Research on Inter Frame Motion Optimization of Visual SLAM

Yibo Wang, Weishan Liang*

Liuzhou Institute of Technology, Liuzhou, Guangxi 545616, China

*Corresponding author’s e-mail: liangwsh@zju.edu.cn

Abstract. The key of visual SLAM is VO (front end) algorithm. Aiming at the problems of complex calculation and low robustness of feature point method widely used in VO, an algorithm for solving camera motion parameters based on chicken swarm optimization algorithm is proposed. Based on the camera pinhole model, the relationship between the two-dimensional image coordinates of adjacent frames and the camera motion parameters is deduced. The optimization function based on Hausdorff distance target is established, and the chicken swarm algorithm is defined combined with visual slam. The fitness function of the algorithm is designed. Based on the checkerboard calibration method, the algorithm is verified by experiments.

1. Introduction

Visual slam has the characteristics of low cost and wide application scenarios, but it has not been widely implemented at present, which is mainly limited by its complex algorithm and poor robustness. According to the current general architecture of visual slam, it includes: front-end (visual odometer), back-end, loop detection and mapping, in which the front-end (VO) is an important factor of algorithm performance. At present, the realization methods of VO mainly include feature point method and optical flow method [1-2]. The former has the problems of time-consuming key point extraction, insufficient information utilization and high requirements for scene texture. The advantage of optical flow method is that it not only carries the motion information of moving objects, but also carries rich information of 3D structure of the scene, but also has the problems of time-consuming, real-time and poor practicability. Therefore, it is important to improve the VO processing algorithm. In the process of camera motion, the change of its external parameters is reflected in the image as a combination of translation, rotation, staggering and scaling. If the camera motion model can be established, it is possible to use optimization method to reconstruct the camera motion relationship from two consecutive images. Therefore, this paper will face the application requirements of visual slam, take the image matching evaluation function as the objective function, and use the chicken swarm algorithm to optimize the camera motion parameters, so as to improve the efficiency of VO processing.

2. Visual slam principle based on image matching

2.1. Camera imaging model

Camera imaging is to project three-dimensional space points into two-dimensional image points. At present, the processing accuracy of the lens is getting higher and higher. The pinhole imaging model derived from the lens imaging principle can achieve a good approximation of the camera. If the lens distortion is not considered, the model is linear.
Suppose that the world coordinate of any point \( P \) in space is \((x_w, y_w, z_w)\). The coordinate in the camera coordinate system is \((x_\gamma, y_\gamma, z_\gamma)\). The physical coordinates of the corresponding image point \( P \) is \((x, y)\). The pixel coordinates are \((u, v)\), and \( F \) is the focal length of the camera.

\[
\begin{align*}
    x &= f \frac{x_\gamma}{z_\gamma} \\
    y &= f \frac{y_\gamma}{z_\gamma}
\end{align*}
\]  

(1)

In order to represent rotation and translation as matrix transformation, space points and image points are represented as homogeneous coordinates \((x_\gamma, y_\gamma, z_\gamma, 1)\) and \((u, v, 1)\), then the relationship between the world coordinates of the space point and the pixel coordinates of the image point is as follows:

\[
\begin{bmatrix}
    u \\
    v \\
    1
\end{bmatrix} = K \begin{bmatrix} R & t \end{bmatrix}
\begin{bmatrix}
    x_\gamma \\
    y_\gamma \\
    z_\gamma \\
    1
\end{bmatrix}
\]  

(2)

\( K = \begin{bmatrix} f_u & 0 & u_0 \\
                        0 & f_v & v_0 \\
                        0 & 0 & 1 \end{bmatrix} \) is the internal parameter matrix of the camera. \((u_0, v_0)\) is the coordinate of the origin of the image physical coordinate system in the pixel coordinate system, \( f_u = f/dx \) and \( f_v = f/dy \) which is the equivalent focal length of U-axis and v-axis respectively, can be obtained by calibration. \([R, t] \) represents the camera external parameter matrix, where \( R \) is the rotation matrix of the camera coordinate system relative to the world coordinate system, and \( t \) is the translation vector of the camera coordinate system relative to the world coordinate system, which dynamically reflects the camera attitude and position.

