ISSN 2706-655X Vol.5, Issue 2: 39-45, DOI: 10.25236/IJFET.2023.050207

A monocular visual inertial SLAM algorithm with point-line feature fusion

Mingli Sun, Cheng Hu*

College of Engineering, Zhejiang Normal University, Jinhua, China
**Corresponding author: 1093972837@qq.com*

Abstract: *A monocular visual inertial SLAM algorithm with point-line feature fusion is proposed to solve the problem that the system fails to achieve high accuracy because of the few effective feature points extracted in weak scenes and the tracking failure of the algorithm in rapid movements. With reference to the open-source VINS-Mono system, a modified LSD algorithm is applied to the front end odometer to detect line features so as to extract more feature of the environment while balancing the accuracy and efficiency. We evaluate the performance of our algorithm in public dataset EUROC MAV as well as compare it with VINS-Mono, the experimental results show that the localization accuracy of our algorithm improves about 21.5%.*

Keywords: *VIO; Point-line fusion; loop detection; localization*

1. Introduction

Recently, Simultaneous Localization and Mapping (SLAM) is considered to be the main component of autonomous navigation, which has become a hot trend in the field of mobile robotics. The monocular camera is widely used in VSLAM because of its small size, less power consumption, lower cost and ability to capture abundant information in the environment. However, there are some drawbacks of the monocular VSLAM system such as scale of uncertainty, which limits its application. It is difficult to accommodate sophisticated scenes with one sensor, but multi-sensor fusion can take advantage of the complementary advantages between sensors to adapt to more sophisticated scenes. In particular, the fusion of an inertial measurement unit (IMU) and a camera. It is possible to achieve greater performance with low-cost IMU-assisted monocular cameras. Because the IMU belongs to the internal sensor with high frequency sampling, which is independent of the imaging quality. The integration of IMU measurements significantly enhances the motion tracking performance and complements the camera for the loss of visual trajectory during varying illumination and lack of texture scenes. The fusion of the both sensors can effectively improve the robustness and accuracy of the VSLAM.

A line segment is a collection of points with stable characteristics in its environment. With the above advantages, we propose a monocular visual inertial SLAM algorithm with point-line feature fusion.

The paper contributes as follows:

1) A point-line fusion, tightly coupled optimization-based real-time monocular visual inertial odometry, which is performed to balance accuracy and time efficiency, with a modified LSD algorithm.

2) With a new key-frame choice mechanisms, it improves the quality of front-end tracking as well as decreases the amount of computation by reducing data redundancy and enhances the robustness of the system.

3) To validate performance of the algorithm in our paper, we performed experimental validation by 11 sequences from EUROC MAV [1] dataset and compared with VINS-Mono[2]. The experimental results showed that the localization accuracy and robustness of the algorithm in this paper are improved.

2. Relevant work

The visual SLAM is divided into feature point method and direct method depending on the front-end. The direct method is based on the grayscale invariance hypothesis. Because of no feature extraction and matching, it is fast to run with a small computational effort. However, it is susceptible to lighting conditions and other effects, such as its representative algorithms LSD-SLAM[4] and DSO[3]. The Feature

