now contain various sensors because of the limited perception of external signals. Radar, laser and vision sensor are commonly used sensors. Traditional pose measurement needs vision sensor to obtain the information of the object to be measured. Its accuracy is mainly affected by the resolution and the distance of the object to be measured [2]. When the resolution is low or the distance of the object to be measured is long, the economy and accuracy of vision sensor will be reduced. In computer vision, the image information of the object can be obtained by placing cameras with different angles and different positions. Based on the multi-objective fusion technology [3], the spatial position and posture estimation is realized. Compared with the visual sensor, the camera has higher economy. The multi-eye vision has better accuracy compared with monocular vision and binocular vision [4].

© Springer Nature Switzerland AG 2022
Y. Pan et al. (Eds.): AIMS 2021, LNCS 12987, pp. 89–98, 2022.
https://doi.org/10.1007/978-3-030-96033-9_8

This paper presents a method of robot arm space pose estimation based on three vision [5]. Based on the visual label detection system, three CMOS cameras are placed in different positions and angles respectively in space to obtain three real-time images of the moving manipulator. The label information includes four corner pixels, center point pixels, and the other is based on the visual label Homography matrix and ID of each tag. The external parameters of the camera are obtained by using the known pixel coordinates and spatial coordinates of the known image by PnP (perspective-n-point) algorithm [6], and the external parameters of the camera can be obtained by the four known points and the internal and external parameters of the calibrated camera, and the origin of the world coordinate system can be determined. According to the label [7] attached to the robot arm, the real-time pixel coordinates of the label can be obtained during the movement of the robot arm. The 3D position and posture in the robot arm space can be reconstructed according to the three-dimensional vision measurement system. The three-dimensional vision measurement system can cover a larger measurement area, and compared with the single binocular vision, the measurement system with three binocular vision has better robustness, and has a wider application in the actual complex application scenarios [8].

2 Position and Pose Measurement System Based on Binocular Vision

2.1 Pose Measurement Model Based on Binocular Vision

The multi eye pose positioning system used in this paper is composed of three cameras equipped with Sony Exmor R COMS sensor, manipulator, visual tag and computer (see Fig. 1).

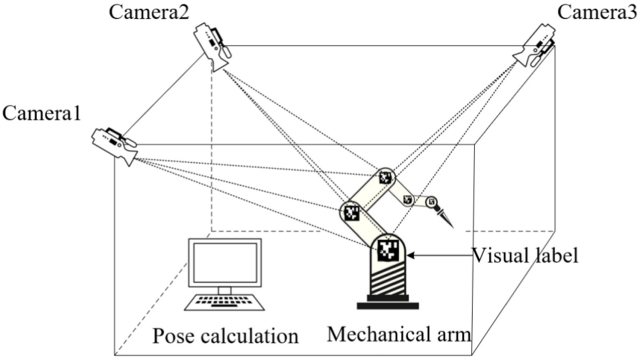


Fig. 1. Three dimensional schematic diagram of position and pose measurement system based on binocular vision.

The system consists of two parts (see Fig. 2), which are camera calibration system and multi camera pose measurement system. The functions of the system include: calibrating the camera to obtain the camera parameters, collecting the data set synchronously, detecting the visual label, transforming the pixel coordinates to the spatial coordinates, and calculating the optimal pose of the multi camera fusion.

