

Robust Edge-Direct Visual Odometry based on CNN edge detection and Shi-Tomasi corner optimization

Kengdong Lu^{1,3,*}, Jintao Cheng^{1,*}, Yubin Zhou¹, Juncan Deng¹, Rui Fan², Kaiqing Luo^{1,3}

Abstract—In this paper, we propose a robust edge-direct visual odometry (VO) based on CNN edge detection and Shi-Tomasi corner optimization. Four layers of pyramids were extracted from the image in the proposed method to reduce the motion error between frames. This solution used CNN edge detection and Shi-Tomasi corner optimization to extract information from the image. Then, the pose estimation is performed using the Levenberg-Marquardt (LM) algorithm and updating the keyframes. Our method was compared with the dense direct method, the improved direct method of Canny edge detection, and ORB-SLAM2 system on the RGB-D TUM benchmark. The experimental results indicate that our method achieves better robustness and accuracy.

I. INTRODUCTION

The visual odometry [1] (VO), which can estimate the camera motion between the adjacent images and the planning of local maps, and plays a vital role in the synchronous positioning and map construction technology [2], [3], [4]. It has many practical applications in robotics, such as autonomous driving, navigation, augmented reality, and three-dimensional reconstruction.

In the past few decades, there has been a lot of work on VO [5], [6]. The idea of VO was first proposed by Moravec et al. [7] Follow-up related work can be divided into feature methods and direct methods. Feature-based (indirect) methods are to extract some salient features from dense image data for calculation. The VO system

Corresponding: Kaiqing Luo. (email: kqluo@scnu.edu.cn)

This work was supported in part by the National Science Foundation of China (NSFC) - Guangdong big data Science Center Project under Grant U1911401, National Natural Science Foundation of China (61975058); Natural Science Foundation of Guangdong Province (2019A1515011401); the Science and Technology Program of Guangzhou (2019050001); and the Science and Technological Plan of Guangdong Province, China (2019B090905005), and in part by the South China Normal University National Undergraduate Innovation and Entrepreneurship Training Program under Grant 202010574050 and 202110574048.

1. K. Lu, J. Cheng, Y. Zhou, J. Deng, K. Luo are with the School of Physics and Telecommunication Engineering, South China Normal University, Guangzhou 510006, P. R. China.

2. R. Fan is with the College of Electronic and Information Engineering, Tongji University, Shanghai 201804, P. R. China. as well as Shanghai Research Institute for Intelligent Autonomous Systems, Shanghai 201210, P. R. China. (email: rui.fan@ieee.org).

3. K. Lu, K. Luo are with the Guangdong Provincial Engineering Research Center for Optoelectronic Instrument, South China Normal University, Guangzhou 510006, P. R. China.

*The authors contributed equally to this work.

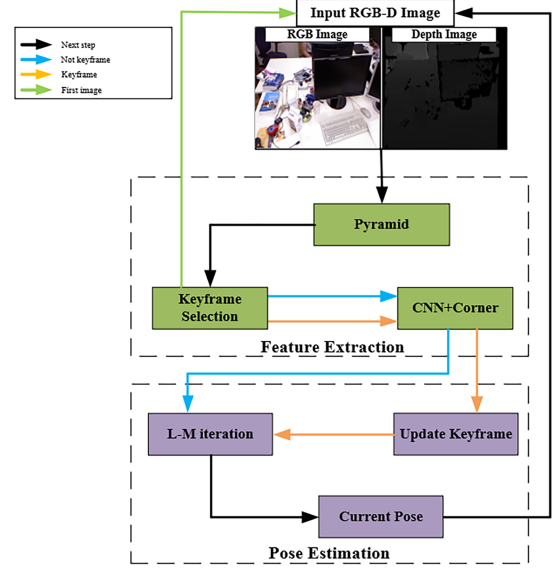


Fig. 1. Flowchart of the our method. Pyramid denote extracts four-layer pyramid, CNN+Corner denote CNN edge detection and Shi-Tomasi corner optimization. L-M iteration denote Levenberg-Marquardt to calculate the pose of two frames iteratively.

using the feature method runs stably with a low computational cost, and it is robust to factors such as illumination, image noise, etc. The famous ORB-SLAM[8] and ORB-SLAM2[9] are based on the feature point method. However, the VO system using the feature method is not suitable for scenes that lack features[10], such as gradual images. Compared with the indirect method, the direct method uses the gray information of all pixels in the image or a specific sub-region to calculate the camera's motion. The VO system using the direct method uses pixel gradients and does not need to use feature points in the image[11]. It makes full use of image information, which is conducive to the realization of visual applications such as building dense maps[12]. However, the direct method has a large amount of calculation and is unsuitable for large motions. Meanwhile, the direct method requires that the image must meet the assumption that the gray pixel value is constant, and this assumption will be destroyed due to illumination and other reasons. According to the number of pixels used, the direct method can be divided into three types: sparse, semi-dense[13], and dense.

