Journal of System Simulation

Volume 34 | Issue 1 Article 12

1-14-2022

Adaptive Optimization in Feature-based SLAM Visual Odometry

Yanan Yu
School of Information Technology and Engineering, Tianjin University of Technology and Education, Tianjin 300222, China;

Dunhuang Shi
School of Information Technology and Engineering, Tianjin University of Technology and Education, Tianjin 300222, China;

Chunjie Hua
School of Information Technology and Engineering, Tianjin University of Technology and Education, Tianjin 300222, China;

Adaptive Optimization in Feature-based SLAM Visual Odometry

Abstract

Abstract: Aiming to reduce the impact of dynamic environments on simultaneous localization and mapping (SLAM) of mobile robots, *an adaptive optimization method in a feature-based visual odometry is proposed. The method helps to improve the invariance of image feature in illumination changing situation and to extract features effectively in areas where the texture information is not sufficient to make contributions to feature matching.* Meanwhile, down sampling is applied to establish image pyramids and each scaled image is divided into cells based on a defined rule. Illumination adaptive nonlinear adjustments for each cell are applied to increase the image details, and low-texture area is removed by computing the image gray level probability distribution. Based on the proposed method, a visual odometry of SLAM system is built and verified on TUM dataset. The results show that, compared with the original system, the proposed method can reduce the trajectory errors of a mobile robot and also improve the performance of robot visual odometry in the unstable dynamic environments.

Keywords

mobile robot, visual odometry, SLAM, feature-based method, low-texture area

Recommended Citation

Yu Yanan, Shi Dunhuang, Hua Chunjie. Adaptive Optimization in Feature-based SLAM Visual Odometry[J]. Journal of System Simulation, 2022, 34(1): 104-112.

第 34 卷第 1 期 系统仿真学报© Vol. 34 No. 1
2022 年 1 月 Journal of System Simulation Jan. 2022

Adaptive Optimization in Feature-based SLAM Visual Odometry

Yu Yanan, *Shi Dunhuang*, *Hua Chunjie*

(School of Information Technology and Engineering, Tianjin University of Technology and Education, Tianjin 300222, China)

Abstract: Aiming to reduce the impact of dynamic environments on simultaneous localization and mapping (SLAM) of mobile robots, *an adaptive optimization method in a feature-based visual odometry is proposed. The method helps to improve the invariance of image feature in illumination changing situation and to extract features effectively in areas where the texture information is not sufficient to make contributions to feature matching.* Meanwhile, down sampling is applied to establish image pyramids and each scaled image is divided into cells based on a defined rule. Illumination adaptive nonlinear adjustments for each cell are applied to increase the image details, and low-texture area is removed by computing the image gray level probability distribution. Based on the proposed method, a visual odometry of SLAM system is built and verified on TUM dataset. The results show that, compared with the original system, the proposed method can reduce the trajectory errors of a mobile robot and also improve the performance of robot visual odometry in the unstable dynamic environments.

Keywords: mobile robot; visual odometry; SLAM; feature-based method; low-texture area

特征点法 SLAM 视觉里程计自适应优化算法

于雅楠，史敦煌，华春杰

(天津职业技术师范大学 信息技术工程学院，天津 300222)

摘要：为减少动态环境对移动机器人同时定位与地图构建(simultaneous localization and mapping, SLAM)的影响，*提出了一种特征点法视觉里程计自适应优化算法。该算法有助于改善光照条件变化情况下图像特征的不变性，有效提取纹理信息不充分区域的特征用于图像匹配。*采用降采样法建立图像金字塔，将每个缩放后的图像根据预先设定规则划分为多个图像块。在每个图像块上进行光照非线性调整来增加图像细节，通过计算图像灰度概率分布来剔除无纹理区域。基于提出的方法建立了 SLAM 系统视觉里程计，并在 TUM 数据集上进行了验证。结果表明：该算法可以减小移动机器人运动轨迹误差，改善机器人在不稳定动态环境下视觉里程计的性能。

关键词：移动机器人；视觉里程计；SLAM；特征点法；弱纹理区域

中图分类号：TP391 文献标志码：A 文章编号：1004-731X (2022) 01-0104-09

DOI: 10.16182/j.issn1004731x.joss.20-0424E

Introduction

SLAM (simultaneous localization and mapping) technology helps mobile robots to confirm positions and construct a surrounding map through continuous observations in unstructured environments, particularly the GPS-denied environments. SLAM provides necessary support for the robot's autonomous navigation, positioning, path planning

