

# Research on the Algorithm for Solving the Indoor Vision Positioning Model of Mobile Robot

Guangbing Zhou<sup>1,\*</sup>, Baosheng Shen<sup>2</sup> and Jie Yan<sup>2</sup>

<sup>1</sup>Guangdong Institute of Intelligent Manufacturing, GuangZhou, China

<sup>2</sup>Midea Group Co., Ltd., FoShan, China

\*Corresponding author

**Abstract**—In the analysis of the features and properties of camera based on the data, constructs the projection model and error model positioning calculation model, expanded beam method traditional adjustment model is only for image observations are optimized, constructed model with two or three dimensional positioning information integration RGB-D camera. The design experiments verified the accuracy of detection and matching, and combined with the depth measurement characteristics, we constructed two weight matrixes of observation information, thus improving the mathematical model of the whole location calculation.

**Keywords**—RGB-D camera; indoor location; mathematical model

## I. INTRODUCTION

According to the US Environmental Protection Agency (EPA), 70% - 90% of the time is in the indoor<sup>[1]</sup>. To the growing demand for indoor positioning of large shopping malls, offices and other transportation hubs, this paper collected visual positioning signal RGB-D camera based on geometric imaging theory of RGB-D camera two kinds of observation information, imaging characteristics and texture projection relation information and depth information of the mobile robot, imaging geometric model RGB-D camera, analysis of two kinds of original observations and measurement error characteristics, extends the traditional adjustment model, the texture information and depth information as the observed value to optimize the exact position of the mobile robot.

## II. INITIALIZATION POSITION CALCULATION

The purpose of initialization location calculation is to calculate the motion platform position and posture of each image frame, and use it as the initial position of position and gesture to estimate the initial value for the subsequent adjustment adjustment model<sup>[2-4]</sup>. After obtaining two dimensional and three dimensional feature points, we can get the projection of feature points with texture information and depth information into the object space to get the 3D coordinates corresponding to the feature points through the strict imaging geometry model of the camera<sup>[5-6]</sup>. It is assumed that the three dimensional point set  $\{d_i\}$  represents the three dimensional points obtained by the current frame feature points, and  $\{m_i\}$  represents the three dimensional points obtained by the feature points of the last continuous frame. The transformation

relationship between two sets of points can be expressed by the transformation of the rigid body:

$$d_i = Rm_i + T + V_i \quad (1)$$

Among them, R represents the rotation relationship between two positions, which is a 3\*3 matrix. T represents the translation relation between two locations. It is a 3\*1 matrix and  $V_i$  represents the error matrix. When the  $V_i$  error is the hourly, the value of R and T is the optimal value. The least squares method can be used to solve the optimal solution, and the formula (1) least square solution model, such as formula (2), is used to decompose the singular value.

$$\sum V_i^2 = \sum_{i=1}^n \|d_i - Rm_i - T\|^2 \quad (2)$$

## III. CONSTRUCTION OF LOCATION MODEL

In the continuous process of movement, position and attitude of each frame image are calculated by the corresponding image sequence, but due to the presence of measurement error, tracking error<sup>[7-10]</sup> and other factors, will inevitably produce errors in calculation, and with the accumulation of time and distance, the prediction error will accumulate rapidly. In order to eliminate or slow down the accumulation of this error, we must take into account the error of the original observation information of the camera, and its mathematical structure should match the imaging characteristics of the camera as much as possible<sup>[11-14]</sup>. The RGB-D camera positioning model<sup>[15]</sup>, is constructed by the beam adjustment model. In the continuous movement of the moving platform, the image sequence obtained by constructing images, feature points in the image as the adjustment of the connection point, texture and depth information of feature points contained in the two RGB-D camera original observations as the measuring value adjustment model. At the same time considering the imaging characteristics of RGB-D camera, with strict imaging geometric model of RGB-D camera as the projection model of adjustment model, error analysis of depth information and texture information by means of observation, so as to construct error model adjustment model strictly imaging geometric model. The adjustment model is solved by the optimization constraints such as texture

information, depth information, camera position and attitude measurement. The adjustment model uses the camera's strict imaging geometry model as the projection model, and the model formula is as follows:

$$\begin{cases} x - x_0 = -f_x \frac{a_1(X - X_S) + b_1(Y - Y_S) + c_1(Z - Z_S)}{a_3(X - X_S) + b_3(Y - Y_S) + c_3(Z - Z_S)} \\ y - y_0 = -f_y \frac{a_2(X - X_S) + b_2(Y - Y_S) + c_2(Z - Z_S)}{a_3(X - X_S) + b_3(Y - Y_S) + c_3(Z - Z_S)} \\ d = -[a_3(X - X_S) + b_3(Y - Y_S) + c_3(Z - Z_S)] \end{cases} \quad (3)$$

