Proceeding of the IEEE
International Conference on Robotics and Biomimetics
Dali, China, December 2019

An Modified Real-time Photometric Calibration Algorithm Based on ORB Features*

Chuanzheng Liu, Jiahui Qi, Chengjin Zhang, Peixin Liu and Xianfeng Yuan
Department of Mechanical Electrical and Information Engineering
University of Shandong
Weihai, Shandong Province, China
Corresponding author: Chengjin Zhang, Email: cjzhang@sdu.edu.cn

***Abstract* - In order to improve the accuracy of direct visual odometry and SLAM algorithm under illumination changes, we propose an modified real-time photometric calibration algorithm based on ORB features. Firstly, we use more robust ORB features as input to a optimization framework to obtain scene points correspondences. Then, we establish the reprojection pixels intensity residual equation based on photometric image formation and feature matching results of the camera. Finally, the incoming frames are corrected by the camera response function, exposure time and vignetting estimation. The proposed approach is tested with the EuRoC dataset and TUM dataset respectively, and experimental results indicate that the exposure compensation and time cost performances of the proposed method are promising.**

Index terms - Visual SLAM; Real-time photometric calibration; ORB features; Reprojection pixels intensity error.

I. INTRODUCTION

Recently, a number of visual SLAM (Simultaneous Localization and Mapping, SLAM) systems have been proposed [1], since it is a fundamental building block for many emerging technologies, e.g., autonomous cars, virtual reality, augmented reality and etc. While for a long time the field was dominated by feature-based (indirect) methods, in recent years a number of different approaches, namely direct formulations [2], have gained more and more attention. Compared with the indirect approaches, the direct and semi-direct methods, such as LSD-SLAM [3], SVO [4][5], DSO [2], Stereo DSO [6] and LDSO [7], optimize the photometric error equation instead of the geometric error to obtain a better mapping performance. However, direct methods rely on more stable visual sensor inputs, that scene pixels' intensities are with constant brightness. Therefore, it is necessary to use a pre-processing step that can provide calibrations for input sequences, to make VO (Visual Odometry, VO)/SLAM more robust to illumination changes.

D. B. Goldman et al.[8] provide a model containing the response function, vignetting and exposure time, and update the nonlinear response function using Gauss-Newton algorithm iteratively. The model can output the photometric parameters after many rounds of optimization which cost long time. S. K. Kim et al. [9] propose a joint system with feature tracking to compute the spatial relationship between images and compute radiometric relationships with the radiometric response function and exposure estimation. The system unifies radiometric calibration, feature tracking and exposure estimates into a common framework, and solves the problem based on the assumption of KLT tracker that the brightness of the scene points cannot stay constantly in high dynamic scene. M. Grundmant et al. [10] present a radiation self-calibration model. Compared with the single camera response model, this approach considers the camera response function as time variant, which changes with scene content and exposure. J. Engel et al. [1] present a monocular VO/SLAM evaluation benchmark to estimate for vignetting by taking images of a uniformly color surface and estimate for response function by images taken of a static scene with different exposure time. J. Engel et al. [2] introduce the photometric calibration files to compensate pixel intensity and propose an additional affine brightness transfer function parameters that can be updated in front-end, to get better tracking performance. However, the photometric calibration files only provide a constant calibration. P. Bergmann et al. [11] propose a photometric error equation based on KLT tracker to update vignetting parameters and response function. However, the performance of KLT tracker can be affected by illumination changes [10], which means that the accuracy of the viginetting and response function may be degraded.

Inspired by [11], we propose an improved photometric calibration algorithm which retains the strategy of establishing a least-squares error equation to update the viginetting and response function. Exposure time is estimated by two continuous frames. Flow chart of photometric calibration is described in Fig. 1.

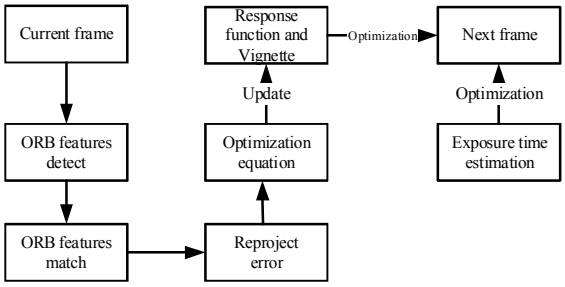


Fig. 1 Flow chart of photometric calibration.