Received date: 2020-06-29 Revised date: 2021-03-22

Fundation: Natural Science Foundation of Tianjin (18JCYBJC84900)

First author: Yu Yanan (1984-), female, PhD, Lecture, research area: computer control technology, photoelectric detection and visual measurement. E-mail: jesuisyyn@126.com

and obstacle avoidance. Its application has been extended to various fields such as drones, unmanned boat, driverless cars, virtual reality and augmented reality[1-3]. According to the different types of sensors of the input channel of a robot, SLAM technology is usually divided into laser-based SLAM and visual SLAM. The former provides users with a range of two or three-dimensional measurement, while the latter provides more abundant image data to assist robots to complete the tasks of scene recognition and semantic perception.

In visual SLAM, in addition to the traditional monocular and binocular cameras for capturing RGB images, the depth-sensing devices such as Kinect, RealSense and Xtion provide the data with both color and depth information (RGBD data). Depth-sensing cameras measure a distance mainly based on two principles of structured light and TOF (time of flight). Structured-light sensor relies on laser speckle images to compute the object's three-dimensional shape, and TOF sensor resolves the target object distance by calculating the time difference or the light phase shift between the emission and the reflection. With the help of the depth information, it is convenient to achieve the three-dimensional scene reconstruction and recover three-dimensional structure of objects or their three-dimensional motion. At present, the depth-sensing cameras have some performance limitations, but, compared to the scale uncertainty of the monocular camera and computing resource demands of the binocular camera, their advantages and potential values in SLAM are obvious.

Indirect method and direct method are two approaches of the state-of-the-art visual odometry. Their main difference is that the traditional indirect method optimizes the reprojection error of features extracted from two images to estimate a camera pose while the direct method optimizes photometric error of two images and uses whole pixel information directly for the estimate. In the direct method, despite the image information being heavily depended to maximize the utilization, an assumption of unchanged pixel gray value for the same space point measured under different angles of view is adopted. When an application is in diffuse reflection, shading, light changes or other special environments, the direct method using the corresponding pixel as the objective optimization is not established. In addition to the above discussion, the environmental requirement in the application is also a research challenge of SLAM. The system might be fragile in dynamic environments, which is not only because of the moving of obstacles, elements and people, but also the environmental change such as the illumination and texture information[4-5]. Robustness, drift as well as the latency in a SLAM system for a long time and large-scale real-time application is highly demanded, especially[6].

In this paper, in indirect visual odometry, an illumination adaptive method is proposed to increase the image details and enhance the contrast of brightness. In order to improve the image feature extraction ability, Gamma nonlinear adjustment is adopted in different regions of input image.

1 Related Works

In this section, the developments and existing works on visual SLAM is reviewed, and the calculation basis of robot’s position and orientation in visual SLAM is introduced. Finally, based on the problem in practical application, the optimization method is proposed.

https://dc-china-simulation.researchcommons.org/journal/vol34/iss1/12
DOI: 10.16182/j.issn1004731x.joss.20-0424E

第 34 卷第 1 期 系统仿真学报 Vol. 34 No. 1
2022 年 1 月 Journal of System Simulation Jan. 2022

1.1 SLAM Framework

SLAM refers to an overall framework of robots' localization and map construction. With RGBD data obtained by depth-sensing cameras, such as Kinect, a complete visual SLAM system includes four major parts: visual odometry, back-end optimization, loop detection, and map construction. Up to the present, there are many excellent and mature visual SLAM systems. Mono-SLAM[7] developed by Davison et al is a successful implementation of the visual system, by which 3D trajectory of a monocular camera is recovered in real-time. Klein et al built a unique innovation named PTAM(Parallel Tracking And Mapping)[8], a SLAM framework with parallel threads in tracking and mapping. Algorithms of feature analysis and pose tracking using key frames are introduced to make a major breakthrough in real-time and stability. In addition, the RGB-D SLAM[9] system provided by Endres et al and the ORB-SLAM[10] open source developed by Mur-Artal et al are both excellent extension of PTAM. Among them, ORB-SLAM2[11] completes image sequence matching and tracking based on ORB features[12], even in loop detection and relocalization, and also provides interface for implementation in monocular, binocular, RGBD, and robot operating systems (ROS). Due to a sparse point clouds built in ORB-SLAM2, obstacle avoidance and navigation in practical tasks could not be used directly. Based on this system, Lv et al used octree structure and spatial prediction to extend an intensive 3D map[13]. Elvira et al proposed ORBSLAM-Atlas to deal with dynamic scenes. It brings the wide-baseline matching detection and exploitation to the multiple map domains with general and robust result[14].