In the model,  $(a_1, a_2, a_3, b_1, b_2, b_3, c_1, c_2, c_3)$  represents the camera rotation matrix. The rotation angle of the camera can be expressed by Euler angle  $(\omega, \varphi, \kappa)$ . The rotation system used in this paper is a  $\omega - \varphi - \kappa$  system with X axis as the spindle. The X axis is the main axis to revolve the Omega angle, then revolves the Y angle around the Y axis, and finally rotates the kappa angle around the Z axis. Then the rotation matrix R can be expressed as:

$$\begin{aligned} R &= R_\omega R_\varphi R_\kappa = \\ &\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \omega & -\sin \omega \\ 0 & \sin \omega & \cos \omega \end{pmatrix} \begin{pmatrix} \cos \varphi & 0 & \sin \varphi \\ 0 & 1 & 0 \\ -\sin \varphi & 0 & \cos \varphi \end{pmatrix} \begin{pmatrix} \cos \kappa & -\sin \kappa & 0 \\ \sin \kappa & \cos \kappa & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (4) \\ &= \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix} \end{aligned}$$

With the expansion of the formula (4), the relationship between the attitude parameters of the camera and the rotation angle can be obtained.

$$\begin{cases} a_1 = \cos \varphi \cos \kappa \\ a_2 = -\cos \varphi \sin \kappa \\ a_3 = \sin \varphi \\ b_1 = \cos \omega \sin \kappa + \sin \omega \sin \varphi \cos \kappa \\ b_2 = \cos \omega \cos \kappa - \sin \omega \sin \varphi \sin \kappa \\ b_3 = -\sin \omega \cos \varphi \\ c_1 = \sin \omega \sin \kappa - \cos \omega \sin \varphi \cos \kappa \\ c_2 = \sin \omega \cos \kappa + \cos \omega \sin \varphi \sin \kappa \\ c_3 = \cos \omega \cos \varphi \end{cases} \quad (5)$$

The rotation angle  $(\omega, \varphi, \kappa)$ , the position and attitude parameters  $(X_S, Y_S, Z_S)$  and the point coordinates of the connection point  $(X, Y, Z)$  are taken as the corrected values, and the point coordinates  $(x, y)$  and the depth value D are taken into consideration. According to the projection model of formula (3), we construct the error model, that is, the residual expression between the observed value and the theoretical value is constructed for the original point of view  $(x, y)$  and the depth value D in the above models. Since there is a nonlinear

relationship between the measurement in the model and the amount to be corrected, we need to linearize the formula in order to establish the residual expression between the observed and the corrected quantity. By using the Taylor series expansion and preserving the first order term, the form of the linearized residual expression is as follows:

$$\begin{cases} v_x = (x) - x + \frac{\partial x}{\partial X_S} \Delta X_S + \frac{\partial x}{\partial Y_S} \Delta Y_S + \frac{\partial x}{\partial Z_S} \Delta Z_S + \\ \frac{\partial x}{\partial \omega} \Delta \omega + \frac{\partial x}{\partial \varphi} \Delta \varphi + \frac{\partial x}{\partial \kappa} \Delta \kappa + \frac{\partial x}{\partial X} \Delta X + \frac{\partial x}{\partial Y} \Delta Y + \frac{\partial x}{\partial Z} \Delta Z \\ v_y = (y) - y + \frac{\partial y}{\partial X_S} \Delta X_S + \frac{\partial y}{\partial Y_S} \Delta Y_S + \frac{\partial y}{\partial Z_S} \Delta Z_S + \\ \frac{\partial y}{\partial \omega} \Delta \omega + \frac{\partial y}{\partial \varphi} \Delta \varphi + \frac{\partial y}{\partial \kappa} \Delta \kappa + \frac{\partial y}{\partial X} \Delta X + \frac{\partial y}{\partial Y} \Delta Y + \frac{\partial y}{\partial Z} \Delta Z \\ v_d = (d) - d + \frac{\partial d}{\partial X_S} \Delta X_S + \frac{\partial d}{\partial Y_S} \Delta Y_S + \frac{\partial d}{\partial Z_S} \Delta Z_S + \\ \frac{\partial d}{\partial \omega} \Delta \omega + \frac{\partial d}{\partial \varphi} \Delta \varphi + \frac{\partial d}{\partial \kappa} \Delta \kappa + \frac{\partial d}{\partial X} \Delta X + \frac{\partial d}{\partial Y} \Delta Y + \frac{\partial d}{\partial Z} \Delta Z \end{cases} \quad (6)$$