In this paper, we propose a robust edge-direct visual odometry based on CNN edge detection and Shi-Tomasi corner optimization. In the evaluation results of the stan-

dard TUM RGB-D benchmark data set, we compared this method with the ORB-SLAM2 system. RPE and ATE data results show that this method has good accuracy in multiple data set sequences.

The contributions of the work:

- Our method uses advanced CNN edge detection algorithms. Combining CNN edge detection and Direct-VO into Edge-Direct VO achieves excellent performance compared to other algorithms in evaluating the TUM RGB-D benchmark.
- In our Edge-Direct VO, we propose an improved Shi-Tomasi angle optimization for edge maps, which optimizes the pose estimation of VO to make the whole method more accurate and robust.
- Keyframes can effectively reduce accumulated errors. Our method uses a dual mechanism combining periodicity and motion amplitude to update keyframes, which significantly improves the accuracy of experimental results.

II. RELATED WORK

In this section, we will briefly introduce edge detection and outline some of the vital work in the field of VO.

A. Edge Detection

Edge detection is a focused area in computers, and it is a challenging subject. The traditional Sobel operator and Canny operator edge detectors have a wide range of applications [15], [16], [17]. However, these detectors only take into account sharp local changes and look for edges from these features, especially sharp changes in color, brightness etc. In 2013, people began to use machine learning-based methods to learn how to combine the features of color, brightness, and gradient for edge detection, such as gPb[14] and StructuredEdge[18]. The Berkeley research group has established an internationally recognized evaluation set called Berkeley Segmentation Benchmark to evaluate the edge detection algorithm better. The BSDS500 data set[19] consists of 500 natural images with hand-labeled edges, which are used to train learning-based edge detection techniques[14], [18]. With the rapid development of deep learning, neural network-based edges detection has become important. This method essentially regards the edge detection process as an edge pattern recognition process. Due to its excellent performance in various edge test data sets, more and more edge algorithms based on deep learning have been considered. In this article, we choose a new CNN edge detection algorithm DexiNed[20], which consists of a bunch of learning filters, receives the image as input, and then predicts the edge map with the same resolution. It shows a better performance.

B. Visual Odometry(VO)

- 1) Feature-based VO: The traditional feature-based (indirect) method extracts features (such as SIFT,

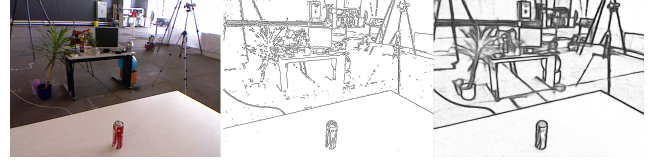


Fig. 2. Original picture (left), Canny edge detection(middle),CNN edge detection (right).

SURF, ORB), finds the correspondence between images, and then tracks them in the sequence to estimate the camera's movement. Due to the high inaccuracy in feature extraction and matching, this algorithm must calculate the basic or homography matrix in the RANSAC loop. In terms of using indirect methods, better-known systems are ORB-SLAM, ORB-SLAM2, and Parallel Tracking and Mapping (PTAM)[21]. In addition, RGBD-SLAM[22] is also a feature-based (indirect) mapping system.

- 2) Direct VO: The direct method can optimize the geometry of the image intensity without the need for feature extraction. Therefore, it can work in some environments without texture. Direct method has been widely used in different sensors, such as DVO[23] of RGB-D camera and LSD SLAM[24] of monocular camera. The core idea is to maintain the semi-dense mapping of keyframes and then minimize the luminosity error. This is a highly non-convex function, so a good initialization is required. However, the direct (feature-free) method estimates camera motion from image data, thus ignoring feature extraction and robust correspondence matching. That's the reason why they are usually restricted to small inter-frame motion, which can only be circumvented to a certain extent by image pyramids. Most of these methods rely on the optical consistency assumption[25], which makes them particularly vulnerable to changes in lighting conditions.
- 3) Edge based VO: Edge is another vital feature besides points. Compared with points, edges can retain more information and are more robust to changes in light. The edge based methods are indirect method and direct method, but the camera motion estimation does not need correspondence, but the camera motion estimation does not need to correspond. The general idea is to minimize the distance between the edge of one frame and the reprojected edge of another frame.

