

Editorial Articles

PC-VINS-Mono: A Robust Mono Visual-Inertial Odometry with Photometric Calibration

Yao Xiao¹, Xiaogang Ruan¹, Xiaoqing Zhu^{1*}

¹ Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

ABSTRACT

Feature detection and Tracking, which heavily rely on the gray value information of images, is a very importance procedure for Visual-Inertial Odometry (VIO) and the tracking results significantly affect the accuracy of the estimation results and the robustness of VIO. In high contrast lighting condition environment, images captured by auto exposure camera shows frequently change with its exposure time. As a result, the gray value of the same feature in the image show vary from frame to frame, which poses large challenge to the feature detection and tracking procedure. Moreover, this problem further been aggravated by the nonlinear camera response function and lens attenuation. However, very few VIO methods take full advantage of photometric camera calibration and discuss the influence of photometric calibration to the VIO. In this paper, we proposed a robust monocular visual-inertial odometry, PC-VINS-Mono, which can be understood as an extension of the opens-source VIO pipeline, VINS-Mono, with the capability of photometric calibration. We evaluate the proposed algorithm with the public dataset. Experimental results show that, with photometric calibration, our algorithm achieves better performance comparing to the VINS-Mono.

Keyword: Photometric Calibration; Visual-Inertial Odometry; Simultaneous Localization and Mapping; Robot Navigation.

ARTICLE INFO

Received: Jan 14, 2019
Accepted: Jan 20, 2019
Available online: Jan 21, 2019

*CORRESPONDING AUTHOR

Xiaoqing Zhu, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China; alex.zhuxq@bjut.edu.cn ;

CITATION

Y Xiao, XG Ruan, XQ Zhu. PC-VINS-Mono: A Robust Mono Visual-Inertial Odometry with Photometric Calibration *Journal of Autonomous Intelligence* 2018; 1(2): 29-35. doi: 10.32629/jai.v1i2.33

COPYRIGHT

Copyright © 2018 by author(s) and Frontier Scientific Publishing. This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC4.0). <https://creativecommons.org/licenses/by-nc/4.0/>

1. Introduction

Monocular Visual-Inertial Odometry (VIO) attracts significant attentions from large number of researchers and is gaining the popularity in various potential applications, such as robotics and augmented reality due to the miniaturization in size and low cost in price. In recent years, a number of VIO frameworks have been proposed, such as MSCKF[1], [2], OKVIS[3], [4], ROVIO[5] and VINS-Mono[6] and demonstrate tremendous improvement in accuracy, robustness and efficiency.

VIO is seen as the extension of VO with fusing IMU and vision measurements to estimate the pose of camera, based on the Bayes' theorem. As a result, many components of VIO share the same ones with VO. Similarly to VO, VIO can be classified into three formulations: indirect method [7], semi-direct method [8]–[10] and direct method [11]–[13], depending how to use the information of image. Both these three formulations use direct or indirect of the image information, the gray value of images, to track the project location of the same scene point across different images to estimate the 6-D camera motion. Therefore, all these methods are sensitive to illumination change, especially the direct and semi-direct method, which assume the brightness of the same scene point projecting into different images appear with constant values, also known as brightness consistency assumption. For images captured by real camera, it is inevitable that pixels corresponding to the same scene point have different intensities across images due to lens attenuation, auto gain and exposure control, which cause the scene point tracking more difficulty. However, no

so many researches have been reported the influence of photometric calibration to VIO

A breakthrough in this filed is the TUM monoVO dataset [14], which provide 50 real-world sequences with photometrically calibrated. A novel, simple approach to non-parametric vignette calibration also proposed in the corresponding paper. Later, in [11] the author proposed Direct Sparse Odometry (DSO), the first fully direct method take full advantage of photometric camera calibration, which showed impressive results. With the combination of IMU, VI-DSO was proposed in [11]. However, the evaluation experiments were run with EuRoc dataset [15], of which, the camera is not photometrically calibrated. As a result, the improvement with photometric calibration is not reported. In [16], the influence of photometric calibration to DSO, ORB-SLAM[17], [18], Semi-Direct Visual Odometry (SVO), three representative, the-state-of-the-art of direct, indirect and semi-direct methods, is evaluated quantitatively with the TUM monoVO dataset. The experiment results drawn a counter-intuitive results that photometric calibration may reduce the overall performance of semi-direct based method, SVO, and for ORB-SLAM, the performance decline even larger. The possible reason is that photometric calibration reduces the contrast of dark areas while increase it for bright areas, which, as a result, decline the performance of feature detection and descriptor extracting algorithm. However, no VIO framework is evaluated in this paper. Moreover, the exposure time t is not integrated into the formulation of ORB-SLAM and SVO. In [19], The TUM VI Benchmark for Evaluating VIO have been proposed and public available. This Benchmark including a diverse set of sequences in different scenes and the camera it used to record is photometric calibration. To the best of our knowledge, still there is not any researches that take full advantage of photometric camera calibration into the VIO framework and evaluate the influence to the performance.

