Research on 3D Pose Measurement Algorithm Based on Binocular Vision

Fu Qiang Shan¹, Feng Wang² and Hai Feng Zhang¹*

¹Xi'an Institute of Optics and Precision Mechanics of UCAS, University of Chinese Academy of Sciences, Xi'an, Shanxi, 710119, China
²Xi'an Institute of Optics and Precision Mechanics of UCAS, University of Chinese Academy of Sciences, Xi'an, Shanxi, 710119, China
*Corresponding author’s e-mail: Zhanghf@opt.ac.cn

Abstract: Based on binocular vision, this paper proposes an algorithm for fast and accurate three-dimensional measurement. The binocular 3D measurement system is mainly divided into four parts: binocular calibration, binocular correction, binocular matching and pose calculation. In this paper, an international software is used to achieve camera calibration, the FAST algorithm is used to achieve the extraction of feature points, and the matching method that combines gray-level correlation coefficient (cc) and epipolar constraint is selected. Experiments have verified that this algorithm can accurately achieve three-dimensional pose measurement.

1. Introduction

Binocular 3D measurement technology is a method of calculating the 3D pose of a target object based on the principle of parallax. In computer vision, monocular vision has the advantages of simple structure and low cost. However, the monocular camera cannot calculate the three-dimensional coordinates of the target point based on the principle of parallax in the static state. Multi-eye vision can calculate the parallax of a single pixel multiple times, thereby reducing calculation errors. However, the calculation amount of multi-eye vision is too large, which causes the running speed of the algorithm to decrease. Moreover, the structure of multi-eye vision is relatively complicated and the cost is relatively high. In contrast, the binocular vision has a simple structure, low cost, small amount of calculation, and takes into account the advantages of high speed and high precision.

In recent years, with the rapid development of aerial surveying and mapping technology, the rapid and accurate realization of three-dimensional pose measurement has become the primary task of researchers. In response to this problem, this paper proposes an algorithm based on the FAST algorithm, which combines the cross-correlation coefficient and the epipolar constraint. This algorithm can accurately calculate the pose of the target object.

2. Camera calibration

The 3D space point imaging model is shown in Figure 1. The projection of the target point $P$ on the left camera is $p_l$, and the projection point on the right camera is $p_r$. $O_l$ is the origin of the left camera coordinate system, and $O_r$ is the origin of the right camera coordinate system.
In the binocular 3D measurement system, the internal parameter matrix and the external parameter matrix are obtained by camera calibration, and the three-dimensional world coordinates can be obtained according to the pixel coordinates of the points. The conversion relationship between image coordinates and world coordinates is shown in formula (1). Among them, $Z_c$ is the scale factor, $f$ is the focal length of the camera, $dx$ and $dy$ is the distance between adjacent pixels on the $X$ and $Y$ axes in the image coordinate system. $(X_W, Y_W, Z_W)$ is the world coordinate of the point $P$ and $(u, v)$ is the image coordinate of the projection point. $M_1$ is the internal parameter matrix, which reflects the parameters of the camera itself. $M_2$ is the external parameter matrix, which reflects the relative position relationship of the left and right cameras.

$$
\begin{bmatrix}
    \frac{f}{dx} & 0 & u_0 & 0 \\
    0 & \frac{f}{dy} & v_0 & 0 \\
    0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
    X_W \\
    Y_W \\
    Z_W
\end{bmatrix}
= M_1
\begin{bmatrix}
    X_W \\
    Y_W \\
    Z_W
\end{bmatrix}
\begin{bmatrix}
    f \times dx + u \\
    f \times dy + v
\end{bmatrix}
\begin{bmatrix}
    0 & R & 0 \\
    0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
    X_C \\
    Y_C
\end{bmatrix}
$$

Camera calibration can calculate the distortion coefficient of the camera. There are three kinds of distortions in the camera lens: radial distortion, tangential distortion and barrel distortion. Since radial distortion and tangential distortion have a greater impact on images, generally only these two distortions are considered. Among them, as shown in the correction relation of radial distortion (2), $k_1$, $k_2$ and $k_3$ are the radial distortion coefficients.

$$
\begin{align*}
\hat{u} &= u \left[ 1 + k_1 r^2 + k_2 r^4 + k_3 r^6 \right] \\
\hat{v} &= v \left[ 1 + k_1 r^2 + k_2 r^4 + k_3 r^6 \right]
\end{align*}
$$