Different from the indirect method which is based on sparse features, Newcombe et al proposed DTAM(Dense Tracking and Mapping in Real-Time)[15]. DTAM maximally utilizes the whole information of image with fewer features and build an intensive map. In order to reduce the computational cost in dense SLAM and to improve the system stability, as a representative of semi-dense types, LSD-SLAM[16] and DSO visual odometry[17] were released by Engel. They only focus on estimation by high gradient pixels and rich texture areas to deal with the system performance of robot localization and dense map creation in a large-scale space. SVO-SLAM[18] developed by Forster et al used a sparse image alignment algorithm, which was combined with feature correspondence and photometric error minimization. SVO-SLAM is a novel semi-direct visual odometry that faster and more accurate than others. Zubizarreta et al presented Direct Sparse Mapping, a full direct monocular visual SLAM based on photometric bundle adjustment[19]. This method reduces both estimated trajectory and map error at the same time to handle reobservations.

1.2 Visual Odometry

Visual odometry is for calculating the position and orientation of the robot by analyzing any two images and determining a global pose through all transformations. Image alignment is usually accomplished by geometric consistency and photometric consistency. Geometric consistency is to minimize geometric reprojection error of features coordinates in the indirect method, and photometric consistency is to minimize photometric reprojection error of pixel intensity in the direct method. The principle is shown in Fig. 1~2.

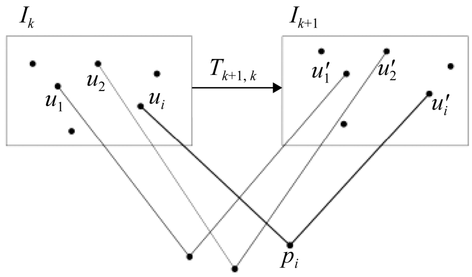


Fig. 1 Feature-based geometric reprojection diagram

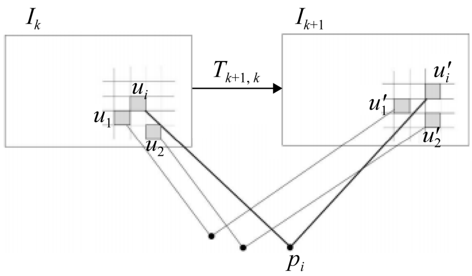


Fig. 2 Pixel-based photometric reprojection diagram

In Fig. 1, feature points u_1, u_2, ··· , u_n and u_1', u_2', ··· u_n' are extracted from adjacent images I_k and I_{k+1} respectively. Feature correspondences are determined through a descriptor matching between two image projection coordinates such as u_i and u_i', both of which are the projection of the same space point p_i in a camera coordinate system. If the transformation matrix of the two images is $\boldsymbol{T}_{k+1,k}$ then

$$u_i' = \boldsymbol{T}_{k+1,k} \cdot u_i \tag{1}$$

However, $T_{k+1,k}$ is an estimated value, the reprojection error minimization is constructed as the objective optimization:

$$\boldsymbol{T}_{k+1,k} = \arg\min \sum_{i=1}^{n} || u_i' - \boldsymbol{T}_{k+1,k} \cdot u_i ||^2 \tag{2}$$

Where n presents the number of image features that participate in the optimization.

In Fig. 2, the gray value of any pixel u_i in I_k is presented as $I_k(u_i)$, and its space projection coordinate in the camera system is p_i. If the transformation matrix of two images is $\boldsymbol{T}_{k+1,k}$, the reprojection coordinate of p_i in $\boldsymbol{T}_{k+1}$ is u_i' and its gray value is presented as $I_{k+1}(u_i')$. In the condition that the pixel gray value of the same space point measured under different view angles is invariant, then

$$I_{k+1}(u_i') = I_k(u_i) \tag{3}$$

If equation (3) is not satisfied, that means the estimation of $\boldsymbol{T}_{k+1,k}$ is not accurate, and the photometric error minimization is constructed as the objective optimization to calculate the optimal transformation matrix:

$$\boldsymbol{T}_{k+1,k} = \arg\min \sum_{i=1}^{n} || I_k(u_i) - I_{k+1}(\boldsymbol{T}_{k+1,k} \cdot u_i) ||^2 \tag{4}$$

where n presents the number of image pixels that participate in the optimization.