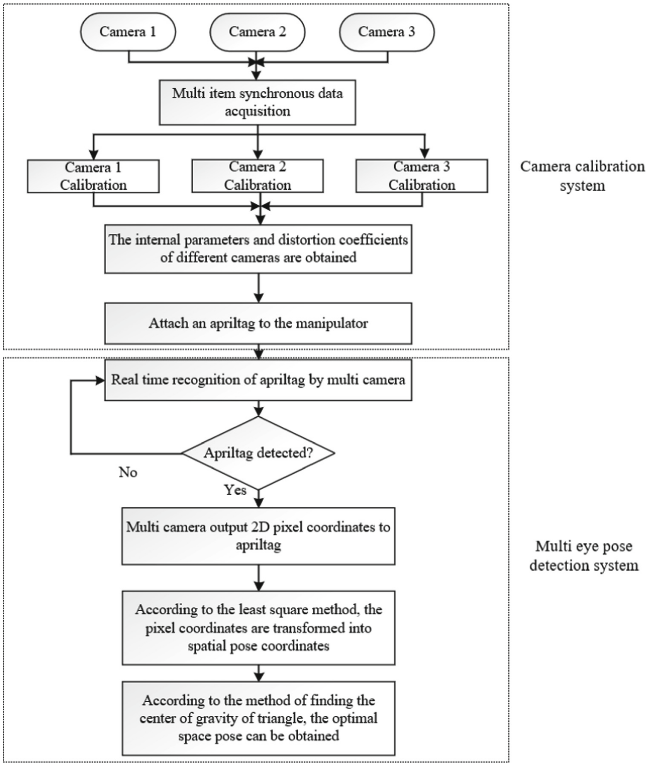

Camera 1
Camera 2
Camera 3
Multi item synchronous data acquisition
Camera 1 Calibration
Camera 2 Calibration
Camera 3 Calibration
The internal parameters and distortion coefficients of different cameras are obtained
Attach an apriltag to the manipulator
Camera calibration system
Real time recognition of apriltag by multi camera
Apriltag detected?
No
Yes
Multi camera output 2D pixel coordinates to apriltag
According to the least square method, the pixel coordinates are transformed into spatial pose coordinates
According to the method of finding the center of gravity of triangle, the optimal space pose can be obtained
Multi eye pose detection system

Fig. 2. System flow chart

3 Position and Pose Measurement System Based on Binocular Vision

3.1 Pinhole Camera Model

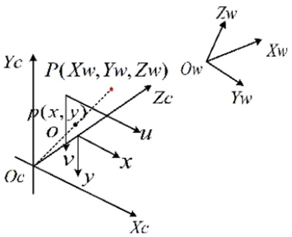


Fig. 3. Pinhole camera model

Image processing involves the following coordinate system: $O_W - X_W Y_W Z_W$: World coordinate system, which describes the position of the camera in m. $O_C - X_C Y_C Z_C$:

Camera coordinate system, optical center as the origin, unit m. $o-xy$: image coordinate system, the optical center is the center point of the image, and the unit is mm. uv: pixel coordinate system, the origin is the upper left corner of the image, the unit is pixel. P: A point in the world coordinate system is a point in real life. p: For the imaging point of point P in the image, the coordinates in the image coordinate system are (x, y), and the coordinates in the pixel coordinate system are $(u, v).f$: Camera focal length, equal to o and O_C [9].

In binocular vision, the origin of the world coordinate system is usually set at the midpoint of the x-axis direction of the left camera or the right camera or both (see Fig. 3).

The next point is about the transformation of these coordinate systems. In other words, how a real object is imaged in the image.

$$\begin{aligned} Z_C\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} &= \begin{bmatrix} f_x\ 0 & u_0\ 0 \\ 0\ f_y & v_0\ 0 \\ 0\ 0 & 1\ 0 \end{bmatrix}\begin{bmatrix} \boldsymbol{R}\ \boldsymbol{T} \\ \boldsymbol{0}\ \boldsymbol{1} \end{bmatrix}\begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \\ &= \boldsymbol{M}_1\boldsymbol{M}_2\boldsymbol{X} = \boldsymbol{MX} \end{aligned} \tag{1}$$

In the above formula, f_x, f_y is called the normalized focal length on the u-axis and v axis in the pixel coordinate system, and $\boldsymbol{M}$ is 3 × Projection matrix of 3. $\boldsymbol{M}_1$ is the internal parameter matrix, $\boldsymbol{M}_2$ is the external parameter matrix [8].

If the pixel coordinates (u, v) of the space point p are known, the internal parameters can be obtained by camera calibration. At this time, the real world cannot be determined (X_W, Y_W, Z_W), Because the $\boldsymbol{M}$ matrix is irreversible and the origin of the world coordinate system needs to be established. Therefore, this paper proposes to construct multiple linear equations by binocular or multi-objective cameras, determine the world origin and get the external parameters of the camera according to PNP, and then estimate the three-dimensional coordinates of the space point P.

3.2 Pinhole Camera Model

The visual label used in this paper is apriltag visual reference system, which can be used in robot, AR, camera calibration and other fields. The system can detect the sign in real time and calculate the relative position quickly. The system consists of the following main parts: the visual detector detects the edge of the input image and the two-dimensional coordinate information of the four corners in the label through gradient detection, establishes the corresponding world coordinate system according to the decoder of the system, and detects the pixel coordinates of the four corners one by one corresponding to the world coordinate system through the visual label system, The monasteric matrix can be obtained.

