ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0017

An improved Perspective-n-Point Algorithm for Bundle Adjustment

Yujie Chen[1, 2, a], Guoliang Zhang[1, 2, b], Junxue Li[1, 2], Lihong Deng[3]

[1]Department of Automation and Information Engineering, Sichuan University of Science & Engineering, Sichuan, China

[2]Sichuan Key Laboratory of artificial intelligence, Sichuan, China

[3]Department of Information Science and Technology, Chengdu University of Technology, Sichuan, China

[a]1050591258@qq.com, [b]zhgl@sohu.com

Abstract

In order to solve the nonsingular and ill conditioned problems of the coefficient matrix of linear equations and the slow calculation speed when solving incremental equations in Bundle Adjustment optimization process, an improved gradient descent algorithm based on Gauss Newton algorithm and Levenberg Marquardt algorithm is proposed. The algorithm mainly finds the best increment by adding a dynamic trust region to the increment. In this effective region, the increment is calculated by the size of Lagrange multiplier to find the best point in the region. The global consistency and real-time performance of slam are improved by optimizing the camera pose and feature points of adjacent key frames in real time. The algorithm avoids the nonsingular and ill conditioned problems of the coefficient matrix of linear equations to a certain extent, corrects the stability problem of Gauss Newton algorithm, and improves the calculation speed of Levenberg Marquardt method. Experimental results based on datasets and real scenes show that the performance of the algorithm is better than the mainstream algorithms such as Gauss Newton algorithm and Levenberg Marquardt algorithm in many real scenes.

Keywords

Simultaneous Localization and Mapping; Bundle Adjustment optimization; Levenburg-Marquadt algorithm; Gauss-Newton algorithm; Trust Region.

1. INTRODUCTION

Robot Simultaneous Localization and Mapping [1, 2] refers to a mobile robot equipped with sensors that builds an environment description model while moving [3] and estimates its own position [4]. SLAM [5] includes two problems of positioning [6, 7, 8] and mapping at the same time. They are one of the key problems to realize robot autonomy and have important research significance in the fields of robot navigation and task planning [9]. The PnP (perspective-n-point) optimization algorithm is considered to be a very important method for solving the 3D to 2D point-to-motion pose [10] estimation method in the positioning [11] problem. The main methods to solve it are: P3P, direct linear transformation (DLT), EPnP (Efcient PnP), UPnP, and so on.In addition, the PnP problem can be constructed as a nonlinear least squares problem defined on Lie algebra [12] and iteratively solved by means of nonlinear optimization, which is the Bundle Adjustment [13] problem.However, in the process of BA optimization, the stability of the algorithm is not enough, which makes the results difficult to converge and the speed of

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0017

the algorithm needs to be improved. Therefore, the problem of improving the PnP optimization of BA [14] has attracted more and more researchers' attention. In [15], the Sida Peng team studied the problem of estimating 6-degree-of-freedom pose from a single RGB image under severe occlusion or truncation.It is proposed to introduce a pixel-level voting network (PVNet) to return the pixel-level unit vectors pointing to key points, and use these vectors to vote for the key point positions using RANSAC [16, 17].This will create a flexible representation for locating key points that are occluded or truncated.The article [18] is a new method of visual refocusing based on direct image alignment, LM-Reloc, proposed by Lukas von Stumberg's team.Compared with the previous work on feature-based formula processing problems, this method does not rely on feature matching and RANSAC, and mainly uses corner points or arbitrary regions of gradient images for visual relocation.Different from [18], the algorithm in this paper uses RANSAC for preprocessing before BA optimization to filter out mismatches generated by feature matching.In [19], the Lipu Zhou team proposed an effective algorithm for solving the least squares problem using point-to-plane cost, which aims to jointly optimize the attitude and plane parameters of the depth sensor for 3D reconstruction.They introduced a simplified Jacobian matrix and a simplified residual vector, and proved that they can replace the original Jacobian matrix and residual vector in the commonly used Levenberg-Marquardt algorithm, which greatly reduces the computational cost.

