Study on the ORB algorithm in the application of Monocular SLAM

Jiwu Wang[1]
School of Mechanical, Electronic and Control Engineering, Beijing Jiaotong University
Beijing 100044, China[1]

Shunkai Zheng[1]
School of Mechanical, Electronic and Control Engineering, Beijing Jiaotong University
Beijing 100044, China[1]

Sugisaka Masanori[2]
Alife Robotics Corporation Ltd, Japan and Open University, United Kingdom[2]
E-mail: jwwang@bjtu.edu.cn[1]*; ms@alife-robotics.co.jp*[2]

Abstract

In view of reducing the accumulative error, we perform loop closing based on PTAM in our Monocular SLAM.As this method relies on extracting natural environment features, we chose ORB algorithm as the feature extraction and matching. We demonstrate that ORB features have enough recognition power to enable place recognition from severe viewpoint change and they are so fast to extract and match (without needing multi-threading or GPU acceleration) that enable real time accurate tracking and mapping. Through to outdoor scene experiment, we validate the algorithm performance.

Keywords: SLAM; ORB; Monocular Vision；Loop closing

1. Introduction

Monocular SLAM has been developed from the initial filtering approaches to the most modern key frame-based SLAM systems. The key frame-based SLAM system is more useful and accurate in the large environment. One of the most representative keyframe-based systems is Parallel Tracking and Mapping, PTAM[1].The map points of PTAM correspond to FAST corners matched by patch correlation. This makes the points only useful for tracking but not for place recognition. In fact PTAM does not detect large loops, and the relocalization is based on the correlation of low resolution thumbnails of the keyframes, yielding a low invariance to viewpoint.

Using just one camera is a more complex problem because the depth of observed features in the image is unknown and multi view geometry is necessary to solve the problem. Stereo and RGB-D cameras have a range in which they can recover the depth but in certain situations such as sunny day in the outdoor they are not not practical and monocular techniques are still necessary.

In this work we focus on ORB and loop closing, two open problems that are essential for real SLAM applications.

ORB features are oriented multi-scale FAST corners with a 256 bits descriptor associated[2]. As binary features, they are extremely fast to compute and match, while they are highly invariant to viewpoint. This allows matching them from wide baselines, boosting the accuracy of BA.

Loop closing is the task of detecting when a robot is

Revisiting a previously mapped area, in order to correct the error accumulated in the robot trajectory during exploration [3]. So, how to determine whether a new frame in the image sequence has occurred? One is based on the robot's position, which is adjacent to the previous position; the two is the appearance of the image, which is similar to the previous key frame. In this paper, we choose the last one.

2. System Overview

The system is composed of three main tasks: tracking, local mapping and loop closing that run in parallel in a multi-core machine

Fig 1 shows a scheme of the different building blocks of the system

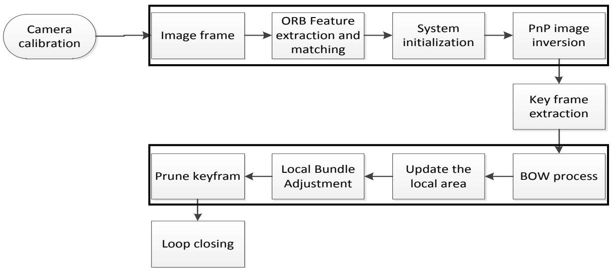

Fig.1. System Overview

2.1. Tracking

The map initialization is performed finding a planar scenes or an essential matrix from two near frames in an automatic process. Once an initial map exists, the tracking estimates the camera pose with every incoming frame. Next we describe each step of the tracking

2.1.1 ORB extraction and tracking

The first step is to extract ORB features in the frame. At first a scale image pyramid is computed and FAST key points are detected at each level, then only a subset of these key points will be described. If tracking was successful in the last frame, the tracking tries to get a first estimation of the camera pose from the last frame. Each ORB from the last frame, which has a map point associated, searches a match in the current frame in a small area around its previous position.

In the large scale environment, feature extraction and matching are influenced by Gaussian blur, rotation, scaling, illumination and processing speed. Therefore design the following experiment to compare the ORB features extraction performance.

