A Novel SLAM Algorithm based on Region Focusing Image Segmentation

Chenshuo Tan
Faulty of Information Technology
Beijing University of Technology
Beijing, China
tanchenshuo1025@163.com

Xiaogang Ruan
Faulty of Information Technology
Beijing University of Technology
Beijing, China
adrxg@bjut.edu.cn

Xiaoqing Zhu*
Faulty of Information Technology
Beijing University of Technology
Beijing, China
alex.zhuxq@bjut.edu.cn

2020 Chinese Automation Congress (CAC) | 978-1-7281-7687-1/20/$31.00 ©2020 IEEE | DOI: 10.1109/CAC51589.2020.9326761

***Abstract*—In order to solve the problem that tracking drift and localization failure caused by pose estimation error in orb-slam (simultaneous localization and mapping), an improved orb feature point extraction algorithm called RFIS-slam(Region Focusing Image Segmentation slam) is proposed. Firstly, when extracting feature points, the image is divided into several areas and feature points which to be extracted is equally distributed into each area to improve the feature points to be more representative. Secondly, an evaluation system is added to evaluate each area, and some areas with low contrast that are not easy to extract highly identifiable feature points are excluded to obtain more robust feature points. Finally, the experimental results were verified that the RFIS-slam improves the feature point matching by 4.1% and 5.8% respectively compared to orb-slam and orb-slam2; in terms of initialization, 111 and 65.4 frames were improved respectively. Better results have also been achieved in position estimation and real-time positioning, so as to improve the precision of feature point matching and positioning without losing speed.**

Keywords—orb-slam, region focusing, feature points extraction, tracking

I. INTRODUCTION

At present, indoor positioning technology shows a trend of flourishing, such as WIFI positioning technology, RFID positioning technology, infrared technology, ultrasonic positioning technology, visual slam technology, etc. In view of the particularity and limitations of indoor positioning, there are special requirements for positioning technology. WIFI positioning technology has become mature, but it still has the disadvantages of being greatly affected by the different environment and causing serious tracking drift. It's also easy to be interfered from co-frequency signals. RFID positioning technology has a short operating distance, and does not have communication capabilities. It is not easy to integrate into other systems, and it is difficult to achieve accurate positioning. Infrared positioning technology is relatively expensive to produce and is greatly affected by noise, making it difficult to popularize. Ultrasonic waves are easy to attenuated during transmission due to the characteristics of their waves then affect positioning accuracy. Visual slam technology has gradually play a leading role of indoor positioning technology because of its small size, low cost, strong versatility, and rich ability to get information.

SLAM[1] aims to put the robot into a completely unknown environment and place for movement. During this process, it uses sensor (cameras) to locate its own position and detect the environment, and finally completes the construction of the unknown environment map to achieve autonomous positioning and navigation. The visual slam is divided into three types according to the type of camera: Monocular visual slam[2], Binocular visual slam[3] and RGB-D slam[4].Among them, the monocular visual slam is widely used by people due to its advantages of low price, flexibility and convenience.

The extraction and matching of image feature points is an important part of the slam technology. The results of the extraction and matching will directly affect the quality of the frame matching. Whether or not the pose estimation is truly realized will ultimately affect the results of tracking and localization. The feature point method first needs to extract the feature points of the image for feature matching, and then use the Epipolar Geometry Estimation [5], PnP [6-7] or ICP [8-9] algorithm to estimate the camera pose.

As traditional SIFT[10] and SURF[11] feature point extraction algorithms consume a lot of time, real-time tracking cannot be achieved in slam. Rublee et al. proposed the orb[12] feature extraction algorithm in 2011, which greatly improved the detection and matching speed of feature points.Mur-artal et al. combined the idea of orb with slam to create orb-slam[13] system with both speed and robustness in 2015. In 2017, he proposed orb-slam2[14], which improved the adaptability of the algorithm in various of environment at the cost of slightly reducing the matching feature point. In addition, there are some slam systems combined with other algorithms[15-17], However, the feature points are too concentrated and many invalid feature points are extracted, which leads to the problem that the tracking and positioning cannot be performed normally.

Based on the above analysis, in order to improve the current effect of localization, we proposes a RFIS-slam algorithm based on orb-slam. Firstly, the image is divided into small areas for feature point extraction to ensure that the feature points are evenly distributed in the image; Secondly, each small area is evaluated to exclude areas with low contrast and poor extraction of feature points. Experimental results show that the proposed algorithm achieves better pose estimation and localization effect than the original algorithm.