ISSN 2706-655X Vol.5, Issue 2: 39-45, DOI: 10.25236/IJFET.2023.050207

point method relies on artificially designed partial features, such as SIFT, SURF and ORB algorithms, to track the co-visual points in two frames by extracting corner points and matching the descriptors. the ORB-SLAM2[5] is a well-known open-source algorithm that balances real-time and accuracy, which is composed of three parallel threads and achieves centimetre-level localization accuracy. The group of points with obvious luminance changes, adjacent positions and similar pixel gradient directions in an Image make up the line features. In contrast to the discrete points, the match between line segments is more remarkable, which is beneficial to improve the robustness of feature detection and tracking. Firstly Smith P[6] expressed line segments by two endpoints in their paper. Sola J[7] uses long line segments in large scene to extract features. The results show that long lines are beneficial for matching. At present, the line features are often combined with point features and simultaneously implemented in SLAM. Ruben Gomez-Ojeda [8] proposes PL-STVO, a bimanual visual odometer based on point-line features. Later, the semi-direct method PL-SVO[10] was proposed based on SVO[9] . A complete SLAM algorithm PL-SLAM[11]is proposed again with point-line features based on the ORB-SLAM2 framework. The addition the line features proved to be effective in enhancing the localization accuracy, however, it is difficult to guarantee the real-time performance. All the above algorithms have a good localization accuracy in the static scene. However, its performance is poor in the fast motion. Especially, the features decrease obviously with the image blurring and texture missing, which causes the cumulative drift. The IMU and camera sensors are complementary in acquiring data, therefore many scholars fuse both of them. The VINS-Mono and the VINS-Fusion[12] are well-known algorithms open-sourced by the Shen, which support multiple types of cameras and IMU fusion. The Shi-Tomas corner points are extracted and the KLT optical flow is used to track them. The tightly coupled nonlinear optimization method obtains a highly robust and localization accuracy by using the bag-of-words model for loopback detection. In comparison to OKVIS[13], the localization accuracy of VINS-Mono is more than doubled. However, its performance also decreases sharply when the environment has fewer point features. The VINS-Mono adopts optical flow to track the corner points, which is of good real-time performance. In addition, the LSD algorithm also has a good real-time performance, so that adding LSD line segment detection increases the localization accuracy.

3. Methods

3.1 System Framework

The framework of the system is shown in Figure 1. The LSD segment detection modules are added in the VIO with the LBD description matched. The error matches are rejected with the RANSAC algorithm. The point feature extraction module remains nearly unchanged, and reduces feature concentration by a uniform extraction method.

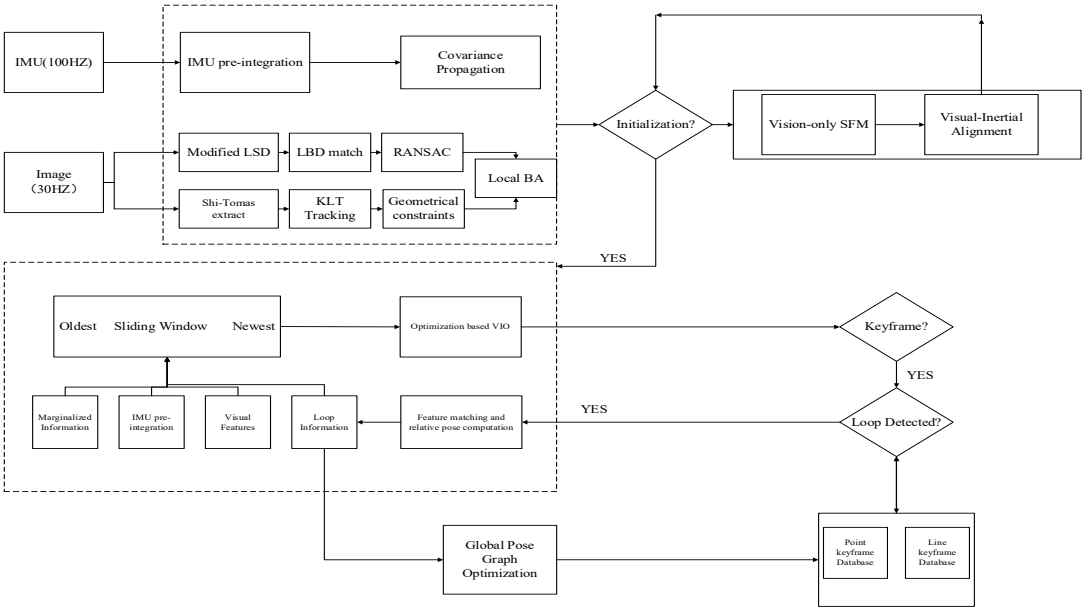

Figure 1: System Framework

ISSN 2706-655X Vol.5, Issue 2: 39-45, DOI: 10.25236/IJFET.2023.050207

3.2 VIO

In the initialization module, the initialization of the monocular is firstly performed with SFM, and then aligned with the results of IMU pre-integration [2]. The zero bias of scale, gravitational acceleration, velocity, and gyroscope are jointly optimized through visual inertial loose coupling. For the point features, an image pyramid is constructed, and the KLT optical flow is applied to track the feature. Then the RANSAC algorithm is to remove the anomalous points. The VIO module extracts point line features. For line features, the LSD algorithm from OpenCV that provides sub-pixel accuracy for line segment detection works in real time. However, the LSD algorithm tends to split a long line segment into multiple short line segments for cases such as occlusion, intersection, and blurred edges, because of the allowance of only one line per pixel point. The short line segments are easy to cause mis-matching and the long matching time leads to higher computational effort, which limits its application. The geometric length threshold and density threshold are attached to the image pyramid of LSD algorithm, effectively reducing the extraction of short line segments and the centralization of line segments. The IMU pre-integration module follows the work of VINS-Mono and makes a median integration of the angular velocity and acceleration of the IMU, aligned with the image frames.