$(v_x, v_y, v_z)$  is the residual of the observational measurements,  $(x), (y), (z)$  is the initial value calculated by taking the initial values of the corrected quantities into the formula (3). By using the parameters of some columns instead of the deviations in the formula (6), the formula is rewritten as:

$$\begin{cases} v_x = a_{11} \Delta X_S + a_{12} \Delta Y_S + a_{13} \Delta Z_S + a_{14} \Delta \omega + a_{15} \Delta \varphi + \\ a_{16} \Delta \kappa + a_{17} \Delta X + a_{18} \Delta Y + a_{19} \Delta Z - l_x \\ v_y = a_{21} \Delta X_S + a_{22} \Delta Y_S + a_{23} \Delta Z_S + a_{24} \Delta \omega + a_{25} \Delta \varphi + \\ a_{26} \Delta \kappa + a_{27} \Delta X + a_{28} \Delta Y + a_{29} \Delta Z - l_y \\ v_d = a_{31} \Delta X_S + a_{32} \Delta Y_S + a_{33} \Delta Z_S + a_{34} \Delta \omega + a_{35} \Delta \varphi + \\ a_{36} \Delta \kappa + a_{37} \Delta X + a_{38} \Delta Y + a_{39} \Delta Z - l_d \end{cases} \quad (7)$$

For each coefficient, it can be further expressed by the amount to be corrected. For each parameter, we need to start with the imaging geometry model, namely the projection model.

$$\begin{cases} \bar{X} = a_1(X - X_S) + b_1(Y - Y_S) + c_1(Z - Z_S) \\ \bar{Y} = a_2(X - X_S) + b_2(Y - Y_S) + c_2(Z - Z_S) \\ \bar{Z} = a_3(X - X_S) + b_3(Y - Y_S) + c_3(Z - Z_S) \end{cases} \quad (8)$$

For the expression of the observation value  $x$  of the image point:

$$x - x_0 = -f_x \frac{a_1(X - X_S) + b_1(Y - Y_S) + c_1(Z - Z_S)}{a_3(X - X_S) + b_3(Y - Y_S) + c_3(Z - Z_S)} \quad (9)$$

The results of the coefficients associated with the observed value  $x$  are obtained as follows:

$$\begin{cases}
 a_{11} = \frac{\partial x}{\partial X_S} = -f_x \frac{1}{Z} \frac{1}{Z} (\bar{X}a_3 - \bar{Z}a_1) \\
 a_{12} = \frac{\partial x}{\partial Y_S} = -f_x \frac{1}{Z} \frac{1}{Z} (\bar{X}b_3 - \bar{Z}b_1) \\
 a_{13} = \frac{\partial x}{\partial Z_S} = -f_x \frac{1}{Z} \frac{1}{Z} (\bar{X}c_3 - \bar{Z}c_1) \\
 a_{14} = \frac{\partial x}{\partial \omega} = f_x \frac{1}{Z} \frac{1}{Z} \left\{ \bar{X}[b_3(Z - Z_S) - c_3(Y - Y_S)] - \right. \\
 \left. \bar{Z}[b_1(Z - Z_S) - c_1(Y - Y_S)] \right\} \\
 a_{15} = \frac{\partial x}{\partial \varphi} = f_x \frac{1}{Z} \frac{1}{Z} \left\{ \bar{X}[\cos \varphi(X - X_S) + \right. \\
 \left. \sin \omega \sin \varphi(Y - Y_S) - \right. \\
 \left. \cos \omega \sin \varphi(Z - Z_S)] - \right. \\
 \left. \bar{Z}[-\sin \varphi \cos \kappa(X - X_S) + \right. \\
 \left. \sin \omega \cos \varphi \cos \kappa(Y - Y_S) - \right. \\
 \left. \cos \omega \cos \varphi \cos \kappa(Z - Z_S)] \right\} \\
 a_{16} = \frac{\partial x}{\partial \kappa} = f_x \frac{\bar{Y}}{Z} \\
 a_{17} = -a_{11} = \frac{\partial x}{\partial X} = f_x \frac{1}{Z} \frac{1}{Z} (\bar{X}a_3 - \bar{Z}a_1) \\
 a_{18} = -a_{12} = \frac{\partial x}{\partial Y} = f_x \frac{1}{Z} \frac{1}{Z} (\bar{X}b_3 - \bar{Z}b_1) \\
 a_{19} = -a_{13} = \frac{\partial x}{\partial Z} = f_x \frac{1}{Z} \frac{1}{Z} (\bar{X}c_3 - \bar{Z}c_1)
 \end{cases} \quad (10)$$

For the expression of the observation value y of the image point:

$$y - y_0 = -f_y \frac{a_2(X - X_S) + b_2(Y - Y_S) + c_2(Z - Z_S)}{a_3(X - X_S) + b_3(Y - Y_S) + c_3(Z - Z_S)} \quad (11)$$

The results of the coefficients associated with the observed value y are obtained as follows:

$$\begin{cases}
 a_{21} = \frac{\partial x}{\partial X_S} = -f_y \frac{1}{Z} \frac{1}{Z} (\bar{Y}a_3 - \bar{Z}a_2) \\
 a_{22} = \frac{\partial x}{\partial Y_S} = -f_y \frac{1}{Z} \frac{1}{Z} (\bar{Y}b_3 - \bar{Z}b_2) \\
 a_{23} = \frac{\partial x}{\partial Z_S} = -f_y \frac{1}{Z} \frac{1}{Z} (\bar{Y}c_3 - \bar{Z}c_2) \\
 a_{24} = \frac{\partial x}{\partial \omega} = f_y \frac{1}{Z} \frac{1}{Z} \left\{ \bar{Y}[b_3(Z - Z_S) - c_3(Y - Y_S)] - \right. \\
 \left. \bar{Z}[b_2(Z - Z_S) - c_2(Y - Y_S)] \right\} \\
 a_{25} = \frac{\partial x}{\partial \varphi} = f_y \frac{1}{Z} \frac{1}{Z} \left\{ \bar{Y}[\cos \varphi(X - X_S) + \right. \\
 \left. \sin \omega \sin \varphi(Y - Y_S) - \right. \\
 \left. \cos \omega \sin \varphi(Z - Z_S)] - \right. \\
 \left. \bar{Z}[-\sin \varphi \sin \kappa(X - X_S) - \right. \\
 \left. \sin \omega \cos \varphi \cos \kappa(Y - Y_S) + \right. \\
 \left. \cos \omega \cos \varphi \cos \kappa(Z - Z_S)] \right\} \\
 a_{26} = \frac{\partial x}{\partial \kappa} = f_y \frac{\bar{X}}{Z} \\
 a_{27} = -a_{21} = \frac{\partial x}{\partial X} = f_y \frac{1}{Z} \frac{1}{Z} (\bar{Y}a_3 - \bar{Z}a_2) \\
 a_{28} = -a_{22} = \frac{\partial x}{\partial Y} = f_y \frac{1}{Z} \frac{1}{Z} (\bar{Y}b_3 - \bar{Z}b_2) \\
 a_{29} = -a_{23} = \frac{\partial x}{\partial Z} = f_y \frac{1}{Z} \frac{1}{Z} (\bar{Y}c_3 - \bar{Z}c_2)
 \end{cases} \quad (12)$$

