Research on the Effectiveness of Monocular Visual SLAM Depth Estimation Base on Improved ORB Algorithm

Jiwu Wang
School of Mechanical and Electronic Engineering, Beijing Jiaotong University
Beijing, Haidian District, China

Weipeng wan
School of Mechanical and Electronic Engineering, Beijing Jiaotong University
Beijing, Haidian District, China
E-mail: jwwang@bjtu.edu.cn, 13014641154@163.com
www.bjtu.edu.cn

Abstract

The application of monocular vision to measure the depth information of image feature points is one of the important tasks of monocular vision slam. Triangulation is a method often used to measure the depth information of feature points, but in the actual application process, the uncertainty of feature point matching will cause greater depth uncertainty. This paper proposes an improved ORB feature point extraction strategy, combined with the quad-tree model to achieve the homogenization of feature points. The feature points matching method uses the brute force matching method, RANSAC filters the matching point pairs, and obtains better matching results for depth estimation. Experiments show that the improved feature point extraction and matching method can effectively obtain camera pose estimation value and depth estimation value. And in this way, the accuracy of the estimated value is improved .

Keywords: orb；monocular vision；depth estimation; vision slam.

1. Introduction

Measuring the depth information of image feature points is a very important part of orb-slam or other monocular vision slam. In the visual odometer based on feature point method, the motion of the camera is estimated with the extracted image feature points.[1] The basic process includes image feature point extraction, image feature point matching between frames, camera motion solution by epipolar constraint . The camera rotation matrix R and translation matrix T obtained at this time are used to estimate the image feature point depth. The feature point extraction and matching in the above process affects the estimation of the rotation matrix R and the translation matrix T, and also affects the estimation of the image depth information.

In view of the above problems, we can find that a good feature point extraction and matching method will bring better pose estimation and better depth estimation. Among the many feature point extraction strategies, the ORB algorithm has the advantages of smaller calculation amount and simple feature point information description. It is widely used in slam based on the feature point method, but the accuracy and stability of traditional ORB detection algorithm in slam need to be further improved.

2. Principle

2.1. *The principle of ORB feature point extraction*

The ORB algorithm is improved on the basis of the FAST algorithm. ORB adds a description of the direction to FAST to ensure rotation invariance and uses the Steer BRIEF descriptor to describe the feature points.[2] The principle steps of the algorithm are as follows:

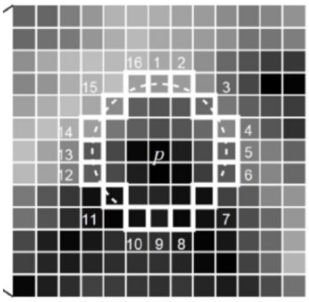


Fig. 1. FAST key points

Step1:Extract FAST key points. As shown in Figure 1. We arbitrarily select a pixel point p in the image, and the gray value of point p is Ip. The gray value of pixel p is compared with the gray value of 16 pixels on a circular window with a pixel radius of 3. Set a threshold t. If there are consecutive N (set to 9 or 12) pixels on the circle whose gray value is greater than Ip+t or less than Ip-t, then the pixel p can be considered as a FAST key point.

Step 2:Calculate the direction of feature points. Redefine the coordinate system with the key point p as the center O as shown in Figure 2.

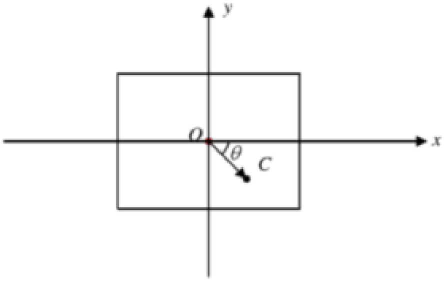


Fig. 2. Key point direction

The direction of the key point can be calculated according to the following equation. In the equation, x and y represent the coordinates of the pixel, I represents the gray value of the pixel, B represents pixel block, C represents the centroid position of the pixel block near the feature point p, and the vector **OC** represents the direction of the feature point p.

