Research on Visual Odometer of Wheeled Robot with Motion Constraints

Wang Chenggong¹, Li Gen¹*, Wang Xi¹, Zhang Youyi²

¹College of Artificial Intelligence and Big Data, Guangdong Business and Technology University, Zhaoqing, Guangdong, China
²Computer institute, Guangdong Business and Technology University, Zhaoqing, Guangdong, China

*email: 2019023217@stu.imust.edu.cn

Abstract. In this paper, aiming at the planar motion characteristics of wheeled robots, the Red Green Blue Depth Camera (RGBD camera) was used as an image acquisition device to reduce the motion of wheeled robots to two-dimensional plane processing, which simplified the calculation method of visual mileage of planar wheeled robots. In addition, according to the two-dimensional motion characteristics of the robot camera, the contour constraint condition of feature point matching between frames was proposed, and the linear constraint condition was proposed by linearizing the motion equation of the camera between frames. The research showed that the contour constraint condition and linear constraint condition could be used to screen the mismatched feature points, this condition could not only screen the mismatched point pairs of color images, but also screen the matching point pairs with correct color image matching but large depth error, which provided high quality matching point pairs for inter-frame motion estimation of visual odometer. Finally, the filtered matching point pairs were reduced in dimension and combined with the two-dimensional Iterative Closest Point (ICP) algorithm to estimate the trajectory of the robot camera. The experimentation results showed that compared with the original three-dimensional ICP algorithm, the combination of contour constraint condition, linear constraint condition and two-dimensional ICP algorithm could significantly improve the computational speed of visual odometry.

1. Introduction
Simultaneous Localisation and Mapping (abbreviated as SLAM) has been an established issue in the territorial domain of mobile robots for at least the last thirty years [1-3], which also remains one of the essential technologies for automation in the 21st century [4]. SLAM is a necessary capability for robots to explore unknown environments [5]. It is widely used in service robots, sweeping robots,
inspection robots, unmanned aerial vehicles and other fields. Visual SLAM (abbreviated as VSLAM) refers to the special case where the only exteroceptive sensors available are cameras, the inherent nonlinearity of the VSLAM problem remains challenging and fulling difficult; fortunately, there are many advantages, for example, VSLAM can carry rich scene information, which is a hot research topic. Its basic framework includes camera information reading, visual odometer, back-end optimization, loopback detection and mapping. The visual odometer is used to estimate the camera motion between adjacent images, which is the core part of visual SLAM and the content of this paper.

The visual odometer consists of image processing and inter-frame motion estimation. At present, feature point extraction and matching are commonly used image processing algorithms. The more mature feature point extraction algorithms include Scale Invariant Feature Transform (SIFT) algorithm [6], Speed-up Robust Feature Transform (SURF) algorithm [7] and Oriented FAST and Rotated BRIEF (ORB) algorithm [8]. The ORB algorithm proposed by RubLee et al. on ICCV2011 ensures that the operation speed could be improved on the basis of invariant scale and rotation, which could basically meet the real-time requirements; this algorithm was based on FAST feature point detection [9] and BRIEF feature point description [10], which was much faster than SIFT algorithm and SURF algorithm in calculation speed.

The current mainstream inter-frame motion estimation algorithms include Random Sample Consensus algorithm (abbreviated as RANSAC algorithm) [11] and ICP algorithm [12]. RANSAC algorithm consists of three parts: sampling modeling, model validation and finding the optimal model. However, the operation time of the algorithm was relatively long and there would be a large calculation error in a few cases [13]. The other was the ICP algorithm proposed by Besl and McKay, Method for registration of 3-D shapes; meanwhile, the central idea was to iterate the solution of the nearest point; for the visual odometer, the correspondence between the two sets of spatial points was known, and there was no need to iterate repeatedly [14]. The experimental verification in this paper shows that the operation speed and calculation accuracy of visual SLAM using ICP algorithm are better than those of RANSAC algorithm. Therefore, this paper uses ICP algorithm for inter-frame motion estimation, and reduces the dimension of 3D ICP algorithm according to the motion characteristics of planar wheeled robot. The results show that the improved algorithm improves the overall operation efficiency.