II. ORB-SLAM2

Orb-slam2 (Oriented FAST and Rotated BRIEF) combines FAST algorithm(Features from Accelerated Segment Test)[18] and BRIEF algorithm(Binary Robust Independent Elementary Features)[19] to extract and describe feature points.

A. Feature point extraction

The idea of the FAST algorithm is that if the gray value of a pixel is large or small than most of the continuous pixels within a certain range, it can be regarded as a feature point to be extracted. When traversing image pixels, first select the current pixel P and set its pixel value to Ip. Set a circle with radius r equal to 3, and 16 pixels will pass on the circumference boundary, as shown in Figure 1.

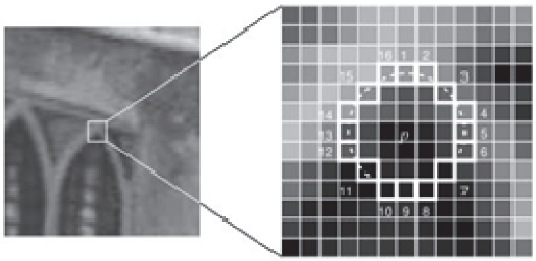

Fig.1 Schematic diagram of FAST feature points

Set a threshold t to detect the 16 pixels. If there are more than 12 continuous (FAST-12) pixel points with pixel values greater than Ip+t or less than Ip-t at the same time, the pixel point P can be considered as a feature point.

Compared with traditional algorithms such as SIFT and SURF, the FAST algorithm can greatly improve the efficiency of feature point extraction. Instead of this, the extracted feature points do not have scale invariability and rotation invariability.

Orb-slam combines the image pyramid to solve the problem of scale invariability. The idea is to down-sample the input image at a ratio of 1.2 to obtain images with the same 8-layer window but different scales. The FAST feature point detection was performed on the 8-layer images, and the feature points obtained had 8 different scales.

To solve the problem of rotation invariability, this algorithm combines the center of gravity method. The idea is to connect the feature point O to the center of mass of the gray value in the range with the circle center and radius R as the direction of the feature point, thereby obtaining rotation invariability.

First get the region B with the feature point as the center of the circle and the radius R, and the area of the region B is:

$$M_{\mathrm{pq}} = \sum_{x,y \in B} x^p y^q I(x,y), \quad p,q = \{0,1\} \tag{1}$$

$I(x,y)$ is the gray value of point (x,y). Then the centroid of area B can be obtained:

$$C = \left(\frac{M_{10}}{M_{00}}, \frac{M_{01}}{M_{00}}\right) \tag{2}$$

Connect the feature point O and the centroid C to get the direction vector $\overrightarrow{OC}$, then the feature point direction can be defined as:

$$\theta = \arctan(M_{01} / M_{10}) \tag{3}$$

The above two methods make the feature points have scale invariability and rotation invariability and enhance the robustness of the representation of feature points between different images.

B. Feature point description

According to the orb algorithm, 1000 feature points will be extracted for each image by default, and each feature point is described. This causes the descriptor to simultaneously describe the feature points as detailed as possible without occupying a large amount of memory. The BRIEF descriptor is a binary number, which can greatly reduce the amount of contrast between pixels while ensuring a high descriptor efficiency. The idea of the BRIEF algorithm is to randomly obtain n pairs of points in the square of the feature points and compare the gray values:

$$\tau(p:x,y): \begin{cases} 1 & if\ p(x) < p(y) \\ 0 & otherwise \end{cases} \tag{4}$$

Where p(x) and p(y) are the gray value of the selected point.

A set of point pairs constitutes a description value.After n operations between point pairs, a series of binary descriptors of length n are obtained:

$$f_n(p) := \sum_{1 \le i \le n} 2^{i-1} \tau(p; x_i, y_i) \tag{5}$$

n is the feature dimension, which is 256 in the orb algorithm. In order to make the descriptor more descriptive, the point pairs x, y all conform to Gaussian distribution of $(0, \frac{1}{25} S^2)$.

During the matching of the feature points, the descriptors of the two points are matched with Hamming distance. When the matching similarity exceeds the threshold, it can be considered that the two feature points match successfully.

