IEEE Robotics and Automation Letters (RAL) paper presented at
2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

LIVER: A Tightly Coupled LiDAR-Inertial-Visual State Estimator with High Robustness for Underground Environments

Tianci Wen[1], Yongchun Fang[1], Biao Lu[1], Xuebo Zhang[1], and Chaoquan Tang[2]

***Abstract*— In this paper, we propose a tightly coupled LiDAR-inertial-visual (LIV) state estimator termed LIVER, which achieves robust and accurate localization and mapping in underground environments. LIVER starts with an effective strategy for LIV synchronization. A robust initialization process that integrates LiDAR, vision, and IMU is realized. A tightly coupled, nonlinear optimization-based method achieves highly accurate LiDAR-inertial-visual odometry (LIVO) by fusing LiDAR, visual, and IMU information. We consider scenarios in underground environments that are unfriendly to LiDAR and cameras. A visual-IMU-assisted method enables the evaluation and handling of LiDAR degeneracy. A deep neural network is introduced to eliminate the impact of poor lighting conditions on images. We verify the performance of the proposed method by comparing it with the state-of-the-art methods through public datasets and real-world experiments, including underground mines (see our attached video at https://youtu.be/0wjXEz3K3ng). In underground mines test, tightly coupled methods without degeneracy handling lead to failure due to self-similar areas (affecting LiDAR) and poor lighting conditions (affecting vision). In these conditions, our degeneracy handling approach successfully eliminates the impact of disturbances on the system.**

I. INTRODUCTION

SLAM, as one of the most fundamental modules, remains undoubtedly at the center of robotics research. After more than thirty years of development, SLAM has become a relatively mature research field with a wide range of applications. However, existing results have focused more on urban and indoor office scenes. Related research is still very challenging in extreme conditions, such as underground environments [1], [2]. The underground environments have some unfriendly characteristics for SLAM. First, the lighting conditions in the underground environments are poor, which brings significant challenges to the visual SLAM. Secondly, there are self-similar areas in underground environments, in which LiDAR SLAM is degenerate normally. Fortunately, despite these challenges, there has been some progress in recent years. The recent *DARPA Subterranean (SubT) Challenge* has promoted the development of underground SLAM [3]. A series of loosely coupled multi-robot SLAM algorithms has been developed based on LiDAR and IMU, supplemented by visual and thermal vision. It indicates that multi-sensor fusion is a feasible solution for underground space detection. However, most works are loosely coupled methods. In contrast, tightly coupled methods have higher robustness due to the fusion of more aspects of sensor information [4].

Unfortunately, several issues exist in current tightly coupled LIV SLAM methods. Specifically, The first one is that the current tight coupled algorithms do not consider the initialization process [4]–[10]. Due to the lack of distance measurements, it is difficult to directly fuse the monocular visual structure with LiDAR and IMU measurements. Hence, the performance of the entire LIV system is limited. Another issue is sensor degeneracy. Existing works ignore camera degeneracy in poor lighting conditions or use thermal vision instead [11], [12]. The former reduces the system's performance, while the latter significantly increases the cost of the system. Regarding the degeneracy of LiDAR in the self-similar area, existing works only use LiDAR information to evaluate degeneracy. Visual-IMU-assisted LiDAR degeneracy processing is necessary.

To address the above problems, this paper proposes LIVER, a tightly coupled LiDAR-inertial-visual state estimator with high robustness for underground environments. The main contributions of the method are as follows:

1. An online procedure fusing LiDAR, visual, and IMU information initializes IMU intrinsic parameters and visual metric scale, which improves the accuracy of pose estimation.
2. A visual-IMU-assisted LiDAR degeneracy handling approach and an image-enhanced deep network improve the robustness of the system, which is able to provide pose estimation in extreme underground environments.
3. An extensive evaluation of the proposed method across both public datasets and real-world environments demonstrates superior robustness when compared with state-of-the-art methods.

II. RELATED WORK

Early scholarly works on underground environments estimation/odometry/SLAM focus on LiDAR mapping. Ebadi *et al.* [11] develop a multi-robot SLAM system LAMP. The local front-end includes a LiDAR front-end that computes odometry estimates, and a vision front-end for artifact detection and localization. Furthermore, Chang *et al.* [13] propose the scalable closed-loop detection module, making LAMP

*This work was supported in part by the National Natural Science Foundation of China under Grant 62233011, Grant 62203235, and Grant 62073328; and in part by the Joint Fund of Guangdong Basic and Applied Basic Research Fund under Grant 2022A1515110046.