$$m_{pq} = \sum_{x,y\epsilon B} x^p y^q \, I(x,y),\ p,q = \{0,1\}. \quad (1)$$

$$\text{C}=\left(\frac{m_{10}}{m_{00}},\frac{m_{01}}{m_{00}}\right). \quad (2)$$

$$\theta = arctan(\frac{m_{01}}{m_{10}}). \quad (3)$$

Step3:Information description of key point P. Use the "Steer BREIEF" descriptor for information description. Randomly select 256 pairs of pixel points in the pixel block near the key point P, each pair of points is represented as (pi, qi), pi and qi represent the pixel points, if the gray value of pi is greater than qi, the value is 1; otherwise, it is 0."Steer BREIEF" uses the rotation information to calculate the descriptor. 256 pairs of pixels form a matrix Q. After the image is rotated by an angle α, the matrix used to describe the feature point information becomes Q1.[3]

$$Q1 = R_\alpha Q. \quad (4)$$

2.2. *Camera pose estimation and feature point depth estimation*

Assuming that a feature point P can be detected in the space, the projection of point P on the image at camera position 1 is p1, and the projection on the image at camera position 2 is p2. From the epipolar constraint, the following formula can be obtained:

$$x_2^T t^\wedge R x_1 = 0. \quad (5)$$

$$x_1 = K^{-1}p_1\,, x_2 = K^{-1}p_2\,. \quad (6)$$

x1 and x2 are the projected coordinates of point P on the normalized plane of the camera, T is the translation matrix, and R is the rotation matrix.

$$s_2 x_2 = s_1 R x_1 + t. \quad (7)$$

$$s_2 \hat{x}_2 x_2 = 0 = s_1 \hat{x}_2 R x_1 + \hat{x}_2 t. \quad (8)$$

s1 and s2 represent the depth of point P at camera position 1 and camera position 2. Observing the following equations 5 and 6 and Figure 3, we can find that the position error of the points p1 and p2 on the image will lead to the estimation error of the R and T matrices, as well as the depth estimation error.

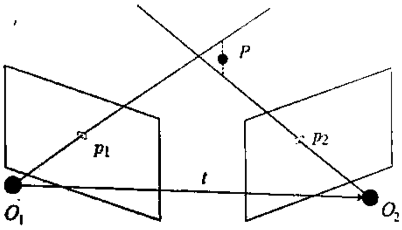


Fig. 3. Projection model

3. Improved ORB feature extraction and matching method

This article considers that the selection of the key point P is not only related to the peripheral pixels of the circular window, but also needs to pay attention to the pixel information inside the circle, and believes that the pixels closer to the key point need higher weights.

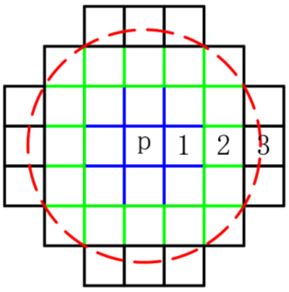


Fig. 4. Improve key point extraction

Set a judgment threshold t, divide the pixels in a circular window with a pixel radius of 3 into three layers, and subtract the gray value of p point from the gray value of each pixel on the inner to outer layer. The grayscale difference of the pixels of the first layer is denoted as vi, the difference of the second layer is denoted as vj, and the difference of the third layer is denoted as vk. Calculate the sum of the absolute value of the weighted difference of each layer. If there are N (take 18) differences whose absolute value is greater than the threshold t and satisfy that sum/37 is greater than t, then point p can be considered as the key point.

$$sum = w1 * |vi| + w2 * |vj| + w3 * |vk| \qquad (9)$$

Use the gray-centroid method to calculate the feature point direction, and use "Steer BRIEF" to determine the feature point descriptor.

In the process of feature point extraction, the quad-tree model is combined to realize the homogenization of feature points and avoid the problem of accuracy reduction caused by the clustering of feature points. In the matching process, combined with the RANSAC algorithm to remove mismatches.