Our algorithm extracts ORB features from the current frame and matched them directly between frames. Then the minimum

*This work is supported by The National Natural Science Foundation of China (Grant No. 61803227, 61973184, 61773242), Independent Innovation Foundation of Shandong University (Grant No. 2018ZQXM005).

978-1-7281-6321-5/19//$31.00 ©2019 IEEE

residual optimization equation is established by calculating the pixel intensity residual between the reprojected feature points and the matched points. Since the estimation of vignetting and response function requires multiple frames, and the real-time calibration cannot capture a large number of frames to optimize the parameters, we will retain the current vignetting and response function. Finally, we use Levenberg-Marquardt (L-M) algorithm to solve the residual equation iteratively and update the photometric parameters. Our method decouples the exposure time from other photometric parameters. Besides, we use our algorithm in parallel to a direct visual SLAM method and test it with EuRoC dataset [12] and TUM dataset [13]. The experiment results proves that our method can improve its performance.

II. PHOTOMETRIC IMAGE FORMATION

A scene point is illuminated by a light source and reflects the light back into space [11]. The total light intensity is called radiance of the scene point that independent of the observers' viewing angle.

When the vision sensor captures the scene as image, the radiance B of scene points is converted into irradiance by lens. For the formation of image each time, the total energy received by the sensor depends on the irradiance through the camera shutter during the exposure time t. Finally, the energy converts into pixels intensity by response function G. The flow chart of photometric image formation process is shown in Fig. 2.

The effective incident off-axis light of the front lens is changed by the size of the aperture and exposure time, causing the pixels intensity of the images gradually weaken from the center to the surround, which is called vignetting. In the automatic exposure mode, different scenes have different duration. The process that a pixel processes a received photon and converts it into a brightness value nonlinearly called response function.

The imaging model of the photometric image formation is defined as follows [1]:

$$I_i = \mathrm{G}_i(t_i V_i B_i) \tag{1}$$

I_i is the pixel intensity observed on frame i, B_i is the irradiance of the sensor, t_i is the exposure time, V is the lens attenuation (vignetting), and G is the response function.

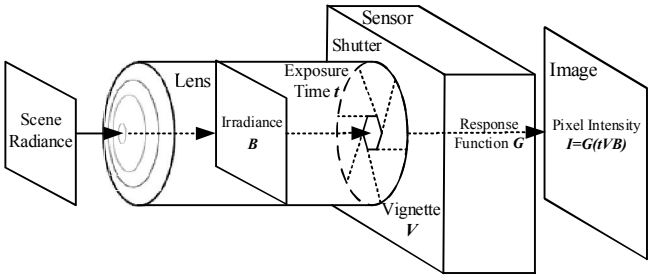


Fig. 2 Flow chart of photometric image formation process

Assuming that the pixel intensity attenuation factors are symmetric around the centre of the image, vignetting can be described as (2):

$$V = 1 + \sum_{n=1}^{3} v_n R^{2n} \tag{2}$$

Vignetting $V: \Omega \to [0,1]$. Where $R(x)$ is the normalized radius of pixel x respect to the centre of the image.

Response function $G: \mathbb{R} \to [0,255]$. When the frames are underexposed and overexposed, their brightness values are 0 and 255 respectively. Linearization is applied to G:

$$G = g_0 + \sum_{k=1}^{4} c_k h_k \tag{3}$$

Where g_0 is the main response function obtained by PCA (Principal Component Analysis, PCA).

III. PHOTOMETRIC EQUATION BASED ON INDIRECT METHOD

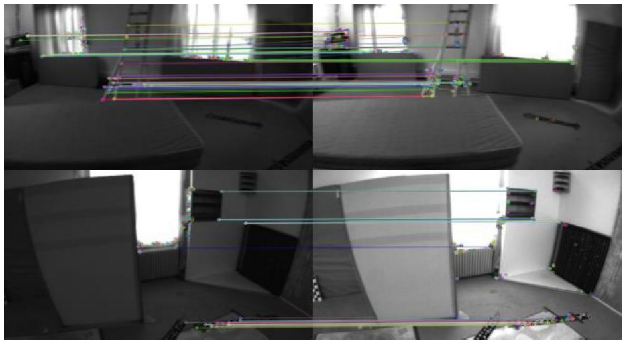
According to the photometric image formation of the camera, the optimization equation is established by the corner points tracker to update viginetting and response function in time. Then the vignetting and response functions are used to compensate the exposure of input sequences.