III. RFIS-SLAM

According to the above original orb algorithm steps, it can be concluded that when traversing each pixel of the entire frame, the extracted feature points will be too concentrated, which is not conducive to subsequent frame pose estimation and tracking, and ultimately leads to poor localization effect. And when the original algorithm enters the tracking process during initialization, it is often impossible to initialize normally because the successful matching point between two consecutive frames is less than 100.Based on the above problems, this paper proposes a image segmentation evaluated scheme. First, the image is divided into several square small areas, and the number of feature points to be extracted per frame is evenly distributed into each area to ensure that the feature points are cover the image; Then, the threshold idea is used to eliminate a part of the areas with low contrast and the extracted feature points are not conducive to matching, and pixel traversal is performed in the remaining areas. It is possible to obtain more robust feature points that are more representative of the image without increasing the extraction time. The improved algorithm steps are as follows:

- Convert the input frame into a gray-scale image, and use Gaussian filtering to perform weighted average processing on the image.

- Divide the processed image into 15 × 15 small areas similar to the original image.

- Calculate the gray standard deviation of each area:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} P(x_i) \tag{6}$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2} \tag{7}$$

- Set a threshold t_1 .Eliminate all areas which $\sigma \le t_1$.

- Sort the remaining areas of each row according to σ , and find the area S with the smallest σ in each row:

$$\sigma_{\min} = \min\{\sigma_1, \sigma_2, \cdots \sigma_n\} \tag{8}$$

- Eliminate the area S where $\sigma_{\min}$ is.

- Define the remaining areas as search domains for FAST feature point extraction.

The algorithm flow is shown in the figure 2:

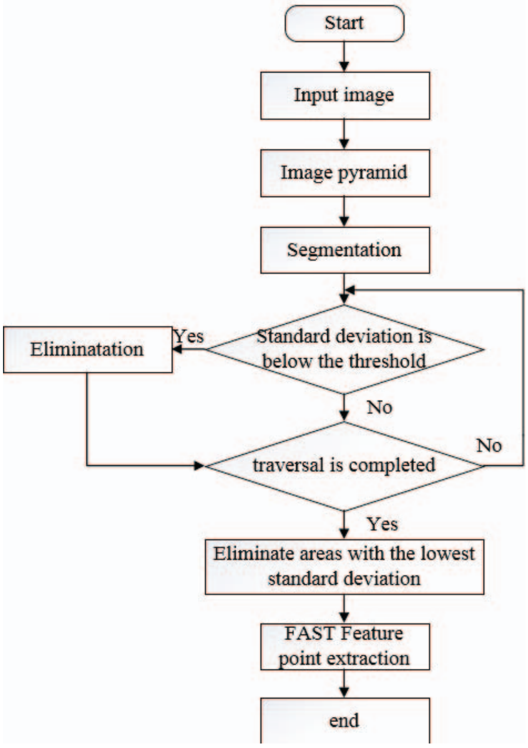

Fig.2 RFIS-slam algorithm

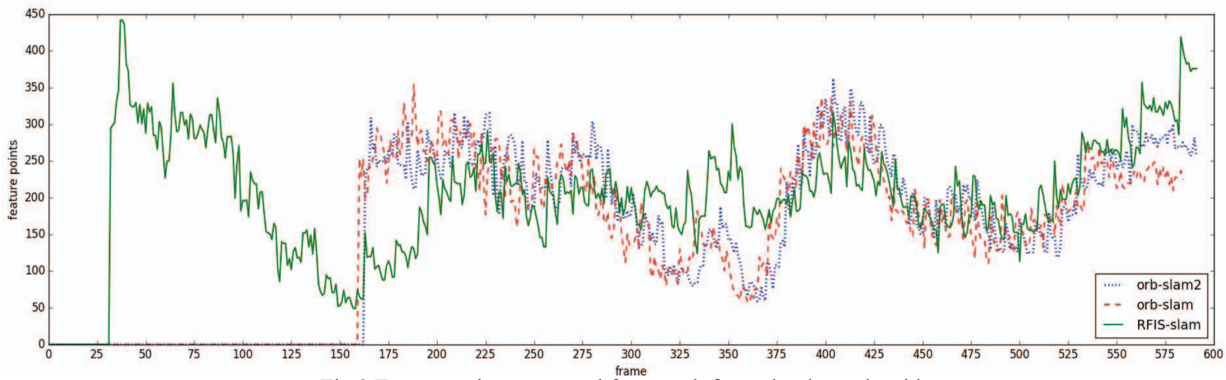

Fig.3 Feature points extracted from each frame by three algorithm

IV. Experiment

To evaluate the effect of the RFIS-slam, we conducted three sets of contrast experiments with the original orb-slam and orb-slam2 in the Linux system, Ubuntu 16.04 operating system. The data set named rgbd_dataset_freiburg1_desk is used. All experiments were completed on a desktop computer with Intelr CoreTM i7-4720HQ CPU, clocked at 2.60GHz, quad-core, and 8 GiB of memory.