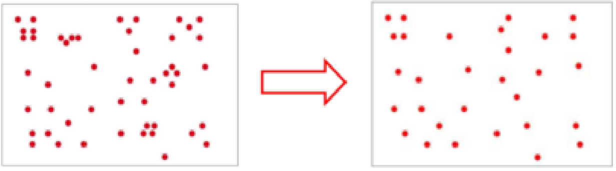


Fig. 5. Quad-tree homogenization

4. Experimental results and analysis

4.1. *Comparison of feature point extraction strategy*

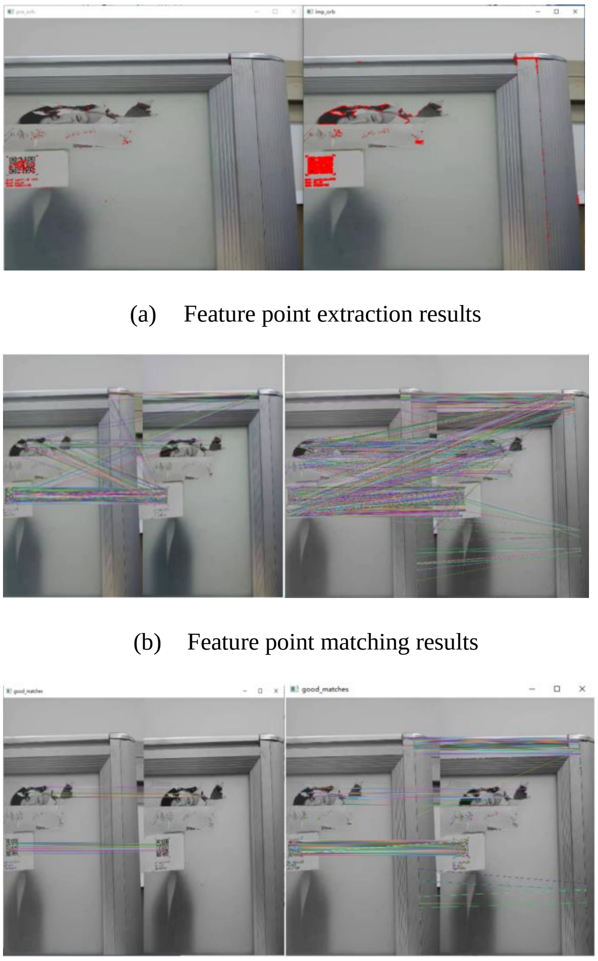
In this section, the following experiments are carried out to verify the effectiveness of the improved algorithm. The experimental results are shown in Table 1

Table 1. The planning and control components.

	ORB	Improved ORB
Feature point extraction number	503	5042
Maximum Hamming Distance	103	99
Number of matches with Hamming distance less than 40	51	671

In the feature point extraction process, the threshold t is set to 40. The original feature point extraction strategy can obtain 503 feature points. Among the matching results, the maximum Hamming distance is 103, and there are 51 matching results with Hamming distance less than 40. The improved feature point extraction strategy can get 5042 feature points. Among the matching results, the maximum Hamming distance is 99, and there are 671 matching results with hamming distance less than 40.

The two feature point extraction methods do not use non maximum suppression.



(a) Feature point extraction results

(b) Feature point matching results

(c) Matching results with Hamming distance less than 40

Fig. 6. Comparison of feature point extraction effects

4.2. *Quad-tree homogenization and RANSAC mismatch removal experiment*

Experimental results show that the improved orb feature point extraction strategy, combined with quad-tree and RANSAC algorithm can effectively manage feature points, quad-tree homogenization can effectively improve feature point matching results, and RANSAC can filters out most of the mismatched results.[4]

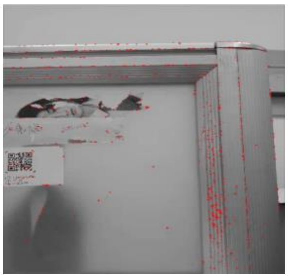


Fig. 7. Homogenization of feature points

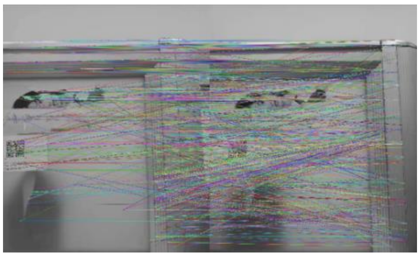


Fig. 8. Homogenization matching results

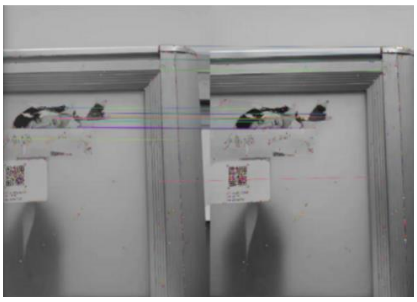
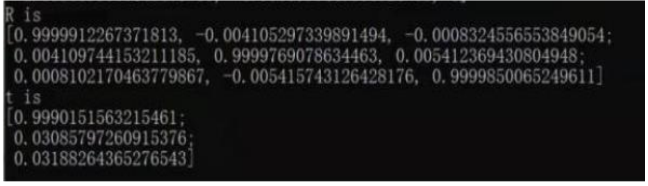


Fig. 9. RANSAC removes mismatches

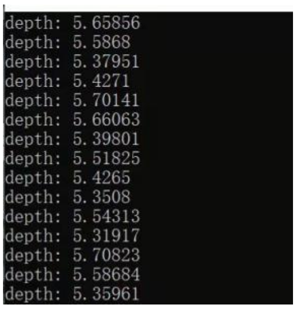
4.3. *Camera pose and feature point depth estimation*