3.3 Key-frame selection

Some representative frames in the image frame become key frames. The key frame selection in VINS-Mono only considers disparity, which has the advantage of less calculation, but it is easy to produce data redundancy, reduce the quality of the mathematical model of back-end optimization, and lead to an increase in the number of iterations and finally increase the time cost. To design a keyframe selection mechanism, three criteria are mainly referred to: 1. There is a certain disparity between adjacent keyframes, which reduces data redundancy [2]; 2. When the motion speed exceeds the threshold, the number of key frames is increased to reduce the number of iterations to ensure that the tracking is not lost; 3. Good image frame tracking quality, as the main selection criteria.

3.4 Optimization

To balance computational efficiency and localization accuracy, a nonlinear optimization method based on a slide-window is adopted to guarantee the real-time performance by restricting the computation. To balance computational efficiency and localization accuracy, a nonlinear optimization method based on a slide window is adopted to guarantee the real-time performance by restricting the computation volume. The state vectors are dynamically added and removed through the sliding window to optimize only the key frame of data in a period of time. The full state variables in the sliding window at moment i are defined as:

$$X = [x_n, x_{n+1} \dots x_{n+N}, \lambda_m, \lambda_{m+1} \dots \lambda_{m+M}, O_l, O_{l+1}, O_{l+L}] \tag{1}$$

$$x_i = \left[p_{wb_i}, q_{wb_i}, v_{wb_i}, b^a_{b_i}, b^g_{b_i}\right]^T, i \in [n, n+N] \tag{2}$$

The above equation contains the IMU state vectors (position q_{wb_i},attitude p_{wb_i}, velocity v_{wb_i}, accelerometer bias $b^a_{b_i}$ and gyroscope bias $b^g_{b_i}$) of the sliding window at the N+1 moment. The inverse depth λ_m of M+1 3D points and the orthogonal representation of L+1 line features in the world coordinate system O_l where the sub-index n, m, l indicates the starting index of the state, point marker and line marker, respectively. M and L mean the number of the observed map points and lines for all keyframes in the sliding window [2]. The optimization function is constructed with four types of residual terms: the a priori residual term of marginalization, IMU measurement, visual reprojection (point reprojection and line reprojection errors), and loopback detection. The form of the optimization function is as follows:

$$\begin{aligned}\min_x \Bigg\{ &\|r_p - H_p x\|^2 + \sum_{k\in\mathcal{B}} \left\| r_{\mathcal{B}}\left(z^{b_k}_{b_{k+1}}, x\right)\right\|^2_{P^{b_k}_{b_{k+1}}} + \sum_{(l,j)\in e} \rho\left(\left\| r_e\left(z^{c_j}_l, x\right)\right\|^2_{P^{c_j}_l}\right) \\ &+ \sum_{(r,j)\in\mathcal{L}} \rho\left(\left\| r_{\mathcal{L}}\left(z^{c_j}_\gamma, x\right)\right\|^2_{L^{c_j}_\mu}\right) \\ &+ \sum_{(l,\gamma,j)\in\mathcal{F}} \rho\left(\left\| r_{\mathcal{F}}\left(z^{c_j}_l, z^{c_j}_\gamma, \mathcal{H}, x\right)\right\|^2\right) + \sum_{(r_i,\gamma_j)\in\mathcal{P}} \rho\left(\|r_{\mathcal{P}}(X)\|^2\right)\Bigg\}\end{aligned} \tag{3}$$

In the formula above, $\{r_p, H_p\}$is the priori information after marginalization; $H_p$is the information