The visual reference system has two-dimensional spatial information, which is easier to calculate than two-dimensional code. As shown in Fig. 4, the visual label includes multiple categories, and the technology relies on alignment of multiple locating points and auxiliary points. Therefore, the visual reference system can detect a longer distance, and it can also detect in dark or poor detection environment, with high robustness (see Fig. 4).

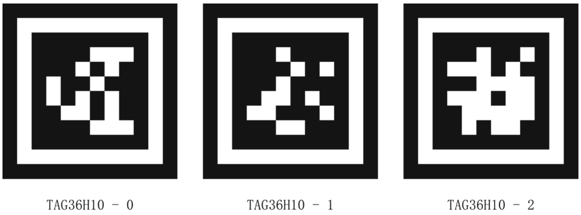


Fig. 4. Some families of apriltag

4 The Model of Three Vision Measurement System

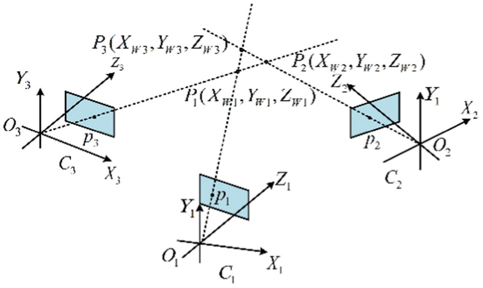


Fig. 5. The model of three vision measurement system

Through the different positions of the three cameras, the image of the visual label on the manipulator is obtained at the same time and in the same scene (see Fig. 5). Through the knowledge of machine vision, a series of two-dimensional coordinates are transformed into three-dimensional coordinates, so as to obtain the pose state of the manipulator [10].

In practical application, because the data is always noisy, as shown in Fig. 5, through the three-dimensional vision fusion, we can get that the three-dimensional coordinates of the measured object are disjoint points P_1, P_2, P_3. After the least square method, the optimal three-dimensional coordinates can be obtained [3]. Let the world coordinate system of P be (X_W, Y_W, Z_W). The optimal objective function should be satisfied.

$$F = \min(P - P_i)\ i = 1, 2, 3 \tag{2}$$

The solution process is as follows:

$$Z_C\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} \boldsymbol{R} & \boldsymbol{T} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}\begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \tag{3}$$

Among: $\boldsymbol{R}$: 3 × 3, $\boldsymbol{T}$: 3 × 1:

$$Z_i\begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \boldsymbol{M}_i\begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} i = 1, 2, 3 \tag{4}$$

The results are as follows:

$$Z_i\begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11}^i & m_{12}^i & m_{13}^i & m_{14}^i \\ m_{21}^i & m_{22}^i & m_{23}^i & m_{24}^i \\ m_{31}^i & m_{32}^i & m_{33}^i & m_{34}^i \end{bmatrix}\begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \tag{5}$$

Where $\boldsymbol{M}_i$ is $\boldsymbol{A}[\boldsymbol{RT}]$, and $\boldsymbol{A}$ is the internal parameter matrix of the camera. According to binocular vision, the coordinates of the least square method [11, 12] can be calculated.

$$\begin{bmatrix} X_W \\ Y_W \\ Z_W \end{bmatrix} = \left(\boldsymbol{C}_i^T\boldsymbol{C}_i\right)^{-1}\boldsymbol{C}_i\boldsymbol{D}_i, i = (1, 2, 3) \tag{6}$$

$$\boldsymbol{C}_1 = \begin{bmatrix} u_1m_{31}^1 - m_{11}^1 & u_1m_{32}^1 - m_{12}^1 & u_1m_{33}^1 - m_{13}^1 \\ v_1m_{31}^1 - m_{13}^1 & v_1m_{32}^1 - m_{13}^1 & v_1m_{33}^1 - m_{23}^1 \\ u_2m_{31}^2 - m_{11}^2 & u_2m_{32}^2 - m_{12}^2 & u_2m_{33}^2 - m_{13}^2 \\ v_2m_{31}^2 - m_{13}^2 & v_2m_{32}^2 - m_{13}^2 & v_2m_{33}^2 - m_{23}^2 \end{bmatrix}$$