On the other hand, [6] proposes VINS-Mono, a VIO pipeline with loop closure and global map optimization and demonstrates the state-of-the-art results compared to the existing VIO method. VINS-Mono uses the optic-flow method, which is based on the brightness consistency assumption, to detect and track the features. As a result, it is sensitive to illumination change.

In this paper, we propose a robust mono visual-inertial

odometry with photometric calibration, PC-VINS-Mono, which can be understood as an extension of VINS-Mono, with the capability of photometric calibration. As a result, PC-VINS-Mono is more robust to the illumination change and shows improvement performance.

The remainder of this paper is as follows. In Section 2, we give an overview of the complete system pipeline. In Section 3, we detail how to extend the VINS-Mono with photometric calibration. In Section 4, we evaluate the proposed VIO with the public available dataset and provide quantity analysis of the performance compared to VINS-Mono. We conclude the paper in Section 5.

2. System Overview

As our work can be understood as an extension of VINS-Mono [6]. For the sake of completeness, we briefly review every stage of VINS-Mono firstly. Then we focus on the feature detection and tracking procedure and details integrating the capability of photometric calibration into this module.

The structure of VINS-Mono is depicted in Fig. 1. The system starts with measurement preprocessing, in which features are extracted and tracked, and IMU measurements between two consecutive frames are preintegrated. The coming frame with its corresponding future tracking information and the preintegrated results would be added to the sliding window for feature optimization. As the monocular camera is incapable of recovering the metric scale information as well as its attitude with respect to the gravity, an initialization procedure is needed to provide all the necessary initial values, including pose, velocity, gravity vector, gyroscope bias, and three-dimensional (3-D) feature location, for bootstrapping the subsequent nonlinear optimization-based VIO. The tightly-coupled nonlinear optimization would be implemented when every frame is added to the sliding window after the system has been successfully bootstrapping. Keyframe management module determines which frame would be removed from the sliding window and marginalizes it out to bound the computation complexity of the optimization procedure. The map management module also removes the map points corresponding to the removed frame. As we concern on the VIO front-end, the loop closure and global pose graph optimization is not included in Fig. 1.

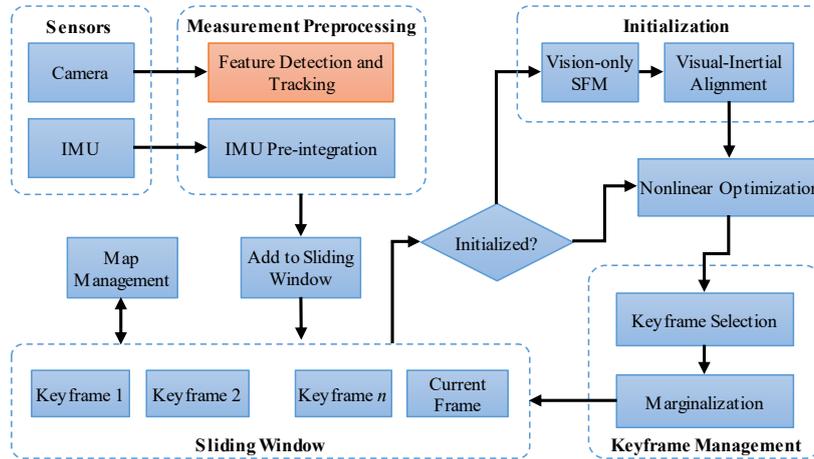


Fig. 1 VINS-Mono Framework. Our work extends the feature detection and tracking procedure with photometric calibration, which is marked as orange block in the figure.