Generally, feature-based (indirect formulation) matching strategies are considered to be more robust to illumination changes and dynamic objects [14]. Firstly, our algorithm uses the ORB feature [15] to match points between the dominant frame and the current frame. Secondly, we calculate the pixels error equation of re-projected pixels intensity between dominant frame and current frame based on photometric image formation. Finally, the re-projection error equation is optimized iteratively to update photometric parameters.

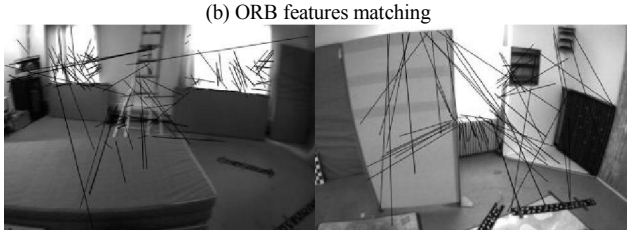
A. Indirect features matching

The ORB features matching calculates the distance between the surrounding BRIEF descriptors of FAST corner points, that extracted in the image pyramid, to obtain a good rotation and scale invariance. The matching results of ORB features and KLT tracker [16] under different exposure conditions are shown in Fig. 3.

(a) Original images of EuRoC



(b) ORB features matching



(c) KLT tracking

Fig. 3 ORB features matching and KLT tracking on EuRoC

Fig. 3(a) includes the original continuous images randomly selected from EuRoC sequence Vicon room 1_03_difficult with changing illumination. The features, tracked by ORB matching and KLT tracker, are shown in Fig. 3(b) and Fig. 3(c) respectively. We can see that the black trajectories of KLT tracker are obviously chaotic while matching of ORB features can be observed clearly.

Since the KLT tracker is based on the gray-scale invariant assumption, the tracking performance of the LK optical flow is degraded more than ORB matching under the illumination changes.

B. Pixel intensity error equation based on photometric image formation

The proposed algorithm establishes the minimum residual equation of the scene irradiation by calculating the pixels intensity residual between the matched points and the reprojected points of the features.

The co-visible point in space, called P, can be observed in host frame I_j and current frame I_i. P is projected by the point p in I_j, and reproject onto I_i as p'. Our algorithm matches the points between I_j and I_i by calculating the Hamming distance of ORB feature, and then minimums the residual between the reprojected pixels intensity and the matched pixels intensity as shown in the Fig. 4.

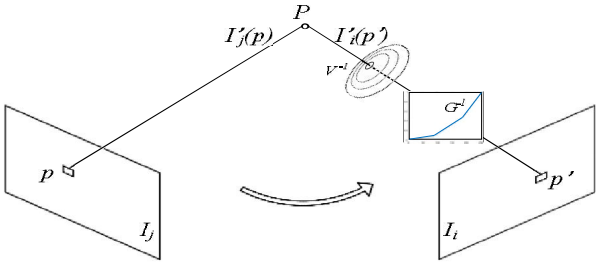


Fig. 4 Reprojection pixel intensity residual

The pixel intensity residuals r' between p and p' on the I_i is defined as follows.

$$r' = \sum_{p\in\mathcal{P}} \left\| I_j'^{p} - I_i'^{p'} \right\|^2 \tag{4}$$

Where $I_i^{p'}$ on I_i is the back projection intensity of point p of I_j. According to the image formation model in (1), it is assumed that the scene irradiation energy B_i^p of the point P is uniform between frames I_i and I_j, the irradiation energy of the frame I_i can be defined as (5).

$$t_i B_i^p = \frac{G^{-1}(I_i^p)}{V_i^p} \tag{5}$$

For a set of tracked points $\mathcal{P}$ and their projections, the total radiance energy function residuals defined in successive inter-frame scenes can be described as (6) and (7).