The algorithm in this paper refers to the advantages of the above algorithm, and regards the BA [20] problem as a problem of minimizing Reprojection error, that is, considering both the camera pose and spatial point position as optimization variables, and put them together for optimization. By adding a dynamic trust region to the increment to correct the ill-conditioned problem of the linear equations, and improve the LM algorithm to reduce the running time of the algorithm, and improve the real-time performance of SLAM.

2. BUNDLE ADJUSTMENT OPTIMIZATION ALGORITHM

In the process of SLAM [21], the robot state process needs to be constructed. Through the calculation of its probability, the problem is transformed into the problem of seeking maximum likelihood estimation. Finally, the problem can be transformed into a problem of solving least squares. Assuming n three-dimensional space points P and their projections P, the pose R, t of the camera needs to be calculated, and the Lie algebra is expressed as ξ . Suppose the coordinates of a certain space point are Pi=[X_i， Y_i， Z_i]^T ,Its projected pixel coordinates are u_i=[u_i,v_i]^T .The relationship between pixel position and spatial point position is as follows:

$$s_i u_i = K \exp(\xi\hat{})P_i \tag{1}$$

In the formula, Si is the scale factor, and K is the camera internal parameter.

Due to the unknown camera pose and the noise generated by the observation point, there is an error in the equation. Usually, the least squares problem is constructed by summing the errors, and then finding the best camera pose to minimize it:

$$\xi^* = \arg\min_{\xi} \frac{1}{2}\sum_{i=1}^{n} \| u_i - \frac{1}{s_i} K \exp(\xi\hat{})P_i \|_2^2 \tag{2}$$

The following text is written for the convenience of writing this style:

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0017

$$\mathrm{K}x^{*}=\min_{x}\frac{1}{2}\|f(x)\|_2^2 \tag{3}$$

The error term of this problem is the error obtained by comparing the pixel coordinates (observed projection position) with the position obtained by projecting the 3D point according to the current estimated pose, and this error is the reprojection error. Using non-homogeneous coordinates, the error is only 2 dimensions. As shown in Figure 1:

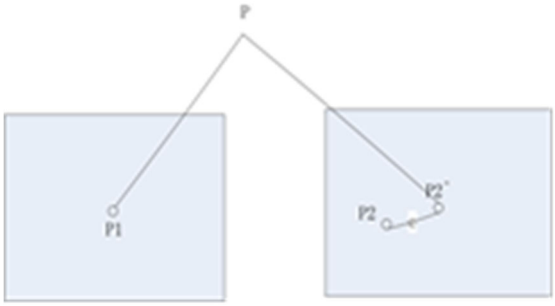


Figure 1. Re projection error diagram

It can be known that P1 and P2 are projections of the same spatial point P. After initialization, there is a certain distance between the projection P2 of P and the actual P2, so the pose of the camera needs to be adjusted to make this distance smaller.However, this adjustment needs to consider many points, so the error of each point in the end is usually not exactly zero.

For the solution of the nonlinear least squares problem in (2), the iterative method is generally used to continuously find the decreasing increment $\mathrm{K}x_k$ .In the BA optimization algorithm, GN and LM algorithms are generally used to solve the increment.

2.1. Gauss-Newton Algorithm

To find an increment $\mathrm{K}x_k$ through the Gauss-Newton algorithm so that the error $\|f(x_k+\mathrm{K}x_k)\|_2^2$ reaches the minimum value, it is necessary to perform a first-order Taylor expansion on $f(x+\mathrm{K}x)$:

$$f(x+\mathrm{K}x)\approx f(x)+J(x)^{T}\mathrm{K}x+o(\mathrm{K}x) \tag{4}$$

Convert the above equation into a nonlinear least squares problem:

$$\mathrm{K}x^{*}=\arg\min\frac{1}{2}\|f(x+\mathrm{K}x)\|_2^2$$

$$\approx\arg\min\frac{1}{2}\|f(x)+J(x)T\mathrm{K}x\|_2^2 \tag{5}$$

Derivation:

$$m(x)=\frac{1}{2}\|f(x)+J(x)^{T}\mathrm{K}x\|^2$$

$$=\frac{1}{2}(f(x)+J(x)^{T}\mathrm{K}x)^{T}(f(x)+J(x)^{T}\mathrm{K}x)$$

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0017

$$=\frac{1}{2}(\| f(x)\|^2 +2f(x)J(x)^T \mathrm{K}x+\mathrm{K}x^T J(x)J(x)^T \mathrm{K}x) \tag{6}$$