2. Constraint conditions of feature point matching
The camera of planar wheeled robot usually conforms to the following two motion characteristics: (1) the camera moves on the horizontal plane, and the vertical displacement is ignored; (2) the camera rotates around the axis perpendicular to the ground, and the rotation of the other two axes is ignored.

2.1 Feature point matching under equal height constraint conditions
The constraint condition of this paper was based on the motion characteristics of wheeled robots. Suppose the camera moved from coordinate system 1 to coordinate system 2, and two different images were taken at two different positions, as shown in Figure 1. Its motion was also shown in Figure 1.
In this paper, the $O_1X_1Y_1Z_1$ camera coordinate system in Figure 1 was defined as the reference coordinate system, and the corresponding pose matrix was a $4 \times 4$ unit's matrix. The camera coordinate system $O_2X_2Y_2Z_2$ was derived from the camera coordinate system $O_1X_1Y_1Z_1$. According to the constraint condition of robot motion, the values of $O_1$ and $O_2$ in the vertical direction were approximately equal, that is, the motion on the $Y$ axis was a mainly the rotation, and the translation was almost zero. According to this motion characteristic, the transfer matrix between two coordinate systems was expressed in the form of transfer matrix, as shown in Eq. (1). Where $\theta$ represents the rotation angle of the robot around the $Y_1$ axis in the reference coordinate system, $t_x$ denotes the translation of the robot on the $X_1$ axis, and $t_z$ represents the translation of the robot on the $Z_1$ axis.

\[
\begin{pmatrix}
\cos \theta & 0 & \sin \theta & t_x \\
0 & 1 & 0 & 0 \\
-\sin \theta & 0 & \cos \theta & t_z \\
0 & 0 & 0 & 1 \\
\end{pmatrix}
\]  

(1)

Assuming that the homogeneous coordinate corresponding to point A in coordinate system 1 was $A_1 = (x_1, y_1, z_1, 1)$, and the homogeneous coordinate corresponding to point A in coordinate system 2 was $A_2 = (x_2, y_2, z_2, 1)$, the relationship between $A_1$ and $A_2$ could be obtained from Equation (1), as shown in Eq. (2). Because the camera height WAS equal, so $A_1$ and $A_2$ ordinates were equal, as shown in formula (3). The camera internal parameter matrix could be obtained by calibrating the camera as shown in Eq. (4). The corresponding inverse matrix is obtained, as shown in Eq. (5).

\[
\begin{pmatrix}
x_1 \\
y_1 \\
z_1 \\
1 \\
\end{pmatrix}
= \begin{pmatrix}
\cos \theta & 0 & \sin \theta & t_x \\
0 & 1 & 0 & 0 \\
-\sin \theta & 0 & \cos \theta & t_z \\
0 & 0 & 0 & 1 \\
\end{pmatrix}
\begin{pmatrix}
x_2 \\
y_2 \\
z_2 \\
1 \\
\end{pmatrix}
\]

(2)

\[y_1 = y_2\]

(3)

\[
K = \begin{pmatrix}
f_x & 0 & c_x \\
0 & f_y & c_y \\
0 & 0 & 0 \\
\end{pmatrix}
\]  

(4)
The equation (6) could be obtained from the camera imaging model. At this time, the point A1 was normalized, and the corresponding coordinate was \( A_1 = (x_1/z_1, y_1/z_1, 1) \). Equation (6) was expressed as Equation (7). Substituting the pixel coordinates \( B_1 \) and \( B_2 \) are corresponding to \( A_1 \) and \( A_2 \) into Formula (7), Formula (8) could be obtained. Similarly, Formula (9) could also be obtained.

\[
A_1 = K^{-1}B_i
\]  

(6)

(7)

\[
\begin{pmatrix}
\frac{1}{f_x} 0 -c_x \\
0 \frac{1}{f_y} -c_y \\
0 0 1
\end{pmatrix}
\begin{pmatrix}
M_i \\
N_i \\
1
\end{pmatrix}
\]

\[ y_1 = z_1 \frac{N_i - c_y}{f_y} \]  

(8)

\[ y_2 = z_2 \frac{N_2 - c_y}{f_y} \]  

(9)