$$E_{Radi} = \sum_{j\in M}\sum_{i\in\mathcal{F}_p}\sum_{p\in\mathcal{P}} w_i^p \left(t_i B_i^p - \frac{G^{-1}(I_i^p)}{V_i^p}\right)^2 \tag{6}$$

$$r = \left(\frac{G^{-1}(I_i^p)}{V_i^p} - t_i B_i^p\right)^2 \tag{7}$$

In equation (6), M is the set of host frames, and $\mathcal{F}_p$ is the set of frames in which all of the points $\mathcal{P}$ are observed. The photometric parameters are updated by minimize (6), which will be described in section IV.

Since the estimation of vignetting and response functions requires multiple frames, while real-time calibration does not capture a large number of images. Therefore, the state of the current vignetting and response function are maintained to optimize the current frame until new frame arrives.

IV. Optimization

For the reprojection pixels intensity error, the L-M algorithm is applied to solve the least-squares problem. Let $\boldsymbol{x}$ be the total variable to be optimized:

$$\boldsymbol{x} = [t, c, v]^T \tag{8}$$

Where $\boldsymbol{c} = (c_1, c_2, c_3, c_4)$ is the vector of response function G, and $\boldsymbol{v} = (v_1, v_2, v_3)$ is the vector of vignetting V. Considering that the exposure time t can be estimated by two continuous frames, we suggest to decouple the exposure time t estimation from other parameters in the optimization equation [11]. According to (6) and (7), the total radiance energy function is introduced as (9):

$$minE_{Radi} = \frac{1}{2} arg\min_x \sum_{j\in M}\sum_{i\in\mathcal{F}_p}\sum_{p\in\mathcal{P}} w_i^p r \tag{9}$$

Equation (10) is iterated by the L-M method to obtain the Jacobian matrix of the residual r.

$$J_{x_i} = \frac{\partial r_i}{\delta x_i} = \left(\frac{\partial r_i}{\delta c_i}, \frac{\partial r_i}{\delta v_i}\right) \tag{10}$$

$$H\delta x_i = -b \tag{11}$$

$$H = J_{x_i}{}^T W_{x_i} J_{x_i} + \lambda I \tag{12}$$

$$b = J_{x_i}{}^T W_{x_i} r_i \tag{13}$$

(11), (12) and (13) describes the iterative solution process of Eq. (10) based on L-M approach. Where the weight matrix W_{x_i} is inversely proportional to the image gradient of $\boldsymbol{x_i}$.

(a) The original images with exposure changes of EuRoC V1_03_difficult sequence

(b) The images processed by our photometric calibration algorithm

Fig. 5 The experimental results of EuRoC V1_03_difficult sequence

By solving $\delta \boldsymbol{x} = (\delta c, \delta v)$, the vignetting V and the response function G are updated to realize real-time photometric calibration of the input sequence. When a new frame arrives, its response function and vignetting based on current estimates are removed.

The incoming frames are corrected based on the current response function, vignetting estimation and the exposure time estimated by continuous frames. System maintains the vignetting and response function state, which can be calculated from the previous M frames, to compensate the current frame, and simultaneously updates the state of the vignetting and response function.

V. EXPERIMENTS

Our algorithm works on a notebook with Inter i7-8750H CPU, 16G memory, and Ubuntu16.04. GPU acceleration is not adopted during the experiments.

The proposed algorithm is tested with the EuRoC and TUM datasets. The EuRoC dataset [12] , made by the Federal Institute of Technology in Zurich, using stereo camera and IMU as sensors to collect images and poses information. TUM dataset [13] contains 50 real-world sequences comprising over 100 minutes of video, recorded across different environments ranging from narrow indoor corridors to wide outdoor scenes of Technische Universität München.

Firstly, we test our algorithm with EuRoC dataset and partial experimental results are shown in Fig. 5.