ISSN 2706-655X Vol.5, Issue 2: 39-45, DOI: 10.25236/IJFET.2023.050207

matrix of previously optimized;p_{bj}^{bi} is the covariance of the IMU pre-integrated noise term,p_{fi}^{cj}、$p_{li}^{cj}$is the noise covariance of the visual observation.J_p is the Jacobian matrix of the state quantities;$r_b(z_{b_i b_{i+1}}, \chi)$ is the IMU residual;$r_f\left(z_{r_j}^{c_i}, \chi\right)$ is the point eigen residuals;$r_l\left(z_{L_i}^{c_i}, \chi\right)$ is the line eigen residuals. B, F, and L are the sets of IMU pre-integrated values, point feature measurements, and line feature measurements in the sliding window. The Cauchy kernel function Cauchy Loss is chosen as the robust kernel function $\rho(\cdot)$ for rejecting abnormal measurements. The pucker coordinates are used to replace the line coordinates, and the line re-projection residuals are represented by the distances from the two end-points to the line. At last, the residuals are minimized by iteratively updating the pucker coordinates with the orthogonal representation of the minimum four parameters.

3.5 Closed-loop detection

The inevitable existence of noise in the procedure of extraction and association of data leads to accumulated errors in the trajectory estimates, which makes SLAM impossible to construct a globally consistent map. To remove the accumulated errors, the closed-loop detection to recognize similar scenes is implemented to provide back-loop constraint for the optimization of the SLAM. The dictionary tree is constructed by clustering descriptors based on the kmeans algorithm, with words weighted by TF-TDF. The system runs before loading the offline trained vocabulary. The front-end of VINS-Mono tracks the optical flow with fewer than 150 Shi-Tomasi corner points, so that key frames extract extra 500 FAST corner points and calculate BRIEF descriptors, which is converted into vectors online. The similarity between images is determined with L1 parametric scale. Therefore, there is no closed-loop detection for the previous 5 frames before the current key frame, to reduce annoyances and time cost. To raise the accuracy of the closed-loop detection, the algorithm also incorporates the verification of temporal consistency and geometric consistency.

4. Experiment and analysis

We validate the feasibility and effectiveness of the proposed monocular visual inertial SLAM algorithm with point-line feature fusion through the public EUROC MAV dataset.

4.1 Experimental Environment and platform

We used hardware configuration of I7-8750H, 8G RAM and no GPU acceleration. The software is based on Ubuntu 18.04 system and ROS (Melodic), and the program is coded based on third-party libraries such as OpenCV, Ceres-Slover, and DBOW, with RVIZ as the visualization tool. The left-eye and IMU data of the EUROC MAV dataset were selected as input data and compared with VINS-Mono to run the respective systems independently. The EVO tool was used to statistically analyze the experimental results, and the absolute translation error was used as the evaluation criterion.

4.2 Experimental results

In Figure 2 and Figure 3, the system runs on the difficult scenes MH05 and VR13.Hhowever, the algorithm can still extract effective and stable number of point and line features. As a result, the fusion of point and line features enables the SLAM front-end to acquire richer and more effective data associations.

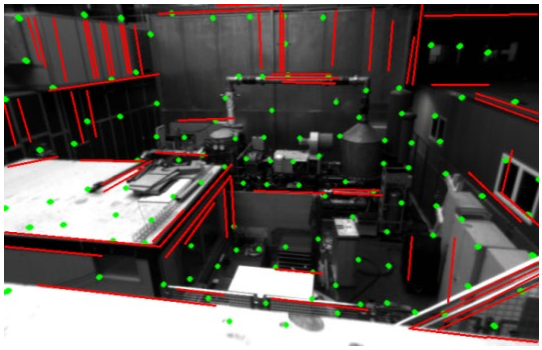

Figure 2: Point and line features of MH05

ISSN 2706-655X Vol.5, Issue 2: 39-45, DOI: 10.25236/IJFET.2023.050207

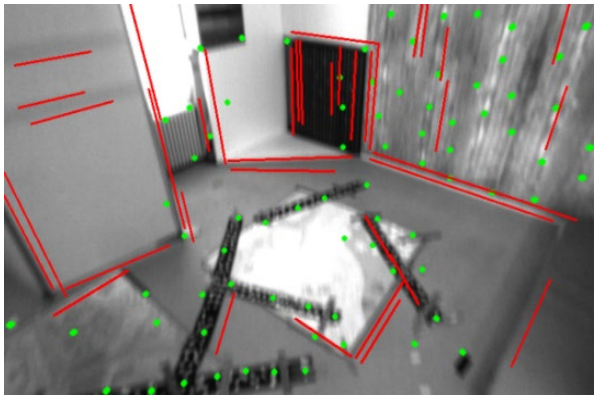

Figure 3: Point and line features of VR13

The MH03 scene is selected as the analysis target. As shown in Figure 4, the trajectory error of our algorithm has a higher overlap compared to VINS-Mono. As shown in Figure 5, the numerical characteristics of the absolute error of our algorithm on MH03 are smaller.