Eade and Drummond[7], Klein, and Murray[21] have mentioned edge-based SLAM. They gathered edge pixels into edges, parameterized and processed them as features. However, there was a problem that the edges are difficult to match. Tarrio and Pedre[26] proposed an edge-based

monocular camera VO, and they searched along the normal direction to match the edge. It also speeds up the matching step by pre-calculating the distance transform (DT) in a frame[27] to speed up this matching step. This idea was adapted by Kuse and Shen[28] applied in their RGB-D Direct Edge Alignment (D-EA). They observed and optimized the camera motion based on gradient estimation. Fabian Schenk and Friedrich Fraundorfer[29] proposed REVO, a real-time robust RGB-D VO method based on the edge. They use the distance transformation on the edge to reduce the Euclidean geometric error, and use the edges of Se and hed to move relative to the camera. Yang and Scherer[30] proposed direct odometry based on points and lines, where for a textured environment, the estimated camera pose is comparable to ORB SLAM[25]. Yi and Laurent's Canny-VO's[31] RGB-D visual odometer proposed two alternative methods of distance transformation commonly used in edge registration: Approximate Nearest Neighbour Fields and Oriented Nearest Neighbour Fields, enhanced the efficiency and accuracy of 3D-2D edge alignment. Mingsu[32] proposed a monocular vision method based on edge feature detection and deep recurrent convolutional neural networks, embedding traditional geometric algorithms into virtual reality based on deep learning to enhance the influence of image edge feature information.

C. Corner Detection

The corner point is a critical geometric element in space, which can retain the image characteristics. There will be corner points around the high rate of change of its gradient value. The current mainstream corner detection algorithm is a corner detection algorithm based on grayscale changes. Moravec et al[8]. proposed to calculate the grayscale difference and select a minor grayscale variance as the corner response value for non-maximum value suppression to determine the angle. Point. Harris et al[33]. based on the Moravec algorithm proposed to extract the corner points through the differential operation and autocorrelation function. Shi and Tomasi et al[34]. improved the corner response function of the Harris operator, proposed to extract The Shi-Tomasi[34] operator with a more uniform and reasonable feature point distribution can significantly improve the corner extraction effect. Rosten et al[35]. proposed the classic FAST and FAST-ER image corner detection methods. It assumes the corner points and performs threshold calculation to verify whether it is an actual corner point.

III. METHOD

In this section, we briefly outline the proposed method. The flow chart of our entire method is shown in Figure 1.

A. CNN edge detection

On the VO based on the sparse method[36], a framework needs to learn geometric feature representation to

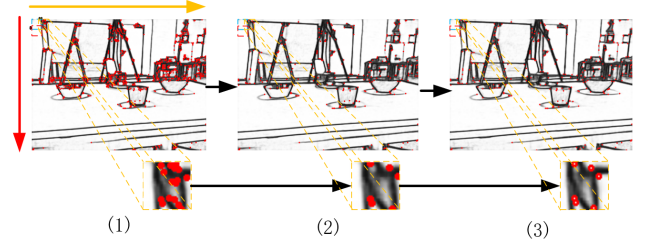


Fig. 3. The process of Shi Tomasi corner optimization.

solve the pose estimation. Compared with the classical edge detection algorithm such as Canny and the machine learning algorithm SE[22], [23], the extracted images have the problems of too much noise so that it cannot obtain a high accuracy pose estimation. Therefore, an edge detection method based on CNN edge detection[20] is utilized in this paper, which is shown in Fig 2. The CNN edge detection has the advantages of extracting high precision features to solve the process of motion.

The CNN edge detection[20] takes the original monocular image as input then predicts an edge-map with the exact resolution. It can be seen as two sub-networks, including the dense extreme inception network and the up-sampling block.

In this work, the CNN edge detection[20] is used by extracting edge features in RGB-D images. We subtract the mean RGB values of the original RGB-D images and study various edge detection methods and depicted in Figure 2. It can be seen that many edges with a significant distance(outliers) are extracted using the method. Thus, we decide to choose the CNN edge detection.