$$m'(x)=J(x)f(x)+J(x)J(x)^T \mathrm{K}x \tag{7}$$

At this point, it can be transformed into a linear solution problem.make $m'(x)=0$,then $J(x)J(x)^T \mathrm{K}x=-J(x)f(x)$ .defined $J(x)J(x)^T$ as $H(x)$, $-J(x)f(x)$ is $g(x)$ .Then the above formula becomes:

$$H\mathrm{K}x=g \tag{8}$$

Compared with the traditional least squares solution method, the GN algorithm avoids finding the H matrix and greatly reduces the amount of calculation. But there are some disadvantages:

(1) In the solution of the incremental equation, the approximate H matrix used is required to be invertible (and positive definite), but the calculated $J^T J$ in the actual data is only positive semi-definite. Therefore, when it is a singular matrix, the stability is poor and the algorithm does not converge;

(2) If the calculated step size $\mathrm{K}x_k$ is too large, the local approximation will be inaccurate. In severe cases, iterative convergence may not be guaranteed;

(3) Like the gradient descent method, it is easy to get into a jagged shape, resulting in a longer number of iterations.

2.2. Levenburg-Marquadt Algorithm

The Levenberg algorithm proposed the Trust Region Method on the basis of the Gauss Newton method, which is used to solve the problem of the Gauss Newton method that is easy to fall into the jagged shape.

In the process of updating and iteration of the algorithm, in order to judge the quality of the approximation, first set a judgment approximation index:

$$\rho=\frac{f(x+\mathrm{K}x)-f(x)}{J(x)^T \mathrm{K}x} \tag{9}$$

According to this approximate index, the results can be divided into the following situations:

(1) ρ is close to 1, the approximation is good, no need to change;

(2) If ρ is too small, the actual reduced value is smaller than the approximate reduced value, which is approximately larger, and the approximate range needs to be reduced;

(3) If ρ is too large, the actual reduction value is greater than the approximate reduction value, and the approximate reduction is small, and the approximate range needs to be expanded.

Through the approximate index, the size of the trust zone can be set. If it is not close to the set threshold, the dynamic area is continuously adjusted until a good approximate result is found. When an approximate result that meets the requirements is found, subsequent normal iterative updates can be performed.

For the Kth iteration, the trust region is added on the basis of the Gauss-Newton method:

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0017

$$\min \frac{1}{2} \| f(x) + J(x)T\text{K}x \|, s.t. \| D\text{K}x \|^2 \leq \mu \tag{10}$$

Construct a Lagrangian function, where λ is the coefficient factor:

$$L(\text{K}x, \lambda) = \frac{1}{2} \| f(x) + J(x)T\text{K}x \|^2 + \frac{\lambda}{2} (\| D\text{K}x \|^2 - \mu) \tag{11}$$

By simplification and derivation, we can get:

$$(JJ^T + \lambda D^T D)\text{K}x = -Jf \tag{12}$$

in the text, make $H = JJ^T$, $g = -Jf$, Under normal circumstances, D is taken as the identity matrix I, then the above formula becomes:

$$(H + \lambda I)\text{K}x = g \tag{13}$$

The LM method solves the problem that $J^T J$ is easily ill-conditioned in the Gauss-Newton method by introducing a damping term, and can switch between the gradient method and the Newton method by adjusting the damping. However, the calculation time is increased compared to the higher S-Newton method.

3. IMPROVED LM ALGORITHM BASED ON BA

In view of the problems in the Gauss Newton algorithm and Levenberg algorithm, the improved algorithm in this paper avoids the non-singular and ill-conditioned problems of the coefficient matrix of the linear equation system in the Gauss Newton method by adding a trust region to a certain extent. It is judged whether to update the radius of the trust area by calculating the size of the criterion factor, which reduces the calculation time compared with the Levenberg-Marquardt algorithm. Overall provides a more stable and accurate increment $\text{K}x$.

3.1. Algorithm Improvement Ideas

The improved algorithm in this paper combines the advantages of the Gauss-Newton method and the Levenberg method, and improves on this basis to correct their existing problems to a certain extent.