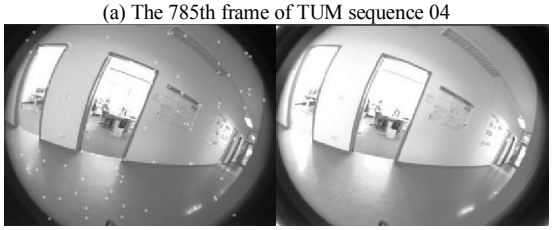
The subfigures (a) of Fig. 5 are the original images with exposure changes of EuRoC V1_03_difficult sequence. While the subfigures (b) of Fig. 5 are the images processed by our photometric calibration algorithm. It can be seen from the second to the forth images in Fig. 5 (b) that our algorithm can compensate brightness value of EuRoC dataset V1_03 _difficult sequence.

In order to further test the performance of the proposed algorithm, we run our algorithm on the TUM dataset sequence 04. The experimental results are illustrated in Fig. 6.

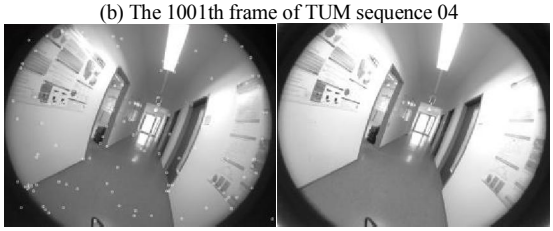
The tracked frames and the calibrated frames of TUM 04 sequence are shown respectively in Fig. 6. We randomly select the vignetting and response function of three frames to compare with the photometric calibration files proposed in [1].

Subfigures (a) and (b) of Fig. 7 respectively show the photometric calibration files proposed in [1]. This response function model maps all possible sensor irradiance to pixel intensity, such as $G(0) = 0$ and $G(1) = 255$. The vignetting model simulates the lens effect, that more light entering into the center and less into the edge. It can be seen in Fig. 7(b) that the vignetting factors are small around the center and large around the edge. These models are applied to restore the scene radiance.

(a) The 785th frame of TUM sequence 04

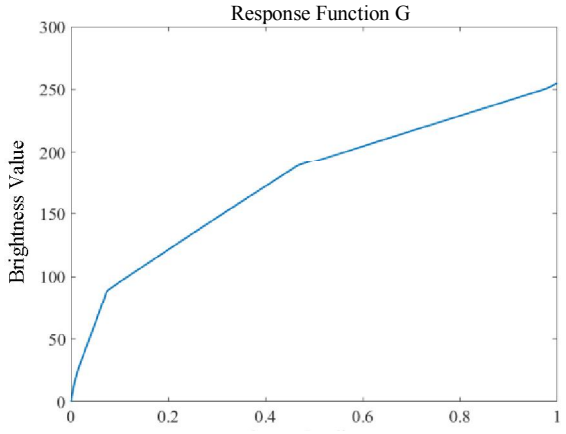


(b) The 1001th frame of TUM sequence 04

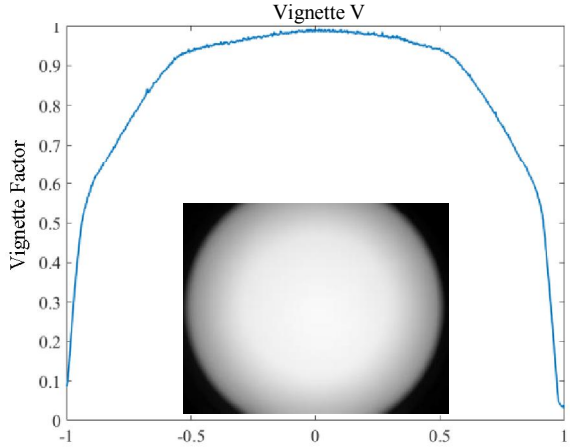


(c) The 1340th frame of TUM sequence 04

Fig. 6 Experimental results of TUM sequence 04

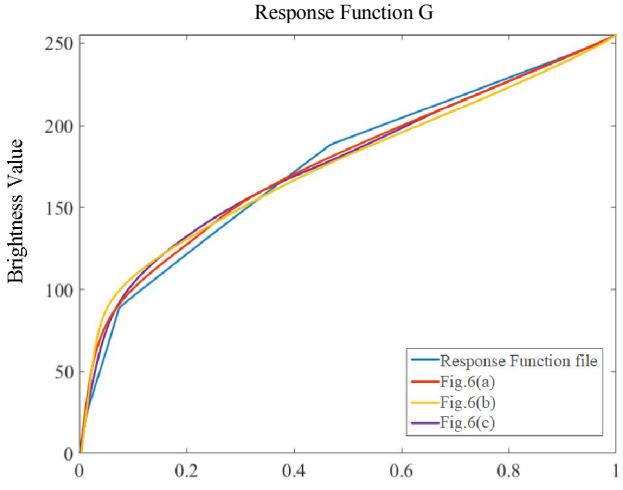


(a) Response function file proposed in algorithm [1]