[1]Tianci Wen, Yongchun Fang, Biao Lu, and Xuebo Zhang are with Institute of Robotics and Automatic Information System, College of Artificial Intelligence, and Tianiin Key Laboratory of Intelligent Robotics, Nankai University, Tianjin, 300353, China wentc@mail.nankai.edu.cn, fangyc@nankai.edu.cn, lubiao@mail.nankai.edu.cn, zhangxuebo@nankai.edu.cn

[2]Chaoquan Tang is with School of Mechanical and Electrical Engineering, China University of Mining and Technology, Xuzhou, Jiangsu 221116, China tangchaoquan@cumt.edu.cn

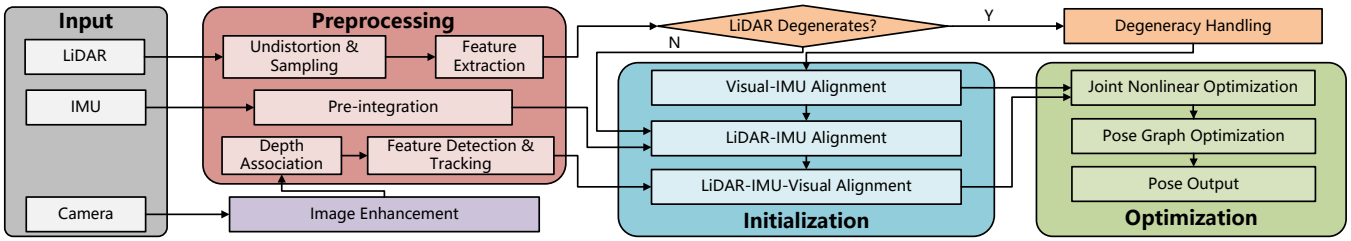

Fig. 1. Diagram above shows the full pipeline of the proposed LIVER. The system receives input from a 3D LiDAR, a monocular camera, and an IMU. The system first enhances the quality of the image in poor lighting conditions. Then, planar features and edge features of LiDAR point cloud are extracted as in LOAM [20]. Visual features are detected by the FAST corner detector and tracked by the KLT sparse optical flow algorithm. Then, the system evaluates the degeneracy of the LiDAR to decide whether to adopt LIV initialization or visual-IMU initialization. Furthermore, the visual-IMU-assisted LiDAR degeneracy handling approach processes the degeneracy of the LiDAR. Lastly, the tightly coupled LIVO is obtained by jointly optimizing the residuals of all sensors. We utilize DBoW2 [21] for visual loop detection and ScanContext++ [22] for LiDAR loop detection. A global pose graph optimization is executed by jointly adding the tightly coupled LIVO constraints, visual loop closure constraints, and LiDAR loop closure constraints to a factor graph using iSAM2 [23].

2.0 more robust in large-scale mapping. To improve the robustness of the system, visual sensors and IMU are adopted to assist LiDAR in positioning. Khattak *et al.* [12] introduce a thermal camera and utilize visual-thermal odometry as the initial estimate of the LiDAR SLAM of a single robot. Most of the existing algorithms for underground environments are LiDAR-IMU coupled SLAM, and some works are loosely coupled LIV SLAM, tightly coupled LIV methods are more desirable because of their superior robustness [4]. On the other hand, few works consider the degeneracy of LiDAR and cameras in underground environments, resulting in less robust systems.

Early efforts on LIV SLAM are loosely coupled algorithms [14]. However, researchers have discovered the limitations of loosely coupled algorithms. Some scholars have tried to use the Multi-State Constraint Kalman Filter (MSCKF) [15] to achieve tightly coupled LIV SLAM, such as LIC-Fusion [5] and LIC-Fusion 2.0 [6]. Recently, LIV tightly coupled systems FAST-LVIO [7] and R3LVIE [8] are proposed based on the error-state iterated Kalman filter. On the other hand, some algorithms based on nonlinear optimization emerge. Shan *et al.* [9] introduce LVI-SAM, where the visual-inertial system (VIS) is integrated with the LiDAR-inertial system (LIS) based on a factor graph. However, the VIS only leverages LIS estimation to facilitate initialization. Wisth *et al.* [4] use factor graphs to jointly optimize LIV information to achieve a more compact LIV system and propose a new method for LiDAR line and surface extraction. They test the performance of the algorithm in underground mines. However, sensor degeneracy is not specifically addressed. Lang *et al.* [10] propose a continuous-time LiDAR-Inertial-Camera Odometry, utilizing non-uniform B-splines to fuse measurements.