VINS-Mono is a tightly-coupled VIO framework, which minimize the sum of prior and the Mahalanobis norm

$$\min_x \left\{ \|r_p - H_p \mathcal{X}\|^2 + \sum_{k \in B} \left\| r_B \left(\hat{z}_{b_{k+1}}^{b_k}, \mathcal{X} \right) \right\|_{P_{b_{k+1}}^{b_k}}^2 + \sum_{k \in B} \rho \left(\|r_C(\hat{z}_i^c, \mathcal{X})\|^2 \right) \right\} \quad (1)$$

where x is the full system state. ρ is the Huber norm. r_B and r_C are the residuals for IMU and visual measurements, respectively. For more details, please refer to [6]. As the feature detection and tracking results were integrated in the objective function to jointly optimization with the IMU measurements, the accuracy and robust of the results of this procedure would significantly affect the performance of the VIO.

Now we focus on the feature detection and tracking procedure. VINS-Mono detects the Shi-Tomasi corners [20] in the images and track the corners in the next image using Lucas-Kanade method [21]. As Lucas-Kanade method is an optic-flow method based on brightness consistency assumption, it is sensitive to illumination change. To relieve this problem, the author uses Contrast Limited Adaptive Histogram Equalization to improve contrast in images before detecting and tracking features. Nonetheless, VINS-Mono still affect significant by high-contrast lighting

of all measurement residuals to obtain a maximum posteriori estimation as

condition environment with auto exposure cameras, in which, the exposure time is vary from frame to frame. Therefore, we extend the feature detection and tracking procedure of VINS-Mono with photometric calibration, which is marked as orange block in Fig 1.

3. Photometric Calibration

3.1 Image formation model

We used the image formation proposed in [8], which accounts for a non-linear response function $G: \mathbb{R} \rightarrow [0, 255]$, as well as lens attenuation (vignetting) $V: \Omega \rightarrow [0, 1]$. The combined model is given by

$$I(x) = G(tV(x)B(x)) \quad (2)$$

Where, I is the observed pixel value, t is the exposure time, B the irradiance image. An illustration of the camera model is showed as Fig 2.

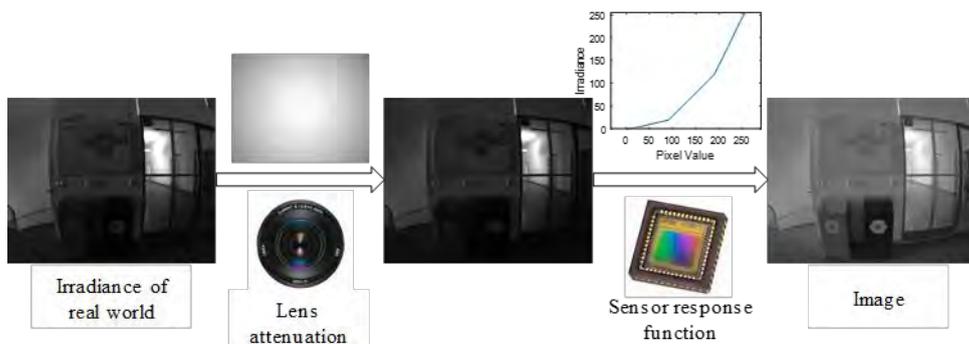


Fig. 2 Illustration of camera model. Image is extracted from TUM monoVO dataset [14].

3.2 Apply photometric calibration to two consequent frames

We try to photometrically calibrate two consequent image frames to satisfy the brightness consistency assumption as much as possible. Suppose two consequent images $k - 1$, k . we calibrate these two images by:

$$I'_{k-1}(x) = \frac{t_k}{t_{k-1}} \frac{G^{-1}(I'_{k-1}(x))}{V(x)} \quad (3)$$

$$I'_k(x) = \frac{G^{-1}(I'_k(x))}{V(x)} \quad (4)$$

Here, we choose to calibrate the previous image with the exposure time as we assume that the new coming image would have the better contrast. Fig. illustrates the comparison of two consequent images with and without photometric calibration.

Eq.(3) - (4) assume that the camera is photometric uncalibrated and the non-linear response function G and lens attenuation V are known. However, it is often the case that only the exposure time of image is known in reality. In this case, we can set the G and V to 1 and calibrate the image with only the exposure time. We will simulate this case in the experiments.

As reported in [16], photometric calibration may

reduce the intensity contrast of the image, which may cause the feature detection algorithm works worse. To avoid this problem, we implement the contrast limited adaptive histogram equalization (CLAHE) algorithm to the current images before detecting the new features.



(a) Two consequent raw images from sensor



(b) Two consequent images after photometric calibration

Fig. 3 Example of photometric calibration with two consequent images. Images are extracted from TUM VI Datasets [19].

