

Stereo UV-SLAM in Hilti SLAM Challenge 2023

Hyunjun Lim¹, *Student Member, IEEE*, and Hyun Myung*, *Senior Member, IEEE*

Abstract—The datasets in the Hilti SLAM Challenge 2023 include fast-moving and dark environments in construction site environments. Under these conditions, methods using cameras have difficulty in estimating an accurate position. To tackle this problem, stereo UV-SLAM is represented. The proposed algorithm is a visual-inertial odometry (VIO) algorithm using only cameras and an inertial measurement unit (IMU). The proposed algorithm uses point features similar to the general VIO algorithm. However, point features alone are not enough to solve the difficulties of datasets. Thus, line features that can be used as rich information in construction site environments are used. Furthermore, vanishing points are used to ensure observability of line features.

I. STEREO UV-SLAM

The proposed method extends UV-SLAM [1] to use a stereo camera to utilize as much information as possible from datasets. The proposed method is based on VINS-Mono [2], and the way IMU and point measurements are used is similar to it. The point features are extracted from Shi-Tomasi [3] and are tracked by KLT [4]. In addition, the IMU measurement model is defined by the pre-integration method [5]. Finally, our optimization-based method employs two-way marginalization with Schur complement [6].

To add line features to the monocular visual-inertial odometry (VIO) system, the line features are extracted from enhanced line segment drawing (ELSED) [7] and are tracked by line binary descriptor (LBD) [8]. In addition, the UV-SLAM can determine whether the extracted lines have structural regularities or not. For vanishing point detection, the proposed algorithm use J-linkage [9], which can find multiple instances in the presence of noise and outliers.

In this paper, $(\cdot)^w$, $(\cdot)^c$, and $(\cdot)^b$ represent the world coordinate, camera coordinate, and body coordinate, respectively. In addition, $(\cdot)_b^w$ reflects the coordinate transformations of a rotation matrix, quaternion, or translation from the body coordinate to the world coordinate. The state vector used in our system is as follows:

$$\begin{aligned} \mathcal{X} = & [\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{I-1}, \\ & \lambda_0, \lambda_1, \dots, \lambda_{J-1}, \\ & \mathbf{o}_0, \mathbf{o}_1, \dots, \mathbf{o}_{K-1}], \quad (1) \\ \mathbf{x}_i = & [\mathbf{p}_{b_i}^w, \mathbf{q}_{b_i}^w, \mathbf{v}_{b_i}^w, \mathbf{b}_a, \mathbf{b}_g], \quad i \in [0, I-1], \\ \mathbf{o}_k = & [\psi_k, \phi_k], \quad k \in [0, K-1], \end{aligned}$$

where \mathcal{X} represents the entire state, and \mathbf{x}_i represents the body state in the i -th sliding window, which is made up

H. Lim is with School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea tp02134@kaist.ac.kr

*H. Myung is with School of Electrical Engineering, KI-AI, and KI-R, KAIST, Daejeon, Republic of Korea hmyung@kaist.ac.kr

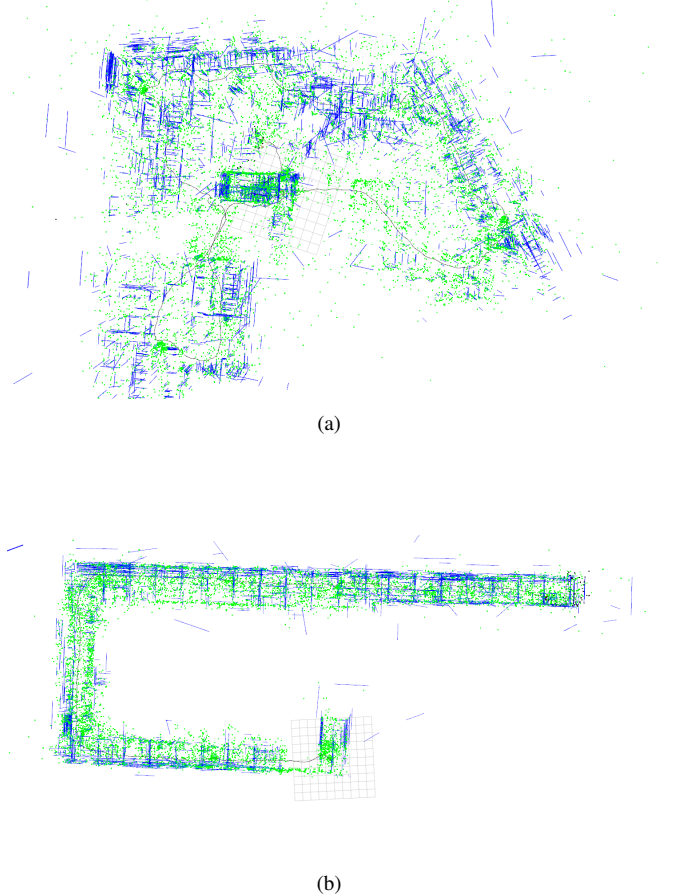


Fig. 1. Top views of 3D point and line mapping results in (a) site1_handheld_3 and (b) site3_handheld_3

of the following parameters: position, quaternion, velocity, and biases of the accelerometer and gyroscope. In addition, the entire state includes the inverse depths of point features, which are represented as λ_j , $j \in [0, J-1]$. In this paper, lines expressed in the orthonormal representations are newly added as \mathbf{o} . I , J , and K are the numbers of sliding window, point features, and line features, respectively.