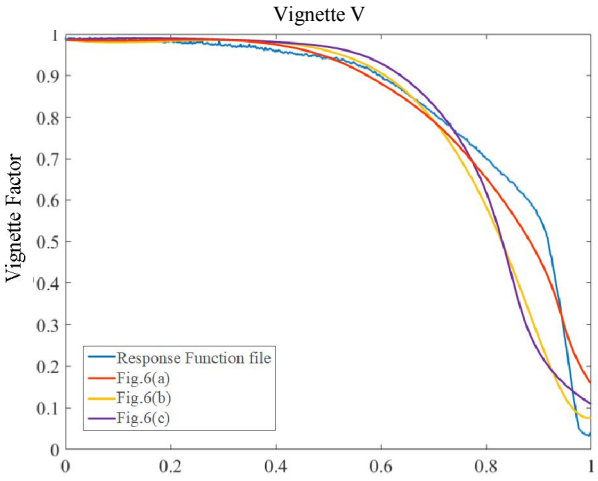


(b) Vignetting file proposed in algorithm [1]

Fig. 7 Photometric calibration files proposed in algorithm [1]



(a) Response Function G



(b) Vignette V

Fig. 8 Comparation between our algorithm and trained files in [1]

Adaptive vignetting and response function are illustrated respectively in Fig. 8. It can be seen from the figures that the vignettng and response function calculated by our algorithm are dynamically adjusted in a reasonable range. In Fig. 6 (b), the global pixels are dark, especially at the edge of fisheye camera. Therefore, the vignetting factors near the edge of image are significantly adjusted and the brightness response values at weak irradiance are increased. In this way, the contrast of the image is enhanced to be more recognizable.

Compare to the intrinsic compensation proposed in algorithm [1] , the adaptive vignetting and response function make the brightness value of input images continuous as Fig. 5. But the intrinsic compensation can only restore pixels values to scene radiance, which has poor robustness in challenge illumination.

TABLE I

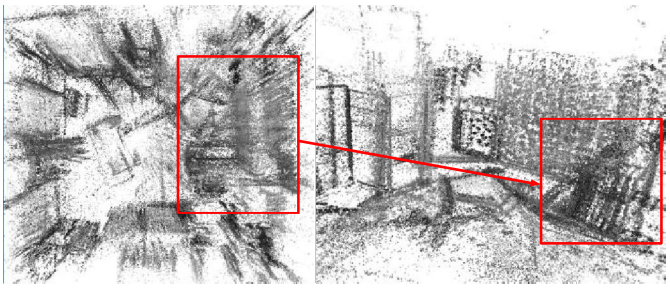
TIMING RESULTS FOR THE IMPLEMENTED ALGORITHM

Module	Execution time
Extract and match ORB features	0.32 s
Photometric parameters optimization	1.75 s
Update exposure situation	1.02 s

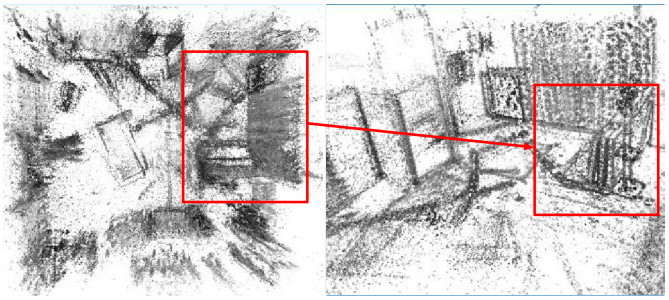
Table I shows the average processing time of the proposed algorithm for a single frame image in the TUM sequence 04. Each module of the pipeline has been timed independently.