With the discussion above, an overview of the new feature detection and tracking procedure with photometric calibration is given in Algorithm 1.

Algorithm 1 Feature detection and Tracking with Photometric Calibration

For each new image with its corresponding exposure time received:

 Apply photometric calibration to the previous image with Eq.(3).

 Apply photometric calibration to the new input image with Eq.(4).

 Track the keypoints of the previous image in the new image using Lucas-Kanade method

 Remove outliers with RANSAC algorithm.

 Improve brightness contrast using Contrast Limited Adaptive Histogram Equalization (CLAHE) method.

 Detect new keypoints with Shi-Tomasi method.

 Save the new raw image and its exposure time as previous image.

End

4. Experiments and Results

We evaluate the proposed VIO algorithm using TUM VI Datasets [19]. The datasets provide a diverse set of sequence in different scenes, with 1024x1024 image resolution camera with at 20 Hz, and known exposure times, linear function and vignette calibration. An IMU measures accelerations and angular velocities on 3 axes at 200 Hz, while the cameras and IMU sensors are time-synchronized

in hardware. Only the 5 corridor and 6 room sequences are used in our experiments as the exposure time of these sequences vary frequently which poses challenge to the optic-flow-based tracking algorithm. All the experiments run with quarter resolution images (512x512) as [19] did, using a laptop computer equipped with an Intel Core(TM) i7-7700 CPU @ 3.60GHz CPU and 8GB RAM. we also run the algorithm with photometric calibration with only the exposure time to simulate the case of camera with unknown

response function and lens attenuation factor. To verify the necessary of CLAHE for feature detection, we run the algorithm with the same config with skipping the CLAHE step.

The root-mean-square error (RMSE) of all the estimation results in shown in Table 1, which is evaluated by an absolute trajectory error (ATE) [22].

Table 1 RMSE ATE[22] of the estimation results in meters. t only means calibrate the image with t only. No CLAHE means skipping the CLAHE step before feature detection.

Sequence	VINS-Mono	Ours	Ours t only	Ours No CLAHE	Trajectory Length[m]
corridor1	0.62	0.57	0.50	0.57	305
corridor2	1.17	0.85	1.12	0.68	322
corridor3	1.31	1.59	1.82	1.26	300
corridor4	0.31	0.21	0.17	0.32	114
corridor5	0.67	0.54	0.55	0.62	270
room1	0.09	0.05	0.07	0.04	146
room2	0.05	0.04	0.05	0.05	142
room3	0.15	0.06	0.09	0.08	135
room4	0.04	0.03	0.03	0.04	68
room5	0.20	0.12	0.18	0.13	131
room6	0.06	0.05	0.06	0.08	67

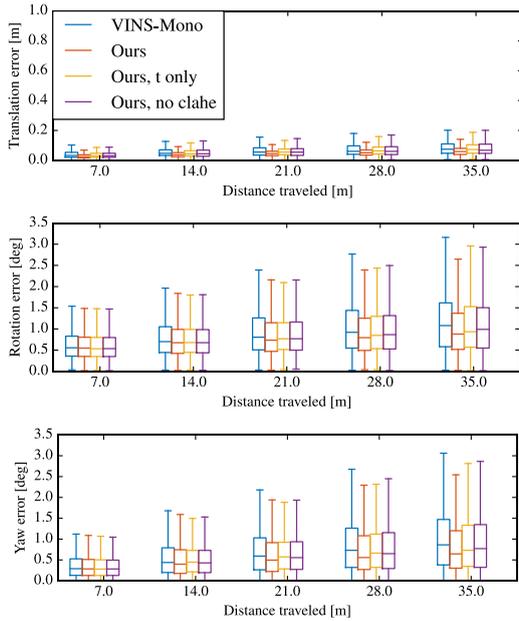


Fig. 4 Overall relative pose error [23] in sequence room1-room6. Three plots are relative errors in translation, rotation and yaw, respectively.

To further evaluate the results, we compute the relative pose error [23] using the toolbox provided by [24], with the results show in Fig. . Please note that the overall relative pose error only computed with the results of sequence room1-room6 as only the ground truth of corridor sequences only available for the start and end segment and in this case the toolbox cannot work.

Fig. and Fig. show the relative pose error and the estimated trajectory as well as the ground truth for the sequence room1 for more details, as a supplement results of

the above table and figure.

