

# LiDAR-Visual-Inertial SLAM Robust in Structurally and Visually Degenerate Environments

○\*Junwoon Lee (The University of Tokyo) Mitsuru Shinozaki (KUBOTA Corporation)  
Toshihiro Kitajima (KUBOTA Corporation) Qi An (The University of Tokyo)  
Atsushi Yamashita (The University of Tokyo)

## 1. Introduction

In recent years, 3D simultaneous localization and mapping (SLAM) has played an important role in the automation of mobile robots in global navigation satellite system (GNSS) denied environments such as tunnel, underground, indoor, and forest environments.

Light detection and ranging (LiDAR) SLAM is one of the robust and accurate methods for solving SLAM problems, as the 3D structural information of the environments, which is measured with LiDAR, can be directly used to construct the 3D map. However, structure-less environments, such as corridors, vast planes, and long tunnels, force LiDAR SLAM to degenerate, as most of the LiDAR SLAM [1–3] solely relies on measured structural information. On the other hand, visual SLAM [4–6], which relies on textural and color information from images, not structural information, has the potential to solve these structural degenerate issues. However, visual SLAM easily suffers with texture-less, lightening changing environments and aggressive motion of the SLAM agent.

To solve these issues originating from LiDAR and visual SLAM, various LiDAR-visual SLAM [7–11] have been proposed. However, as most of these methods rely on maximum a posteriori (MAP) fusion of the information of each sensor, the entire system can experience failure when either method degenerates largely. On the other hand, a hard-switching based method [12] between LiDAR and visual SLAM has been proposed to deal with each sensor degenerate environment. Although this method can have potential to solve the situation with the failure of either sensor, the entire system severely relies on thresholds for the switch.

To address the above problems, in this paper, we propose a novel LiDAR-visual-inertial SLAM using a hybrid method between MAP-based and hard-switching sensor fusion method. The main contributions of our method are as follows:

- **Active factor graph:** Active factor graph is newly proposed to fuse MAP-based and hard-switching-based method. Each factor is selectively connected or disconnected along with switching conditions.
- **Degeneracy-aware optimization:** To deal with structurally degenerate situations, the optimization of feature-to-map matching will vanish along directions of degenerate degree of freedom (DOF).

## 2. Proposed Method

### 2.1 System Overview

The overview of the proposed method is described in Fig. 1. The proposed approach consists of a visual inertial odometry (VIO) module, 3-D feature extraction, active factor graph, and degeneracy-aware optimization.

Firstly, VINS-MONO [6] is used for our VIO module to derive VIO results and visual residual that contains visual-inertial bundle adjustment formulation. Then, LOAM [1] is used for feature extraction to extract plane features and LiDAR residual calculated as point-to-plane distance. After the initial guess is selected between VIO and IMU preintegration [13] along with degeneracy detection results, an active factor graph is constructed for sliding window optimization. Finally, degeneracy-aware optimization is performed to selectively optimize only directions of well-conditioned DOF.

### 2.2 Structural Degeneration Detection

To detect structurally degenerate situations, we utilize a method as proposed in [14], which evaluates the distribution of LiDAR point cloud data based on the eigenvalues of the Hessian matrix used in scan-to-map matching.

Following [1, 2], the scan-to-map matching cost  $\mathbf{f}_1$  can be defined as follows:

$$\mathbf{f}_1(\mathbf{x}) = \mathbf{d}_p, \quad (1)$$

where  $\mathbf{x}$  is a 6-DOF pose of SLAM agent, and  $\mathbf{d}_p$  is a vector stacked with  $d_p$ , which represents the distances between a planar feature of current scan  $\mathbf{p}_j^p$  and corresponding map planar features  $\mathbf{p}_{i,1}^p$ ,  $\mathbf{p}_{i,2}^p$  and  $\mathbf{p}_{i,3}^p$ . Note that  $d_p$  is denoted as follows:

$$d_p = \frac{\|(\mathbf{p}_j^p - \mathbf{p}_{i,1}^p)((\mathbf{p}_{i,1}^p - \mathbf{p}_{i,2}^p) \times (\mathbf{p}_{i,1}^p - \mathbf{p}_{i,3}^p))\|}{\|(\mathbf{p}_{i,1}^p - \mathbf{p}_{i,2}^p) \times (\mathbf{p}_{i,1}^p - \mathbf{p}_{i,3}^p)\|}. \quad (2)$$