The above equations (8) and (9) represented the coordinate values of the camera optical center in the two coordinate systems. Theoretically, these two values were equal. However, in the process of actual image acquisition, due to the existence of image distortion, depth measurement error and poor ground level, \( y_1 \) and \( y_2 \) could not be completely equal. Therefore, the definition of contour constraint conditions was shown in Equation (10).

\[
\Delta y = y_1 - y_2 = z_1 \frac{N_1 - c_y}{f_y} - z_2 \frac{N_2 - c_y}{f_y}
\]  

(10)

The threshold range of equal height constraint could be dynamically adjusted according to the actual situation. The contour constraint condition could effectively eliminate the mismatch feature point pairs by simple calculation. However, in some cases, the constraint condition of equal height could not work: (1) mismatch feature point pairs with the same vertical height; (2) matching points close to the midline of the image level: due to the high frequency of camera image acquisition, the pixel coordinates between adjacent frame matching points were very close. As shown in formula (11), when \( N_1 \) and \( N_2 \) values were very close to cy, even if the depth value and \( Z_i \), \( Z_2 \) error was large, the value of \( \Delta y \) was still small. At this time, the threshold in the equal height constraint could not be used to screen out such mismatching point pairs.
2.2 Line constraint condition
Equation (11) could be obtained from Equation (2). Because the frame rate of robot image acquisition is high and the motion speed was slow, the camera rotation angle \( \theta \) between two adjacent frames was very small, so Eq. (11) could be approximated as Eq. (12). The approximate formula (12) was two linear equations. The parameters of the first linear equation were \( \theta \) and \( t_s \), and the parameters of the second linear equation were \( \theta \) and \( t_r \). Using the matched point pairs filtered by the equal height constraint condition, the two straight lines in equation (12) were fitted separately, and the \( t_s, t_r \) and \( \theta \) values corresponding to each straight line could be obtained [15]. The fitted straight line could be used as the screening condition of mismatch feature point pairs, and only the points close to the straight line were retained. Using the fitted two straight lines as the conditions for screening the mismatched feature points, this paper defined them as the straight-line constraint conditions. According to the motion characteristics of planar motion wheeled robots equipped with RGBD cameras, the contour constraint conditions and linear constraint conditions proposed in this paper could effectively screen the mismatched feature points, including the matching point pairs with correct color map matching but large depth error, so as to provide high-quality matching point pairs for visual odometry.

\[
\begin{align*}
\begin{cases}
x_t = x_2 \cos \theta + z_2 \sin \theta + t_x \\
z_1 = -x_2 \sin \theta + z_2 \cos \theta + t_z
\end{cases}
\end{align*}
\tag{11}
\]

\[
\begin{align*}
\begin{cases}
z_2 \theta + t_x = x_1 - x_2 \\
-x_2 \theta + t_z = z_1 - z_2
\end{cases}
\end{align*}
\tag{12}
\]

3. ICP algorithm

3.1 Dimension Reduction of ICP Algorithm
This paper verifies that ICP algorithm was faster and more accurate than RANSAC algorithm, so the ICP algorithm was used for inter-frame motion estimation in this paper. Because the robot camera moved in three-dimensional space, the general ICP algorithm was to solve the problem according to three-dimensional. However, the planar wheeled robot belonged to two-dimensional motion. To solve its motion parameters, only the vertical method angle and horizontal translation were considered, that is, only the solution \( \theta \), \( t_s \) and \( t_r \) were required. On this basis, the three-dimensional ICP algorithm was reduced, and the algorithm was verified by the first 100 images collected by the wheeled robot in the TUM (rgbd_dataset_freiburg2_pioneer_slam3) data set in the warehouse [16]. The TUM dataset provided a standard pose for the camera. According to the standard pose, this paper calculated the inter-frame motion of the first 100 images, and the formula was shown in Formula (13). Where \( T_0 \) represents the camera pose corresponding to the first frame of image, \( T_1 \) denotes the camera pose corresponding to the second frame of image, \( T \) represents the motion of the camera from the first frame to the second frame of image, including three-dimensional rotation matrix and three-dimensional translation component.

\[
T = T_0^{-1} T_1
\tag{13}
\]

The camera 's motion along the Y-axis (vertical) direction was shown in Figure 2. It could be seen from Figure 2 that the corresponding numerical change on the Y axis was almost zero and negligible. Therefore, this paper reduced the dimension of the three-dimensional space ICP algorithm and
estimated the parameters $\theta, t_x, t_z$ of the two-dimensional motion of the camera [17]. This could simplify the operation and further improved the calculation speed of inter-frame motion estimation.