B. Shi-Tomasi Corner Optimization

The corner extraction of image is easily affected by the image environment, so we use the edge image as a method. A large number of corners will be generated where the gray value changes drastically. This will significantly increase the number of post-processing calculations and affect the accuracy of the subsequent LM iterative calculation of pose. Then Shi-Tomasi corner optimization is performed on areas with too dense corner points to improve the quality of corner points and the speed of computation. The specific steps are as follows:

- 1) Perform Shi-Tomasi[34] corner detection on the RGB edge images in the TUM dataset. A fixed-size window $W(x,y)$ is set, and its pixel gray value is $I(x,y)$. The window is moved to the x and y directions by a small displacement u , v , and the pixel corresponding to the new position is gray the degree value is $I(x+u,y+v)$. The change value of the gray value of this movement can be obtained $[I(x+u,y+v) - I(x,y)]$. Let Gaussian kernel The window functions whose function $\omega(x,y)$ is $W(x,y)$ represents the weight of each pixel in the window. The resulting grayscale value change $E(u,v)$ can be

expressed as:

$$\begin{aligned}
E(u, v) &\approx \sum_{(x,y)} \omega(x, y) [I_x u + I_y v]^2 \\
&= \sum_{(x,y)} \omega(x, y) [u, v] \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \quad (1) \\
&= [u, v] M \begin{bmatrix} u \\ v \end{bmatrix}
\end{aligned}$$

$$M = \sum_{(x,y)} \omega(x, y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \quad (2)$$

- 2) It can be seen from the Eq (2). that the magnitude of the gray value change depends on the matrix M . To find the window that causes a large gray value change, the eigenvalues λ_1 and λ_2 of the M matrix can be used to calculate the corner response function R corresponding to each window. Setting a threshold τ_c , and the corner will meet the following conditions:

$$R = \min(\lambda_1, \lambda_2) > \tau_c \quad (3)$$

Set a 20x20 pixel window, slide the image from left to right and from top to bottom, and set the sliding step to 20. Count the number of corner points in the sliding window. If the corner point is greater than the preset threshold, the corner point in the window is cleared, and the preset Shi-Tomasi corner point extraction scheme is set.

- 3) Draw a circle at the optimized corner coordinates and assign pixel points at the center of the circle to form a small concentric circle area. The process is shown in Figure 3.

Through our practice, Through our practice, when we use LM algorithm to iteratively optimize the front and rear frames and collect the Shi Tomasi corner improvement scheme proposed earlier, the accuracy of RPE and ate has been significantly improved.

C. Keyframe Selection

Our proposed VO method introduced keyframes as reference frames for some new frames, thereby reducing accumulated errors. The selection of keyframes generally depends on the type of VO algorithm. Based on feature points[25] usually restrict that a significant number of frames pass, on the order of tens of frames. The method we adopt is to update the keyframe by combining the dual mechanism of periodicity[35] and motion amplitude[38]. Among them, the updated keyframe is shown in Figure 4.

D. Pose estimation based Levenberg-Marquardt

To solve the pose estimation, we employ a coarse-to-fine approach Levenberg-Marquardt (LM)[37] minimization to avoid the pose optimization falling into a locally optimal solution. Meanwhile, the selection of the image pyramid plays a vital role in the system performance

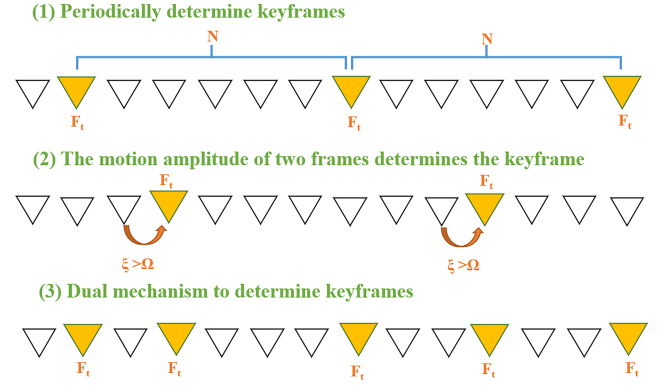


Fig. 4. (1) Our method determines the keyframe through a fixed period. (2) Our method determines the keyframe through the motion amplitude. (3) The final keyframe is determined by a dual mechanism. F_t represents the key frame, ξ represents the camera movement, Ω and represents the preset movement amplitude.

liked [35] and [38]. Thus, we choose a three levels image pyramid because it performs well in most data sets.