As a result, the nonlinear optimization eq. (1) is solved using Gauss-Newton as follows:

$$\delta\mathbf{x} = -\mathbf{H}_1^{-1}\mathbf{J}_1^\top \mathbf{d}_1, \quad (3)$$

where  $\mathbf{J}_1 = \frac{\partial \mathbf{f}_1}{\partial \mathbf{x}}$  denotes the Jacobian matrix, and  $\mathbf{H}_1 = \mathbf{J}_1^\top \mathbf{J}_1$  denotes the Hessian matrix. If the minimum eigenvalue of  $\mathbf{H}_1$  is smaller than a predefined threshold, our system treats the current state as structurally degenerate.

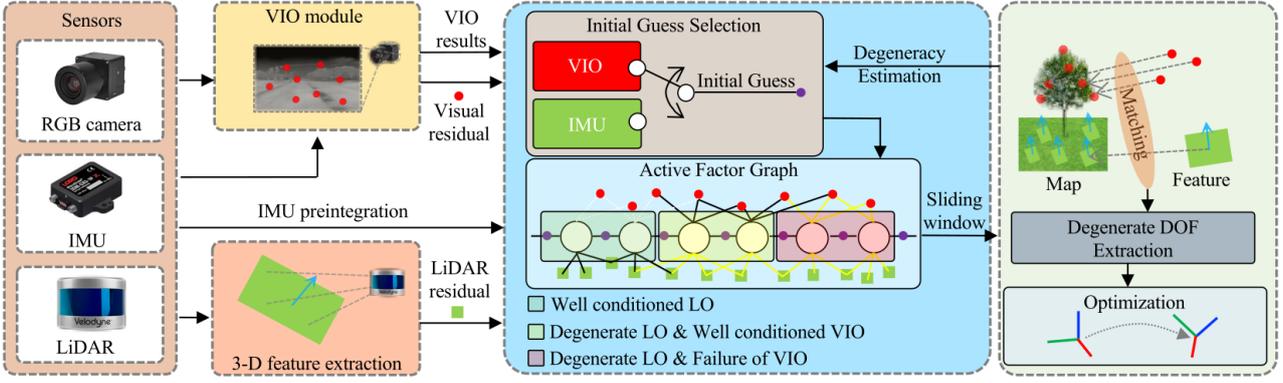


Fig.1: The system structure of proposed method

### Algorithm 1 Active Factor Graph

**Input:** Prior status  $\mathbf{T}_{k-1}$ ,  $\mathbf{H}_1$  in eq. (3), status of LO  $S_1$ , status of VIO  $S_{vio}$ , differential state of VIO  $\delta\mathbf{T}_{k-1,k}^v$ , differential state of IMU pre.  $\delta\mathbf{T}_{k-1,k}^i$ , LiDAR residual factor  $\mathbf{F}_k^l$ , and visual residual factor  $\mathbf{F}_k^v$

**Output:** Active factor graph  $\mathbf{F}_k^f$

- 1: **if**  $S_1 == 1$  //Well conditioned LO **then**
- 2:  $\mathbf{F}_{k-1,k}^{init} = \mathbf{T}_{k-1} \boxplus \delta\mathbf{T}_{k-1,k}^i$
- 3: Insert  $\mathbf{F}_{k-1,k}^{init}$  and  $\mathbf{F}_k^l$  to  $\mathbf{F}_k^f$
- 4: **else if**  $S_1 == 0 \vee S_{vio} == 1$
- //Degenerate LO & Well conditioned VIO **then**
- 5:  $\mathbf{F}_k^{init} = \mathbf{T}_{k-1} \boxplus \delta\mathbf{T}_{k-1,k}^v$
- 6: Insert  $\mathbf{F}_{k-1,k}^{init}$ ,  $\mathbf{F}_k^v$ , and  $\mathbf{F}_k^l$  to  $\mathbf{F}_k^f$
- 7: **else if**  $S_1 == 0 \vee S_{vio} == 0$
- //Degenerate LO & Failure of VIO **then**
- 8:  $\mathbf{F}_k^{init} = \mathbf{T}_{k-1} \boxplus \delta\mathbf{T}_{k-1,k}^i$
- 9: Insert  $\mathbf{F}_{k-1,k}^{init}$  and  $\mathbf{F}_k^l$  to  $\mathbf{F}_k^f$
- 10: **end if**
- 11: **return**  $\mathbf{F}_k^f$