For the degeneracy of LiDAR, Zhang *et al.* [16] propose an online method to mitigate degeneracy in optimization-based problems, through analysis of the geometric structure of the problem constraints. Tuna *et al.* [17] propose a geometry-based localizability-awareness framework, X-ICP, which enables fine-grained localizability detection. In order to eliminate the dependence of the LiDAR SLAM algorithm on geometry, a neural network-based estimation approach is proposed [18], which enables early detection of failure. To make full use of the advantages of a multi-sensor system, Han *et al.* [19] propose a lightweight iEKF-based LiDAR-inertial odometry with a degeneration-aware and modular sensor-fusion pipeline.

III. OVERVIEW AND NOTATION

An overview of the proposed LIVER, which receives input from a 3D LiDAR, a monocular camera, and an IMU, is shown in Fig. 1. The system first enhances the quality of the image in poor lighting conditions. Then, measurements are preprocessed, in which visual features are extracted and tracked, LiDAR features are extracted, and IMU measurements between two consecutive frames are pre-integrated. After the LiDAR features are extracted, the system detects the degeneracy of the LiDAR. It adopts the two-stage LIV initialization strategy when the LiDAR is not degenerate. Otherwise, the system executes the visual-IMU initialization process. Lastly, the tightly coupled LIVO is obtained by jointly optimizing visual reprojection residuals, IMU pre-integration residuals, and LiDAR feature residuals. The pose and loop detection results are fed to the pose graph, which performs global optimization to eliminate the drift.

We now define the notations utilized throughout this paper. Denote $(\cdot)^w$ as the world frame and $(\cdot)^l$ as the LiDAR frame. $(\cdot)^c$ is the camera frame. We use $(\cdot)^b$ to represent the body frame, which is the same as the IMU frame. l_k is the LiDAR frame when the k-th LiDAR scan is received. b_k and c_k are the body frame and the camera frame while receiving the k-th LiDAR scan, respectively. We use both rotation matrices $\mathbf{R}$ and Hamilton quaternions $\mathbf{q}$ to represent rotation. $\boldsymbol{t}$ denotes translation vector.

IV. SYSTEM INITIALIZATION

An accurate initial guess is necessary for tightly coupled LIVO, which is a highly nonlinear system. Hence, we propose a coarse-to-fine LIV Initialization process. In Section IV.A, we first get a coarse value by loosely aligning IMU pre-integration with the LiDAR odometry. In Section IV.B, an accurate initial value is obtained by tightly aligning

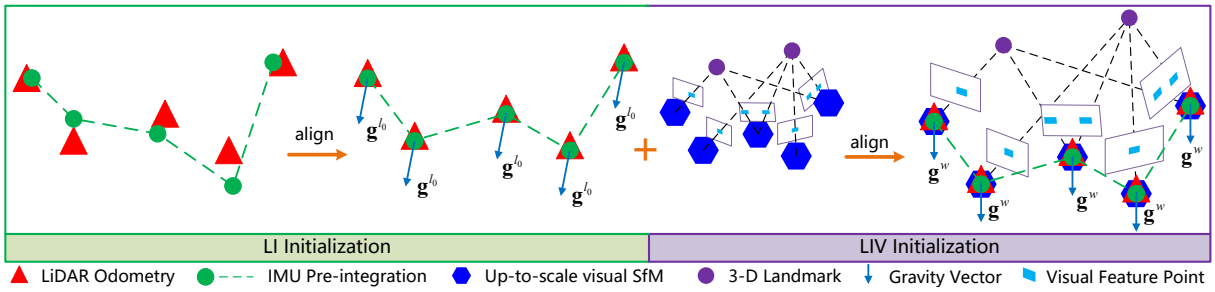

Fig. 2. Illustration of the LiDAR-inertial-visual alignment process of initialization. The process starts with LiDAR-inertial (LI) initialization to adjust the intrinsic parameters of the IMU so that the result of IMU pre-integration matches the LiDAR odometry. Then, the process is finished by LIV initialization, the basic idea of which is to match up-to-scale visual structure with LiDAR odometry and IMU pre-integration.