Before solving PnP, the RANdom SAmple Consensus algorithm is used for preprocessing to filter out mismatches after matching between adjacent frames of VSLAM. This algorithm is an iterative algorithm that correctly estimates the parameters of the mathematical model from a set of data containing outliers. Outliers generally refer to the noise in the data.

The number of iterations K is:

$$K = \frac{\log(1-P)}{\log(1-t^n)} \tag{14}$$

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0017

Among them, P is the probability of RANSAC getting the correct model, and t is the proportion of inliers in the data. After the preprocessing is completed, the initial trust area radius and the evaluation criterion u are set, and the value is also determined according to the difference between the approximate model and the actual function.

$$u = \frac{f(x+\mathrm{K}x) - f(x)}{J(x)\mathrm{K}x} * \gamma \tag{15}$$

γ is the scale factor.

First, an initial value of x_0 and an initial trust area radius of μ are given. Then, in a spherical region centered on the current value and μ as the radius, the true radius is obtained by finding the best point of an approximate function of the objective function. After the radius is obtained, the objective function value is calculated again. If it makes the decrease of the objective function value meet a certain condition, then the radius is reliable, so continue to iteratively calculate according to this rule. As long as it is in the spherical area, the approximation is valid, and if it is outside this area, the approximation will have problems.

After calculating the criterion factor, perform BA optimization calculation. For the Kth iteration, solve:

$$\min_{\mathrm{K}x} \frac{1}{2} \| f(x) + J(x)\mathrm{K}x \|^2, s.t. \|\mathrm{K}x\|^2 \le \mu \tag{16}$$

It is transformed into an unconstrained optimization problem by Lagrangian multiplier λ :

$$\min_{\mathrm{K}x} \frac{1}{2} \| f(x) + J(x)\mathrm{K}x \|^2 + \frac{\lambda}{2} \|\mathrm{K}x\|^2 \tag{17}$$

After expanding the above equation, the core of the problem is still to calculate the linear equation of the increment, that is, to solve: $(H + \lambda I)\mathrm{K}x = g$ .When λ is relatively small, H is dominant. The model is better in this range, that is, the algorithm is closer to the Gauss-Newton method; When λ is relatively large, it occupies the main position. The algorithm is equivalent to a step-descent method, indicating that the second approximation is not good enough.So choose to discard this matching pair in the calculation to reduce the calculation time.

3.2. Algorithm overall optimization framework

At this point, an improved nonlinear optimization framework can be constructed:

step 1. RANSAC pretreatment, screen out mismatch;

step 2. Given the initial value x_0 , set the initial trust zone radius μ ;

step 3. For the Kth iteration, solve: $\min_{\mathrm{K}x} \frac{1}{2} \| f(x) + J(x)\mathrm{K}x \|^2, s.t. \|\mathrm{K}x\|^2 \le \mu$;

step 4. Calculate the criterion factor u;

step 5. When $u \ge 1$,Approximately feasible. Make $\mu = 3\mu$, $x_{k+1} = x_k + \mathrm{K}x$.

step 6. When $u < 1$,think that approximation is not feasible, end the process;

step 7. Determine whether the algorithm has converged. If it does not converge, return to step 2, otherwise end.

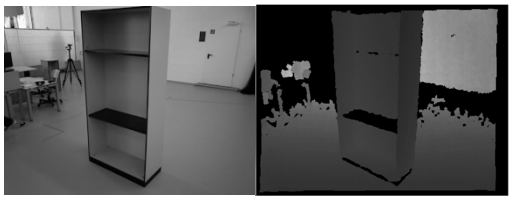
ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0017

4. EXPERIMENTAL VERIFICATION AND ANALYSIS

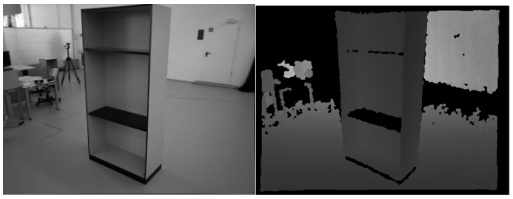
This section is mainly to verify the calculation speed of the improved algorithm. The experiment is first tested on the TUM data set. In the experiment, the time used for posture estimation between two adjacent frames is estimated, and the improved algorithm in this paper is compared with the current commonly used gradient descent methods, including the GN algorithm and the LM algorithm. In addition, the experimental plan is evaluated through actual scene data in the laboratory. The CPU of all experimental running platforms is Inter i5-8500 processor, the main frequency is 3.00GHz, the memory is 8G, GPU acceleration is not used, the system is Ubuntu16.04, and the ROS version is Kinect.