A. Feature point matching

From Fig.3, it is obvious that the RFIS-slam has made great development in initialization than the original two orb-slam. Because the system's initialization condition is that there are more than 100 feature points extracted in two continuous frames and the number of feature points successfully matched between the two frames is also more than 100, so the algorithm in this paper can quickly capture the robust feature points in the first two hundred frames to match and maintain a certain number of successful matching points to keep the system tracking.

Because the RFIS-slam is the same as the original two orb-slam in terms of FAST extraction algorithm, the results obtained are not much different from the original two orb algorithms in most of the frames where the lens is stable and the image features are rich; However, in the interval between 300 frames and 375 frames in which the camera movement is relatively fast, the new system can extract more stable feature points and enhance the robustness of tracking for the parts that are easy to locate and lose.

B. *Algorithm performance*

In order to evaluate the real-time performance and matching effect of RFIS-slam, the results of ten experiments are continuously counted under the same conditions and the average value is calculated as shown in Table 1 and 2.

TABLE I. THE RESULTS OF TIME

The Results of Time	**Algorithm**		
	Orb-slam	*Orb-slam2*	***RFIS-slam***
Median time/s	0.0202	0.0251	**0.0283**
Mean time/s	0.0232	0.0268	**0.0298**

In Table 1, the average extraction time of RFIS-slam is increased by about 3ms per frame compared to orb-slam2, and increased by about 6.6ms per frame compared to orb-slam due to the addition of an image evaluation step during feature point extraction.

The time taken by the system to operate each frame has increased, but it is still within the acceptable range for rapid extraction of 30ms per frame, so it can be regarded as no problem in terms of detection time.

TABLE II. THE RESULTS OF PERFORMANCE

The Results of Performance	**Algorithm**		
	Orb-slam	*Orb-slam2*	***RFIS-slam***
matched percentage	16.0%	14.3%	**20.1%**
Initialization time of frame	188.6	143	**77.6**

It can be seen from Table 2 that RFIS-slam is 4.1% better than orb-slam, 5.8% better than orb-slam2 in terms of feature point matching, and 111 frames earlier than orb-slam, 66.4 frames earlier than orb-slam2 in initialization, which proving that our system can achieve good results in both aspects.

C. *localization and tracking*

Fig.4(a),(b),(c) are the comparison between the real trajectory (red line)graph obtained on the rgbd_dataset_ freiburg1 _ desk by orb-slam, orb-slam2 and RFIS-slam(blue line).

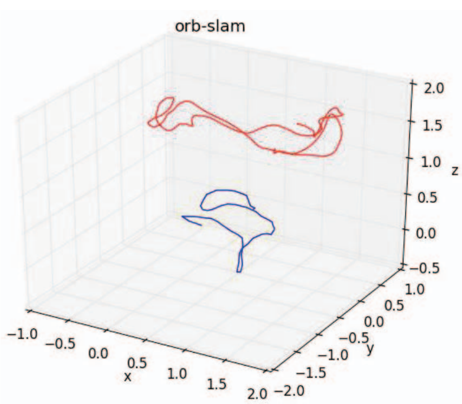

Fig.4(a)

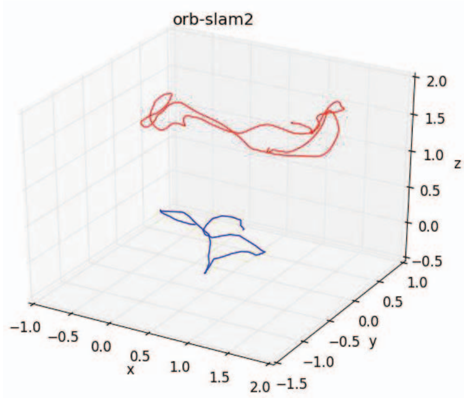

Fig.4(b)

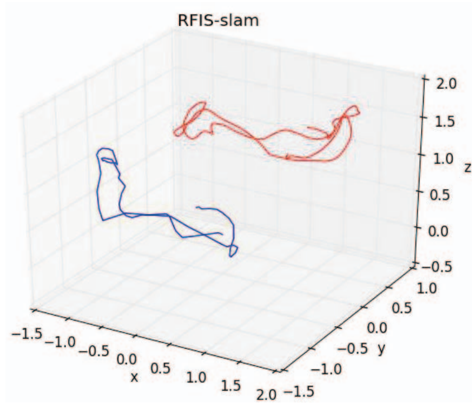

Fig.4(c)

From these results, it can be clearly concluded that the trajectory of RFIS-slam has the highest similarity with the real trajectory, and it can successfully estimate a lot of motion such as the rapid lateral movement of the lens and the rotation of the small area, so as to obtain a better pose estimation and localization effect.