![Fig. 2 Changes in Y value of ordinate](image)

### 3.2 Two-dimensional ICP algorithm

Let the point set of the three-dimensional space point corresponding to the pixel matching point of the first frame image projected onto the X and Z plane was $P$, and the point set corresponding to the second frame image was $P'$, which was expressed as Equation (14). Where any point in $P$ was $(P_{ix}, P_{iz})$ and any point in $P'$ was $(P_{ix}', P_{iz}')$. The camera motion includes rotation matrix $R$ and translation vector $t$. Both $(P_{ix}, P_{iz})$ and $(P_{ix}', P_{iz}')$ satisfied Eq. (15). The rotation matrix $R$ was a two-dimensional rotation matrix, and the translation vector $t$ was composed of the translation components in the X-axis and Z-axis, which was shown in formula (16). The construction error in formula (16) was shown in formula (17), where $e_i$ represents the coincidence error of the corresponding matching points in the two frames in the two-dimensional space [18]. The least square problem was constructed to solve the rotation matrix $R$ and the translation vector $t$, as shown in Equation (18).

$$\forall i, p_i = Rp_i' + t$$  \hspace{1cm} (14)

$$R = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}, \quad t = (t_x, t_z)'$$  \hspace{1cm} (15)

$$e_i = p_i - (Rp_i' + t)$$  \hspace{1cm} (16)

$$\arg \min_{R, t} \frac{1}{2} \sum_{i=1}^{n} \| (p_i - (Rp_i' + t)) \|_2^2$$  \hspace{1cm} (17)

The centroids corresponding to the two sets of points in Eq. (14) were obtained and expressed by $pc$ and $pc'$ as shown in Eq. (19). Formula (20) obtained from formula (18) and formula (19).

$$p_c = \frac{1}{n} \sum_{i=1}^{n} (p_i)$$

$$p_c' = \frac{1}{n} \sum_{i=1}^{n} (p_i')$$  \hspace{1cm} (18)

The first term in (20) was only related to the rotation matrix $R$, so that the first term was minimized to find the value corresponding to the rotation matrix $R$. Then the second term was substituted to make the second term zero, and the translation vector $t$ could be obtained [19]. The specific solving steps
were as follows: (1) Defined matrix $W$ as shown in (21); (2) SVD decomposition of matrix $W$ could obtain Equation (22), where was a diagonal matrix composed of singular values, and the diagonal elements were arranged from large to small; $u$ and $V$ were diagonal matrices; (3) The rotation matrix $R$ is solved as shown in Eq. (23); (4) The translation vector $t$ was solved as shown in (24).

$$\arg \min_{R,t} J = \frac{1}{2} \sum_{i=1}^{n} \| p_i - p_c - R(p_i' - p_c') \|^2 + \frac{1}{2} \sum_{i=1}^{n} \| p_c - Rp_c' - t \|^2$$

$$W = \sum_{i=1}^{n} p_i p_i'^\top$$

$$W = U \Sigma V^\top$$

$$R = U V^\top$$

$$t = p_c - Rp_c'$$

3.3 Constraint conditions are combined with two-dimensional ICP algorithm

In this paper, the contour constraint, line constraint and two-dimensional ICP algorithm were combined to estimate the motion trajectory of inter-frame camera. The results showed that the mismatch feature point screening method based on the constraint conditions in this paper could effectively screen out the mismatch feature point pairs generated by violent matching. The obtained high-quality matching point pairs were combined with the two-dimensional ICP algorithm. Under the premise of ensuring the accuracy of inter-frame motion calculation, the calculation speed was significantly improved compared with the three-dimensional ICP algorithm. The specific process was as follows: (1) Feature points were extracted from two adjacent frames of images and violent matching was performed; (2) Using contour constraint and linear constraint to screen out mismatched feature point pairs; (3) Formula (8) was used to calculate the three-dimensional space coordinates corresponding to the matching feature points; (4) The three-dimensional space points of two frames of images were projected onto the X, Z plane, that is, the obtained matching points were reduced; (5) Using two-dimensional ICP algorithm to solve the camera inter-frame motion.