The results of rotation matrix R, translation matrix T and feature point depth obtained by using the improved feature point extraction strategy combined with quad-tree and RANSAC algorithm are shown in Figure 10 and Figure 11.



```
R is
[0.9999912267371813, -0.004105297339891494, -0.0008324556553849054;
 0.004109744153211185, 0.9999769078634463, 0.005412369430804948;
 0.0008102170463779867, -0.005415743126428176, 0.999985006524961l]
t is
[0.9990151563215461;
 0.03085797260915376;
 0.03188264365276543]
```

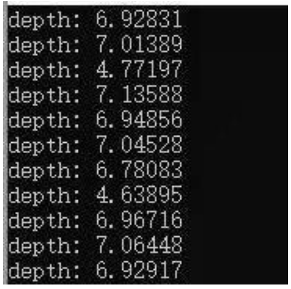
Fig. 10. depth information



```
depth: 5.65856
depth: 5.5868
depth: 5.37951
depth: 5.4271
depth: 5.70141
depth: 5.66063
depth: 5.39801
depth: 5.51825
depth: 5.4265
depth: 5.3508
depth: 5.54313
depth: 5.31917
depth: 5.70823
depth: 5.58684
depth: 5.35961
```

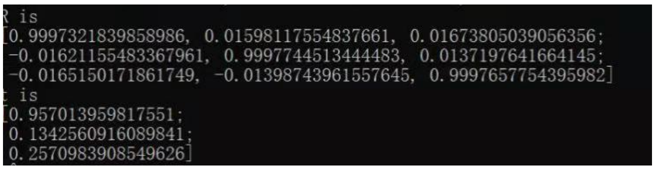
Fig. 11. Rotation matrix R and translation matrix T

The results without any optimization strategy are shown in the figure 12 and figure 13.



```
depth: 6.92831
depth: 7.01389
depth: 4.77197
depth: 7.13588
depth: 6.94856
depth: 7.04528
depth: 6.78083
depth: 4.63895
depth: 6.96716
depth: 7.06448
depth: 6.92917
```

Fig. 12. Without optimized depth information



```
R is
[0.9997321839858986, 0.01598117554837661, 0.01673805039056356;
 -0.01621155483367961, 0.9997744513444483, 0.0137197641664145;
 -0.0165150171861749, -0.01398743961557645, 0.9997657754395982]
t is
[0.957013959817551;
 0.1342560916089841;
 0.2570983908549626]
```

Fig. 13. Without optimized Rotation matrix R and translation matrix T

In the image acquisition process, only let the camera translate along the x direction, and the feature points are basically distributed in the same plane. The experimental results show that the optimized rotation matrix R is similar to the original result. But the result of the optimized translation matrix T is [0.999,0.031,0.032], and the result of the Original translation matrix is [0.957,0.134,0.257], which shows that the optimization effect is better. And the optimized feature point depth is basically distributed around 5 unit distances, which is more in line with the experimental environment.

5. Conclusion

This paper proposes an improved ORB algorithm for monocular vision slam to measure the depth information of image feature points and estimate the pose of the camera. Experimental results show that this method can meet the measurement accuracy requirements. Compared with the unimproved method, the method proposed in this paper improves the estimation accuracy of pose and depth. This method can provide good support for the posture tracking of the robot. However, due to the increase in the computational complexity of the algorithm, it takes longer to run. In order to meet the real-time requirements of slam, the speed of the algorithm needs further improved.

References

1. Guanci Yang, Zhanjie Chen, Yang Li, etc.Rapid Relocation Method for Mobile Robot Based on Improved ORB-SLAM2 Algorithm [J].Remote Sens, 2019,11,149.
2. Yang Hong-fan, LI Hang, etc.Image feature points extraction and matching method based on improved ORB algorithm [J].JOURNAL OF GRAPHICS, 2020, 41(4) : 548-555.
3. Leng Xue,Yang Jinhua. Improvement of ORB algorithm [J]. AER-Advances in Engineering Research, 2012, 28:903-906.
4. Jiangguo She, Rentong Xu, Ning Chen. Image stitching technology based on orb and improved RANSAC algorithm. Journal of Jiangsu University of Science and Technology (Natural Science Edition), 2015,29(2):164-169.

Authors Introduction

Dr. Jiwu Wang

He is an associate professor, Beijing Jiaotong University. His research interests are Intelligent Robot, Machine Vision, and Image Processing.

Mr. Weipeng Wan

He is a postgraduate in Beijing Jiaotong University. His research interests are vision slam and image processing.