The correction relationship of tangential distortion is shown in equation (3). $\rho_1$ and $\rho_2$ is the tangential distortion coefficient.

$$
\begin{align*}
\hat{u} &= u + 2\rho_1 v + \rho_2 \left[ r^2 + 2u^2 \right] \\
\hat{v} &= v + 2\rho_2 u + \rho_1 \left[ r^2 + 2v^2 \right]
\end{align*}
$$

This paper uses a camera with a resolution of 2048*2048 to collect 17 pairs of checkerboard images, and obtains the distortion coefficients, the parameters of the internal parameter matrix and the external parameter matrix through the software M. Among them, the distortion coefficient is
represented by a matrix \( D = \begin{bmatrix} k_1, k_2, k_3, p_1, p_2 \end{bmatrix} \). The calibration results are as follows:

- **Distortion coefficient:**
  - left camera: \( D_l = [-0.0356, 0.0845, -0.0445, -0.5774, 0.6520] \)
  - right camera: \( D_r = [-0.0356, 0.0845, -0.0445, -0.5774, 0.6520] \)

- **Internal parameter matrix:**
  - left camera: \( M_l = \begin{bmatrix} 1556.5177 & 1.6161 & 1051.7793; 0, 1553.5464, 969.7014; 0, 0, 1 \end{bmatrix} \)
  - right camera: \( M_r = \begin{bmatrix} 1557.0674 & 1.0549 & 1064.9025; 0, 1553.7551, 976.5633; 0, 0, 1 \end{bmatrix} \)

- **External parameter matrix:**
  \( R = \begin{bmatrix} 0.999, -0.0018, 0.0104; -0.0018, 1.0000, 0.0063; 0.0104, 0.0063, 0.9999 \end{bmatrix} \)
  \( T = [-240.2000, -0.7900, 6.6568] \)

### 3. FAST feature point extraction

In binocular three-dimensional measurement, the real-time performance of the algorithm has always been a hot research topic. This paper uses FAST algorithm to extract the feature points of the image, which greatly improves the running speed of the entire algorithm. In the FAST algorithm, if the gray value of a certain pixel is different from that of enough pixels in its domain, then this point is defined as a feature point. As shown in Figure 2, take any point \( q \) in the image and the gray value is \( I(q) \). With the point \( q \) as the center, a Bresenham circular area with a radius of 3 is formed, and the point \( q \) is surrounded by 16 pixels on the circle. Setting a threshold \( \Delta \) to indicate the degree of difference between the gray values of the circle pixels \( q \) and the dots. If the difference between the gray value of a point on the circle and the pixel value of the point \( q \) is greater than \( \Delta \), the center point \( q \) is the candidate feature point. This article sets the value of \( \Delta \) to 9.

![Figure 2. Schematic diagram of FAST feature points](image-url)
points are selected. After the above three steps, the candidate feature points will be clustered at the edge of the image. This is because the real feature point and its surrounding points are similar in nature. In order to remove these similar points, all candidate feature points need to be screened again by the maximum value suppression method to obtain the true feature points.

4. Polar constraint and NCC matching

The gray-scale correlation coefficient matching method is a high-precision matching algorithm. Combining this algorithm with the epipolar constraint can reduce the search range of the matching point to the entire two-dimensional plane to the corresponding epipolar line, thereby improving the speed of the matching algorithm. The epipolar constraint relationship is as follows:

$$p_r^T F p_l = 0$$ (4)

Among them, $p_l$ is the feature point of the left image, $p_r$ is the right feature point, and $F$ is the basic matrix. According to the coordinates of the characteristic points in the left picture and the basic matrix, the coordinates of the characteristic points in the right picture can be obtained.

The gray-scale correlation coefficient matching method first calculates the cross-correlation coefficient (Cross Correlation Coefficient, cc) between the feature point neighborhood window in the left image and the pixel point neighborhood window in the right image. The definition of the correlation coefficient is shown in the following formula:

$$S(x,y) = \frac{\sum_{x,y\in W} [I_l(x,y) - \overline{I}_l] [I_r(x,y) - \overline{I}_r]}{\sqrt{\sum_{x,y\in W} [I_l(x,y) - \overline{I}_l]^2} \sqrt{\sum_{x,y\in W} [I_r(x,y) - \overline{I}_r]^2}}$$ (6)

Among them, $\overline{I}_l$ is the average of the gray values of the pixels in the correlation window on the left, and $\overline{I}_r$ is the average of the gray values of the pixels in the correlation window on the left. According to the NCC matching algorithm, the pixel corresponding to the maximum value of the correlation coefficient is selected as the matching point of the feature point.