When structural degeneration is detected, the state of LiDAR odometry, denoted as  $S_1$ , is set to “0” (indicating structural degeneration; otherwise, it is set to “1”).

### 2.3 Visual Failure Detection

The number of tracked features and positional changes are used to detect visually failed situations as proposed in [6]. When visual degeneration is detected, the state of VIO, denoted as  $S_{vio}$ , is set to “0” (indicating visual degeneration; otherwise, it is set to “1”).

### 2.4 Active Factor Graph

To mitigate degenerate situations involving either individual sensors or all sensors, we propose an active factor graph. The active factor graph utilizes three types of factors: the initial guess factor, the LiDAR residual factor,

and the visual residual factor. The factor graph is actively constructed along with  $S_1$  and  $S_{vio}$ .

When LO is well-conditioned, the initial guess, propagated with IMU preintegration and LiDAR odometry, is used for the initial guess factor. Moreover, only the LiDAR residual in eq. (1) and the initial guess factor are used to construct the factor graph, excluding the visual residual. Note that visual odometry is relatively less accurate compared to LiDAR odometry in well-structured environments, which makes the multimodal fusion less accurate in such environments.

When LO is degenerate and VIO is well-conditioned, the VIO result, propagated at the IMU rate, is used for the initial guess factor. Furthermore, both the LiDAR and visual residuals are inserted into the factor graph to help escape from LiDAR degeneration. Note that visual residual is denoted in [6].

When LO is degenerate and VIO experiences failure simultaneously, the IMU preintegration value is utilized for the initial guess factor. The LiDAR residual and initial guess factor are then utilized to construct the factor graph. In this case, the initial guess factor has a significant impact on the directions of the degenerate DOF in optimization process, reducing drift even in structurally and visually degenerate situations.

### 2.5 Degeneracy-aware optimization

After constructing the factor graph, the sliding window optimization considering the directions of structurally degenerate degrees of freedoms (DOFs) is formulated as follows:

$$\delta\mathbf{x} = \underset{\delta\mathbf{x}}{\operatorname{argmin}} ( \|\delta\mathbf{x} + (\mathbf{U}\mathbf{\Lambda}_p\mathbf{U}^{-1})^{-1}\mathbf{J}_1^\top\mathbf{d}_1\|^2 + \|\delta\mathbf{x} + (\mathbf{H}_v^{-1}\mathbf{J}_v^\top\mathbf{d}_v)\|^2 + \|\delta\mathbf{x} - \delta\mathbf{x}_0\|^2 ) \quad (4)$$

Here,  $\mathbf{\Lambda}_p$  denotes the matrix with eigenvalues removed corresponding to degenerate DOFs from  $\mathbf{\Lambda}$ , where eigen-decomposition of  $\mathbf{H}_1$  is  $\mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}$ . Moreover,  $\delta\mathbf{x}_0$  is a vector of initial guesses,  $\mathbf{J}_v$  denotes the Jacobian matrix of VIO, and  $\mathbf{H}_1 = \mathbf{J}_1^\top\mathbf{J}_1$  denotes the corresponding Hessian matrix. Note that VIO constraints in eq. (4) are used only

Table1: Comparison of Absolute Translational Errors (Maximum, RMSE) on Prepared Datasets. The units are in meters.