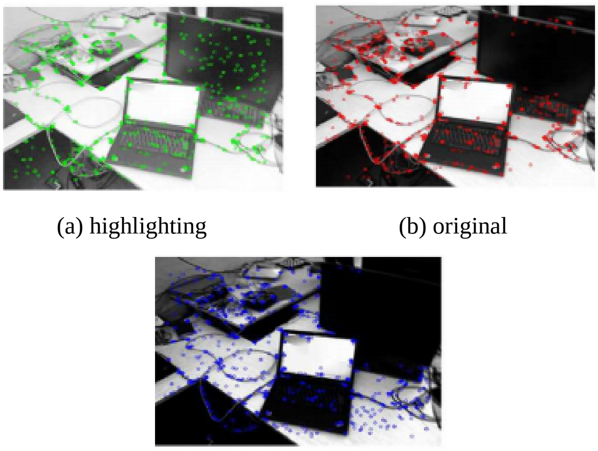
1.3 Visual Odometry Optimization

Due to the strong assumption in the direct method, algorithms of visual odometry have lower tolerance to shadows, poor illumination conditions, and exposure parameters changing. On contrast, the indirect method has certain tolerance for illumination, but the performance of sensor is differently influenced by light intensity. For the same scene, according with illumination condition, the ability of feature extraction is fixed, which may affect the performance of feature detection algorithm. As shown in Fig. 3, compared to the original image (original), feature points extracted from highlighting image (highlighting) and darkening image (darkening) are discrepant, and the details of different local areas are more visible even in the area with low and high brightness.

Instead of the standardized pre-processing for image sequences, this paper proposes an illumination adaptive method in visual odometry. In the available literature, Gamma correction is usually used to modify the whole image, which cannot be called an adaptive adjustment method. However, in the paper, Gamma correction is adopted for different regions in image pyramids to complete the nonlinear adjustment. This way, in which the process reduces the local shadow caused by light changes, enhances the

第 34 卷第 1 期　　系统仿真学报　　Vol. 34 No. 1
2022 年 1 月　　Journal of System Simulation　　Jan. 2022

contrast of brightness, increases the image details, and benefits the noise suppression. In addition, in order to improve efficiency for feature extraction and descriptor calculation, a gray level probability distribution is used to remove low-texture regions and even the area of no obvious gradient. On the basis of the proposed method, features with scale invariance and rotation invariance are extracted from image and are matched among frames to estimate motion and generate the trajectories of the camera/robot.



(a) highlighting　　(b) original

(c) darkening

Fig. 3　Feature comparation before and after illumination adjustment

2　System Development and Analysis

2.1 System Framework

System framework of the proposed illumination adaptive method in visual odometry is shown in Fig. 4. In the algorithm, the part of adaptive sparse feature extraction is used to improve the systematic performance of visual odometry.

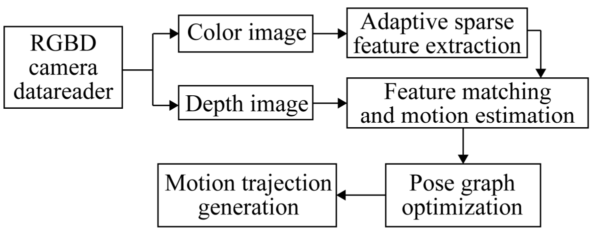


Fig. 4　Illumination adaptive visual odometry system

Adaptive sparse feature extraction algorithm is explained as follows. Firstly, before establishing a n level image pyramids, each frame in input sequences is converted to grayscale for better subsequent processing. Then, on each level of pyramids, the whole image is divided into grid cells with exactly the same size c. For each cell, change in illumination is computed and adjusted adaptively and nonlinearly by equation $G(x, y)$. Local illumination correction here could enhance the contrast of the region, increase the image details and suppress the noise interference. In order to reduce sub-regions with very poor texture, low-texture area is removed by analyzing the gray level probability density distribution. The removal of low-texture area could improve the efficiency for feature extraction without performance loss. FAST corners are extracted on retained cells. The corner coordinates extracted from each pyramid level are restored to the original image according to the scale factor. The key points are allocated and managed by quadtree, and the strongest corner is selected as the feature of the node according to the pre-set threshold to achieve the uniform distribution of image features. After corners screening, finally, orientation and descriptor for each feature are calculated and output.

