An improved ORB-SLAM2 algorithm based on image information entropy

Xiaokang Ren*, Yongye Wang*, Hongxiang Wang, Xingzhen Li, Xingxing Liu

School of computer science and Engineering, Northwest Normal University, Lanzhou Gansu, 730070, China
1872985675@qq.com
*renxk@nwnu.edu.cn

Abstract. Aiming at the feature point extraction process in visual ORB-SLAM2, a visual SLAM method based on information entropy is proposed. In the process of ORB-SLAM2, due to the uncertainty of illumination and surrounding environment information, when the environment information is reduced, the number of feature points is likely to decrease dramatically, which results in the failure of SLAM process. In order to solve the above problems, a visual SLAM method based on image information entropy is proposed. According to the information entropy of the image, the amount of information is determined, and the image with low contrast and small gradient change is screened. The image is enhanced, and the feature points that can represent the image information are extracted as much as possible as the correlation basis of adjacent frame matching and key frame matching. In order to improve the performance of the mobile robot system, the local and global optimization of cumulative error of position and attitude is carried out by combining the optimization method of attitude graph.

1. Introduction

Simultaneous Localization and Mapping (SLAM) technology allows the robot to start from an unknown location, locate its position and posture through the map features repeatedly observed by the mounted vision, laser and other sensors during the movement, to then according to its own position and incrementally construct maps to achieve the goal of simultaneous positioning and map construction, and provide necessary support for robot autonomous positioning, obstacle avoidance, navigation, path planning and other tasks [1-3]. In SLAM applications, compared with laser sensors that can directly obtain the depth of the environment, the visual SLAM with image sensors [4] is cheaper and more popular, so it has a wider application prospect.

There are still many challenges in the research of visual SLAM. In the indirect method using sparse features, there are some problems, such as how to design an optimal feature to represent image information, how to make the feature more robust in the environment of illumination change, camera parameter change or insufficient texture information. In the direct method, although the image information can be used to the maximum, if there are special circumstances such as occlusion, diffuse reflection, or light changes, the premise that the optimized pixel value is the objective function is not true. In addition to the key technologies discussed above, visual SLAM research also has special requirements for the application environment. In complex and multi-dynamic target motion environments, SLAM systems are more vulnerable. Real-time applications for long or long distances require robustness of the system higher. In addition, the visual SLAM posture tracking results are extremely...
This paper proposes a visual SLAM enhancement optimization algorithm based on image information entropy, which improves the problem of image matching failure and tracking loss caused by insufficient feature extraction during the operation of the visual SLAM system, and improves the visual positioning accuracy and system robustness of mobile robots. This paper first introduces the key technology and feasibility of visual SLAM. Combined with the practical problems in application, such as the image texture or structural features are not rich, tracking loss occurs when moving fast or rotating at a large angle, etc., an image feature extraction method based on information entropy is proposed. Then, slam system is constructed, and the open source data set verification and analysis are carried out compared with orb-slam2 system. Finally, the relevant conclusions are drawn.

2. Improved orb-slam system

2.1. ORB-SLAM2 overall process
ORB-SLAM2[5] completes the matching and tracking of image sequences based on ORB features, even loop detection and relocation, and provides interfaces that can be used under monocular, binocular, RGB-D and ROS (robot operating system). The algorithm is divided into three threads: tracking, local map construction, and loop detection. The ORB feature point detection method is used, and the Bundle adjustment (BA) is used to perform nonlinear iterative optimization on the three processes of tracking, composition and closed-loop detection, so as to obtain accurate camera position information and three-dimensional map points. The overall process is shown in Figure.1.

![Figure.1 ORB-SLAM2 system framework](image)

After the image is collected by the camera, the feature points are first extracted. Feature point extraction can convert image information into feature point information. The preprocessing process includes three stages: the detection and recognition process of feature points, which solves the problem of where the feature points are; the description and expression process of feature points, solves the problem of what the feature points are; the classification and matching process of feature points, To solve the problem of where the feature points belong.
2.2. Algorithm improvements

In the actual environment, due to the uncertainty of illumination and surrounding environment information, when the contrast of the environment decreases, it will easily cause the number of feature points to decrease sharply and cause the failure of the SLAM process. In order to solve such problems, this paper proposes a visual SLAM optimization algorithm with relatively high image correlation accuracy. It proposes to filter images based on the image information entropy value, and enhance the details of the filtered images to improve the characteristics of visual SLAM. Insufficient problems such as image matching failures improve the visual positioning accuracy and system robustness of mobile robots. The overall process is shown in Figure.2.