In our method, the edges E_t will be detected in each original frame F_t from the intensity I_t as:

$$E_t = E(I_t) \quad (4)$$

We estimate the relative pose P_{KC} from a current frame F_c to keyframe F_t by minimizing the sum over all edge distance errors r :

$$\xi^* = \arg \max_{P_{KC}} \sum r^2 \quad (5)$$

Meanwhile, we use an iteratively re-weighted residual error function with the Huber weights function. The Huber weights are defined as:

$$\delta_H(r) = \begin{cases} 1 & r \leq \theta_H \\ \frac{\theta_H}{r} & r > \theta_H \end{cases} \quad (6)$$

The error function becomes:

$$\xi^* = \arg \max_{P_{KC}} \sum \delta_H(r) r^2 \quad (7)$$

We use an iteratively re-weighted Levenberg-Marquardt Method[37] to optimize Eq.(7). The optimization process from coarse to fine prevents the loss function from falling into local extremes to a great extent and makes the whole VO system more robust.

IV. RESULTS AND DISCUSSION

This section tests our method on the RGB-D TUM benchmark provided by the Technical University of Munich. The RGB-D TUM benchmark is widely used by various visual mileage calculation methods. Each sequence contains RGB images, depth images, accelerometer data. There are several challenging data sets in this benchmark. For example, the duration, trajectory, translation, and rotation speed of each sequence are different. We use eight sequences to benchmark their system performance to achieve a direct comparison with ORB-SLAM2.

A. Results on the TUM RGB-D Benchmark

The methods we tested include CNN edge detection and Shi-Tomasi corner optimization (CNN+Corner), Canny edge detection and Shi-Tomasi corner optimization (Canny+Corner), dense direct method (Origin) to follow ORB-SLAM2 system comparison. In our practice, Shi-Tomasi corner optimization can significantly improve the overall accuracy of the method. When we test the data set in the Origin group, tracking loss will occur, resulting in the RPE and ATE data obtained completely deviating from the ground truth. After we added keyframes, the accuracy of the data results has also been improved. Similarly, the Canny+Corner group is also due to the addition of Shi-Tomasi corner optimization and keyframes, and the accuracy and robustness of the data results have been greatly improved. In CNN+Corner, the main method we recommend is that the test performance of data sets is excellent. The best results in multiple data sets, the sequence of estimated trajectories, and the reference trajectories of the results are shown in Figures 5, 6, and 7. On the other hand, the two groups Origin and Canny+Corner can only achieve the best results in a particular data set, and the data results in the test are not stable enough. For the ORB-SLAM2 system, because the system cannot be initialized well in some data sets, we only changed the parameter of more than 500 feature points required during initialization, and the other entire systems did not change. This means that our method just is VO to compare with the complete SLAM system.

B. Evaluation Metrics

To measure the local accuracy of the VO method, Sturm et al.[19]. proposed relative pose error (RPE) and absolute trajectory error (ATE). RPE measures the drift of Δt in a fixed time interval between a set of attitude Q from the ground true trajectory and a set of attitude P from the estimated trajectory, and the time step i is defined as:

$$RPE_i = (Q_i^{-1} Q_{i+\Delta t})^{-1} (P_i^{-1} P_{i+\Delta t}) \quad (8)$$

where Δt is the time distance between poses. The ATE at a time step i is given as:

$$ATE_i = Q_i^{-1} S P_i \quad (9)$$

where Q and P are aligned by a rigid body transformation S . As suggested by Sturm et al[39]. we evaluate the root mean squared error (RMSE) of the translational component of the RPE and ATE.

C. Discussion

From our experiments, Table I can know that the RPE results of the CNN+Corner method on all data set almost entirely exceed the Canny, original image, ORB-SLAM2 system. We attribute this to the use of relatively accurate and highly robust CNN edge detection, and improve Shi Tomasi's unique corner design. At the same time, it has dual judgment mechanism key frames. From Table I, we

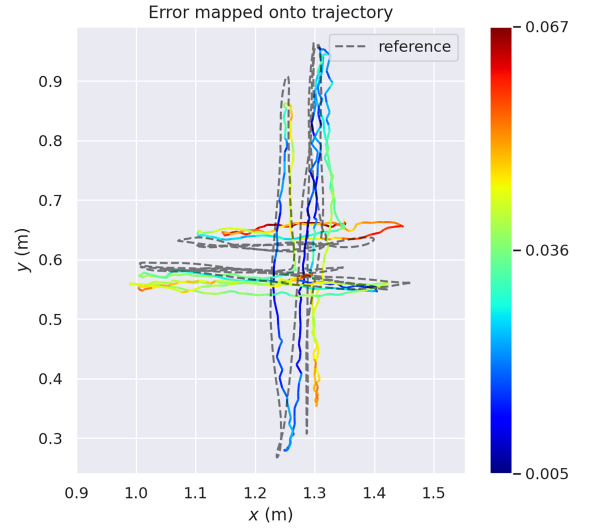


Fig. 5. The result of the sequence fr1xyz estimated trajectory and reference trajectory