2.2 Illumination Adaptive Model

A generic illumination correction function for image processing is given:

$$G(x, y) = (I(x, y)/255)^{\gamma} \times 255 \tag{5}$$

Where, $I(x, y)$ is the gray value of a pixel (x, y); $G(x, y)$ is the corrected gray value of the pixel (x, y); γ is a correction parameter. As shown in Fig. 5, the original image without adjustment is on red line. It can be brightened when γ is less than 1, such like the image on green line with Gamma 0.25 and the image on blue line with Gamma 0.5. On contrast, the same image can be darkened when γ is more than 1, such like the image on dark line with Gamma 2 and the image on purple line with Gamma 5.

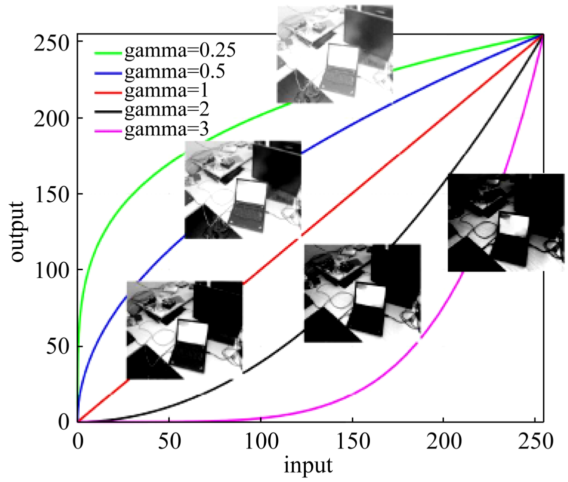


Fig. 5 Nonlinear adjustment under different parameters

In order to adjust different regions in image, here γ is determined upon the mean gray value of each cell adaptively:

$$\gamma = \left[\lg \frac{I_{\text{mean}}}{255} / \lg \mu \right]^{-1} \tag{6}$$

In equation (6), μ is the grayscale correction threshold with the range (0, 1). In this paper, 0.5 is used as the value of empirical parameter μ. In addition, except the impractical pure black or pure white in actual application, I_{mean} cannot be 0 or 255 to contradict the definition of logarithmic computation.

On each level of image pyramids, in order to improve the efficiency for extracting corners, a pixel gray level probability density distribution of grid cell is established. r_k is the gray value corresponding to the peak value. If the pixel number n_k in the interval $[r_k-m, r_k+m]$ around r_k is more than a certain proportion p of the total pixel number N, this region of image is considered as the area with low-texture or no obvious gradient. So that it can be removed due to less contribution to corner extraction. Here, threshold m and p are both empirical values.

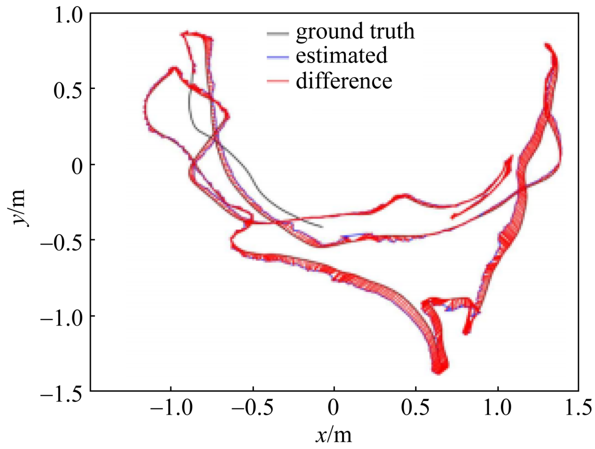
3 Experimental Results

An experimental validation is provided to the proposed method. In the experiment, datasets containing kinect RGB-D data and ground-truth are provided by the computer vision group from TUM[20]. The system adopts ubuntu 16.04 and the reading speed for the image sequence is set to 30 fps. All testing is completed on desktop (Intel Core™ i5-3470 CPU@3.20 GHz×4, 15.6 GiB memory). When running the algorithms of visual odometry, the back-end of the system adopts g2o to optimize a pose graph and generate a motion trajectory of the camera. For evaluation of the proposed method, three datasets are selected as benchmarks and compare results with an open-source SLAM system ORB-SLAM2.