4.1. Data set test

This article uses the cabinet_big data set in the public TUM data set for simulation experiments. This data set is obtained by the ASUS Xtion sensor rotating around a large office cabinet. The cabinets have no texture and structure, and are mostly flat and right-angled. This paper selects two sets of data in the data set for testing, as shown in Figure 2, the data contains rgb maps and depth maps. First, perform feature matching on the two images, as shown in Figure 3.



(a) Indoor scene 1 and its depth map



(b) Indoor scene 2 and its depth map

Figure 2. Sample dataset diagram

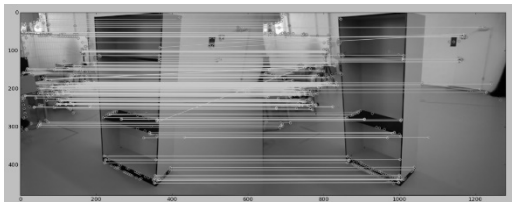


Figure 3. Data set feature matching diagram

First, the RANSAC algorithm is used to filter out the mismatches after the SLAM adjacent frames are matched. After that, the number of matching pairs was reduced from 351 to 199 and then improved BA optimization. Compare the result with traditional GN algorithm and LM algorithm. The results are shown in Table 1.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0017

Table 1. Runnng time table of each algorithm in TUM dataset

algorithm	The algorithm of this article	GN algorithm	GN algorithm
operation hours	0.00094	0.00121	0.00219

4.2. Real Scene Test

The experiment was carried out in a laboratory on the fourth floor of a teaching office building of a university. The testers used the Astra Pro depth camera to shoot. The depth camera is a motion sensing camera of LeEco and Obi Zhongguang, which is compatible with Microsoft Kinect. For 3D reconstruction, SLAM learning, it can also be used as a drive-free UVC camera somatosensory camera.

(C) Laboratory scene map 1 and depth map

(D) Laboratory scene graph 2 and depth graph

Figure 4. Laboratory data acquisition diagram

In the experiment, the camera was used to collect video images. A total of two sets of data are collected, as shown in Figure 4. Compare the result with traditional GN algorithm and LM algorithm. The results are shown in Table 2.

Table 2. The running schedule table of each algorithm in the actual scene

algorithm	The algorithm of this article	GN algorithm	GN algorithm
operation hours	0.00034	0.00038	0.00235

Compared with the GN algorithm, the algorithm in this paper increases the trust region and can correct the stability problem of the Gauss-Newton algorithm. Compared with the LM algorithm, it reduces the gradient descent method and improves the calculation speed.It can also be seen from the above two experimental results that the calculation speed of the algorithm in this paper is improved compared with the LM algorithm and the GN algorithm.

5. CONCLUSION

This paper proposes an improved PnP optimization algorithm for BA. First, the RANSAC algorithm is used to filter out mismatched pairs to improve the accuracy of the algorithm.The most important thing is to set a trust zone for the increment, in which the increment is calculated by the size of the Lagrangian multiplier. This method corrects the stability problem of the Gauss-Newton method.After that, the size of the evaluation factor is used to determine

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0017

whether the approximation is feasible. This method can reduce the calculation time to a certain extent.Finally, the data set test and the actual scene test show that the algorithm avoids the non-singular and ill-conditioned problems of the coefficient matrix of the linear equation system to a certain extent and reduces the calculation time.However, the algorithm is currently limited to the pose optimization between two adjacent frames in PnP. How to further apply it to the back-end BA calculation that contains a large number of feature points and camera poses is the focus of our next step.

REFERENCES

[1] MUR-ARTAL R, TARDOS J D.ORB−SLAM2: An open source SLAM system for monocular, stereo, and RGB-D cameras [J]. IEEE Transactions on Robotics, 2017, 33(5) : 1−8.

