

Research on robot scene recognition based on improved feature point matching algorithm

Yuxia Cui, Zhipeng Sun and Xianlun Wang*

College of Mechanical and Electronic Engineering, Qingdao University of Science and Technology, Qingdao, China

Abstract. A new feature description method based on the fusion of fast retina keypoint (FREAK) and the rotation-aware binary robust independent elementary features (rRBRIEF) is proposed to realize the effective combination of efficiency and accuracy of the two feature descriptions. In addition, in the elimination stage of mismatched point pairs, by setting the base point and its neighborhood, an improved neighborhood parallel random sample consensus (RANSAC) algorithm is proposed to achieve efficient parallel operation of the algorithm in multiple local neighborhoods. The improved feature point matching algorithm and the existing algorithm were tested in different scales, different rotations, different illuminations, and different fuzzy data sets. The experimental results show that the improved algorithm improves the average scene recognition accuracy by 18.21%, improves the efficiency by 15.58%, and shows good robustness.

Keywords: Scene recognition, Feature point matching, Feature description, Random sample consensus algorithm.

1 Introduction

Visual Simultaneous Localization Mapping (VSLAM) technology based on vision has developed rapidly. The robot scene recognition technology based on feature point matching is widely used in the visual front end of VSLAM technology to complete the initial recognition and matching of the mobile robot to the scene. The commonly used feature point matching algorithm is ORB (Oriented FAST and Rotated BRIEF) algorithm, which meets the real-time requirements of mobile robot synchronous positioning and mapping. It is often divided into four stages including image feature point extraction, feature description calculation, feature matching and mismatch point pair elimination. The SLAM framework based on ORB feature matching algorithm has good performance in many VSLAM frameworks. Raúl and Tardós proposed the ORB-SLAM2[1] system framework in 2015. The framework homogenizes the distribution of feature points by quadtree decomposition on the basis of ORB algorithm. Their fellow team member Campos C

* Corresponding author: xlwang@qust.edu.cn

proposed ORB-SLAM3 [2] system framework on the basis of ORB-SLAM2 system framework in July 2020.

ORB algorithm has high real-time performance, but poor robustness and matching recognition accuracy. An improved ORB feature extraction algorithm based on local region adaptive threshold was proposed [3]. Zhang et al. proposed an image matching algorithm based on the combination of ORB and clustering algorithm[4]. Wang et al. proposed a feature description algorithm for extracting feature points under different affine transformations[5]. Secondly, the ORB feature matching algorithm uses the violent matching algorithm in the matching stage. Violent matching is bound to have some mistakes, and the RANSAC algorithm is commonly used for eliminating mismatch point pairs. RANSAC algorithm has some limitations in accuracy and efficiency. In order to solve its limitations, Barath et al. proposed to optimize and improve the interior point judgment method of RANSAC with graph cutting algorithm [6]. Chum et al. proposed the progressive sampling consensus (PROSAC) [7]. Chum et al. proposed a locally optimized RANSAC algorithm [8].

Inspired by the above researches, aiming at the low recognition accuracy and efficiency of ORB feature matching algorithm in the process of robot scene recognition, two ideas are proposed to improve the original algorithm in the stage of feature description calculation and error matching point elimination, in order to improve the feature recognition accuracy and efficiency of feature point matching algorithm in the process of robot scene recognition.

2 FREAK-rBRIEF feature description

The expressions of FREAK and rBRIEF are the same. After analysis, a FREAK-rBRIEF feature description based on the fusion of FREAK and rBRIEF is proposed to realize the effective combination of the accuracy and efficiency of the two feature descriptions.

Through the analysis of FREAK feature description, it is found that FREAK feature description is a binary coding feature description similar to rBRIEF feature description. It not only reduces the demand for storage space, improves the generating speed of feature description, but also reduces the time required for feature matching. FREAK can screen 90% of irrelevant candidate points through the coarse information of the first 16 bytes, and runs in parallel, with higher accuracy and efficiency than rBRIEF. However, FREAK has poor robustness to scale changes. When the image scale changes greatly, FREAK has a significant impact on the Fovea fine information region, resulting in a large number of mismatches in the matching stage. By analyzing the advantages and disadvantages of the two feature descriptions, the first 128 bits coarse precision feature description of FREAK is used to replace the first 128 bits of rBRIEF, and a new FREAK-rBRIEF feature description is formed after fusion. Based on the FREAK-rBRIEF feature description, cascade violence matching is carried out. Firstly, the first 128 bits of FREAK are used for coarse matching, and the candidate points are quickly selected, for what is lower than the threshold, the rBRIEF with scale invariance is used for accurate matching to realize the combination of accuracy and efficiency of the two feature descriptions, so as to improve the recognition accuracy and efficiency of the matching algorithm based on this feature description in the process of robot scene recognition.

3 Neighborhood parallel RANSAC algorithm

RANSAC algorithm is implemented on the basis of hypothesis and test framework. It can obtain accurate global optimal model parameters from the sample dataset containing

outliers by continuous value iteration, and filter out the outliers in the sample dataset. It is commonly used in feature matching algorithm to remove mismatched feature points.