1. Generating and matching method of test data:

Taking image sequence Fig2 as original image that photoed in Beijing Jiaotong University mechanical building Z706

Fig2 Left: original frame Right: next frame of original frame

Then using the affine transformation of the specified argument and mathematical model including Gaussian blur, rotation, scaling, illumination for Right frame to generate the sequences， Analog cameras in the complex scene pictures taken in the process of moving, close to the performance of the real data

2.Correct judgment of matching points：

Use RANSAC find out the Homography Matrix between the generated sequences and original frame, then the sequences frame of the feature points projection to original frame, calculate the size of the projection error, if less than the threshold value of set (this experiment using threshold for 2), argues that the point is the right match, otherwise it is a false matching points. The relation between the calculated is:

$$ratioMatch = \frac{correctMatch}{Match} \times 100\% \qquad (2\text{-}1)$$

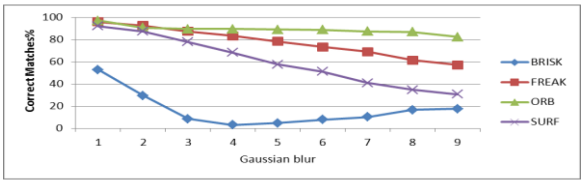

Fig3 Gaussian blur

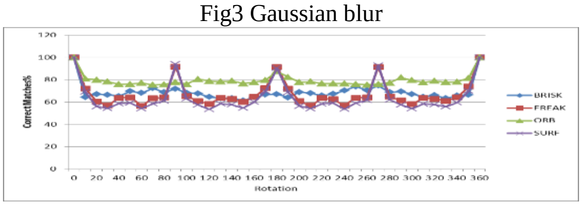

Fig4 Rotation Angle (0-360, per 20)

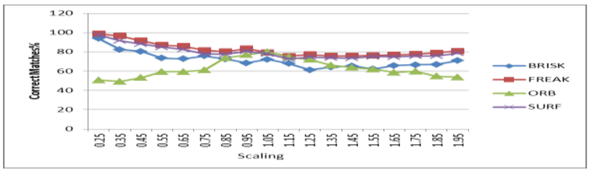

Fig5 Scaling (0.25-1.95 per 0.1)

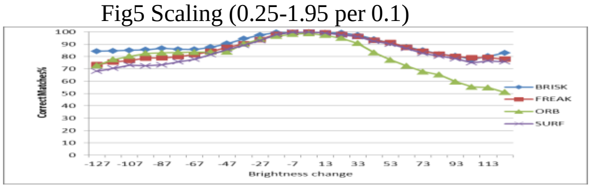

Fig6 Additional brightness values

Tab1 Processing time

Algorithm	Average time per Frame(MS)	Average time Per Key point(MS)
BRISK	58.3473	0.201757
FREAK	845.55	0.282354
ORB	132.27	0.267296
SURF	2244.56	0.61098

Fig 3 show that the feature matching performance of ORB under the condition of fuzzy is the best, it can greatly improve the stability of the algorithm; Fig4 show that ORB is the most stable for rotation .Fig 5and Fig6 show that the several feature extraction algorithms are similar to the change of the scale and brightness in a certain range.

Tab1 show that extracting all the feature points of the whole image, the speed of ORB is the fastest and 17 times faster than SURF， for every feature point of the image, ORB is 3 times faster than SURF which is the slowest.

The above results and analysis show that the excellent performance of the ORB feature extraction algorithm. We demonstrate that ORB features have enough recognition power to enable place recognition from severe viewpoint change and they are so fast to extract and match (without needing multi-threading or GPU acceleration) that enable real time accurate tracking and mapping.

2.2. Loop Closing

Fig7 is a loop closing detection of visual SLAM based on the bag-of-words .A node in the map corresponding to the position of the robot, represented by 1 - 7, each node with a key frame scene at the location of the image description. The main idea of the bag-of-words model is to extract features from the image, the feature clustering (K-means) to the vocabulary tree, get the image of the visual word description vector, then calculate the similarity between the image and the matching strategy, complete closed-loop detection.

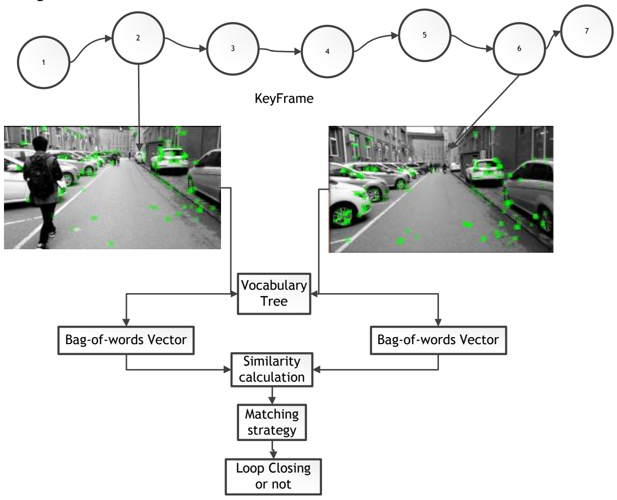

Fig7 loop closing detection based on bag-of-words model

TF-IDF (Frequency Inverse Document Frequency Term) model is used to evaluate the similarity of images. The similarity between two bag of words vectors v_1 and v_2 can described as [4]:

$$s(v_1, v_2) = 1 - \frac{1}{2}\left| \frac{v_1}{|v_1|} - \frac{v_2}{|v_2|} \right| \qquad (2\text{-}1)$$

The loop closing thread takes K_t , the last key frame processed by the local mapping, and tries to detect and close loops. The steps are next described.