$$\boldsymbol{D}_1 = \begin{bmatrix} m_{14}^1 - u_1m_{34}^1 \\ m_{24}^1 - v_1m_{34}^1 \\ m_{14}^2 - u_2m_{34}^2 \\ m_{24}^2 - v_2m_{34}^2 \end{bmatrix} \tag{7}$$

$$\boldsymbol{C}_2 = \begin{bmatrix} u_2m_{31}^2 - m_{11}^2 & u_2m_{32}^2 - m_{12}^2 & u_2m_{33}^2 - m_{13}^2 \\ v_2m_{31}^2 - m_{21}^2 & v_2m_{32}^2 - m_{22}^2 & v_2m_{33}^2 - m_{23}^2 \\ u_3m_{31}^3 - m_{11}^3 & u_3m_{32}^3 - m_{12}^3 & u_3m_{33}^3 - m_{13}^3 \\ v_3m_{31}^3 - m_{21}^3 & v_3m_{32}^3 - m_{22}^3 & v_3m_{33}^3 - m_{23}^3 \end{bmatrix}$$

$$\boldsymbol{D}_2 = \begin{bmatrix} m_{14}^2 - u_2m_{34}^2 \\ m_{24}^2 - v_2m_{34}^2 \\ m_{14}^3 - u_3m_{34}^3 \\ m_{24}^3 - v_3m_{34}^3 \end{bmatrix} \tag{8}$$

$$\boldsymbol{C}_3 = \begin{bmatrix} u_1m_{31}^1 - m_{11}^1 & u_1m_{32}^1 - m_{12}^1 & u_1m_{33}^1 - m_{13}^1 \\ v_1m_{31}^1 - m_{21}^1 & v_1m_{32}^1 - m_{22}^1 & v_1m_{33}^1 - m_{23}^1 \\ u_3m_{31}^3 - m_{11}^3 & u_3m_{32}^3 - m_{12}^3 & u_3m_{33}^3 - m_{13}^3 \\ v_3m_{31}^3 - m_{21}^3 & v_3m_{32}^3 - m_{13}^3 & v_3m_{33}^3 - m_{23}^3 \end{bmatrix}$$

$$\boldsymbol{D}_3 = \begin{bmatrix} m_{14}^1 - u_1m_{34}^1 \\ m_{24}^1 - v_1m_{34}^1 \\ m_{14}^3 - u_3m_{34}^3 \\ m_{24}^3 - v_3m_{34}^3 \end{bmatrix} \tag{9}$$

So I got it,

$$\begin{aligned} F &= \min(\|P - P_1\| + \|P - P_2\| + \|P - P_3\|) \\ &= \sum_{i=1}^{3} \left\{ \begin{array}{c} (X_W - X_{Wi})^2 + (Y_W - Y_{Wi})^2 + \\ (Z_W - Z_{Wi})^2 \end{array} \right\} \end{aligned} \tag{10}$$

To get the optimal objective function, the following conditions should be satisfied at the same time:

$$\begin{array}{c} f_i = \left\{ \begin{array}{c} (X_W - X_{Wi})^2 + (X_W - X_{Wi})^2 + \\ (X_W - X_{Wi})^2 \end{array} \right\} \\ i = 1, 2, 3 \end{array} \tag{11}$$

According to the method of finding the triangle center of gravity, the optimal 3D pose coordinates can be obtained:

$$X_W = \frac{1}{3}\sum_{i=1}^{3} X_{Wi}, Y_W = \frac{1}{3}\sum_{i=1}^{3} Y_{Wi}, Z_W = \frac{1}{3}\sum_{i=1}^{3} Z_{Wi} \tag{12}$$

In the above formula, X_{Wi}, Y_{Wi}, Z_{Wi} $(i = 1,\ 2,\ 3)$ is the pose output of three groups of binocular views, X_W, Y_W, Z_W is the pose output of the tricular view.

5 Experiment

5.1 Camera Calibration

Camera calibration [13] is to obtain camera internal parameters, distortion coefficient and other parameters. The common methods of camera calibration include linear calibration, nonlinear calibration, camera self calibration and Zhang Zhengyou calibration method [14]. The experiment adopts Zhang Zhengyou calibration method, and the specific calibration process is shown in Fig. 6. The method is simple and robust. In this experiment, 20 calibration plate images with different angles and positions were collected, and the camera internal parameters were obtained according to the calculated single stress matrix (see Fig. 6).