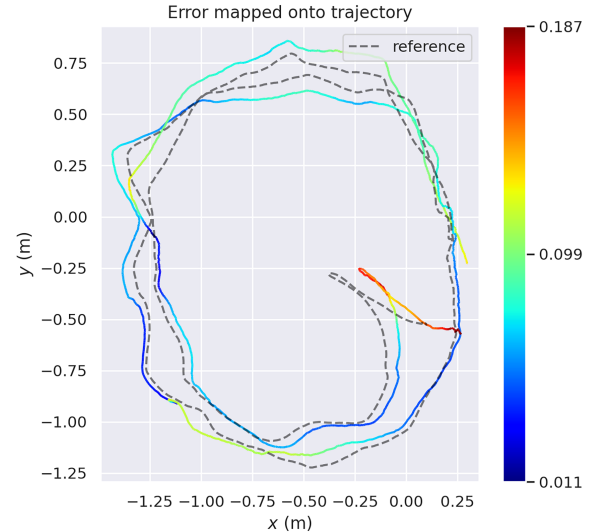


Fig. 6. The result of the sequence fr2coke estimated trajectory and reference trajectory.

can see that the RPE data results achieve the best accuracy in multiple sequences of the data set. Even if the accuracy does not reach the best accuracy of the data results, it is only slightly inferior to the highest accuracy. This shows that our method CNN+Corner is highly accurate and robust and can work well in a variety of scenarios. From Table II, we can see that the ORB-SLAM2 entire system performs very well in ATE, achieving the best accuracy in multiple data set sequences. Even though we do not use any Bundle Adjustment or global optimization as employed by ORB-SLAM2, we once again performed well in all non-SLAM methods, still showing excellent robustness and competitiveness.

TABLE I
RELATIVE POSE RMSE (R: DEG/S, T:M/S) OF TUM DATASETS

Seq.	ORB-SLAM2		Origin		Canny+Corner		CNN+Corner	
	RMSE(R)	RMSE(T)	RMSE(R)	RMSE(T)	RMSE(R)	RMSE(T)	RMSE(R)	RMSE(T)
fr1/xyz	0.978583	0.015260	1.568226	0.027317	1.472533	0.024445	0.724896	0.025371
fr2/360/hemhere	1.145817	0.127742	2.872298	0.390793	-	-	1.021171	0.080454
fr2/dishes	2.402205	0.098421	0.700566	0.016443	0.959770	0.023250	0.787372	0.018816
fr2/coke	5.759803	0.104471	1.168573	0.621661	1.783891	0.046894	1.052839	0.025705
fr3/cabinet	-	-	4.200325	0.129285	4.214291	0.110508	3.462629	0.081417
fr3/large/cabinet	0.805515	0.051905	2.483856	0.271248	1.747718	0.169891	0.718709	0.048065
fr3/str/texture/far	0.540715	0.014733	0.454445	0.013473	0.781303	0.019318	0.455929	0.013280
fr3/str/noture/far	0.723704	0.027528	2.782048	0.216976	1.227557	0.045992	0.607414	0.021371

TABLE II
ABSOLUTE TRAJECTORY RMSE(M) OF TUM DATASETS

Seq.	ORB-SLAM2	Origin	Canny+Corner	CNN+Corner
	RMSE(ATE)	RMSE(ATE)	RMSE(ATE)	RMSE(ATE)
fr1/xyz	0.009436	0.045409	0.042076	0.038506
fr2/360/hemhere	0.229800	0.920911	-	0.328613
fr2/dishes	0.122247	0.094776	0.091492	0.083078
fr2/coke	0.419077	8.952361	0.148833	0.087646
fr3/cabinet	-	0.389984	0.447204	0.443365
fr3/large/cabinet	0.090161	0.346849	0.267330	0.117691
fr3/str/texture/far	0.016497	0.027270	0.041634	0.034999
fr3/str/noture/far	0.027106	0.300592	0.115235	0.036792

V. CONCLUSION

We propose robust edge-direct visual odometry that combines CNN edge detection and Shi-Tomasi corner optimization. The LM algorithm minimizes the photometric error between the two images frames, thereby determining the relative pose of the two frames of images. Through our experiment, the method proposed in this paper is combined with direct method of Canny edge detection and ORB-SLAM2 system. Among the eight sequences, the rum rgb-d dataset has the best RPE accuracy, and the other sequences are only slightly lower. Regarding the accuracy of ATE, our method also has an excellent performance in non-slam methods. Although compared with the entire system of SLAM, our method can still achieve the best in two sequences without back-end optimization. Our method can run well in multiple scenarios, demonstrating the accuracy and robustness of our method. There is still a gap between the accuracy of our proposed method in ATE and the entire ORB-SLAM2 system. We will add back-end optimization to our method in the future to perfect the whole SLAM system.