Employing defined states in (1), the entire objective for optimization is as follows:

$$\begin{aligned} \min_{\mathcal{X}} & \left\{ \|\mathbf{r}_0 - \mathbf{J}_0 \mathcal{X}\|^2 \right. \\ & + \sum_{i \in \mathcal{B}} \|\mathbf{r}_I(\mathbf{z}_{b_{i+1}}^{b_i}, \mathcal{X})\|_{\Sigma_{b_{i+1}}^{b_i}}^2 + \sum_{(i,j) \in \mathcal{P}} \rho_p \|\mathbf{r}_p(\mathbf{z}_{p_j}^{c_i}, \mathcal{X})\|_{\Sigma_{p_j}^{c_i}}^2 \\ & \left. + \sum_{(i,k) \in \mathcal{L}} \rho_l \|\mathbf{r}_l(\mathbf{z}_{l_k}^{c_i}, \mathcal{X})\|_{\Sigma_{l_k}^{c_i}}^2 + \sum_{(i,k) \in \mathcal{V}} \rho_v \|\mathbf{r}_v(\mathbf{z}_{v_k}^{c_i}, \mathcal{X})\|_{\Sigma_{v_k}^{c_i}}^2 \right\}, \quad (2) \end{aligned}$$

TABLE I
REPORTED SCORES OF OUR FINAL RESULTS

	< 0.5cm	< 1cm	< 3cm	< 6cm	< 10cm	< 40cm	> 40cm	Score
site1_handheld.1	0	0	0	0	1	3	0	7.50
site1_handheld.2	0	0	0	1	0	3	0	10.00
site1_handheld.3	0	0	0	0	1	2	1	6.25
site1_handheld.4	0	0	0	0	1	2	0	8.33
site1_handheld.5	0	1	0	0	0	2	0	20.00
site2_robot.1	0	0	0	0	0	0	0	0.00
site2_robot.2	0	0	0	0	0	3	0	5.00
site2_robot.3	0	0	0	0	0	0	0	0.00
site3_handheld.1	0	0	1	1	2	0	0	42.50
site3_handheld.2	0	0	1	3	2	0	0	45.00
site3_handheld.3	0	0	1	1	1	5	0	23.75
site3_handheld.4	0	0	1	1	0	2	0	32.50
Total	0	1	4	7	8	22	1	200.83

where \mathbf{r}_0 , \mathbf{r}_I , \mathbf{r}_p , \mathbf{r}_l , and \mathbf{r}_v represent marginalization, IMU, point, line, and vanishing point measurement residuals, respectively. In addition, $\mathbf{z}_{b_{i+1}}^{b_i}$, $\mathbf{z}_{p_j}^{c_i}$, $\mathbf{z}_{l_k}^{c_i}$, and $\mathbf{z}_{v_k}^{c_i}$ stand for observations of IMU, point, line, and vanishing point, respectively; \mathcal{B} is the set of all pre-integrated IMU measurements in a sliding window; \mathcal{P} , \mathcal{L} and \mathcal{V} are the sets of point, line, and vanishing point measurements in observed frames; and $\Sigma_{b_{i+1}}^{b_i}$, $\Sigma_{p_j}^{c_i}$, $\Sigma_{l_k}^{c_i}$ and $\Sigma_{v_k}^{c_i}$ represent IMU, point, line, and vanishing point measurement covariance matrices, respectively. ρ_p , ρ_l , and ρ_v mean loss functions of the point, line, and vanishing point measurements, respectively. ρ_p and ρ_l are set to the Huber norm function [10] and ρ_v is set to the inverse tangent function because of the vanishing point measurement model's unbound problem. For the optimization process, Ceres Solver [11] is used.

II. EXPERIMENTAL RESULTS

The experiments were carried out on an Intel Core i7-9700K processor with 32GB of RAM. TABLE I shows the reported scores of our final results. At site 2 we barely scored, the camera input rate was low and the IMU was quite inaccurate in rotational situations.

III. CONCLUSION

We scored about 200 points in the Hilti SLAM challenge 2023 through stereo UV-SLAM. Loop closing through the camera rarely occurred in these datasets. In the next challenge, the loop closing method through multi-camera will be applied.

REFERENCES

- [1] H. Lim, J. Jeon, and H. Myung, "UV-SLAM: Unconstrained line-based SLAM using vanishing points for structural mapping," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 1518–1525, 2022.
- [2] T. Qin, P. Li, and S. Shen, "VINS-Mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.
- [3] J. Shi *et al.*, "Good features to track," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1994, pp. 593–600.
- [4] C. Tomasi and T. Kanade, "Detection and tracking of point," *International Journal of Computer Vision*, vol. 9, pp. 137–154, 1991.
- [5] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, "On-manifold preintegration for real-time visual-inertial odometry," *IEEE Transactions on Robotics*, vol. 33, no. 1, pp. 1–21, 2016.

- [6] G. Sibley, L. Matthies, and G. Sukhatme, "Sliding window filter with application to planetary landing," *Journal of Field Robotics*, vol. 27, no. 5, pp. 587–608, 2010.
- [7] I. Suárez, J. M. Buenaposada, and L. Baumela, "ELSEED: Enhanced Line SEGment Drawing," *Pattern Recognition*, vol. 127, p. 108619, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0031320322001005>
- [8] L. Zhang and R. Koch, "An efficient and robust line segment matching approach based on LBD descriptor and pairwise geometric consistency," *Journal of Visual Communication and Image Representation*, vol. 24, no. 7, pp. 794–805, 2013.
- [9] R. Toldo and A. Fusiello, "Robust multiple structures estimation with J-linkage," in *Proc. European Conference on Computer Vision (ECCV)*. Springer, 2008, pp. 537–547.
- [10] P. J. Huber, "Robust estimation of a location parameter," in *Breakthroughs in Statistics*. Springer, 1992, pp. 492–518.
- [11] S. Agarwal, K. Mierle, and Others, "Ceres solver." Accessed on: Sep. 09, 2021. [Online]. Available: <http://ceres-solver.org>.