Finally, we run LDSO [7] on the EuRoC dataset sequence V1_03_difficult and TUM sequence 04 with our algorithm. The results are shown in Fig. 8 and Fig. 9.

(a) The scene before reconstruction



(b) LDSO without our algorithm

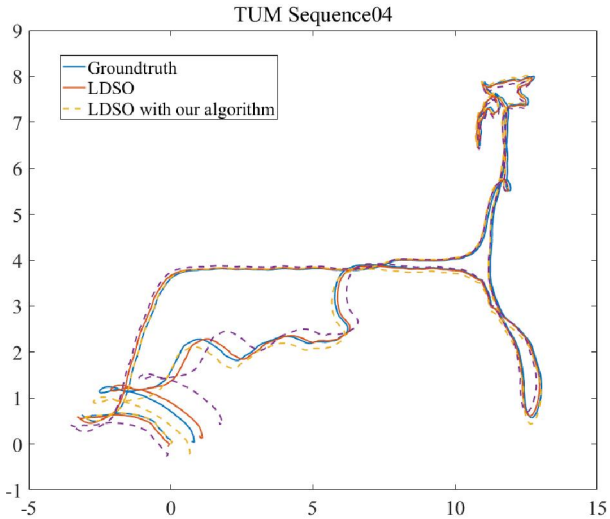


(c) LDSO with our algorithm

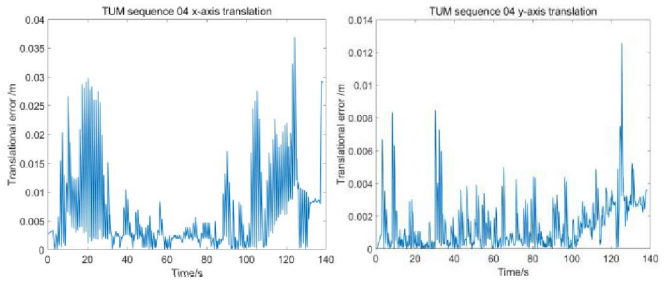
Fig. 9 Experimental results of LDSO on EuRoC sequence V1_03 _difficult

Fig. 9 shows the qualitative results without and with our algorithm, from which we can see the significant improvement of point cloud map reconstructed by LDSO. Compared with the repeated point cloud in Fig. 9 (b), Fig. 9 (c) has better mapping effect and less noise.

The trajectory and ATE error of TUM sequence 04 are shown in Fig. 10. It can be seen that the trajectory with our algorithm is closer to groundtruth. The error along 6-DoF rigid body motion is presented in Fig. 10 (b) to Fig. 10 (g), from which we can conclude that the proposed algorithm can enhance the performance of direct VSLAM because the error of each DoF rigid body motion changes in a reasonable range.