2.2. Camera motion optimization model

Assuming that the camera and the robot are connected by rigid body, when the camera moves with the robot, the pose of the camera coordinate system will change continuously, and the image will be collected at discrete time, and its optical centre can be expressed as \( O_1, O_2, \ldots, O_n \). The camera motion of two adjacent positions can be expressed as \((R_k, t_k)\).

Therefore, the slam motion optimization idea based on image matching is as follows: the camera coordinate system of the previous image acquisition time is \( O_k = x_{ck}y_{ck}z_{ck} \) is regarded as the world coordinate system, the motion of the later time relative to the previous time can be determined by the camera external parameter \((R_k, t_k)\) Description. By searching the external parameter \((R_k, t_k)\) In order to achieve the optimal matching between the image acquired at the next moment and the image acquired at the previous moment, the motion relationship between adjacent moments can be solved. The specific steps are as follows:

1. Gaussian filtering is applied to the images collected at adjacent times.
2. Canny operator is used to extract the filtered image edge, and the image edge feature set \( \Omega_k = \{f_{11}, f_{12}, \ldots, f_{1n}\} \) at the previous time is assumed, the edge feature set \( \Omega_k = \{f_{21}, f_{22}, \ldots, f_{2m}\} \) at the next moment.
3. According to the pinhole model of the camera, the relationship of image points between adjacent frames can be deduced.

\[
\begin{bmatrix}
    u_k' \\
    v_k' \\
    1
\end{bmatrix} = \frac{x_{ck}}{z_{ck}} \begin{bmatrix} 1 \\
                        0 \\
                        0 \\
\end{bmatrix} + \frac{1}{z_{ck}} R_k K^{-1} K t_k
\]  

(3)

\[
t_k = (t_{ck}, t_{yk}, t_{zk}) \quad R_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\
                        0 & \cos \psi_k & \sin \psi_k & 0 \\
                        0 & -\sin \psi_k & \cos \psi_k & 0 \\
                        0 & 0 & 0 & 1 \\
\end{bmatrix}
\]  

(4)

After calculation, the feature set \( \Omega_k' = \{f_{k1}, f_{k2}, \ldots, f_{kn}\} \) of the previous time is obtained from \( \Omega_k' = \{f_{k1}, f_{k2}, \ldots, f_{kn}\} \).

4. The optimization objective function is established.

\[
\min h'_{k,e}(\Omega_k', \Omega_{k+1})
\]  

(5)
The Hausdorff distance between two sets \( \Omega_k \) and \( \Omega_{k+1} \) is denoted by \( H_{HF}(\Omega_k, \Omega_{k+1}) \). Similarly, \( H_{HF}^f(\Omega_k, \Omega_{k+1}) \) denotes the Hausdorff distance of the forward part. \( f_{ki} \) is the i-th edge point component of the set \( \Omega_k \), and \( f_{(k+1)j} \) is the j-th edge point component in \( \Omega_{k+1} \). \( \beta \) is a numerical value between 0 and 1, which means to take some points in the set. Similarly, \( H_{HF}^b(\Omega_k, \Omega_{k+1}) \) denotes the backward part Hausdorff distance.

The optimized parameters include rotation \((\theta_k, \varphi_k, \psi_k)\) and peaceful migration \((t_{xk}, t_{yk}, t_{zk})\).

### 3. Chicken Swarm Motion Optimization Algorithm

Chicken swarm optimization (CSO) is a new bionic algorithm proposed by Meng Xiangbing at the Fifth International Conference on swarm intelligence (ICSI) in 2014[4]. Inspired by the hierarchical structure and foraging behaviour of chicken flock, the algorithm divides the chicken flock into several subgroups led by roosters. Each subgroup is composed of roosters, several hens and chicks, which are updated according to the inherent behaviour rules. The optimal solution is obtained through constant iterations, keeping the correlation coefficient of hens following Roosters and chicks following hens unchanged. Compared with other artificial intelligence optimization algorithms, CSO has the advantages of good convergence speed and high convergence accuracy.