Dataset	Fast rotate		Plane		Farm		Mine		Handheld		Multi Floor		Long Corridor	
	Max	RMSE	Max	RMSE										
LOAM	1.41	0.44	-	-	-	-	0.60	0.28	-	-	17.9	10.6	25.6	12.6
LIO-SAM	0.72	0.21	-	-	-	-	5.84	2.33	-	-	-	-	17.6	7.64
VINS-MONO	-	-	<b>1.17</b>	<b>0.41</b>	-	-	0.65	0.29	21.4	10.3	12.8	6.30	23.8	11.8
LVI-SAM	8.82	1.82	1.82	0.69	28.8	5.75	2.69	<b>0.23</b>	3.27	<b>1.23</b>	-	-	18.4	8.7
R2LIVE	1.67	0.64	19.5	8.53	8.52	4.21	3.95	2.58	-	-	27.2	14.6	-	-
R3LIVE	10.1	6.43	9.01	5.84	58.6	34.7	6.70	4.13	-	-	<b>5.38</b>	3.08	15.5	7.87
FAST-LIVO	11.3	7.12	-	-	51.2	26.5	-	-	-	-	-	-	-	-
Proposed	<b>0.26</b>	<b>0.15</b>	1.47	0.63	<b>1.69</b>	<b>0.70</b>	<b>0.51</b>	0.26	<b>3.02</b>	1.30	5.72	<b>2.13</b>	<b>10.5</b>	<b>4.77</b>

“-” denotes the failure of localization.