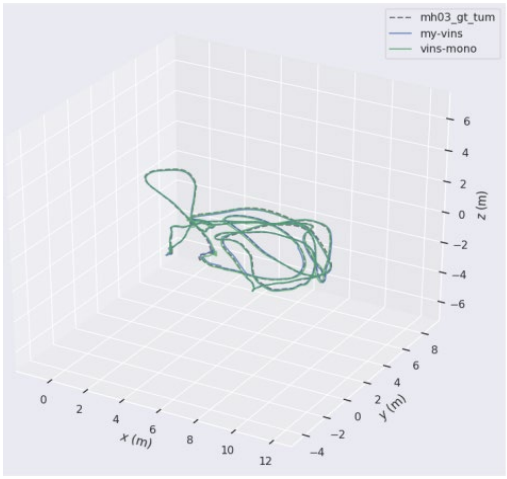

Figure 4: Trajectory error of MH03

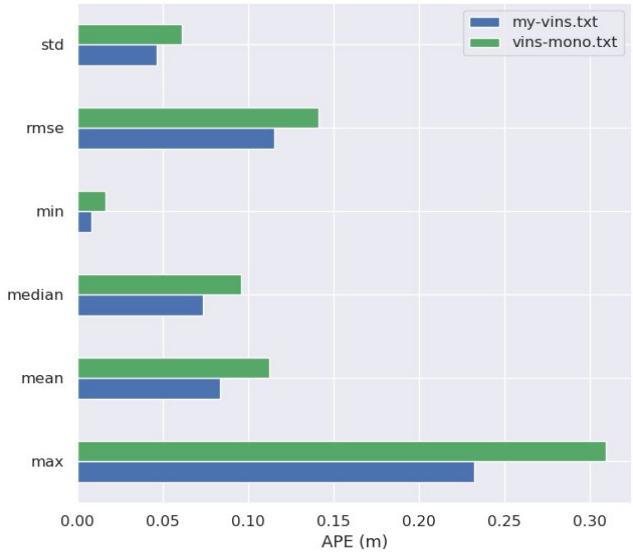

Figure 5: Absolute error of MH03

The EUROC MAV has 11 dataset sequences and the difficulty metrics are given on the official website. Table 1 shows the comparison between our algorithm and VINS-Mono for these 11 dataset sequences. By RMSE as the evaluation criterion, the localization accuracy is improved by about 21.5%.

ISSN 2706-655X Vol.5, Issue 2: 39-45, DOI: 10.25236/IJFET.2023.050207

Table 1: Experimental results on the EUROC MAV dataset

VINS-Mono/My-Vins (m)						
Sequence	MAX	MIN	MEAN	MEDIUM	STD	RMSE
MH_01_easy	0.372/0.296	0.019/0.022	0.137/0.125	0.137/0.119	0.055/0.054	0.142/0.132
MH_02_easy	0.307/0.231	0.013/0.013	0.089/0.075	0.070/0.070	0.052/0.042	0.124/0.106
MH_03_medium	0.309/0.232	0.016/0.007	0.112/0.083	0.095/0.073	0.061/0.046	0.141/0.115
MH_04_difficult	0.573/0.538	0.044/0.034	0.209/0.185	0.191/0.184	0.107/0.089	0.215/0.175
MH_05_difficult	0.518/0.443	0.011/0.035	0.244/0.220	0.225/0.197	0.115/0.089	0.270/0.198
V1_01_easy	0.148/0.120	0.002/0.009	0.054/0.049	0.047/0.046	0.030/0.024	0.069/0.055
V1_02_medium	0.273/0.138	0.016/0.009	0.083/0.070	0.075/0.070	0.042/0.031	0.089/0.076
V1_03_difficult	0.328/0.303	0.007/0.020	0.152/0.141	0.160/0.140	0.081/0.073	0.193/0.159
V2_01_easy	0.349/0.273	0.002/0.013	0.061/0.072	0.057/0.048	0.060/0.049	0.095/0.078
V2_02_medium	0.218/0.210	0.010/0.010	0.094/0.078	0.087/0.077	0.040/0.038	0.152/0.089
V2_03_difficult	0.487/0.374	0.050/0.024	0.220/0.156	0.220/0.152	0.093/0.074	0.239/0.173
Mean	0.353/0.287	0.017/0.018	0.132/0.114	0.124/0.107	0.067/0.055	0.157/0.123