#### 3.1. Fitness Function

According to the principle of chicken colony optimization algorithm, the fitness function is the basis to determine the rooster, hen and chick. Roosters’ fitness value is small, the search vitality is strong, and they are in the dominant position of the subgroup; hens’ fitness value is slightly larger, mainly with the rooster searching for food; chickens’ fitness value is the largest, the search ability is the worst, with the hen searching for food, mainly reflecting the local search.

Appropriate fitness function is very important for optimizing camera motion parameters. Generally, the objective function is taken as the fitness function, so according to equation (5), the fitness function of camera motion optimization is as follows:

\[
F_{fit}(c_i) = -H_{HF}^b(\Omega_k^i, \Omega)
\]  

Among them, \( c_i = (\theta_i, \varphi_i, \psi_i, t_{xi}, t_{yi}, t_{zi}) \) is the i-th chicken using floating-point code. If the interval of image acquisition is small, the changes of rotation and translation are relatively small, which can be considered to occur in a limited range. Let the rotation change interval be \([-\Phi, \Phi]\), and the translation transform interval be \([-\Delta, \Delta]\).

The fitness values are ranked from large to small, and the maximum value of fitness function is taken as the optimal individual in the chicken group, and the optimal individual is output at last.

#### 3.2. Population Initialization

The quality of chicken colony initialization has a great influence on the convergence speed and optimization accuracy of the algorithm. How to generate a chicken colony with good randomness and ergodicity is the focus of attention. Therefore, the random initialization method based on logistic and Chebyshev chaotic system is selected in this paper. The chaotic sequence generated by this method is more uniform and has better scattering and diffusion effect than other random methods in \((-1,1)\)[5].

Assuming that the total number of chickens is \(n\), the initial population generation steps are as follows:

1. Random initial chaotic variable \(u_1, v_1 \in (-1,1)\), the chaotic sequence \(U = \{u_1, u_2, \ldots, u_{6N}\}\), and \(V = \{v_1, v_2, \ldots, v_{6N}\}\) containing 6N chaotic variables is generated according to the following formula.

\[
\begin{align*}
  u_{i+1} & = 1 - \lambda u_i^2, u_i \in (-1,1) \\
  v_{i+1} & = \cos(\rho \cdot \arccos(v_i)), v_i \in (-1,1)
\end{align*}
\]  

(7)
The values of $\lambda$ and $\rho$ are 2.

(2) The new sequences $U'$ and $V'$ are obtained by taking the values of sequences $U$ and $V$ in the limited interval.

$$\begin{align*}
u'_m = u_i, & u_i \in (-0.1, 0.1) \\
u'_n = v_i, & v_i \in (-0.8, 0.8) \quad \text{(8)}
\end{align*}$$

(3) A new random sequence $S = \{s_1, s_2, \ldots, s_{6N}\}$ on sequence (-1,1) is constructed by using sequences $U'$ and $V'$.

$$s_j = \begin{cases} u'_m \times 2 & \text{mod}(j, 5) = 0 \\
v'_n - 0.2 & \text{mod}(k, 5) \neq 0 \text{ and } v'_n < 0 \\
v'_n + 0.2 & \text{mod}(k, 5) \neq 0 \text{ and } v'_n > 0 \quad \text{(9)}
\end{cases}$$

(4) When the random sequence $s$ is transformed into the search space, the $k$-th chicken $C_k$ can be initialized as:

$$\begin{align*}
\theta_k &= s_{6k+1} \Phi \\
\varphi_k &= s_{6k+2} \Phi \\
\psi_k &= s_{6k+3} \Phi \\
t_{xk} &= s_{6k+4} \Delta \\
t_{yk} &= s_{6k+5} \Delta \\
t_{zk} &= s_{6k+6} \Delta \quad , \quad k = 1, 2, \ldots, N.
\end{align*}$$

3.3. Population regeneration

The chickens were renewed in the order of rooster, hen and chick. The rooster position is updated as follows:

$$c_{k+1}^t = c_k^t + (c_{\text{best}}^t - c_k^t) \cdot X \quad \text{(10)}$$

Among them, $X \sim N(0, \sigma^2)$, $\sigma^2 = \left\{ \begin{align*} 1, & \text{if } F(tt(c_k)) \leq F(tt(c_{\text{best}})) \\
\exp\left(\frac{F(tt(c_{\text{best}})) - F(tt(c_k))}{|F(tt(c_{\text{best}})) + \varepsilon|}\right), & \text{otherwise}
\end{align*} \right.$ $\varepsilon$ is a very small positive number, $c_1$ is divided by any rooster other than $c_k$; $c_{\text{best}}^t$ is the global optimal individual in the $t$-th iteration.