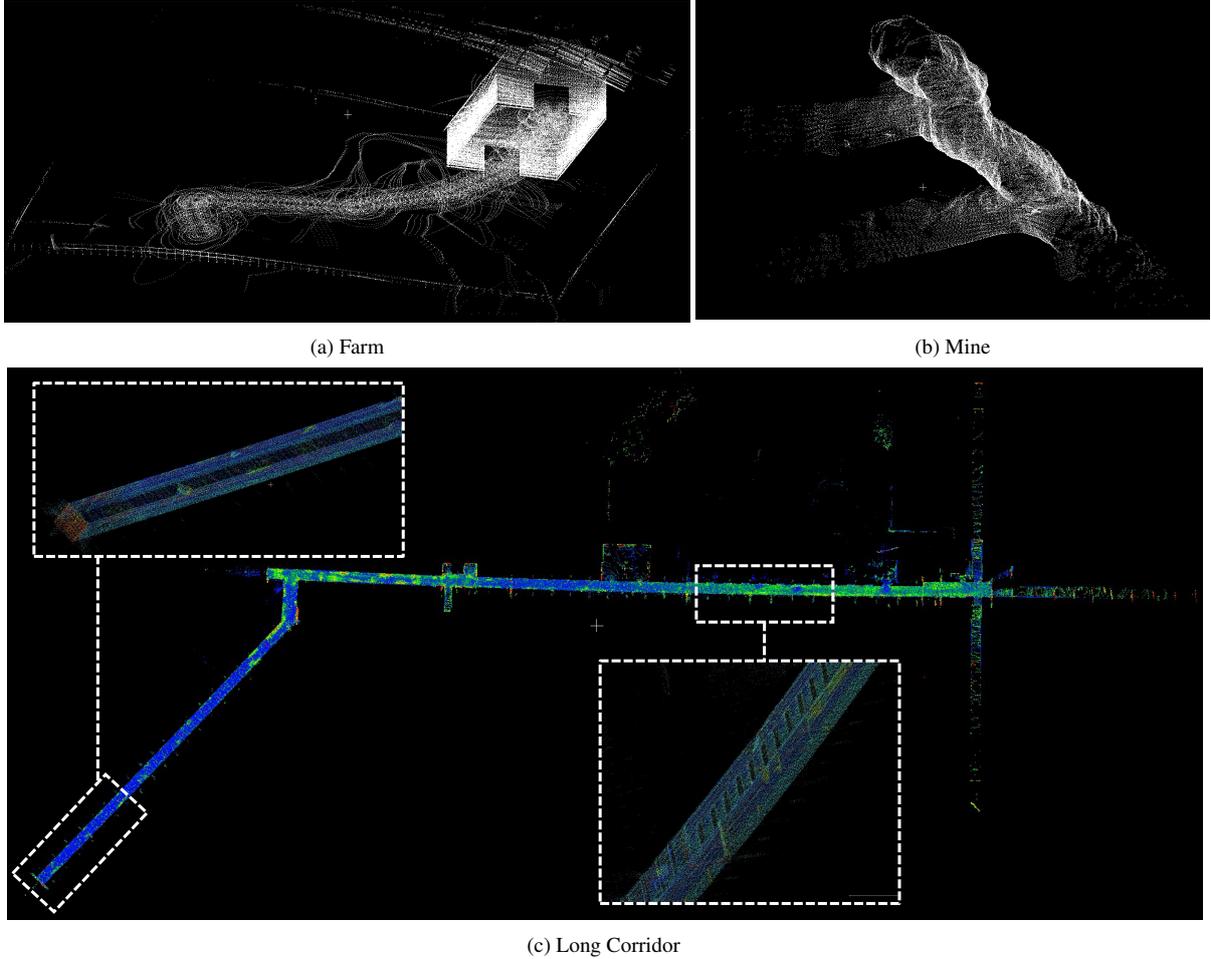


Fig.2: Examples of resulting maps using proposed method.

in structurally degenerate and visually well-conditioned situations to avoid LiDAR odometry degeneracy.

### 3. Experiments

The proposed method is evaluated using both simulated and real-world datasets. The simulated datasets are referred to as Fast Rotate, Plane, Farm, and Mine, which is simulated with ROS Gazebo simulator and sensor suite containing a Velodyne VLP-16 with 10 Hz,  $640 \times 480$  pinhole-based camera with 60 Hz, 9-axis IMU

with 200 Hz. The real-world datasets are referred to as Handheld [8], Multi Floor, and Long Corridor [15]. Compared methods are state-of-the-art LiDAR [1,2], visual [6], LiDAR-visual [8–11] SLAM. The benchmarking results are shown in Table 1. Moreover, The resulting maps using the proposed method are shown in Fig. 2. Overall, the proposed method achieved the lowest root mean square error (RMSE) compared to other LiDAR-visual SLAM methods, except for the Handheld dataset, where our method remains competitive with the best results among them.

In the Fast Rotate and Farm dataset, visual SLAM experienced failure due to aggressive motion, leading to degradation in the performance of the compared LiDAR-visual SLAM methods. In contrast, our method remained robust in visually degenerate scenes, thanks to the active factor graph and degeneracy-aware optimization.

In the Plane, Farm, Mine, and Handheld datasets, LiDAR SLAM experienced structural degeneration due to vast plane or corridor-like structures. Our method is robust in such scenes by using VIO results as the initial guess and fusing LiDAR and visual measurements in the active factor graph. Note that although LVI-SAM outperformed the proposed method in the Handheld dataset, the degeneration regions in the Handheld dataset are relatively short. If the degeneration of LiDAR SLAM is prolonged, as seen in the Plane and Farm datasets, LVI-SAM can also experience degeneration.

In the Multi Floor and Long Corridor datasets, both LiDAR and visual SLAM experienced degeneration simultaneously due to high velocity and corridor-like structures in the Long Corridor dataset, and texture-less white walls and stairs in the Multi Floor dataset. Our method remained robust even in both sensors' degenerate situations by using degeneracy-aware optimization, which conducts feature-to-map matching only along the directions of well-conditioned DOFs, while the degenerate DOFs are determined using the initial guess derived from IMU preintegration.

#### 4. Conclusion

In this paper, we propose a novel LiDAR-visual-inertial SLAM system designed to be robust and accurate in structurally and visually degenerate environments. To tackle the limitations of MAP-based and hard switching-based sensor fusion, the active factor graph, wherein factors are selectively connected based on degeneracy or failure detection, is proposed. Furthermore, to enhance optimization stability in structurally degenerate environments, we propose degeneracy-aware optimization. In our degeneracy-aware optimization, feature-to-map matching is performed only along well-conditioned directions and selectively fuse visual residuals to solve structural degeneracy. Our method was thoroughly tested across various environments, including those with both structural and visual challenges. Through experimental evaluations, our approach demonstrates a high level of robustness and accuracy, surpassing state-of-the-art LiDAR visual SLAM methods.

As a future work, for enhancement of the VIO module, we will integrate sparse depth of LiDAR into the VIO module to achieve fast initialization and enhanced accuracy, which is also switchable.