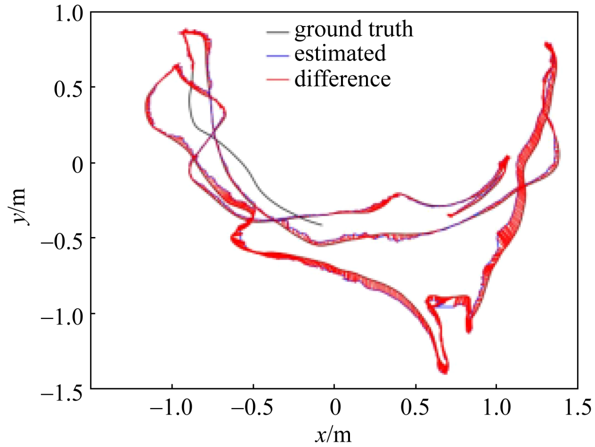
In the experiment, three video sequences from a hand-held camera that fr1/floor, fr1/desk and fr1/room are used. All of those are sequences with several loop closures in the same office scene. The fr1/floor sequence contains a camera sweep over the wooden floor, the fr1/room sequence cover the whole room with slow motion and the fr1/desk sequence cover two tables with higher motion. The motion trajectory of the optimized system is constructed and its accuracy is compared with that generated by ORB-SLAM2. Compared with the benchmark, the constructed trajectories on fr1/room are shown in Fig. 6. From Fig. 6(a) and Fig. 6(b), it can be seen that the estimation result from the proposed system is better than ORB-SLAM2 especially during the sharp turn at the bottom.

https://dc-china-simulation.researchcommons.org/journal/vol34/iss1/12
DOI: 10.16182/j.issn1004731x.joss.20-0424E

第 34 卷第 1 期 系统仿真学报 Vol. 34 No. 1
2022 年 1 月 Journal of System Simulation Jan. 2022



(a) ORB-SLAM2 trajectory



(b) Trajectory from the proposed method

Fig. 6 fr1/room: Comparison of constructed trajectories with the benchmark result

3.1 Absolute Trajectory Error

The root mean square error (rmse) of absolute trajectory errors (ate) is used as a measure to evaluate the accuracy of the trajectory estimation. It is defined as:

$$RMSE(x) = \sqrt{\frac{\sum_{i=1}^{n} ||(x_{e,i} - x_{s,i})||^2}{n}} \tag{7}$$

where, $x_{e,i}$ represents the estimated location of frame i in the image sequence and $x_{s,i}$ represents the standard location (ground truth) of the same frame.

The evaluation result is given in Table 1. Both of the tracking time and ate.rmse values in table is the average of 10 tests. Since the illumination correction process is added in the proposed method, the mean tracking time for each frame is nearly doubled, but the tracking accuracy is obviously increased compared to that from ORB-SLAM2. The relative improvement of tracking accuracy is also listed in last column.

3.2 Tracking Lost Improvement

On some occasions, odometry tracking tends to get lost and the current camera could not be found by a system if there is a fast moving or big rotation. In this situation, there are no sufficient key points visible to match frames and then to construct a map. When running dataset fr1/room, the odometry loses tracking at frame 845 occasionally. This frame and its two adjacent images are shown in Fig. 7. There is a white cabinet with low texture in the frame and it is easy to lose camera locations under fast moving because of the insufficient corresponding features.