The formula for updating the position of hens is as follows:

$$c_{k+1}^t = c_k^t + C_1 \cdot Y \cdot (c_{kk}^t - c_k^t) + C_2 \cdot Y \cdot (c_k^t - c_{\text{best}}^t) \quad \text{(11)}$$

Among them, $C_1 = \exp\left(\frac{F(tt(c_{\text{best}})) - F(tt(c_k))}{|F(tt(c_{\text{best}})) + \varepsilon|}\right)$, $C_2 = \exp(F(tt(c_k)) - F(tt(c_{\text{best}})))$, $Y \sim U(0, 1)$ $\varepsilon$ is a very small positive number, $c_{kk}$ is the cock followed by the $k$th hen, $c_1$ is divided by any hen outside $c_k$.

The chicken position was updated as follows:

$$c_{k+1}^t = \omega \cdot c_k^t + Z_1 \cdot (c_{kh}^t - c_k^t) + Z_2 \cdot (c_{khc}^t - c_k^t) \quad \text{(12)}$$

Among them, $\omega$ is inertia weight, which controls the global and local search ability of the algorithm. When the value of $\omega$ is larger, it is conducive to traverse the search space, and the global search ability is better. When the value of $\omega$ is smaller, it is conducive to local detailed search, and the local search ability is better. We usually want to jump out of the local optimum at the beginning of the algorithm and get a larger value of $\omega$, while at the end of the algorithm, we can accurately search the region and get a smaller value of $\omega$.

For this reason, $\omega = \omega_{\text{min}} \cdot \left(\frac{\omega_{\text{max}} - \omega_{\text{min}}}{\text{maxgen}}\right)^{\gamma} \frac{\text{iter}}{\text{maxgen}}$, $\text{maxgen}$ is the maximum number of iterations $c_{kh}$ is the hen the chick follows, $c_{khc}$ is the rooster that the hen follows, $Z_1 \sim U(0, 2)$, $Z_2 \sim U(0, 2)$.

3.4. Algorithm flow

According to the general steps of chicken swarm optimization algorithm, the flow of camera motion optimization algorithm is designed, as shown in Figure 1.
Figure 1. Algorithm flow  
Figure 2. Checkerboard calibration image

4. Experimental analysis

In order to verify the proposed camera motion optimization algorithm, the checkerboard image calibration method is proposed. Firstly, nine checkerboard images are collected, as shown in Figure 2. The traditional Zhang calibration method is used to calibrate the internal and external parameters of the camera [6]. The results are shown in Table 1 and Table 2.

| Camera calibration parameters | $f_x$ | $f_y$ | $u_0$ | $v_0$ |
|------------------------------|-------|-------|-------|-------|
|                              | 1379.22 | 1375.63 | 732.68 | 533.95 |

Table 2. Camera calibration parameters

| | $\theta$ | $\varphi$ | $\psi$ | $t_x$ | $t_y$ | $t_z$ |
|------------------------------|---------|---------|--------|-------|-------|-------|
| 1 | -0.87 | 2.86 | -0.08 | -3.59 | -125.70 | 899.72 |
| 2 | -1.58 | 2.57 | -0.44 | -52.83 | -30.81 | 948.84 |
| 3 | 1.85 | 2.27 | 0.01 | -48.96 | -23.89 | 1048.43 |
| 4 | -0.19 | 2.82 | -0.22 | -49.73 | -59.03 | 876.36 |
| 5 | -1.11 | 2.92 | -0.12 | 10.43 | -44.31 | 1003.58 |
| 6 | 0.59 | 2.80 | -0.25 | -49.57 | -14.32 | 884.24 |
| 7 | 1.66 | 2.48 | -0.18 | -54.04 | -31.74 | 1038.94 |
| 8 | 2.32 | 1.87 | -0.13 | -76.66 | -13.96 | 1052.15 |
| 9 | 0.03 | 2.92 | 0.00 | -109.73 | 7.45 | 942.77 |
The above external parameters are the changes relative to the same reference position. Due to the mature chessboard calibration method, the re projection error is 0.26597 ~ 0.30165 pixels, so the general application requirements of slam are composite as the reference object.