4. Experimental results and related content analysis

This paper used TUM data set to verify the algorithm. The data set scene was indoor horizontal ground, and the data set also provided the standard pose information of the camera. Figure 3 showed the matching effect of the TUM dataset obtained by combining the contour constraint condition, line constraint condition and 3D ICP algorithm.

Fig. 3 Data set feature point matching result display

After obtaining the matching point pairs, different algorithms were used to calculate the camera pose. The previous contour constraint and line constraint were combined with the three-dimensional ICP algorithm, RANSAC algorithm and two-dimensional ICP algorithm, respectively. The camera motion was solved according to the matching feature points between frames, and then compared with the standard value in the TUM dataset to obtain the error value of camera motion estimation. The results were shown in Figure 4.
Fig. 4 Comparison of calculation results between ICP algorithm and RANSAC algorithm

Figure 4 (a) represented the error value of the vertical rotation angle $\theta$, Figure 4 (b) denoted the X-axis translation component $t_x$ error value, and Figure 4 (c) was the Z-axis translation $t_z$ error value. The ICP2DHL _ err curve was the error value between the constraint condition and the standard value after the combination of the two-dimensional ICP algorithm, the ICP3DHL _ err curve was the error value between the constraint condition and the standard value after the combination of the three-dimensional ICP algorithm, and the RANSACHL _ err curve was the error value between the constraint condition and the standard value after the combination of the RANSAC algorithm. $n$ represented the inter-frame camera motion transfer matrix number, for example, the 0th image to the 1st image inter-frame transferred matrix number $n=0$, the 1st image to the 2nd image inter-frame transfer matrix number $n=1$; the ordinate table error value in meters. It could be seen from Figure 4 that the RANSAC algorithm showed large calculation error in a few cases. Although the ICP algorithm had a certain error value, the overall data were relatively stable and the overall deviation was small. At the same time, the trend of data change was basically the same after dimensionality reduction of the three-dimensional ICP algorithm, and even the accuracy and stability of the three-dimensional ICP algorithm were better at some matching points. The running time corresponding to each algorithm was analyzed, as shown in Table 1. It could be seen from Table 1 that the combination of constraints and RANSANC algorithm taken the longest time, and the combination of constraints and ICP algorithm could greatly improve the computational efficiency. The efficiency of 3D ICP algorithm had been greatly improved after dimension reduction.

| Types               | Exercise time |
|---------------------|---------------|
| RANSAC algorithm    | 0.367209      |
| ICP_3D algorithm    | 0.050629      |
| ICP_2D algorithm    | 0.030401      |

From the perspective of average and variance, based on the constraint conditions of equal height and linear fitting, and combined with the two-dimensional ICP algorithm and the three-dimensional ICP algorithm, the average and variance of angle value $\theta$, X axis and Z axis corresponding to the two
methods were obtained, as shown in Table 2. It could be seen from Table 2 that the dimension reduction of the three-dimensional ICP algorithm had no significant effect on its accuracy. On the whole, the dimension reduction of the three-dimensional ICP algorithm had greatly improved the operation speed.

| Different algorithm | Angle value $\theta$ | X-axis | Z-axis |
|---------------------|----------------------|--------|--------|
|                     | Average value        | Variance | Average value | Variance | Average value | Variance |
| ICP_2D algorithm    | 0.004534             | 0.000105 | 0.004801    | 0.002483 | 0.001670    | 0.000028 |
| ICP_3D algorithm    | 0.005513             | 0.000136 | 0.004245    | 0.001967 | 0.001412    | 0.000165 |

5. Conclusion

Visual odometer was the basis of the whole visual SLAM. Based on the planar wheeled robot equipped with RGBD depth camera, according to its motion characteristics, the contour constraint condition and linear constraint condition were proposed. The constraint condition could effectively filter out the mismatch feature points, and the constraint condition could effectively eliminate the mismatch points with correct color image matching but large error of depth value. On this basis, this paper reduced the dimension of the three-dimensional ICP algorithm, and used the two-dimensional ICP algorithm to estimate the inter-frame motion, which had greatly improved the operation speed. The experimental results showed that the two-dimensional ICP algorithm based on the constraint conditions in this paper showed better effect in estimating the inter-frame motion trajectory of the camera, and the computational efficiency had been significantly improved.