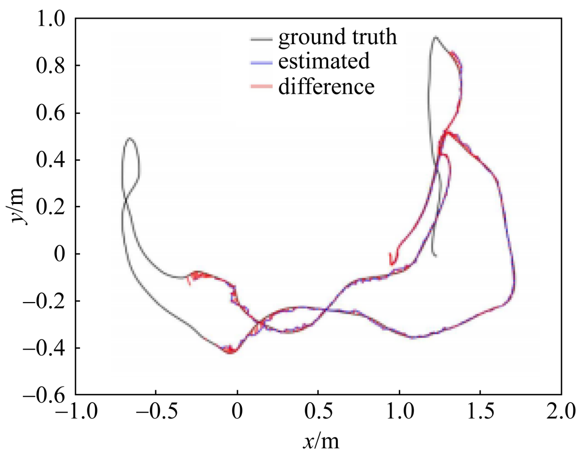
This situation also happens in another dataset fr1/desk. Although the relocalization module can recover the camera pose after a period and continue tracking operation when a cloud map has been previously obtained, the system doesn't have enough information to estimate the camera's position and pose correctly when frames are lost. It is interesting that the proposed method can track these frames correctly. This is because of the adaptive illumination correction improving the feature extraction. Details of image regions are enhanced and made available to the points on image gradients such as corners and edges. The performance improvement is shown in Fig. 8.

Table 1 Evaluation results

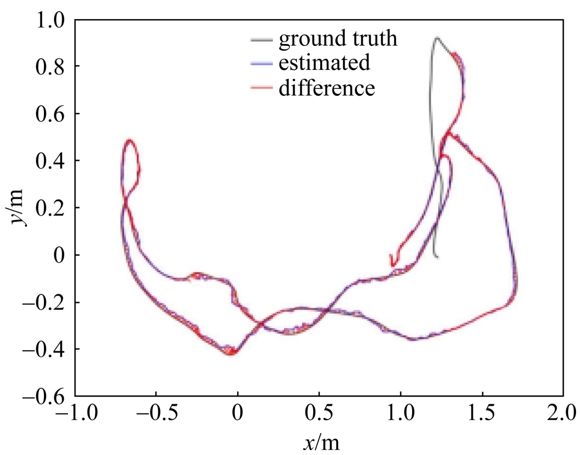
	ORB-SLAM2		optimized system		
dataset	*t*/s	ate.rmse/m	*t* /s	ate.rmse/m	Improvement/%
fr1/room	0.035 955	0.057 292	0.064 367	0.048 388	15.54
fr1/floor	0.028 165	0.015 982	0.055 230	0.014 204	11.13
fr1/desk	0.036 486	0.018 198	0.070 435	0.015 475	14.96

(a) previous frame (b) frame 845 (c) next frame

Fig. 7 Frames of fr1/room, which is easy to get lost in visual odometry



(a) ORB-SLAM2 trajectory



(b) Trajectory from the proposed method

Fig. 8 fr1/desk: Comparison of constructed trajectories with the benchmark result

4 Conclusions

In this paper, an adaptive optimization method based on the indirect method of visual odometry is proposed to add illumination invariance to the sparse features extracted from dynamic environments. It improves the performance of the feature-based visual odometry. The optimized system is evaluated on TUM kinect datasets with ground-truths fr1/room, fr1/floor and fr1/desk. The test results have shown the following advantages of the proposed method compared to original SLAM system: ① The proposed method improves the trajectory accuracy. Compared to ORB-SLAM2, optimized algorithm combined with low-texture remove process relatively reduces the absolute trajectory error with 15.54%, 11.13% and 14.96% respectively; ② The proposed method improves the system robustness. More local details of different image area effectively avoid the tracking lost under camera fast moving and large rotation. At present, however, there is performance improvement potential in next study. Because of the added illumination correction process, the computation cost is slightly high to extract feature

https://dc-china-simulation.researchcommons.org/journal/vol34/iss1/12
DOI: 10.16182/j.issn1004731x.joss.20-0424E

第 34 卷第 1 期 系统仿真学报 Vol. 34 No. 1
2022 年 1 月 Journal of System Simulation Jan. 2022

points and compute features descriptors in the indirect method, which may influence the system real-time performance. In addition, since the frame registration in the method is based on sparse features extracted from relevant images, tracking lost may happen in image areas where the texture information is not sufficient to generate features.

References:

[1] Fink G, Franke M, Lynch A F, et al. Visual Inertial SLAM: Application to Unmanned Aerial Vehicle[J]. IFAC-PapersOnline (S2405-8963), 2017, 50(1): 1965-1970.

[2] Ros G, Sappa A D, Ponsa D, et al. Visual SLAM for Driverless Cars: a Brief Survey[C]//IEEE Workshop on Navigation, Perception, Accurate Positioning and Mapping for Intelligent Vehicles. Los Alamitos: IEEE Computer Society Press, 2012: Article No.3.