LiDAR odometry with IMU pre-integration and the vision-only structure.

A. LiDAR-Inertial Alignment

The left of Fig. 2 illustrates the LiDAR-inertial alignment process. The basic idea is to match the IMU pre-integration with the LiDAR odometry.

1) Acceleration Bias and Gyroscope Bias Calibration: For the l_k frame and l_{k+1} frame LiDAR point cloud, we use LOAM [20] to obtain the poses $(\mathbf{q}_{l_k}^{l_0}, \boldsymbol{t}_{l_k}^{l_0})$ and $(\mathbf{q}_{l_{k+1}}^{l_0}, \boldsymbol{t}_{b_{k+1}}^{l_0})$. Based on the extrinsic parameter $\mathbf{T}_l^b$ between the LiDAR and IMU, the pose is rotated to the IMU frame:

$$\mathbf{q}_{b_k}^{l_0} = \mathbf{q}_{l_k}^{l_0} \otimes (\mathbf{q}_l^b)^{-1}, \ \boldsymbol{t}_{b_k}^{l_0} = \boldsymbol{t}_{l_k}^{l_0} - \mathbf{R}_{b_k}^{l_0} \boldsymbol{t}_l^b \tag{1}$$

According to [24], we can obtain the translation, velocity, and rotation constraints $\boldsymbol{\alpha}_{b_{k+1}}^{b_k}$, $\boldsymbol{\beta}_{b_{k+1}}^{b_k}$, $\boldsymbol{\gamma}_{b_{k+1}}^{b_k}$ between two time consecutive frames by IMU pre-integration. Store the above results in a sliding window $\mathcal{S}$, the optimal gyroscope bias $\mathbf{b}_w$ and acceleration bias $\mathbf{b}_a$ can be obtained by minimizing the following cost function:

$$\min_{\delta b_w, \delta b_a} \sum_{k \in \mathcal{S}} \left\{ \| \mathbf{q}_{b_{k+1}}^{l_0}{}^{-1} \otimes \mathbf{q}_{b_k}^{l_0} \otimes \boldsymbol{\gamma}_{b_{k+1}}^{b_k} \|^2 + \| \boldsymbol{t}_{b_{k+1}}^{b_k} - \boldsymbol{\alpha}_{b_{k+1}}^{b_k} \|^2 \right\} \tag{2}$$

where $\otimes$ represents the multiplication operation between two quaternions, $\boldsymbol{t}_{b_{k+1}}^{b_k} = \boldsymbol{t}_{b_{k+1}}^{l_0} - \mathbf{R}_{b_{k+1}}^{b_k} \boldsymbol{t}_{b_k}^{l_0}$.

2) Velocity and Gravity Vector Initialization: The variables requiring calibration are the velocity $\mathbf{v}_{b_k}^{b_k}$ in the body frame when receiving the k-th LiDAR scan and the Gravity vector $\mathbf{g}^{l_0}$ in the l_0 frame. Consider two consecutive frames b_k and b_{k+1} in the window, we can obtain linear measurement model:

$$\begin{bmatrix} \delta \boldsymbol{\alpha}_{b_{k+1}}^{b_k} \\ \delta \boldsymbol{\beta}_{b_{k+1}}^{b_k} \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{\alpha}}_{b_{k+1}}^{b_k} - \boldsymbol{\alpha}_{b_{k+1}}^{b_k} \\ \hat{\boldsymbol{\beta}}_{b_{k+1}}^{b_k} - \boldsymbol{\beta}_{b_{k+1}}^{b_k} \end{bmatrix} = \mathbf{A}_{b_{k+1}}^{b_k} - \mathbf{B}_{b_{k+1}}^{b_k} \boldsymbol{x}_I \tag{3}$$