3.1 Selection of base point

Firstly, the ratio-test algorithms used to assign a confidence score to the unpurified feature matching points. The violent matching process of the original feature points is based on finding the minimum Hamming distance of the feature description as the final matching result, and the confidence score is defined as follows.

$$S = \frac{d_{min}}{d_{smin}} \tag{1}$$

Where, S represents the confidence score, d_{min} represents the minimum feature description distance, and d_{smin} represents a minor feature description distance. The smaller the S, the greater the difference between d_{min} and d_{smin} , indicates that the robustness of the matching point pair is higher. Local non-maximum suppression in a circle with radius R is applied to a matching point with lower confidence. If the point is the maximum, the point is defined as the base point, and all matching points satisfying this condition are called the base point sets.

3.2 Selection and filtering of local neighborhood of base point

After determining the base point set, find matching pairs that support base point matching, that is, select the matching point in the same area as the base point and meet certain similar constraints in the initial raw matching pairs, and set S_i as the i th base point matching relationship, which is defined as follows

$$S_i = (X_1^{S_i}, X_2^{S_i}) \tag{2}$$

Where, $X_1^{S_i}$ and $X_2^{S_i}$ represent the i th base point matching pairs of the two images that need to be matched. They conform to similar transformations (rotation + scaling), in which the rotation can be shown as $\alpha^{S_i} = \alpha_2^{S_i} - \alpha_1^{S_i}$ and scaling can be shown as $\sigma^{S_i} = \sigma_2^{S_i} / \sigma_1^{S_i}$. So for any matching pair, it meets the following formula:

$$(p_1, p_2) = ((X_1, \mathbf{d}_1, \sigma_1, \alpha_1), (X_2, \mathbf{d}_2, \sigma_2, \alpha_2)) \in M \tag{3}$$

Where, M represents the initial raw matching point pairs, d represents the feature description. Then two constraints are set according to the relative position relationship with the base point and the scale consistency of the matching point, which is defined as follows:

$$\|X_1^{S_i} - X_1\| \leq R_1 \wedge \|X_2^{S_i} - X_2\| \leq R_2 \tag{4}$$

$$|\alpha^{S_i} - \alpha^p| \leq t_\alpha \wedge \left| \ln \left(\frac{\sigma^{S_i}}{\sigma^p} \right) \right| \leq t_\sigma \tag{5}$$

Where, R_1 and R_2 represent the base point diffusion radii of the image. t_α and t_σ represents the confidence threshold of rotation and scaling. $\alpha^p = \alpha_2 - \alpha_1$ and $\sigma^p = \sigma_2 / \sigma_1$ represent the angle and scale difference between two matching points. Formula (4) is used to constrain the relative position relationship between the initial matching point and the base

point. Formula (5) is used to constrain the angle and scale consistency of matching points. If the matching points near the base point meet the two constraints of formula (4) and formula (5), they can be incorporated into the base point matching neighborhood $N_i \subseteq M$. Finally, the RANSAC algorithm with the average optimal fixed iteration number is run in parallel in each neighborhood to complete the elimination of false matching points in the global range. Compared with the original RANSAC algorithm, the neighborhood parallel RANSAC algorithm reduces the number of iterations, increases the number of interior points, avoids the phenomenon of high locality, obtains more accurate transformation matrix, and improves the recognition efficiency and accuracy of matching algorithm in the process of robot scene recognition.

3.3 The overall flow of improving the feature matching algorithm

The basic flow of the improved algorithm, which is mainly divided into feature point detection and screening, calculation and fusion of feature description, matching and improvement of feature point, and algorithm of eliminating mismatch point pairs.

4 Experiment and result analysis

In order to verify the performance of the improved feature matching algorithm in robot scene recognition, the improved feature matching algorithm is compared with the ORB+RANSAC algorithm which is more traditional but increases the scale pyramid. In order to ensure the objectivity of the experiment, different scenes are selected from representative data sets for verification, so as to make the images and scenes closer to reality. The transformation of different scales and perspectives can best reflect the performance of the algorithm. Therefore, the graf sequence of Mikolajczyk dataset and the images with different scales of oxbuild dataset are used for two groups of contrast experiments under different scales. Then, oxbuild dataset is used for a single group of contrast experiments under different rotation conditions, different light and dark conditions and different fuzzy conditions, a total of five groups.

4.1 Output diagram of feature matching recognition effect

It can be seen from the matching recognition comparison diagram of the experimental output that the improved feature point matching algorithm has better robustness and recognition accuracy than ORB + RANSAC under different scale changes, different illumination changes, different rotation changes and different fuzzy changes.

4.2 Comparative analysis of experimental data

The experimental results are analyzed from two aspects of matching recognition accuracy and efficiency. The data are the average value of 10 experimental test results.