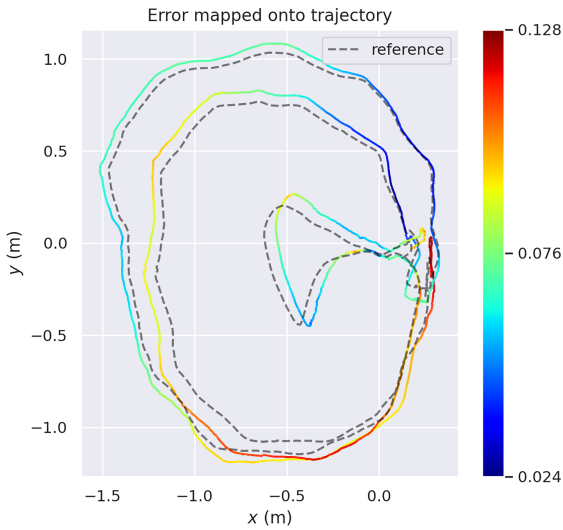


Fig. 7. The result of the sequence fr2dishes estimated trajectory and reference trajectory.

REFERENCES

- [1] Nistér, D., Naroditsky, O., Bergen, J.: Visual odometry. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2004), vol. 1, pp. 1–652 (2004)
- [2] Durrant-Whyte, H., Bailey, T.: Simultaneous localization and mapping: part I. *IEEE Robot. Autom. Mag.* 13(2), 99–110 (2006)
- [3] Bhutta, M.U.M., Kuse, M., Fan, R., Liu, Y. and Liu, M., 2020. Loop-box: Multiagent direct slam triggered by single loop closure for large-scale mapping. *IEEE transactions on cybernetics*.
- [4] Durrant-Whyte, H., Rye, D., Nebot, E.: Localization of autonomous guided vehicles. In: Proceedings of the IEEE International Conference on Robotics and Automation, vol. 7, pp.613–625 (1996)
- [5] Fan, R., Wang, L., Bocus, M.J. and Pitas, I., 2020. Computer stereo vision for autonomous driving. *arXiv preprint arXiv:2012.03194*.
- [6] Huang, H., Fan, R., Zhu, Y., Liu, M. and Pitas, I., 2019, December. A robust pavement mapping system based on normal-constrained stereo visual odometry. In 2019 IEEE International Conference on Imaging Systems and Techniques (IST) (pp. 1-6). IEEE.
- [7] Moravec H P . Obstacle avoidance and navigation in the real world by a seeing robot rover [D]. Palo Alto: Stanford University, 1980
- [8] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardes. Orb-slam: A versatile and accurate monocular slam system. *IEEE Transactions on Robotics*, 31(5):1147–1163, Oct 2015. 2, 5
- [9] R. Mur-Artal and J. D. Tardes. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE Transactions on Robotics*, 33(5):1255–1262, Oct 2017. 2, 6.
- [10] Luo K, Lin M, Wang P, et al. Improved ORB-SLAM2 Algorithm Based on Information Entropy and Image Sharpening Adjustment[J]. *Mathematical Problems in Engineering*, 2020.
- [11] Yin D, Zhou S, Wang P, et al. A Feature Points Extraction Algorithm Based on Adaptive Information Entropy[J]. *IEEE Access*, 2020, 8: 127134-127141.
- [12] C. Kerl, J. Sturm, and D. Cremers. Dense visual slam for rgb-d cameras. In 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 2100–2106, Nov2013. 2, 5, 6, 7
- [13] J. Engel, J. Sturm, and D. Cremers, “Semi-dense visual odometry for a monocular camera,” in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 1449–1456.
- [14] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, “Contour detection and hierarchical image segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 33, no. 5, pp. 898–916, 2011.
- [15] Ozgunalp, U., Fan, R., Ai, X. and Dahnoun, N., 2016. Multiple lane detection algorithm based on novel dense vanishing point estimation. *IEEE Transactions on Intelligent Transportation Systems*, 18(3), pp.621-632.
- [16] Fan, R., Wang, H., Xue, B., Huang, H., Wang, Y., Liu, M. and Pitas, I., 2021. Three-filters-to-normal: An accurate and ultrafast surface normal estimator. *IEEE Robotics and Automation Letters*, 6(3), pp.5405-5412.
- [17] Fan, R. and Dahnoun, N., 2018. Real-time stereo vision-based lane detection system. *Measurement Science and Technology*, 29(7), p.074005.
- [18] P. Dollár and C. L. Zitnick, “Fast edge detection using structured forests,” *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 37, no. 8, pp. 1558–1570, 2015.
- [19] S.” Xie and Z. Tu, “Holistically-nested edge detection,” in *Proceedings of the International Conference on Computer Vision (ICCV)*, 2015.