For the expression of the depth observation value D:

$$d = -[a_3(X - X_S) + b_3(Y - Y_S) + c_3(Z - Z_S)] \quad (13)$$

The results of the coefficients associated with the observed value D are obtained as follows:

$$\begin{cases}
 a_{31} = \frac{\partial d}{\partial X_S} = \sin \varphi \\
 a_{32} = \frac{\partial d}{\partial Y_S} = -\sin \omega \cos \varphi \\
 a_{33} = \frac{\partial d}{\partial Z_S} = \cos \omega \cos \varphi \\
 a_{34} = \frac{\partial d}{\partial \omega} = \cos \omega \cos \varphi(Y - Y_S) + \sin \omega \cos \varphi(Z - Z_S) \\
 a_{35} = \frac{\partial d}{\partial \varphi} = -\cos \varphi(X - X_S) - \\
 \sin \omega \sin \varphi(Y - Y_S) + \cos \omega \sin \varphi(Z - Z_S) \\
 a_{36} = \frac{\partial d}{\partial \kappa} = 0 \\
 a_{37} = -a_{31} = \frac{\partial d}{\partial X} = -\sin \varphi \\
 a_{38} = -a_{32} = \frac{\partial d}{\partial Y} = \sin \omega \cos \varphi \\
 a_{39} = -a_{33} = \frac{\partial d}{\partial Z} = -\cos \omega \cos \varphi
 \end{cases} \quad (14)$$

At this point, the exact expression of the coefficients in the error model can be obtained, and the error model formula (7) is rewritten as a matrix form, as follows:

$$V = AX - L, P \quad (15)$$

Among them:

$$\begin{cases} V = [v_x, v_y, v_d]^T \\ L = [l_x, l_y, l_d]^T \\ A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} & \cdot & -a_{11} & -a_{12} & -a_{13} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} & \cdot & -a_{21} & -a_{22} & -a_{23} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} & \cdot & -a_{31} & -a_{32} & -a_{33} \end{pmatrix} \\ X = [\Delta X_S, \Delta Y_S, \Delta Z_S, \Delta \omega, \Delta \phi, \Delta \kappa, \Delta X, \Delta Y, \Delta Z]^T \end{cases} \quad (16)$$

X represents the value matrix to be corrected, that is, the quantity to be solved, which includes the location of the camera, the attitude angle and the three-dimensional coordinates of the points. The P is the weight matrix between the observation value of the point coordinate and the depth measurement. It reflects the accuracy of the observation. According to the principle of least square indirect adjustment, the formula (15) of the formula can be listed:

$$A^T P A X = A^T P L \quad (17)$$

Thus the expression of the amount to be solved can be obtained:

$$X = (A^T P A)^{-1} A^T P L \quad (18)$$

The matrix P of formula (18) expresses the weight relation between the texture image point information of the RGB-D camera and the depth information of the two original observations. Weight is an index to compare the relative precision between the observed values, and the measurement accuracy is the standard deviation. At the same time, the weight of the observed values is inversely proportional to the variance of the observed values (the square of the standard deviation). Texture image point information is extracted and matched by feature points, and its measurement accuracy depends on the extraction of feature points and the accuracy of matching tracking. According to the measurement accuracy of these two kinds of observation information, the weight matrix P in the location model can be constructed. At this point, a two or three dimensional information integration model for the integration of RGB-D cameras is constructed, such as formula (15) and solution formula (17). The model is based on the original observation information of RGB-D camera, the observation point and depth measurement value of texture, and the weight matrix of observed value is constructed according to the measurement error of the observed value. The optimal constraint of the amount to be solved is obtained through the solution of the model<sup>[15]</sup>.

#### IV. EXPERIMENTAL VERIFICATION

In order to verify the feasibility of the proposed method, build the experimental environment, the design of mobile robot driving a 180 meter long closed path, set a starting point and the point as the origin of the sets of 0 coordinates, from this point of view eventually return to the point. The frame frequency of the acquired and stored images is the same as that of the experiment. Finally, a total of 2800 frames of deep images and RGB images are stored. The depth images and RGB images of the RGB-D cameras are obtained as shown in Figure 1.



FIGURE 1. THE DEPTH IMAGE AND RGB IMAGE OF A MOBILE ROBOT

In this experiment, the traditional BA method is compared with the method proposed in this paper. The accuracy of the method is evaluated with closed error. The result of the experiment is table 1.