4.2.1 Efficiency analysis of matching recognition

The feature matching algorithm was improved in the two stages of feature description calculation and error matching point elimination, respectively. It is difficult to measure the efficiency of the algorithm with single-stage time. Therefore, the overall running time of the algorithm is used to verify the efficiency of the algorithm, which is the total time consumed by the algorithm from the beginning to the end. In order to intuitively represent

the difference between the running time of the improved algorithm and the ORB+RANSAC algorithm, the data in the table is represented in the form of a line graph, as shown in Figure 1.

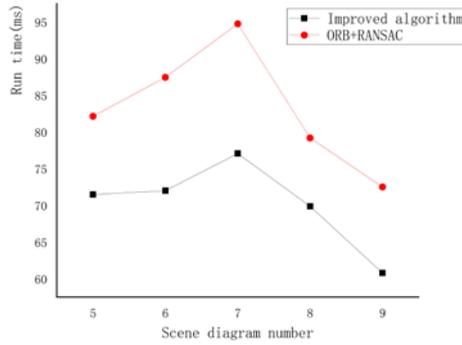


Fig. 1. Run time line graph of algorithm.

By analyzing table 1 and fig.1, it is concluded that the running time of the two algorithms has the same trend under the conditions of different scale changes, different illumination changes, different rotation changes and different fuzzy changes. However, compared with ORB+RANSAC, the average running time of the improved feature point matching algorithm is reduced by 12.96 ms, and the overall matching recognition efficiency is increased by 15.58 %. Although the improved algorithm needs to calculate two kinds of feature description in the calculation feature description stage, which is more time-consuming than ORB + RANSAC, but it takes less time in the matching and mismatch point elimination stages, so the scene matching recognition efficiency is increased by 15.58 %, which has better applicability in robot scene recognition.

4.2.2 Analysis of matching recognition accuracy

The matching recognition accuracy is represented by the total matching pair in the correct matching comparison. In order to intuitively represent the difference of matching recognition accuracy between the improved algorithm and ORB + RANSAC algorithm, the data in the table is represented in the form of a line graph, as shown in Figure 2.

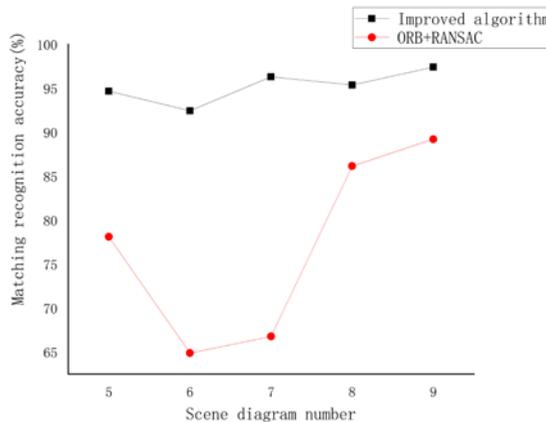


Fig. 2. Line graph of algorithm matching recognition accuracy.

By analyzing Figure 2, it is concluded that under the conditions of different scale changes, different light changes, different rotation changes and different fuzzy changes, the average matching recognition accuracy of the feature point matching algorithm is improved by 18.21% higher than that of the ORB+RANSAC algorithm, and the improved algorithm has higher robustness. The matching recognition accuracy is above 90 % in different scenes, and it has better applicability in robot scene recognition.

5 Conclusion

In view of the fact that the traditional feature point matching algorithm cannot meet the requirements of high precision and high efficiency of robot scene recognition technology, the free rBRIEF feature description and neighborhood parallel RANSAC algorithm are proposed. Compared with the widely used ORB+RANSAC feature point matching algorithm, the improved algorithm shows better robustness, matching recognition efficiency and accuracy IN different scene experimental tests, and has better applicability in robot scene recognition technology, especially in mobile robot VSLAM technology based on feature point method.

References

1. Mur-Artal R, Montiel J M M. ORB-SLAM: A Versatile and Accurate Monocular SLAM System[J]. *IEEE Transactions on Robotics*, 2015, 31(5):1147-1163.
2. Campos C, Elvira R, JG Rodríguez, et al. ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial and Multi-Map SLAM[J]. 2020.
3. J Yao, Zhang P, Wang Y, et al. An Adaptive Uniform Distribution ORB Based on Improved Quadtree[J]. *IEEE Access*, 2019, PP(99):1-1.
4. Zhang L, Cai F, Wang J, et al. Image Matching Algorithm based on ORB and K-Means Clustering[C]// 2020 5th International Conference on Information Science, Computer Technology and Transportation (ISCTT). 2020.
5. Wang Z, Li Z, Cheng L, et al. An improved ORB feature extraction and matching algorithm based on affine transformation[C]// 2020 Chinese Automation Congress (CAC). 2020.
6. Barath D, Matas J. Graph-Cut RANSAC[J]. *IEEE*, 2018.
7. Chum O, Matas J. Matching with PROSAC - progressive sample consensus[C]// *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. *IEEE*, 2005.
8. Chum O, Ji Matas, Kittler J. Locally Optimized RANSAC[C]// *Joint Pattern Recognition Symposium*. Springer, Berlin, Heidelberg, 2003.