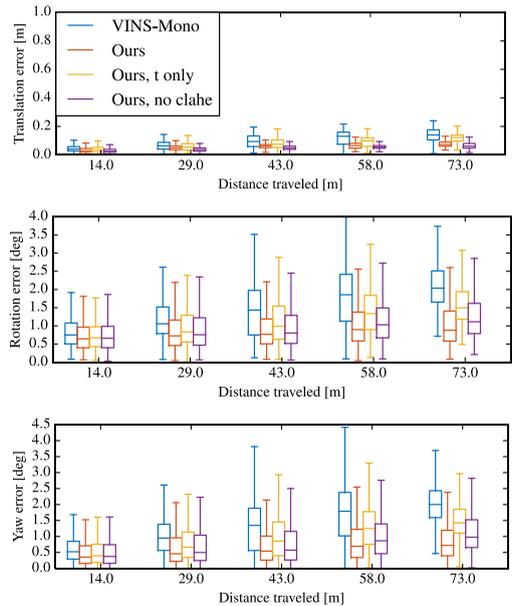


Fig. 5 Relative pose error [23] in sequence room1. Three plots are relative errors in translation, rotation and yaw, respectively.

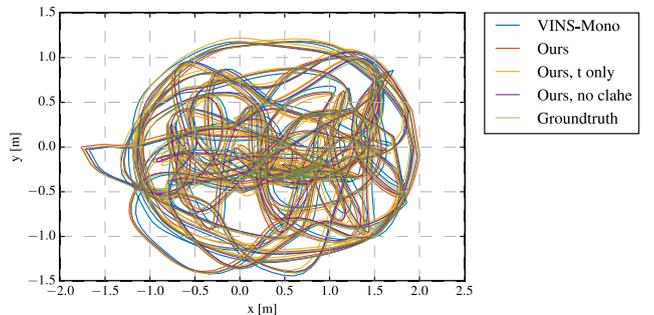


Fig. 6 Trajectory in room 1 sequence.

Table 1 it is easily to conclude that the propose VIO with photometric calibration outperforms the original VINS-Mono except the sequence of corridor3. The Overall relative pose error [23] showed in Fig. and Fig. confirm this conclusion further.

When considered the case that the camera is photometric uncalibrated and only the exposure time of the image is known, experiment results show that even only exposure time calibrated the performance of our algorithm still increase in most cases. This conclude that vary exposure time pose large challenge to the optic-flow based tracking method.

When compare the estimation results of our algorithm with and without CLAHE implementation before feature detection, we can find that the ATE of algorithm without CLAHE is overall worse than the algorithm with CLAHE. This conclude that implementation CLAHE to the image to improve the contrast of image is important for feature detection algorithm. Besides, the algorithm without CLAHE outperforms the original VINS-Mono on most of the sequence, again.

It is interesting that the RMSE ATE of corridor 1 and corridor3 with t only show the best performance compare to other results. And two of the results of our algorithm with skipping the CLAHE implementation, corridor 2 and corridor 3, also shows the best performance compare to other results. This means apply CLAHE algorithm to each image may not be the optimal policy and the algorithm can be further improved by adaptively implementation CLAHE on the selected image.

5. Conclusion

In this paper, we proposed a robust Visual-Inertial Odometry with photometric calibration, PC-VINS-Mono, using the photometric response function, vignetting, and exposure times. The proposed algorithm can be understood as an extension of VINS-Mono with the photometric calibration. With this extension, the proposed algorithm is capable for high-contrast lighting condition environment with auto exposure camera, in which, the exposure time is vary from frame to frame, as a result, violating the brightness consistency assumption. We evaluate the propose algorithm with TUM VI dataset, which including diverse sequence in different scenes with different lighting condition. Comparison experiments showed that with the photometric calibration, the performance of PC-VINS-Mono increase significantly. For camera with unknown

response function and lens attenuation factor, experiment results show that even only exposure time calibrated the performance of our algorithm still increase in most cases. Experiments with skipping the CLAHE step show reduce performance of our algorithm, which confirm that it is necessary applying CLAHE algorithm to improve the contrast of images before feature detection as the photometric calibration may reduce the contrast of the images.

Acknowledgments

We acknowledge support from National Natural Science Foundation of China (No. 61375086); Key Project (No. KZ201610005010) of S&T Plan of Beijing Municipal Commission of Education; Beijing Natural Science Foundation(4174083).