TABLE I. ERROR COMPARISON

Contrast	Closure difference(m)	error(mm)
This paper method	4.45	2.42%
Traditional BA method	7.6	3.86%

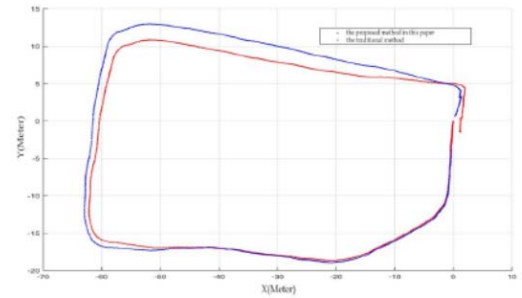


FIGURE II. EXPERIMENTAL RESULT

Can be seen from the experimental results, the error of two or three dimensional integrated positioning model is proposed in this paper for 2.42% full closed full closed, the error of the traditional BA method<sup>[15]</sup> is 3.86%, the accuracy is obviously improved; Figure 2 shows, estimate the path of this method (blue) to estimate path compared with the traditional BA method (red) in the end of the closure is not only smaller, but the position deviation is smaller, the path caused by the deformation of the positioning errors in the estimation of path in the process of the smaller, more in line with the actual situation.

#### ACKNOWLEDGMENT

This research was supported by Guangdong Province Science and Technology Project (Grant No. 2016B090926005 and 2016B090926002).

#### REFERENCES

- [1] Weiser M. The Computer for the 21st Century[J]. Scientific American, 1991, 265(3):94-104.
- [2] Lowe D. G. Distinctive Image Features from Scale-Invariant Keypoints[J]. International Journal of Computer Vision, 2004, 60(2):91-110.
- [3] Bay H, Tuytelaars T, Gool L V. SURF: Speeded Up Robust Features[J]. Computer Vision & Image Understanding, 2006, 110(3):404-417.
- [4] Rublee E, Rabaud V, Konolige K, et al. ORB: An efficient alternative to SIFT or SURF[C]. IEEE International Conference on Computer Vision. IEEE, 2011:2564-2571.
- [5] Olson C F, Matthies L H, Schoppers M, et al. Rover Navigation using Stereo Ego-motion[J]. Robotics & Autonomous Systems, 2003, 43(4):215-229.
- [6] Cheng Y, Maimone M W, Matthies L. Visual odometry on the Mars exploration rovers - a tool to ensure accurate driving and science imaging. Robotics & Automation Magazine IEEE. 2006, 13, 54-62.
- [7] Huang A S, Bachrach A, Henry P, Krainin M, Maturana D, Fox D, Roy N. Visual Odometry and Mapping for Autonomous Flight Using an RGB-D Camera. International Symposium on Robotics Research (ISRR), 2011.
- [8] Pirkner K, Rother M, Schweighofer G, et al. GPSlam: Marrying Sparse Geometric and Dense Probabilistic Visual Mapping[C]. British Machine Vision Conference. 2011:115.1-115.12.
- [9] Glocker B, Izadi S, Shotton J, et al. Real-time RGB-D camera relocalization[C]. IEEE International Symposium on Mixed and Augmented Reality. IEEE, 2013:173-179.
- [10] Hu G, Huang S, Zhao L, et al. A robust RGB-D SLAM algorithm[C]. IEEE/rsj International Conference on Intelligent Robots and Systems. IEEE, 2012:1714-1719.
- [11] Endres F, Hess J, Sturm J, et al. 3-D Mapping With an RGB-D Camera[J]. Robotics IEEE Transactions on, 2014, 30(1):177-187.
- [12] Kerl C, Sturm J, Cremers D. Dense visual SLAM for RGB-D cameras[C]. IEEE/rsj International Conference on Intelligent Robots and Systems. IEEE, 2013:2100-2106.
- [13] Maier R, Sturm J, Cremers D. Submap-Based Bundle Adjustment for 3D Reconstruction from RGB-D Data[M]. Pattern Recognition. 2014:54-65.
- [14] Eggert D W, Lorusso A, Fisher R B. Estimating 3-D rigid body transformations: a comparison of four major algorithms[J]. Machine Vision and Applications, 1997, 9(5):272-290.
- [15] Zhao Qiang. Research on real-time navigation and positioning model and method of motion platform based on RGB-D camera research [D]. University of Chinese Academy of Sciences, 2017.