In the simulation environment of software M, the algorithm of this paper is simulated and verified. The simulation result is shown in Figure 3. Figure (a) is the feature point extracted from the left image, and Figure (b) is the feature point extracted from the right image.
The three-dimensional coordinates of the characteristic points in the figure are shown in Table 1:

| Feature points | X(mm)   | Y(mm)   | Z(mm)   |
|----------------|---------|---------|---------|
| A              | 18.7405 | -23.9251| 763.1967|
| B              | 20.9199 | -64.8850| 771.2382|
| C              | 59.37249| -63.1876| 768.7057|
| D              | 69.24907| -22.2952| 757.8883|
| E              | 71.5927 | -63.0762| 767.3513|
| F              | 76.7256 | -32.1244| 759.5992|
| G              | 78.18486| -52.9876| 764.3235|
| H              | 87.3578 | -46.9993| 762.9529|

The distance between the points can be calculated from the three-dimensional coordinates of the characteristic points, and the distance between the points obtained is compared with the true value, so as to determine the accuracy of the algorithm in this paper. The distance between each point is shown in Table 2. It can be seen from Table 2 that the error percentages of this algorithm are all within 0.6%, which has high accuracy.

| Distance | Truth value | Calculated | Error | Error percentage |
|----------|-------------|------------|-------|------------------|
| LAB      | 42.0000     | 41.7987    | 0.2031| 0.48%            |
| LAC      | 57.0000     | 56.7701    | 0.2299| 0.40%            |
| LAD      | 51.0000     | 50.8129    | 0.1871| 0.37%            |
| LAE      | 66.0000     | 65.9047    | 0.0953| 0.14%            |
| LAF      | 59.0000     | 58.6723    | 0.3277| 0.56%            |
| LAG      | 66.0000     | 66.1780    | 0.1780| 0.27%            |

In this paper, the principal component analysis method is used to obtain the relative posture of the camera coordinate system and the real coordinate system, and the rotation matrix and translation matrix are calculated. The calculation results are as follows:

\[
R = \begin{bmatrix}
62.7633 & 60.1255 & 62.7017 \\
-48.2066 & -45.3236 & -47.6910 \\
791.7594 & 762.7850 & 791.1176
\end{bmatrix}
\]
5. Conclusion
This paper proposes a three-dimensional pose measurement algorithm based on binocular vision. The internal and external parameters of the binocular camera are obtained through the software M, the FAST algorithm is used to extract the feature points of the image, and then the cross-correlation coefficient and epipolar line are used. The algorithm of combining constraints completes image matching. The simulation verified that this algorithm can accurately calculate the pose parameters of the target object.

Acknowledgments
This paper was completed under the guidance of teacher Zhang Haifeng and the help of all the students. I would like to express my gratitude to them. At the same time, I would like to thank the Xi'an Institute of Optics and Precision Mechanics of UCAS for its funding support.

Reference
[1] Ge, F., G. Feng, C. Li, Z. Xu, J. Yun & Q. Yong (2020) An error reduction method of a binocular measurement system under vibrating conditions. Measurement Science and Technology, 31.
[2] Jin, D. & Y. Yang (2018) Sensitivity analysis of the error factors in the binocular vision measurement system. Optical Engineering, 57.
[3] Ju, H., Z. Guiyang, C. Jiashan & Y. Ming (2018) Corrected calibration algorithm with a fixed constraint relationship and an error compensation technique for a binocular vision measurement system. Applied optics, 57.
[4] Shu, A., Pei, H.D., Ding, L., Duan, H.X. & Z, S.S. (2020) Binocular visual pose measurement method for space non-cooperative targets. Chinese Journal of Optics, 40, 107-117.
[5] Yang, Y., Xu, S.X., Fang, J.Z & Cai, Y.Z. (2019) Binocular vision positioning measurement method based on improved ORB algorithm. Journal of Sensing Technology, 32, 1694-1699.
[6] Zhao, H., Sun, D.M., Dong, Q.Q & Wu, X.C. (2019) Research on feature point matching based on binocular vision measurement system. Modern Electronic Technology, 42, 154-157+162.