where

$$\begin{aligned} \mathbf{A}_{b_{k+1}}^{b_k} &= \begin{bmatrix} \hat{\boldsymbol{\alpha}}_{b_{k+1}}^{b_k} + (\mathbf{R}_{l_0}^{b_k} \mathbf{R}_{l_{k+1}}^{l_0} - 1) \boldsymbol{t}_l^b - \mathbf{R}_{l_0}^{b_k} \left(\boldsymbol{t}_{l_{k+1}}^{l_0} - \boldsymbol{t}_{l_k}^{l_0} \right) \\ \hat{\boldsymbol{\beta}}_{b_{k+1}}^{b_k} \end{bmatrix} \\ \mathbf{B}_{b_{k+1}}^{b_k} &= \begin{bmatrix} -\mathbf{I} \Delta \mathcal{T}_k & \mathbf{0} & \frac{1}{2} \mathbf{R}_{l_0}^{b_k} \Delta \mathcal{T}_k^2 \\ -\mathbf{I} & \mathbf{R}_{l_0}^{b_k} \mathbf{R}_{b_{k+1}}^{l_0} & \mathbf{R}_{l_0}^{b_k} \Delta \mathcal{T}_k \end{bmatrix} \\ \boldsymbol{x}_I &= \left[\mathbf{v}_{b_0}^{b_0}, \mathbf{v}_{b_1}^{b_1}, \dots \mathbf{v}_{b_{k+1}}^{b_{k+1}}, \mathbf{g}^{l_0} \right] \end{aligned} \tag{4}$$

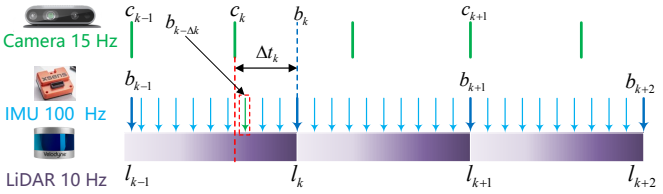

Fig. 3. LiDAR-inertial-visual timestamp synchronization. When receiving the k-th frame of LiDAR scan l_k, we select the first IMU measurement b_k and image c_k before l_k. The first IMU measurement after c_k is defined as $b_{k\ \Delta k}$, Δt_k is the time difference between b_k and c_k.

Finally, we can obtain the optimal $\mathbf{v}_{b_k}^{b_k}$ and $\mathbf{g}^{l_0}$ by solving the following linear least squares problem:

$$\min_{\boldsymbol{x}_I} \sum_{k \in \mathcal{S}} \left\| \mathbf{A}_{b_{k+1}}^{b_k} - \mathbf{B}_{b_{k+1}}^{b_k} \boldsymbol{x}_I \right\|^2 \tag{5}$$

B. LiDAR-Inertial-Visual Alignment

The right of Fig. 2 illustrates the LIV alignment process. The basic idea is to match up-to-scale visual structure with LiDAR odometry and IMU pre-integration.

1) LiDAR-Inertial-Visual Timestamp Synchronization: One of the difficulties in LIV fusion is synchronizing three different sensor data. As shown in Fig. 3, when receiving the k-th frame of LiDAR scan l_k, we select the first IMU measurement b_k and image c_k before l_k. Since the frequency of the LiDAR does not match the camera's frequency, the time difference between c_k and b_k cannot be ignored, as shown in Fig. 3. In this case, we use the IMU measurements between the image c_k and the IMU measurement b_k to obtain the pose of c_k. Specifically, the first IMU measurement after c_k is defined as $b_{k-\Delta k}$, and Δt_k is the time difference between the IMU measurement b_k and the image c_k. The relative transform relationship between $b_{k-\Delta k}$ and c_k is obtained through extrinsic parameter rotation, as $\boldsymbol{t}^{c_k} = \mathbf{T}_b^c \boldsymbol{t}^{b_{k-\Delta k}}$. Finally, the pose $(\mathcal{R}_{b_{k-\Delta k}}^{l_0}, \boldsymbol{t}_{b_{k-\Delta k}}^{l_0})$ of the body frame b_k in the l_0 frame is expressed as:

$$\begin{aligned} \mathcal{R}_{b_{k-\Delta k}}^{l_0} &= \mathbf{R}_{l_k}^{l_0} \mathbf{R}_b^l (\boldsymbol{\gamma}_{b_k}^{b_{k-\Delta k}})^{-1} \\ \boldsymbol{t}_{b_{k-\Delta k}}^{l_0} &= -\mathbf{R}_{l_k}^{l_0} \mathbf{R}_b^l \left((\boldsymbol{\gamma}_{b_k}^{b_{k-\Delta k}})^{-1} \boldsymbol{\alpha}_{b_k}^{b_{k-\Delta k}} + \boldsymbol{t}_b^l \right) + \boldsymbol{t}_{l_k}^{l_0} \end{aligned} \tag{6}$$

where $\boldsymbol{\gamma}_{b_k}^{b_{k-\Delta k}}$ and $\boldsymbol{\alpha}_{b_k}^{b_{k-\Delta k}}$ are the IMU pre-integration results between $b_{k-\Delta k}$ and b_k.

2) Metric Scale Initialization: The graph of up-to-scale camera poses $(\mathbf{q}_{c_k}^{c_0}, \boldsymbol{t}_{c_k}^{c_0})$ can be estimated by a vision-only structure-from-motion (SfM) in [24], as shown in the right of Fig. 2. Given the extrinsic parameter $(\mathbf{R}_c^b, \boldsymbol{t}_c^b)$ between the camera and the IMU, the camera poses can be translated from the camera frame to body frame as

$$\begin{aligned}\mathbf{R}_{b_{k-\Delta k}}^{l_0} &= \mathbf{R}_{c_0}^{l_0}\mathbf{R}_{c_k}^{c_0}(\mathbf{q}_c^b)^{-1} \\ \lambda\bar{\boldsymbol{t}}_{b_{k-\Delta k}}^{l_0} &= \mathbf{R}_{c_0}^{l_0}\left(\lambda\bar{\boldsymbol{t}}_{c_k}^{c_0} - \mathbf{R}_{b_{k-\Delta k}}^{c_0}\boldsymbol{t}_c^b\right) + \boldsymbol{t}_{c_0}^{l_0}\end{aligned} \tag{7}$$

where $\mathbf{R}_{c_0}^{l_0} = \mathbf{R}_b^l(\gamma_{b_{0-\Delta 0}}^{b_0})^{-1}\mathbf{R}_c^b$, λ is the metric scale of the translation to be initialized. $\mathcal{R}_{b_{k-\Delta k}}^{l_0}$ and $\mathbf{R}_{b_{k-\Delta k}}^{l_0}$ represent the same relative rotation. $\mathcal{R}_{b_{k-\Delta k}}^{l_0}$ is solved by the LiDAR odometry and IMU pre-integration results in (6), while $\mathbf{R}_{b_{k-\Delta k}}^{l_0}$ is obtained by vision-only SfM and the extrinsic parameters $\mathbf{R}_c^b$ in (7). Furthermore, to match the relative camera rotations $\mathcal{R}_{b_{k-\Delta k}}^{l_0}$ and $\mathbf{R}_{b_{k-\Delta k}}^{l_0}$, we calibrate the gyroscope bias b_w by solving the following optimization problem:

$$\min_{\delta b_w}\sum_{k\in\mathcal{I}}\left\|{\mathcal{R}_{b_{k-\Delta k}}^{l_0}}^{-1}\mathbf{R}_{b_{k-\Delta k}}^{l_0}\right\|^2 \tag{8}$$

where $\mathcal{I}$ indexes all frames in a sliding window, in which we maintain several frames to ensure accuracy. The variable to be calibrated is defined as

$$\boldsymbol{x}_c = [\mathbf{v}_{b_{0-\Delta 0}}^{b_{0-\Delta 0}}, \mathbf{v}_{b_0}^{b_0}, \mathbf{v}_{b_{1-\Delta 1}}^{b_{1-\Delta 1}}, \mathbf{v}_{b_1}^{b_1}, \ldots \mathbf{v}_{b_{k-\Delta k}}^{b_{k-\Delta k}}, \mathbf{v}_{b_k}^{b_k}, \mathbf{g}^{l_0}, \lambda] \tag{9}$$