![Figure.2 Improved ORB-SLAM2 system framework](image)

First, build an image pyramid to increase the scale-invariant characteristics of features. Because each layer of image is scaled according to the scale factor, in order to achieve uniform extraction of image features and reduce the complexity of subsequent image matching, in different scale spaces, according to the size of the image block definition (width, height) Segment the image. Quickly extract corner features for each image block layer by layer, assign and manage key points through the quadtree method, and select the most influential features (the strongest features) as the node according to the preset threshold Features are retained to achieve uniform distribution of image features. At the same time, the rotation-invariant characteristic of the feature is added, the actual position of the feature point in the image is calculated, and the feature descriptor is generated and saved.

2.3. Image screening based on information entropy

The information entropy was first proposed by the American engineer Shannon. The information entropy of the image can directly represent the amount of information contained in the image, and it is also a measure of the uncertainty of the information in the image. In this paper, in order to more effectively extract the features that characterize the texture or structural information of the image, while reducing the amount of feature extraction calculations, the statistical form of the information features of the image entropy is used to calculate the gray distribution and probability of the image:

\[
H(x) = - \sum_{i=0}^{255} p(x_i) \log_2 p(x_i)
\]  

(1)

Among them, the one-dimensional entropy of the image \(H(x)\) represents the clustering characteristics of the grayscale distribution of the image, and \(p(x_i)\) represents the probability of pixels with grayscale \(i (i = 0, 1 \ldots 255)\) appearing in the image. Conversely, if the entropy value is small, the pixel
gradient of the image is considered small, and the features extracted from the image do not contribute much to the image matching, and image enhancement can be performed. According to the information feature statistics of the image, the closer $H(x)$ is to 0, the smaller the pixel gradient value in the image is, and the texture information of the image is insufficient. Therefore, $H(x)$ is compared with the threshold $R$ to complete the screening. The threshold $R$ here is an empirical value.

2.4. Algorithm improvements

The Laplacian algorithm is a linear quadratic differential operator with rotation invariance, which satisfies the image edge sharpening requirements in different directions, and its obtained borders are relatively thin, including more detailed information. Laplacian operator:

$$\nabla^2 f(x, y) = \frac{\partial^2 f(x, y)}{\partial x^2} + \frac{\partial^2 f(x, y)}{\partial y^2}$$

(2)

The basic formula for enhancing images using Laplacian operator is:

$$g(x, y) = f(x, y) + c[\nabla^2 f(x, y)]$$

(3)

Where $g(x, y)$ is the enhanced image, $f(x, y)$ is the input image, when $c$ is -1, the middle of the convolution kernel is a negative number, and when $c$ is 1, the middle of the convolution kernel is a positive number. Since the laplacian operator is a differential operator, it sharpens the image and makes the constant area 0. Laplacian filtering the image with information entropy less than $R$ to enhance the details of the image.

3. Test Results and Discussions

The system uses the TUM data set fr1_floor for experimental verification. The image sequence reading speed is set to 30 frames per second. The tracking results are calculated with the standard trajectory data. At the same time, they are compared and analyzed with the open source ORB-SLAM2 system. All experiments were completed on a laptop with Intel i5-4210 CPU, clocked at 2.4GHz, quad-core, and 8GB of memory. The back end of the system uses G2O for posture optimization based on posture maps to generate motion trajectories.

3.1. Feature extraction

In this paper, the fr1_floor data set in TUM is used to test the ORB and the improved ORB feature point extraction algorithm. The results are shown in Figure.3. It can be seen from the figure that the improved ORB feature point extraction algorithm can obtain more and more uniform feature points on the image compared to the ORB feature extraction algorithm.

(a) Optimization algorithm

(b) ORB-SLAM2 algorithm

Figure.3 Feature extraction comparison
3.2. Tracking accuracy

In tracking accuracy, the absolute trajectory error reflects the difference between the motion estimation value and the standard value. Taking the fr1_floor data set as an example, the error between the optimized motion trajectory and the real trajectory is shown in Figure.4.

The root mean square error (RMSE) of the absolute trajectory error reflecting the measurement precision is used as the evaluation standard. The absolute trajectory error is defined as follows:

$$RMSE(x) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \| x_{e,i} - x_{s,i} \|^{2}}$$

(4)

Among them, $x$ represents the estimated position value of the i-th frame in the image sequence, and $y$ represents the standard value of the position of the i-th frame in the image sequence.