Reference

- [1] M. Li and A. I. Mourikis, "High-precision, consistent EKF-based visual-inertial odometry," *Int. J. Rob. Res.*, vol. 32, no. 6, pp. 690–711, 2013.
- [2] A. I. Mourikis and S. I. Roumeliotis, "A Multi-State Constraint Kalman Filter for Vision-aided Inertial Navigation," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*, 2007, no. April, pp. 3565–3572.
- [3] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based visual-inertial odometry using nonlinear optimization," *Int. J. Rob. Res.*, vol. 34, no. 3, pp. 314–334, 2015.
- [4] S. Leutenegger, P. Furgale, V. Rabaud, M. Chli, K. Konolige, and R. Siegwart, "Keyframe-Based Visual-Inertial SLAM using Nonlinear Optimization," in *Robotics: Science and Systems IX*, 2013, vol. 34, no. 3, pp. 314–334.
- [5] K. Sun *et al.*, "Robust Stereo Visual Inertial Odometry for Fast Autonomous Flight," pp. 1–8, 2017.
- [6] T. Qin, P. Li, and S. Shen, "VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator," *IEEE Trans. Robot.*, vol. 34, no. 4, pp. 1004–1020, Aug. 2018.
- [7] R. Mur-Artal and J. D. Tardos, "Visual-Inertial

- Monocular SLAM With Map Reuse,” *IEEE Robot. Autom. Lett.*, vol. 2, no. 2, pp. 796–803, Apr. 2017.
- [8] C. Forster, Z. Zhang, M. Gassner, M. Werlberger, and D. Scaramuzza, “SVO: Semidirect Visual Odometry for Monocular and Multicamera Systems,” *IEEE Trans. Robot.*, vol. 33, no. 2, pp. 249–265, Apr. 2017.
- [9] C. Forster, M. Pizzoli, and D. Scaramuzza, “SVO: Fast semi-direct monocular visual odometry,” in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, 2014, pp. 15–22.
- [10] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, “On-Manifold Preintegration for Real-Time Visual-Inertial Odometry,” *IEEE Trans. Robot.*, vol. 33, no. 1, pp. 1–21, Feb. 2017.
- [11] L. von Stumberg, V. Usenko, and D. Cremers, “Direct Sparse Visual-Inertial Odometry using Dynamic Marginalization,” 2018.
- [12] V. Usenko, J. Engel, J. Stückler, and D. Cremers, “Direct visual-inertial odometry with stereo cameras,” *Proc. - IEEE Int. Conf. Robot. Autom.*, vol. 2016–June, pp. 1885–1892, 2016.
- [13] J. Engel, V. Koltun, and D. Cremers, “Direct Sparse Odometry,” *arXiv:1607.02565*, no. October, pp. 1–17, 2016.
- [14] J. Engel, V. Usenko, and D. Cremers, “A photometrically calibrated benchmark for monocular visual odometry,” *arXiv Prepr. arXiv1607.02555*, 2016.
- [15] M. Burri et al., “The EuRoC micro aerial vehicle datasets,” *Int. J. Rob. Res.*, vol. 35, no. 10, pp. 1157–1163, Sep. 2016.
- [16] N. Yang, R. Wang, X. Gao, and D. Cremers, “Challenges in Monocular Visual Odometry: Photometric Calibration, Motion Bias, and Rolling Shutter Effect,” *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 2878–2885, Oct. 2018.
- [17] R. Mur-Artal and J. D. Tardos, “ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras,” *IEEE Trans. Robot.*, vol. 33, no. 5, pp. 1255–1262, Oct. 2017.
- [18] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, “ORB-SLAM: A Versatile and Accurate Monocular SLAM System,” *IEEE Trans. Robot.*, vol. 31, no. 5, pp. 1147–1163, Oct. 2015.
- [19] D. Schubert, T. Goll, N. Demmel, V. Usenko, J. Stückler, and D. Cremers, “The TUM VI Benchmark for Evaluating Visual-Inertial Odometry,” 2018.
- [20] Jianbo Shi and Tomasi, “Good features to track,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition CVPR-94*, 1994, pp. 593–600.
- [21] S. Baker and I. Matthews, “Lucas-Kanade 20 Years On: A Unifying Framework,” *Int. J. Comput. Vis.*, vol. 56, no. 3, pp. 221–255, Feb. 2004.
- [22] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, “A benchmark for the evaluation of RGB-D SLAM systems,” in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012, pp. 573–580.
- [23] A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for autonomous driving? The KITTI vision benchmark suite,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 3354–3361.
- [24] Z. Zhang and D. Scaramuzza, “A Tutorial on Quantitative Trajectory Evaluation for Visual(-Inertial) Odometry,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018.