V. CONCLUSION

In this paper, an improved orb algorithm called RFIS-slam is proposed, which can extract more robust and more representative feature points, and verifies its effective function on pose estimation and positioning through experiments. Moreover, based on the original orb-slam's excellent performance in terms of speed, even if the new algorithm sacrifices some speed for image evaluation, it can still achieve the requirements of slam system real-time tracking.

The part of feature point matching in orb-slam system still has many problems need to be solved urgently, such as simply rotating the lens, tracking loss caused by insufficient contrast in the entire frame, etc., which is the direction of efforts to break through.

REFERENCES

[1] Frese U. A discussion of simultaneous localization and mapping.Autonomous Robots,2006,20(1):25-42

[2] Davison A J, Reid I D,Molton N D,Stasse O. MonoSLAM: real-time single camera SLAM[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2007,29(6):1052-1067

[3] Cvisic I, Cesic J, Markovic I,Petrović I. SOFT-SLAM: Computationally efficient stereo visual simultaneous localization and mapping for autonomous unmanned aerial vehicles*[J]. Journal of Field Robotics,2018,35(4):578-595

[4] Kerl C, Sturm J, Cremers D. Robust odometry estimation for RGB-D on Robotics and Automation. Karlsruhe,Germany: IEEE,2013.3748-3754

[5] Richard H.Multiple view geometry in computer vision [M]. 2nd ed.England: Cambridge University Press,2003

[6] Gao X S, Hou X R, TANG J, et al. Complete solution classification for the perspective-three-point problem[J]. IEEE Transactions on Pattern Analysis & Machine Intelli- gence,2003,25(8) : 930-943.

[7] Lepetit V, Moreno-Noguer F, Fua P. Epnp: an ac- curate o (n) solution to the PnP problem [J]. Interna- tional Journal of Computer Vision,2009,81 (2): 155- 166.

[8] Besl P J, Mckay Neil D. A method for registration of 3- D shapes[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,1992,14(2):239-256.

[9] Rusinkiewicz S, Levoy M. Efficient variants of the ICP algorithm[C]//Proceedings of Third International Conference on 3-D Digital Imaging and Modeling. Quebec: IEEE,2001: 145-152.

[10] David G. Distinctive image features from scale invariant keypoints[J]. International Journal of Computer Vision, 2004,60(2):91-110.

[11] Herbert B, Andreas E ,Tinne T,Luc V G. Speeded-up robust features[J]. Computer Vision and Image Understanding, 2008;110(3):346-359

[12] Rublee E, Rahaud V, Konolige K, Bradski, G. ORB: an efficient alternative to SIFT or SURF. Proceedings of the IEEE International Conference on Computer Vision, Barcelona, Spain,2011:2564-2571

[13] Mur-Artal, Raul Montiel, J.M.M, Tardos, Juan D. ORB-SLAM: A versatile and Accurate Monocular SLAM System. IEEE Transactions on Robotics 31(2015):1147-1163.

[14] Mur-Artal, Raul and Juan D.Tardos. ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras. IEEE Transactions on Robotics 33 (2017):1255-1262.

[15] Bescos,Berta, Facil, Jose M; Civera,Javier; Neira, Jose. DynaSLAM: Tracking,Mappin,and Inpainting in Dynamic Scenes. IEEE Robotics and Automation Letters 3(2018):4076-4083.

[16] Yu,Chao et al. DS-SLAM :A Semantic Visual SLAM towards Dynamic Environments. 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(2018):1168-1174

[17] Lianos, Konstantinos-Nektarios, Schönberger, Johannes L. Pollefeys, Marc; Sattler, Torsten. VSO: Visual Semantic Odometry. ECCV(2018).

[18] Rosten E, Porter R,Drummond T. Faster and better: a machine learning approach to corner detection. [J]. IEEE transactions on pattern analysis and machine intelligence, 2010,32(1).

[19] Gao Xiang, Zhang Tao, Liu Yi, Yan Qin-Rui. Fourteen the Visual SLAM from Theory to Practice. Beijing:Electronic Industry Press,2017.