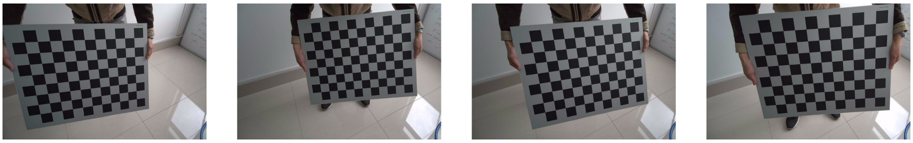


Fig. 6. Calibration picture

The size of the chessboard used in this experiment is 12 × 9. By taking pictures of different angles and distances, the internal parameters and distortion coefficients of the camera are obtained according to the calibration software. According to the principle of camera calibration, the internal parameters of the camera are fixed, and will not change because of the change of the pose of the camera, while the external parameters of the camera will change with the change of the pose of the camera. This experiment uses the PNP method to fix the origin of the world coordinate and obtain the external parameters of the camera.

5.2 Experimental Environment Layout

This experiment is based on the movement of the manipulator, through the auxiliary positioning of the visual label to obtain the spatial pose of the manipulator, by drawing the trajectory of the manipulator, measuring the error between the fixed truth point and the estimated point to test the accuracy and robustness of the three eye vision pose measurement system. The physical diagram of the experimental system is shown in Fig. 7.

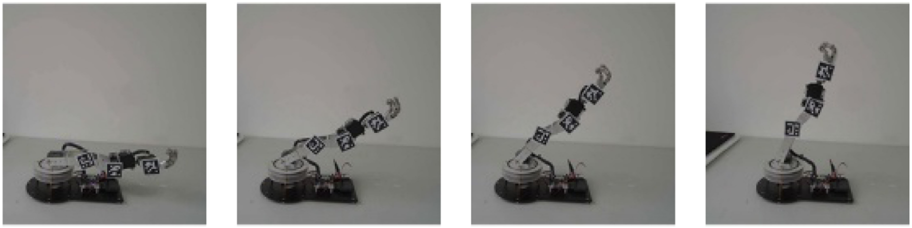
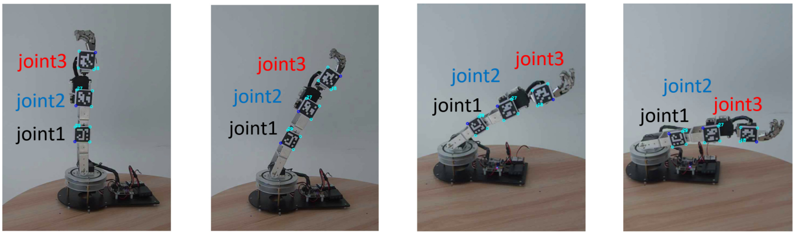


Fig. 7. Motion picture of manipulator

The pose of the manipulator is acquired synchronously by three cameras. We control the manipulator to swing. The three-dimensional trajectory of the manipulator [15] is shown in Fig. 8 below. The measured values are obtained by measuring different fixed points.



(a)Pose1 (b)Pose2 (c)Pose3 (d)Pose4

Fig. 8. Robot fixed point

The position and attitude measurement system of the manipulator based on the multi vision predicts the motion trajectory of the manipulator (see Fig. 8). The attitude estimation based on the three vision is correct. Because there will be a small amplitude of mechanical vibration in the motion process of the manipulator, there will be slight

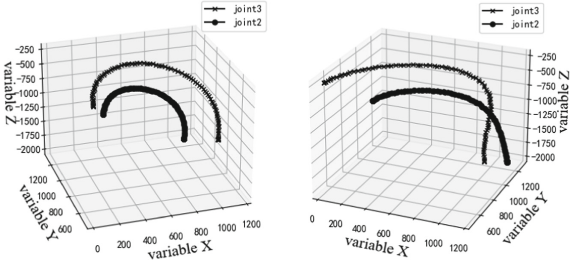


Fig. 9. Fixed point trajectory

frequency in the trajectory drawing. Figure 9 left shows the motion trajectory of joints 2 and 3 under the perspective 1, and Fig. 9 right shows joint 2 under the perspective 2, 3.