[3] H¨oll M, Lepetit V. Monocular LSD-SLAM Integration Within AR System[EB/OL]. ArXiv: 1702.02514 (2017-02-08) [2020-06-01].https://arxiv.org/pdf/1702.02514.pdf.

[4] Cadena C, Carlone L, Carrillo H, et al. Past, Present, and Future of Simultaneous Localization and mapping: Towards the Robust-perception Age[J]. IEEE Transactions on Robotics (S1552-3098), 2016, 32(6): 1309-1332.

[5] Karlsson N, Bernardo E D, Ostrowski J P, et al. The vSLAM Algorithm for Robust Localization and Mapping[C]//IEEE International Conference on Robotics and Automation. Barcelona, Spain: IEEE Press, 2005: 24-29.

[6] Zhao Y, Smith J S, Karumanchi S H, et al. A. Closed-loop Benchmarking of Stereo Visual-inertial SLAM System: Understanding the Impact of Drift and Latency on Tracking Accuracy[C]//IEEE International Conference on Robotics and Automation. Paris, France: IEEE Press, 2020: 1105-1112.

[7] Davison A J, Reid I D, Molton N D, et al. MonoSLAM: Real-time Single Camera SLAM[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence (S0162-8828), 2007, 29(6): 1052-1067.

[8] Klein G, Murray D. Parallel Tracking and Mapping for Small AR Workspaces[C]// 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality. Piscataway, NJ: IEEE Press, 2007: 225-234.

[9] Endres F, Hess J, Sturm J, et al. 3-D Mapping with an RGB-D Camera[J]. IEEE Transactions on Robotics (S1552-3098), 2014, 30(1): 177-187.

[10] Mur-Artal J, Montiel J M M, Tardós J D. ORB-SLAM: a Versatile and Accurate Monocular SLAM System[J]. IEEE Transactions on Robotics (S1552-3098), 2015, 31(5): 1147-1163.

[11] Mur-Artal R, Tardós J D. ORB-SLAM2: an Open-source SLAM System for Monocular, Stereo and RGB-D Cameras[J]. IEEE Transactions on Robotics (S1552-3098), 2017, 33(5): 1255-1262.

[12] Rublee E, Rabaud V, Konolige K, et al. ORB: an Efficient Alternative to SIFT or SURF[C]//IEEE International Conference on Computer Vision (ICCV). Barcelona, Spain: IEEE Press, 2011: 2564-2571.

[13] Lv Q, Lin H, Wang G, et al. ORB-SLAM-based Tracking and 3D Reconstruction for Robot Using Kinect 2.0[C]//2017 29th Chinese Control and Decision Conference (CCDC). Chongqing, China: IEEE Press, 2017: 3319-3324.

[14] Elvira R, Tardós J D, Montiel J M M. ORBSLAM-Atlas: a Robust and Accurate Multi-map System[C]//IEEE/RSJ International Conference on Intelligent Robots and System. Macau, China: IEEE Press, 2019: 6253-6259.

[15] Newcombe R A, Lovegrove S J, Davison A J. DTAM: Dense Tracking and Mapping in Real-time[C]//IEEE International Conference on Computer Vision (ICCV). Barcelona, Spain: IEEE Press, 2011: 2320-2327.

[16] Engel J, Schöps T, Cremers D. LSD-SLAM: Large-scale Direct Monocular SLAM[C]//European Conference on Computer Vision. Zurich, Switzerland: Springer, 2014: 834-849.

[17] Engel J, Koltum V, Cremers D. Direct Sparse Odometry[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence (S0162-8828), 2018, 40(3): 611-625.

[18] Forster C, Pizzoli M, Scaramuzza D. SVO: Fast Semi-direct Monocular Visual Odometry[C]//IEEE International Conference on Robotics and Automation (ICRA). Hong Kong, China: IEEE Press, 2014: 15-22.

[19] Zubizarreta J, Aguinaga I, Montiel J M M. Direct Sparse Mapping[J]. IEEE Transaction on Robotics (S1552-3098), 2020, 36(4): 1363-1370.

[20] Sturm J, Engelhard N, Endres F, et al. A Benchmark for the Evaluation of RGB-D SLAM Systems[C]//IEEE International Conference on Intelligent Robot Systems (IROS). Algarve, Portugal: IEEE Press, 2012: 573-580.