- [20] Poma X S, Riba E, Sappa A. Dense extreme inception network: Towards a robust cnn model for edge detection[C]//*Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2020: 1923-1932.
- [21] Georg Klein and David Murray. Parallel tracking and mapping for small ar workspaces. In *Mixed and Augmented Reality*, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on, pages 225–234. IEEE, 2007.
- [22] F. Endres, J. Hess, J. Sturm, D. Cremers, and W. Burgard, “3-d mapping with an rgb-d camera,” *IEEE Transactions on Robotics*, vol. 30, no. 1, pp. 177–187, 2014.
- [23] F. Steinbrücker, J. Sturm, and D. Cremers, “Real-time visual odometry from dense rgb-d images,” in *Proceedings of the International Conference on Computer Vision Workshops (ICCVW)*, 2011, pp. 719–722.
- [24] J. Engel, T. Schöps, and D. Cremers. LSD-SLAM: Large-scale direct monocular SLAM. In *European Conference on Computer Vision (ECCV)*, September 2014. 2, 4
- [25] M. P . Kuse and S. Shen, “Robust camera motion estimation using direct edge alignment and sub-gradient method,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2016, pp. 573–579.
- [26] Tarrio and S. Pedre. Realtime Edge-Based Visual Odometry for a Monocular Camera. *IEEE International Conference on Computer Vision (ICCV)*, pages 702–710, 2015
- [27] H. Strasdat, J. Montiel, and A. J. Davison. Real-time monocular slam: Why filter? In *Robotics and Automation (ICRA)*, 2010 IEEE International Conference on, pages 2657–2664. IEEE, 2010. 1.
- [28] Kuse M, Shen S. Robust camera motion estimation using direct edge alignment and sub-gradient method[C]//2016 IEEE international conference on robotics and automation (ICRA). IEEE, 2016: 573-579.
- [29] Schenk F, Fraundorfer F. Robust edge-based visual odometry using machine-learned edges[C]//2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2017: 1297-1304.
- [30] Yang S, Scherer S. Direct monocular odometry using points and lines[C]//2017 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2017: 3871-3877.
- [31] Zhou Y, Li H, Kneip L. Canny-vo: Visual odometry with rgb-d cameras based on geometric 3-d-2-d edge alignment[J]. *IEEE Transactions on Robotics*, 2018, 35(1): 184-199.
- [32] Yan M, Shi C, Wang Y. A monocular visual odometry combining edge enhance with deep learning[C]//2019 IEEE International Conference on Robotics and Biomimetics (ROBIO). IEEE, 2019: 370-374.
- [33] Harris C, Stephens M. A combined corner and edge detector[C]//*Alvey vision conference*. 1988, 15(50): 10-5244.
- [34] Shi J. Good features to track[C]//1994 Proceedings of IEEE conference on computer vision and pattern recognition. IEEE, 1994: 593-600.
- [35] Smith S M, Brady J M. SUSAN—a new approach to low level image processing[J]. *International journal of computer vision*, 1997, 23(1): 45-78.
- [36] Engel J, Koltun V, Cremers D. Direct sparse odometry[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2017, 40(3): 611-625.
- [37] Moré J J. The Levenberg-Marquardt algorithm: implementation and theory[M]//*Numerical analysis*. Springer, Berlin, Heidelberg, 1978: 105-116.
- [38] Christensen K, Hebert M. Edge-direct visual odometry[J]. *arXiv preprint arXiv:1906.04838*, 2019.
- [39] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, “A benchmark for the evaluation of rgb-d slam systems,” in *Proceedings of the IEEE/RSJ Conference on Intelligent Robots and Systems (IROS)*, 2012, pp. 573–580.