#### 5. Acknowledgement

This research was conducted under the university-corporate collaboration agreement between Kubota Corporation and the University of Tokyo.

#### References

- [1] Ji Zhang and Sanjiv Singh : "LOAM: Lidar Odometry and Mapping in Real-time" Proceedings of the Robotics: Science and Systems, vol. 2, no. 9, pp. 1–9, 2014.
- [2] Tixiao Shan, Brendan Englot, Drew Meyers, Wei Wang, Carlo Ratti, and Daniela Rus : "LIO-SAM: Tightly-coupled Lidar Inertial Odometry via Smoothing and Mapping" Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 5135–5142, 2020.
- [3] Wei Xu, Yixi Cai, Dongjiao He, Jiarong Lin, and Fu Zhang : "FAST-LIO2: Fast Direct LiDAR-inertial Odometry" IEEE Transactions on Robotics, vol. 38, no. 4, pp. 2053–2073, 2022.
- [4] Raul Mur-Artal, Jose Maria Martinez Montiel, and Juan D Tardos : "ORB-SLAM: A Versatile and Accurate Monocular SLAM System" IEEE Transactions on Robotics, vol. 31, no. 5, pp. 1147–1163, 2015.
- [5] Jakob Engel, Vladlen Koltun, and Daniel Cremers : "Direct Sparse Odometry" IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 3, pp. 611–625, 2017.
- [6] Tong Qin, Peiliang Li, Shaojie Shen : "VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator" IEEE Transactions on Robotics, vol. 34, no. 4, pp. 1004–1020, 2018.
- [7] Ji Zhang and Sanjiv Singh : "Visual-Lidar Odometry and Mapping: Low-Drift, Robust, and Fast" Proceedings of the 2015 IEEE International Conference on Robotics and Automation (ICRA), pp. 2174–2181, 2015.
- [8] Tixiao Shan, Brendan Englot, Carlo Ratti, and Daniela Rus : "LVI-SAM: Tightly-coupled Lidar-Visual-Inertial Odometry via Smoothing and Mapping" Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA), pp. 5692–5698, 2021.
- [9] Jiarong Lin, Chunran Zheng, Wei Xu, and Fu Zhang : "R2LIVE: A Robust, Real-time, LiDAR-Inertial-Visual tightly-coupled state Estimator and mapping" IEEE Robotics and Automation Letters, vol. 6, no. 4, pp. 7469–7476, 2021.
- [10] Jiarong Lin and Fu Zhang : "R3LIVE: A Robust, Real-time, RGB-colored, LiDAR-Inertial-Visual tightly-coupled state Estimation and mapping package" Proceedings of the 2022 International Conference on Robotics and Automation (ICRA), pp. 10672–10678, 2022.
- [11] Chunran Zheng, Qingyan Zhu, Wei Xu, Xiyuan Liu, Qizhi Guo, and Fu Zhang : "FAST-LIVO: Fast and Tightly-coupled Sparse-Direct LiDAR-Inertial-Visual Odometry" Proceedings of the 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4003–4009, 2022.
- [12] Shehryar Khattak, Huan Nguyen, Frank Mascarich, Tung Dang, and Kostas Alexis : "Complementary Multi-Modal Sensor Fusion for Resilient Robot Pose Estimation in Subterranean Environments" Proceedings of the 2020 International Conference on Unmanned Aircraft Systems (ICUAS), pp. 1024–1029, 2020.
- [13] Christian Forster, Luca Carlone, Frank Dellaert, and Davide Scaramuzza : "On-Manifold Preintegration for Real-Time Visual-Inertial Odometry" IEEE Transactions on Robotics, vol. 33, no. 1, pp. 1–21, 2016.
- [14] Ji Zhang, Michael Kaess, and Sanjiv Singh : "On Degeneracy of Optimization-Based State Estimation" Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA), pp. 809–816, 2016.
- [15] Shibo Zhao, Yuanjun Gao *et al.* : "SubT-MRS Dataset: Pushing SLAM Towards All-weather Environments" arXiv preprint arXiv:2307.07607, 2023.