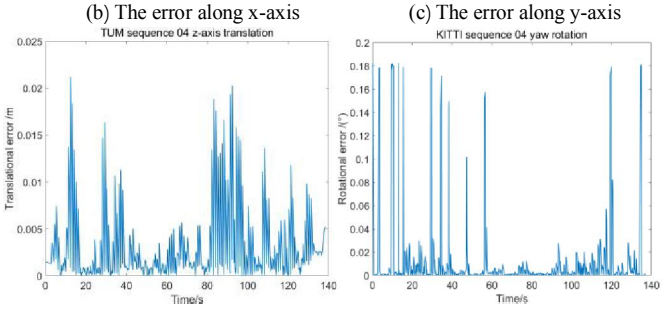


(a) The trajectory of TUM sequence 04



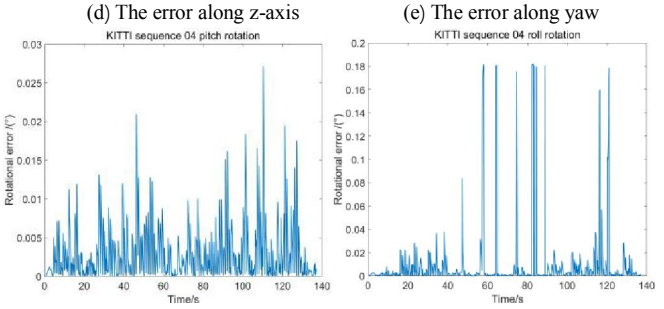
(b) The error along x-axis

(c) The error along y-axis



(d) The error along z-axis

(e) The error along yaw



(f) The error along pitch

(g) The error along roll

Fig. 10 Experimental results of LDSO with our algorithm on TUM sequence04

VI. Conclusion

Based on the indirect feature matching, we propose an real-time photometric calibration strategy for VO/SLAM, specifically the direct method for visual odometer. Firstly, the ORB features are matched directly between frames. Secondly, the reprojection pixels intensity error equation based on photometric image formation and feature matching results between input frames is established. Finally, the photometric parameters are updated by Levenberg-Marquardt (L-M) algorithm. The proposed approach is tested on the EuRoC dataset and TUM dataset respectively. The results prove that our algorithm enhances the precision of the direct method for visual odometer.

In order to improve the performance of real-time photometric calibration which can enhance VO/Visual SLAM, we will use other tracking methods to establish photometric optimization equations in the future.

REFERENCES

[1] J. Engel, V. Usenko, and D. Cremers, "A photometrically calibrated bench-mark for monocular visual odometry," arXiv, October 2016[online].

[2] J. Engel, V. Usenko, and D. Cremers, "Direct sparse odometry," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 40, no. 3, pp. 611-625, March 2018.

[3] J. Engel, T. Schöps, and D. Cremers, "LSD-SLAM: Large-scale direct monocular SLAM," *European Conference on Computer Vision*, vol.8690, pp. 834-849, September 2014.

[4] C. Forster, M. Pizzoli, and D. Scaramuzza, "SVO: Fast semi-direct monocular visual odometry," *IEEE International Conference on Robotics & Automation*, pp. 15-22, June 2014.

[5] C. Forster, Z. Zhang, and M. Gassner, "SVO: Semidirect visual odometry for monocular and multicamera Systems," *IEEE Transactions on Robotics*, vol. 33, no.2, pp. 249-265, April 2017.

[6] R. Wang, M.Schworer, and D. Cremers, "Stereo DSO: Large-scale direct sparse visual odometry with stereo cameras," *IEEE International Conference on Computer Vision(ICCV)*, pp. 1-6, October 2017.

[7] X. Gao, R. Wang, N. Demmel, and D. Cremers, "LDSO: Direct sparse odometry with loop closure," *IEEE International Conference on Intelligent Robots and Systems(IROS)*, pp.1-7, October 2018.

[8] D. B. Goldmannan, "Vignette and exposure calibration and compensation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 12, pp. 2276-2288, December 2010.

[9] S. K. Kim, D. Gallup, J. Frahm, and M. Pollefeys, "Joint radiometric calibration and feature tracking for an adaptive stereo system," *Computer Vision and Image Understandin*, vol. 114, pp. 574-582, May 2010.

[10] M. Grundmann, C. McClanahan, S. B. Kang, and I. Essa, "Post-processing approach for radiometric self-calibration of video," *IEEE International Conference on Computational Photography (ICCP)*, pp. 1-9, April 2013.

[11] P. Bergmann, R. Wang, and D. Cremers, "Online photometric calibration of auto exposure video for realtime visual odometry and SLAM," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 627-634, April 2018.

[12] M. Burri, J. Nikolic, and P. Gohl, et al, "The EuRoC micro aerial vehicle datasets," *International Journal of Robotics Research*, vol. 35, no. 10, pp. 1157-1163, January 2016.

[13] J. Sturm, N. Engelhard, and F. Endres, "A benchmark for the evaluation of RGB-D SLAM systems," *IEEE/RSJ International Conference on Intelligent Robots & Systems*, pp. 1-8, October 2012.

[14] X. Gao, T. Zhang, Y. Liu, Q. R. Yan, "Visual SLAM XIV: From Theory to Practice," Beijing: Electronic Industry Press, 2017, pp.132-134.

[15] R. Mur-Artal, J. D. Tardós, "ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255-1262, October 2017.

[16] M. A. Haque, R. Irani, and K. Nasrollahi, et al, "Facial video-based detection of physical fatigue for maximal muscle activity," *IET Computer Vision*, vol. 10, no. 4, pp. 323-329, June 2016.