References
[1] Jorge F P, José R A and Juan M R M 2015 Visual simultaneous localization and mapping: a survey Artif. Intell. Rev. 43(1) pp 55-81
[2] Goor P V, Mahony R, Hamel T and Trumpf T 2021 Constructive observer design for visual simultaneous localisation and mapping Automatica 132 pp 109803
[3] Cadena C, Carlone L, Carrillo H, Latif Y, Scaramuzza D, Neira J and Reid I, John J 2016 Past, present, and future of simultaneous localization and mapping: towards the robust-perception age IEEE T. Robot 32(6) pp 1309-1332
[4] Mahony R, Hamel T and Trumpf J 2021 An homogeneous space geometry for simultaneous localisation and mapping Annu. Rev. in Control 3(51) pp 254-267
[5] Jia Q, Gao X K, Luo Z X, Fan X and Guo H 2015 Feature point matching algorithm based on geometric relation constraint J. Computer-Aided Design Comput. Graphics 27(08) pp 1388-1397
[6] Lowe D G 2004 Distinctive image features from scale-invariant keypoints Int. J. Comput. vision 60(2) pp 91-110
[7] Song A L and Zhang J P 2010 Remote sensing image registration based on Retracted SURF Algorithm and Trajectories Generated From Lissajous Figures IEEE Geosci. Remote S. 7(3) pp 491-495
[8] Rublee E, Rabaud V, Konolige K and Bradski G 2011 ORB: An efficient alternative to SIFT or SURF IEEE Int. Conf. Comput. Vision (Barcelona) vol 1 IEEE USA pp 2564-2571
[9] Mair E, Hager G D, Burgschka D and Suppa M 2010 Adaptive and generic corner detection based on the accelerated segment test Comput. Vision 63(12) p 183
[10] Mohammad S and Morris T 2017 Binary robust independent elementary feature features for texture segmentation Adv. Sci. Lett. 23(6) pp 5178-5182
[11] Huang C Y, Han X and Han H Y 2015 The application of the improved RANSAC method in the fundamental matrix Comput. Eng. Design 36(03) pp 738-741, 758

[12] Wei L, Xue M Y, Huo J and Zhang J J 2020 Non-cooperative target feature point cloud registration optimization based on ICP algorithm J. System Simulation 32(12) pp 2383-2387

[13] Pathak K, Birk A, Vaskevicius N and Poppinga J 2010 Fast registration based on noisy planes with unknown correspondences for 3-D mapping IEEE T. Robot 26(3) pp 424-441

[14] Norstad I, Fagerholt K and Laporte G 2011 Tramp ship routing and scheduling with speed optimization Transport. Res. C-Emer. 19(5) pp 853-865

[15] Li SQ, Xie X P and Zhuang Y J 2018 Research on the calibration technology of an underwater camera based on equivalent focal length Measurement 122 pp 275-283

[16] Feng B, Liu G H, Zeng W L, Yu Y D and Zhang W K 2021. Research on SLAM algorithm of fusion robot based on stereo pointline feature and inertial navigation J.Chinese Comput. Systems 42(06) pp 1267-1275

[17] Li P, Wang R S, Wang Y X and Gao G 2020 Fast Method of Registration for 3D RGB Point Cloud with Improved Four Initial Point Pairs Algorithm Sensors 20(1)

[18] Potsis A, Reigber A, Mittermayer J, Moreira A and Uzunoglou N 2001 Sub-aperture algorithm for motion compensation improvement in wide-beam SAR data processing Electron. Lett. 37(23) pp 1405-1407

[19] Xiao X W, Guo B X, Li L H and Yang N 2016 Multi-view stereo matching based on self-adaptive patch and image grouping for multiple unmanned aerial vehicle imagery Remote Sensing 8(2) p 89