Based on the results in (6)-(7), the IMU pre-integration terms for the two frames $b_{k-\Delta k}$ and b_k are derived as follows:

$$\begin{aligned}\boldsymbol{\alpha}_{b_k}^{b_{k-\Delta k}} =& \mathcal{R}_{l_0}^{b_{k-\Delta k}}(\boldsymbol{t}_{b_k}^{l_0} - \lambda\bar{\boldsymbol{t}}_{b_{k-\Delta k}}^{l_0} + \frac{1}{2}\mathbf{g}^{l_0}\Delta t_k^2 \\ & - \mathcal{R}_{b_{k-\Delta k}}^{l_0}\mathbf{v}_{b_{k-\Delta k}}^{b_{k-\Delta k}}\Delta t_k) \\ \boldsymbol{\beta}_{b_k}^{b_{k-\Delta k}} =& \mathcal{R}_{l_0}^{b_{k-\Delta k}}(\mathcal{R}_{b_k}^{l_0}\mathbf{v}_{b_k}^{b_k} + \mathbf{g}^{l_0}\Delta t_k - \mathbf{R}_{b_{k-\Delta k}}^{l_0}\mathbf{v}_{b_{k-\Delta k}}^{b_{k-\Delta k}})\end{aligned} \tag{10}$$

Furthermore, by combining (6)-(7) with (10), the error of IMU pre-integration can be calculated as follows:

$$\begin{bmatrix}\delta\boldsymbol{\alpha}_{b_k}^{b_{k-\Delta k}} \\ \delta\boldsymbol{\beta}_{b_k}^{b_{k-\Delta k}}\end{bmatrix} = \begin{bmatrix}\hat{\boldsymbol{\alpha}}_{b_k}^{b_{k-\Delta k}} - \boldsymbol{\alpha}_{b_k}^{b_{k-\Delta k}} \\ \hat{\boldsymbol{\beta}}_{b_k}^{b_{k-\Delta k}} - \boldsymbol{\beta}_{b_k}^{b_{k-\Delta k}}\end{bmatrix} = \mathcal{A}_{b_k}^{b_{k-\Delta k}} - \mathcal{B}_{b_k}^{b_{k-\Delta k}}\boldsymbol{x}_c \tag{11}$$

where

$$\mathcal{A}_{b_k}^{b_{k-\Delta k}} = \begin{bmatrix}\hat{\boldsymbol{\alpha}}_{b_k}^{b_{k-\Delta k}} + \mathcal{R}_{l_0}^{b_{k-\Delta k}}(\boldsymbol{t}_{c_0}^{l_0} - \boldsymbol{t}_{b_k}^{l_0}) - \boldsymbol{t}_c^b \\ \hat{\boldsymbol{\beta}}_{b_k}^{b_{k-\Delta k}}\end{bmatrix},\ \mathcal{B}_{b_k}^{b_{k-\Delta k}} = \begin{bmatrix}-\mathbf{I}\Delta t_k & \mathbf{0} & \frac{1}{2}\mathcal{R}_{l_0}^{b_{k-\Delta k}}\Delta t_k^2 & -\mathcal{R}_{l_0}^{b_{k-\Delta k}}\mathbf{R}_{c_0}^{l_0}\bar{\boldsymbol{t}}_{c_k}^{c_0} \\ -\mathcal{R}_{l_0}^{b_{k-\Delta k}}\mathbf{R}_{b_{k-\Delta k}}^{l_0} & \gamma_{b_k}^{b_{k-\Delta k}} & \mathcal{R}_{l_0}^{b_{k-\Delta k}}\Delta t_k & \mathbf{0}\end{bmatrix} \tag{12}$$

Combining (9)-(11) and (12), the optimal $\boldsymbol{x}_c$ is obtained by solving the following linear least squares problem:[1]

$$\min_{\boldsymbol{x}_c}\sum_{k\in\mathcal{I}}\left\|\mathcal{A}_{b_k}^{b_{k-\Delta k}} - \mathcal{B}_{b_k}^{b_{k-\Delta k}}\boldsymbol{x}_c\right\|^2 \tag{13}$$

Finally, the up-to-scale visual structure matches the LiDAR odometry and IMU pre-integration, as shown in the Fig. 2.

[1]We utilize the gravity vector optimization method in the paper [24] to convert the pose and velocity estimated in LiDAR frame $(\cdot)^{l_0}$ to the world frame $(\cdot)^w$.