[2] S. H. Lee and J. Civera,.Loosely-Coupled Semi-Direct Monocular SLAM [J]. IEEE Robotics and Automation Letters, vol. 4, no. 2, pp. 399-406, April 2019.

[3] Lei Zhou, Zixin Luo, Mingmin Zhen, Tianwei Shen, Shiwei Li, Zhuofei Huang, Tian Fang, Long Quan. Stochastic Bundle Adjustment for Efficient and Scalable 3D Reconstruction. Computer Vision and Pattern Recognition.2020. arXiv:2008.00446.

[4] Larsson V, Kukelova Z, Zheng Y. Making minimal solvers for absolute pose estimation compact and robust[C]//2017 IEEE International Conference on Computer Vision (ICCV). IEEE, 2017: 2335-2343.

[5] ZOU X,XIAO C S,WEN Y Q,YUAN H W. Research on vSLAM based on feature point method and direct method. Research on computer application. 2020.37(5). (in Chinese).

[6] Taira H, Okutomi M, Sattler T, et al. InLoc: Indoor Visual Localization with Dense Matching and View Synthesis[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7199-7209.

[7] Schönberger J L, Pollefeys M, Geiger A, et al. Semantic Visual Localization[J]. ISPRS Journal of Photogrammetry and Remote Sensing (JPRS), 2018.

[8] Huaiyang Huang, Haoyang Ye, Jianhao Jiao, Yuxiang Sun, Ming Liu. Geometric Structure Aided Visual Inertial Localization. 2020. arXiv:2011.04173.

[9] Zhao Y, Vela P A. Good Line Cutting: Towards Accurate Pose Tracking of Line-Assisted VO/VSLAM[C]//European Conference on Computer Vision. Springer, Cham, 2018: 527-543.

[10] Palmér T, Astrom K, Frahm J M. The Misty Three Point Algorithm for Relative Pose[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 2786-2794.

[11] PENG Q Y,XIA L,Y,WU,D,J. Semi direct monocular vision localization algorithm based on fusion of luminosity and point line features. Sensors and Microsystems. 2020.39(4). (in Chinese).

[12] Huu Le, Christopher Zach, Edward Rosten, Oliver J. Woodford. Progressive Batching for Efficient Non-linear Least Squares.2020. arXiv:2010.10968.

[13] Yipu Zhao, Justin S. Smith, Patricio A. Vela. Good Graph to Optimize: Cost-Effective, Budget-Aware Bundle Adjustment in Visual SLAM. Computer Vision and Pattern .Recognition.2020. arXiv:2008.10123.

[14] Liu H, Chen M, Zhang G, et al. ICE-BA: Incremental, Consistent and Efficient Bundle Adjustment for Visual-Inertial SLAM[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 1974-1982.

[15] Sida Peng, Yuan Liu, Qixing Huang, Hujun Bao, Xiaowei Zhou. PVNet: Pixel-wise Voting Network for 6DoF Pose Estimation.2018. arXiv:1812.11788.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0017

[16] Brachmann E, Krull A, Nowozin S, et al. DSAC—Differentiable RANSAC for camera localization[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017: 2492-2500.

[17] CHUNYAN SHAO, CHI ZHANG, ZAOJUN FANG, AND GUILIN YANG. A Deep Learning-Based Semantic Filter for RANSAC-Based Fundamental MatrixCalculation and the ORB-SLAM System[J]. IEEE ACCESS.2019.2962268.

[18] Stumberg, Lukas & Wenzel, Patrick & Yang, Nan & Cremers, Daniel. LM-Reloc: Levenberg-Marquardt Based Direct Visual Relocalization. 2020.arXiv:2010.06323

[19] Lipu Zhou,Daniel Koppel,Hui Ju,Frank Steinbruecker,Michael Kaess. An Efficient Planar Bundle Adjustment Algorithm.2020. arXiv: 2006.00187v2.

[20] Zhang R, Zhu S, Fang T, et al. Distributed very large scale bundle adjustment by global camera consensus[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 29-38.

[21] Maity S, Saha A, Bhowmick B. Edge SLAM: Edge Points Based Monocular Visual SLAM[C]//ICCV Workshops. 2017: 2408-2417.