# S-VIO: Exploiting Structural Constraints for RGB-D Visual Inertial Odometry (Supplementary Material)

Pengfei Gu, Ziyang Meng

***Abstract*—In this supplementary materials, we provide the additional experiment results to show the efficiency improvement of the proposed candidate interval calculation method over the original one [1] as a part of structural line clustering and camera rotation estimation algorithm. Besides, synthetic experiments are carried out to evaluate the accuracy of the proposed line initialization method leveraging the known direction information.**

## I. Evaluation of Candidate Interval Calculation

We utilize the open-sourced code[1] provided by the authors of [1] to demonstrate the computational efficiency of the new candidate interval calculation method. In particular, the proposed candidate interval calculation method is integrated into the Mine-and-Stab algorithm proposed in [1] for structural line clustering and the camera rotation estimation. Here the correctness of our method is reflected by the accuracy of the final rotation estimation precision. Also, the original candidate interval calculation method contained in the open-sourced code is tested to provide a baseline for our method. Both the original and the improved method are implemented in MATLAB and run on a laptop computer with an Intel Core i7 (2.8GHz) CPU and 32 GB RAM.

Table I reports the quantitative evaluation results of two methods on the ICL-NUIM dataset, including the average error of the rotation estimation and the average computation time of the candidate interval calculation per line. Experiment results show that both methods can obtain accurate camera rotation results. Particularly, it is observed that the rotation estimation results of two methods are almost the same, which demonstrates the correctness of our candidate interval calculation method. In terms of the computational speed, the execution time of the proposed candidate interval computation method is an order of magnitude lower than the original method's. As stated in the paper, this is mainly because we take a different way to establish the equation and thus avoid the cumbersome coefficient calculation in the original method. Table II shows the number of operations needed by a single calculation of the equation coefficients. It is observed that our method requires fewer operations than the original method does. Besides, our method only needs to solve a second-order equation, and thus eliminates the need for high-order equation solving.

[1] https://github.com/PyojinKim/MWMS

TABLE I
THE ROTATION ERROR [DEGREE] AND THE COMPUTATION TIME [MS] OF TWO METHODS ON THE ICL-NUIM DATASET

| methods | lr-kt0 | lr-kt1 | of-kt0 | of-kt1 | computation time |
|---|---|---|---|---|---|
| [1] | 0.24 | 0.43 | 0.33 | 0.22 | 0.11 |
| ours | 0.24 | 0.43 | 0.33 | 0.22 | 0.01 |

TABLE II
THE NUMBER OF OPERATIONS FOR A SINGLE CALCULATION

| methods | addition | subtraction | multiplication | division | others[1] |
|---|---|---|---|---|---|
| [1] | 309 | 306 | 2955 | 2 | 1508 |
| ours | 7 | 8 | 24 | 7 | 7 |

1: *Others* represents other complicate operations including exponentiation, square root and trigonometry operations.

## II. Evaluation of Direction-aided Line Initialization

Previous works generally use the intersection of two planes to estimate the initial position of the line feature in the world/reference frame [2], [3]. In this section, we preform a synthetic experiment to compare the accuracy of the proposed direction-aided line initialization method with the commonly used plane-intersection-based method. Particularly, we implement the n-view line initialization method proposed in [4] as the representation of the traditional plane-intersection-based method. Compared with the two-view line initialization method generally used in many odometry systems, the n-view approach avoids the degeneracy problem by aggregating multiple views' measurements. This n-view method is then followed by an intersection of the estimated line and the projection plane to calculate the initial two-parameter expression for the structural line features, which is also the approach adopted by the structural VIO systems [5], [6]. Both methods are implemented in C++ and tested on a desktop computer with an Intel Core i9 (3.6GHz) CPU and 32 GB RAM.

TABLE III
THE CONFIGURATION FOR THE SYNTHETIC LINE ENVIRONMENT

| configuration | | [x, y, z] / m | [yaw, pitch, roll][1] / deg |
|---|---|---|---|
| A | camera 1 | -5, 0, 0 | 0, 0, 0 |
| | camera 2 | -5, 2, 0 | 0, 0, 0 |
| | camera 3 | -5, -2, 0 | 0, 0, 0 |
| | camera 4 | -5, 0, 2 | 0, 0, 0 |
| | camera 5 | -5, 0, -2 | 0, 0, 0 |
| B | camera 1 | -5, 0, 0 | 0, 0, 0 |
| | camera 2 | -5, 2, 0 | 45, 0, 0 |
| | camera 3 | -5, -2, 0 | -45, 0, 0 |
| | camera 4 | -5, 0, 2 | 0, 45, 0 |
| | camera 5 | -5, 0, -2 | 0, -45, 0 |

1: Here the rotation adopts the ZYX Euler angles representation.

TABLE IV
THE MAE [CM] OF LINE INITIALIZATION IN THE SYNTHETIC EXPERIMENT

| configuration | A | | | B | | |
|---|---|---|---|---|---|---|
| & noise std. | $0.1°$ | $0.5°$ | $1.0°$ | $0.1°$ | $0.5°$ | $1.0°$ |
| [4] | 0.93 | 4.66 | 9.53 | 0.93 | 4.66 | 9.52 |
| ours | 0.80 | 4.09 | 8.72 | 0.81 | 4.12 | 8.79 |

As shown in Table III, we generate a synthetic environment which contains (a) 10000 lines generated randomly in terms of both the direction and the position but with a constraint that the distance of the line and the origin of the world frame must be lower than 2 meters, (b) five fixed cameras with known poses. In the experiment, we project the lines onto the image plane of five cameras and perturb the normal vector of the projected line feature by the zero-mean Gaussian noise. Then the two methods are used to calculate the initial two-parameter expression of these line features. It should be noted that the direction of these lines are known for both methods, and all five views' measurements are used in these two methods to initialize the line. After the line initialization completes, we compute the distance between the estimated line and the true line to reflect the accuracy of the line initialization algorithm.

Table IV reports the mean absolute error (MAE) of the distance between the estimated line and the true line. Under the lowest noise level (i.e. $0.1°$), both methods can produce accurate line estimation results very close the true value, which demonstrates the correctness of both methods. Also, it is observed that the proposed direction-aided method obtains a smaller error than the traditional n-view plane intersection-based method does under different noise levels. This shows the effectiveness of our method for structural line initialization. Besides, we also record the average computation time of the two methods. The proposed method consumes 0.18 ms for a single line initialization from five views, slightly slower than the traditional method does (0.12 ms).

## REFERENCES


[1] P. Kim, H. Li, and K. Joo, "Quasi-globally optimal and real-time visual compass in manhattan structured environments," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 2613–2620, 2022.

[2] Y. He, J. Zhao, Y. Guo, W. He, and K. Yuan, "Pl-vio: Tightly-coupled monocular visual–inertial odometry using point and line features," *Sensors*, vol. 18, no. 4, 2018.

[3] Q. Fu, J. Wang, H. Yu, I. Ali, F. Guo, Y. He, and H. Zhang, "Pl-vins: Real-time monocular visual-inertial slam with point and line features," *arXiv:2009.07462*, 2020.

[4] S. J. Lee and S. S. Hwang, "Elaborate monocular point and line slam with robust initialization," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 1121–1129.

[5] H. Wei, F. Tang, Z. Xu, C. Zhang, and Y. Wu, "A point-line vio system with novel feature hybrids and with novel line predicting-matching," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 8681–8688, 2021.

[6] H. Wei, F. Tang, Z. Xu, and Y. Wu, "Structural regularity aided visual-inertial odometry with novel coordinate alignment and line triangulation," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 10 613–10 620, 2022.