On the basis of the above calibration, the motion parameters between adjacent frames are calculated by using the optimization algorithm proposed in this paper, that is, the previous frame is regarded as the reference of the next frame. Here, the population size is 50, the proportion of roosters is 0.1, the proportion of hens is 0.3, the proportion of chicks is 0.5, the role renewal algebra is set to 8, the maximum number of iterations is 120, and the inertia weight is $\omega_{\text{min}} = 0.3 $, $ \omega_{\text{max}} = 0.8 $. The results are shown in Table 3.

|   | $\theta$ | $\varphi$ | $\psi$ | $t_x$ | $t_y$ | $t_z$ |
|---|---|---|---|---|---|---|
| 1-2 | -1.41 | -1.28 | -0.50 | -46.58 | 92.65 | 49.97 |
| 2-3 | 4.08 | 0.25 | 1.27 | 6.82 | 8.72 | 96.83 |
| 3-4 | -1.96 | 1.18 | -0.87 | -3.90 | -33.59 | -169.56 |
| 4-5 | 0.07 | 0.84 | -0.37 | 60.06 | 11.35 | 124.77 |
| 5-6 | 0.86 | -0.95 | -0.84 | -60.54 | 26.18 | -119.28 |
| 6-7 | 0.96 | -0.52 | -0.66 | -3.01 | -17.44 | 156.69 |
| 7-8 | -0.13 | -1.09 | 0.79 | -20.53 | 22.38 | 17.12 |
| 8-9 | -1.37 | 1.65 | 0.29 | -30.52 | 19.81 | -104.79 |

Table 3 shows the relative motion parameters of the camera between two frames. By comparing with the data obtained from calibration, the average translation error of the camera optimization algorithm is 2.23, and the average rotation error is 0.60.

5. Conclusion

Aiming at the problems of feature point method and optical flow method which are widely used in VO of vision slam, this paper seeks to establish a more simple and stable camera motion parameter optimization method. This method makes full use of the characteristics of small sampling interval and limited range of motion parameters, and can achieve effective search through chicken swarm algorithm. The experimental results show that the algorithm can replace the feature point method. At present, the algorithm is still running on the premise of a given scale. If the scale changes greatly, it can not be dealt with. In the future, we will carry out in-depth research around this problem to continuously improve the applicability of the algorithm.

Acknowledgments

This paper is supported by the National Natural Science Foundation of China (no.54105510), the Guangxi science and technology base and talent project (2020ac19115), and the project to improve the basic scientific research ability of young and middle-aged teachers in Guangxi Universities (2020ky60013).

References

[1] C. Cadena, L. Carlone, H. Carrillo, et al. Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age[J]. IEEE Transactions on Robotics,2016,32(6): 1309-1332.

[2] B. Guillaume, A. Zayed, Y. Li, et al. Simultaneous Localization and Mapping: A Survey of Current Trends in Autonomous Driving[J]. IEEE Transactions on Intelligent Vehicles,2017,2(3): 194-220.
[3] D. Huttenlocher, G. Klanderman, W. Rucklidge. Comparing Images Using the Hausdorff Distance[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence. 1993, 15(9):850-863.

[4] Meng X, Liu Y, Gao X, et al. A new bio-inspired algorithm: chicken swarm optimization [C] International Conference in Swarm Intelligence, Hefei: Springer International Publishing, 2014: 86-943.

[5] LI Yuan, GAO Shang, LI Zhaoji, et al. Multi-level sub-population invasive weed optimization algorithm with new differential evolution model [J]. Computer Engineering and Applications, 2018, 54(21): 107-114.

[6] Zhang Z Y. Camera calibration with one-dimensional objects[J].IEEE Trans on pattern Analysis And Machine Intelligence, 2004(7): 892-899.