In addition, the relative pose error is a standard measure of relative pose conversion. Since the relative pose includes a translation component and a rotation component, the relative pose is usually decomposed into a pose translation part and a pose rotation part. Evaluation. The error between the relative posture optimized by the algorithm and the standard relative posture is shown in Figure.5.

Similarly, the root mean square error (RMSE) of relative pose error reflecting the precision of measurement is used as the evaluation criterion, and the root mean square error $RMSE(T)$ of relative pose error is defined as follows:

$$RMSE(T) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \| T_{e,ij} - T_{s,ij} \|^{2}}$$

(5)
Among them, $T_{e,i}^j = T_{s,i}^j - T_{s,j}$ represents the relative transformation of the estimated pose between the i-th frame and the j-th frame in the image sequence, and $T_{s,i}^j = T_{s,i}^j - T_{s,j}$ represents the relative transformation of the standard pose between the i-th frame and the j-th frame in the image sequence.

By comparing the motion trajectory errors of the ORB-SLAM2 algorithm and the optimized SLAM algorithm, the method proposed in this paper has improved the accuracy of motion tracking. The test results based on the TUM data set fr1_floor are shown in Table 1.

| SLAM algorithm | Average processing time | Absolute trajectory error | Relative translation error |
|----------------|-------------------------|----------------------------|---------------------------|
| orb-slam2      | 0.030s                  | 0.0521m                    | 0.0433m                   |
| optimization   | 0.033s                  | 0.0450m                    | 0.0408m                   |

Among them, the average processing time is slightly increased due to the detail enhancement of the local image, but the tracking trajectory error and translation error index have decreased.

### 3.3. Sports robustness

The optimization algorithm proposed in this paper shows strong robustness in the fast motion of visual sensors, and can solve the problem of motion tracking loss caused by the failure of inter-frame matching. Taking fr1_floor data set as an example, its SLAM tracking result is shown in Figure 6. In Figure a, since the inter-frame features cannot be used for effective matching, the tracking can only be continued when the camera is repositioned to the original position. In Fig. b, because the details of the image retained after filtering based on information entropy are enhanced and useful features are added to improve the matching success rate, the system tracking is successful. Multiple test results show that the success rate when using ORB-SLAM2 to track the fr1_floor data set is less than 30%, and the success rate of SLAM motion tracking using optimization algorithms can be increased to more than 50%.

(a) ORB-SLAM2 algorithm tracking results  
(b) Optimization algorithm tracking results  

**Figure 6** Trajectory tracking results

### 4. Conclusion

This paper proposes a visual SLAM optimization algorithm based on image information entropy, which uses image information entropy to judge the size of image information, select images with low contrast and small gradient changes, and enhance the details of the selected images. The enhanced image is subjected to feature extraction to improve the tracking accuracy of the SLAM system. Although this optimization algorithm has an increase in processing time compared with ORB-SLAM2, it can
realize that the mobile robot can still effectively perform motion estimation when the visual sensor rotates at a large angle and the image texture information is insufficient. The optimization algorithm was verified on the TUM data set fr1_floor. Experiments show that after applying the optimization algorithm proposed in this paper, the tracking success rate of the SLAM system can be increased to 50%, while reducing the predicted trajectory error value. This algorithm is robust to the changes of the environment and to the motion of the robot. When the visual sensor moves fast, it can solve the normal trajectory tracking problem of the mobile robot when the texture information is insufficient.

References
[1] CADENA C, CARLONE L, CARRILLO H, 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] Henein M, Zhang J, Mahony R, et al. Dynamic SLAM: The Need For Speed[J]. arXiv: Robotics, 2020.
[3] Liu Q, Wang Z, Wang H, et al. SD-VIS: A Fast and Accurate Semi-Direct Monocular Visual-Inertial Simultaneous Localization and Mapping (SLAM)[J]. Sensors, 2020, 20(5).
[4] Zeng Weilin, Liu Guihua, Chen Hao. Monocular vision SLAM based on probabilistic motion statistical feature matching[J]. Computer Engineering, 2019,45(12):222-231+236(in Chinese).
[5] Mur-Artal R, Tardos J D. ORB-SLAM2: an open-source SLAM system for monocular, stereo and rgb-d cameras. IEEE Transactions on Robotics, 2017, 33(5): 1255–1262.