During the movement of the manipulator, different positions are taken to measure the measurement values of different joint points. As shown in Fig. 8 (pose1-pose4), the measurement values of different joint points (joint points 1–3) of different positions are obtained. The error analysis of the three eye vision posture measurement system and binocular vision pose measurement system is shown in Table 1 below:

Table 1. Error analysis of binocular and tricular pose measurement

Name	Pose1(joint1)	Pose1(joint2)	Pose1(joint3)
True value	(−55, 40, 180)	(−58, 40, 278)	(−70, 40, 360)
Binocular vision	(−57.6, 43.7, 182.1)	(−62.1, 41.2, 260.9)	(−73.9, 43.9, 365.3)
Tricular vision	(−56.2, 41.1, 179.9)	(−60.3, 39.8, 271.3)	(−72.3, 42.3, 363.7)
Binocular relative error	(4.6%, 9.3%, 1.1%)	(7.0%, 3.0%, 6.2%)	(5.6%, 9.8%, 1.5%)
Three eyes relative error	(2.2%, 1.1%, 0.08%)	(2.4%, 0.4%, 2.4%)	(3.2%, 5.8%, 1.0%)

According to Table 1, it can be concluded that the relative error of x-axis is reduced from 7.0% to 2.4% based on the joint 2 of pose1 of the robot arm measured by three eye vision posture measurement. The relative error of y-axis is reduced from 9.8% to 2.8% for the joint 3 of the robot arm pose1. The relative error of z-axis is reduced from 6.2% to 2.4%, The robot arm 3D pose estimation system can effectively improve the results of binocular vision 3D pose measurement, and has a better accuracy.

6 Conclusion

In this paper, a manipulator spatial pose estimation system based on multi vision is proposed. The manipulator spatial pose can be obtained in real time through a low-cost visual tag system. According to the experimental results, the position and pose estimation of the manipulator based on multi vision can be completed under certain occlusion, which verifies the robustness of the system. Through the experimental error analysis of each axis, the experimental error of the three vision position and pose measurement system is less than 4.9 mm, which meets the requirements of the use.

References

1. Krogius, M., Haggenmiller, A., Olson, E.: Flexible layouts for fiducial tags. In: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE (2019)
2. Lee, S., et al.: Vision based localization for multiple mobile robots using low-cost vision sensor. In: 2015 IEEE International Conference on Electro/Information Technology (EIT). IEEE (2015)
3. Yu, C., Cai, J., Chen, Q.: Multi-resolution visual fiducial and assistant navigation system for unmanned aerial vehicle landing. Aerosp. Technol. **67**(aug.), 249–256 (2017)
4. Peng, J., Xu, W., Yuan, H.: An efficient pose measurement method of a space non-cooperative target based on stereo vision. IEEE Access **5**, 22344–22362 (2017)
5. Zarándy, Á., et al.: A real-time multi-camera vision system for UAV collision warning and navigation. J. Real-Time Image Process. **12**(4), 709–724 (2014)
6. Zhou, B., Chen, Z., Liu, Q.: An efficient solution to the perspective-n-point problem for camera with unknown focal length. IEEE Access **PP**(99), 1 (2020)
7. Zhenglong, G., Qiang, F., Quan, Q.: Pose estimation for multicopters based on monocular vision and apriltag. In: 2018 37th Chinese Control Conference (CCC)
8. Wu, L., Zhu, B.: Binocular stereovision camera calibration. In: IEEE International Conference on Mechatronics & Automation, pp. 2638–2642. IEEE (2015)
9. Bračun, D., Sluga, A., et al.: Stereo vision based measuring system for online welding path inspection. J. Mater. Process. Technol. **223**, 328–336 (2015)
10. Li, S., Chi, X., Ming, X:. A robust O(n) solution to the perspective-n-point problem. IEEE Trans. Pattern Anal. Mach. Intell. **34**(7), 1444–1450 (2012)
11. Xia, R., et al.: Global calibration of multi-cameras with non-overlapping field of views based on photogrammetry and reconfigurable target. Meas. Sci. Technol. **29**(6), 065005-1–065005-10 (2018)
12. Shahzad, A., Mu, Y., Gao, X.: Tracking RGB color markers through DLT calibrated monocular vision system. In: IEEE International Conference on Mechatronics & Automation. IEEE (2016)
13. Zhang, J., Zhu, J., Deng, H., et al.: Multi-camera calibration method based on a multi-plane stereo target. Appl. Opt. **58**(34), 9353 (2019)
14. Zhang, Z.: A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. **22**(11), 1330–1334 (2000)
15. Yang, Y., et al.: Multi-camera visual SLAM for off-road navigation. Robot. Auton. Syst. **128**, 103505 (2020)