- At first we compute the similarity between the bag of words vector of K_t and all its neighbors $\{K_c\}$ in the covisibility graph and retain the lowest score $S_{\min}$;
- In the DBoW2 database which save all the key frames to find the key frames whose score is not less than $S_{\min}$, the key frames is $\{K_s\}$, $\{K_s\}$ cannot include the $\{K_c\}$;
- By step 2 to find the key frame K_{t-1} associated with $\{K_{s-1}\}$,and the key frame K_{t-2} associated with $\{K_{s-2}\}$,if there is a key frame K_p in the

$\{K_s\}$ and similar to the key frame in the $\{K_{s-1}\}$ and $\{K_{s-2}\}$ in the covisibility graph, take K_p as a candidate key frame;

- If K_t and K_p is supported by enough inliers, the loop with K_t is accepted.

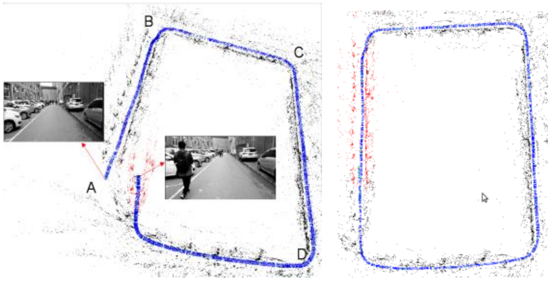

Fig8 Use loop closing to modify the map（Laboratory in Beijing Jiaotong University）

3. Experiment and Conclusion

3.1. *Experiment on public data sets*

The odometry benchmark from the KITTI dataset[5] contains sequences from a car driven around a residential area with accurate ground truth from GPS and a Velodyne laser scanner. This is a very challenging dataset for monocular vision due to fast rotations, areas with lot of foliage, which make more difficult data association, and relatively high car speed, being the sequences recorded at 10 fps.

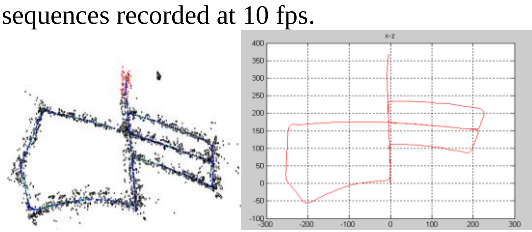

Fig9 Left: points and key frame trajectory Right: map after scale change

Tab2 Results of the KITTI dataset

Sequence	Dimension ($m \times m$)	KFs	RMSE (m)
KITTI 05	478x425	850	8.23

The result of Sequences 05 of the KITTI dataset is shown as Tab2：The RMSE of this experiment is 8.23m.

3.2. *Experiment on our Machine building*

Experiments on the image sequences generated by the road of Mechanical Laboratory of Beijing Jiaotong University, The actual map is as follows:

Fig10 the map of Mechanical Laboratory

The walking path is ABCDA，get the real map by scale transformation and the comparison between the two map：

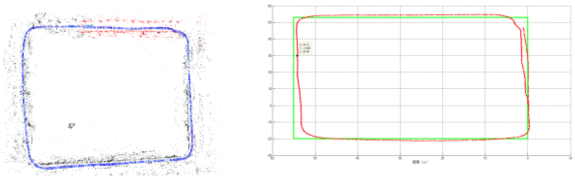

Fig11 Left: Generated map of points and key frame Right: the comparison between the two map

From the above results, the results are more accurate. In the process of the experiment, the characteristics of the corner are easy to be lost, and the error of the distance is inevitable estimated. In the future, we focus on solving the robustness of the map initialization and the processing that easy to lost at the corner.

References

1. G. Klein and D.Murray.Parallel tracking and mapping for small AR workspaces. In International Symposium on Mixed and Augmented Reality (ISMAR), 2007...
2. E. Rublee, V. Rabaud, K. Konolige, and G. Bradski.ORB: an efficient alternative to SIFT or SURF. In IEEE International Conference on Computer Vision (ICCV),2011
3. B. Williams, M. Cummins, J. Neira, P. Newman, I. Reid, and J. D.Tard´os, “A comparison of loop closing techniques in monocular SLAM,” Robotics and Autonomous Systems, vol. 57, no. 12, pp. 1188–1197, 2009.
4. R. Mur-Artal and J. D. Tard´os, “Fast relocalisation and loop closing in keyframe-based SLAM,” in IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, June 2014, pp. 846–853.
5. A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, “Vision meets robotics: The KITTI dataset,” The International Journal of Robotics Research,vol. 32, no. 11, pp. 1231–1237, 2013.