5. Conclusion

We proposes a Monocular visual-inertial SLAM with Point and Line Features. The front-end adds a modified LSD algorithm to detect line features. It enhances the robustness of the front-end and effectively suppresses the problems of unstable performance and low localization accuracy in a weak-texture environment. Besides, a new key-frame strategy may reduce iterations, which also decreases time cost. In the future, we will use binocular and IMU fusion to extract point and line features to improve localization accuracy.

Acknowledgements

Fund Support: supported by the Key Research and Development Program of Zhejiang Province [grant number 2022C01139]; Supported by the industry university research innovation fund of the Ministry of education of China [grant number 2021JQR012]; Zhejiang Collaboration and Innovation Center for Urban Rail Transit Operation Safety Technology & Equipment; Zhejiang Provincial Key Laboratory of Urban Rail Transit Intelligent Operation and Maintenance Technology & Equipment.

References

[1] Burri, Michael, et al. "The EuRoC micro aerial vehicle datasets." The International Journal of Robotics Research 35.10 (2016): 1157-1163.
[2] Qin Tong, Peiliang Li, and Shaojie Shen. "Vins-mono: A robust and versatile monocular visual-inertial state estimator." IEEE Transactionson Robotics 34.4 (2018): 1004-1020.
[3] Wang R, Martin Schwörer, Daniel Cremers. Stereo DSO: Large-scale direct sparse visual odometry with stereo cameras[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy, October 22-29, 2017: 3923-3931.
[4] Jakob Engel, Thomas Schöps, Daniel Cremers. LSD-SLAM: Large-scale direct monocular SLAM [C]//European Conference on Computer Vision. Zurich, Switzerland, September 6-12, 2014: 834-849.
[5] Raúl Mur-Artal, Juan D. Tardós. ORB-SLAM2: An open-source SLAM system for monocular, stereo and RGB-D cameras [J]. IEEE Transactions on Robotics, 2017, 33(5): 1255-1262.
[6] Smith P, Reid I D, Davison A J. Real-time monocular SLAM with straight lines [C]// Proceedings of the British Machine Vision Conference 2006. London: British Machine Vision Conference, 2006, 6: 17-26.
[7] Joan Sola, Vidal-Calleja T, Michel Devy. Undelayed initialization of line segments in monocular SLAM [C]. //2009 IEEE/RSJ International Conference on Intelligent Robotics and Systems (IROS 2009). [v.3]: IEEE, 2009: 1553-1558.
[8] Gomez-Ojeda R, Gonzalez-Jimenez J. Robust stereo visual odometry through a probabilistic combination of points and line segments [C]// IEEE International Conference on Robotics & Automation. IEEE, 2016.
[9] Forster C, Pizzoli M, D Scaramuzza SVO: Fast semi-direct monocular visual odometry [C] //IEEE International Conference on Robotics & Automation. IEEE, 2014.

ISSN 2706-655X Vol.5, Issue 2: 39-45, DOI: 10.25236/IJFET.2023.050207

[10] Gomez-Ojeda R, Briales J, Gonzalez-Jimenez J. PL-SVO: Semi-direct Monocular Visual Odometry by combining points and line segments [C]// 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2016.
[11] Pumarola A, Vakhitov A, Agudo A, et al.PL-SLAM: real-time mo- nocular visual SLAM with points and lines [C]// IEEE Internation- al Conference on Robotics and Automation (ICRA), 2017: 4503-4508.
[12] QIN T, CAO S, PAN J, et al. A general optimization-based framework for global pose estimation with multiple sensors [EB/OL]. [2022-01-12]. https://arxiv.org/pdf/1901.03638.pdf.
[13] Stefan Leutenegge, Simon Lynen, Michael Bosse, et al. Keyframe-based visual-inertial odometry using nonlinear optimization [J]. International Journal of Robotics Research, 2015, 34(3): 314-334.