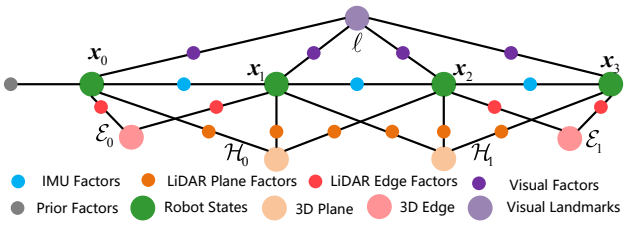

Fig. 4. Factor graph representation of the tightly coupled optimization in our system.

V. TIGHTLY COUPLED LiDAR-INERTIAL-VISUAL ODOMETRY

After initialization, a tightly coupled LIVO is proposed by jointly optimizing visual, LiDAR, and inertial residuals, which fits with the factor graph representation as shown in Fig. 4.

The system state to be estimated is defined as:

$$\boldsymbol{x}_k \triangleq \left[\boldsymbol{t}_{b_k}^w, \mathbf{v}_{b_k}^w, \mathbf{q}_{b_k}^w, \mathbf{b}_a, \mathbf{b}_g, s_1, s_2, \cdots, s_n\right] \tag{14}$$

where s_n is the inverse distance of the n_{th} feature from its first observation. We define the optimal system state $\boldsymbol{x}_k^\star$ as one that maximizes the likelihood of the measurements $\hat{\mathbf{z}}_k$:

$$\begin{aligned}\boldsymbol{x}_k^\star =& \arg\max_{\boldsymbol{x}_k} p(\boldsymbol{x}_k \mid \hat{\mathbf{z}}_k) \propto \arg\max_{\boldsymbol{x}_k} p(\boldsymbol{x}_0)p(\hat{\mathbf{z}}_{k,\mathcal{I}} \mid \boldsymbol{x}_k)\cdot \\ & p(\hat{\mathbf{z}}_{k,\mathcal{C}} \mid \boldsymbol{x}_k)p(\hat{\mathbf{z}}_{k,\mathcal{E}} \mid \boldsymbol{x}_k)p(\hat{\mathbf{z}}_{k,\mathcal{H}} \mid \boldsymbol{x}_k)\end{aligned} \tag{15}$$

where $\mathcal{C}$ is a set of current visual feature points observed at least twice, $\mathcal{E}$ and $\mathcal{H}$ are the sets of current LiDAR planar features and edge features, respectively. All measurements are assumed to be conditionally independent and the noise is zero-mean Gaussian distributed. Therefore, the Maximum-a-Posteriori Estimation of state $\boldsymbol{x}_k$ is obtained by formulating (15) as the following least squares minimization problem:

$$\begin{aligned}\min_{\boldsymbol{x}_k}\Bigg\{ & \left\|\mathbf{r}_{\mathcal{I}}(\hat{\mathbf{z}}_{b_k}^{b_{k-1}}, \boldsymbol{x}_k)\right\|_{\mathbf{P}_{b_k}^{b_{k-1}}}^2 + \sum_{\ell\in\mathcal{C}}\rho(\|\mathbf{r}_{\mathcal{C}}(\hat{\mathbf{z}}_\ell^{c_k}, \boldsymbol{x}_k)\|_{\mathbf{P}_\ell^{c_k}}^2) + \\ & \sum_{m\in\mathcal{E}}\left\|\mathbf{r}_{\mathcal{E}}(\hat{\mathbf{z}}_m^{l_k}, \boldsymbol{x}_k)\right\|_{\mathbf{P}_m^{l_k}}^2 + \sum_{n\in\mathcal{H}}\left\|\mathbf{r}_{\mathcal{H}}(\hat{\mathbf{z}}_n^{l_k}, \boldsymbol{x}_k)\right\|_{\mathbf{P}_n^{l_k}}^2 + \|\mathbf{r}_0\|^2\Bigg\}\end{aligned} \tag{16}$$

where $\mathbf{r}_{\mathcal{I}}(\hat{\mathbf{z}}_{b_k}^{b_{k-1}}, \boldsymbol{x}_k)$, $\mathbf{r}_{\mathcal{C}}(\hat{\mathbf{z}}_\ell^{c_k}, \boldsymbol{x}_k)$, $\mathbf{r}_{\mathcal{E}}(\hat{\mathbf{z}}_m^{l_k}, \boldsymbol{x}_k)$, $\mathbf{r}_{\mathcal{H}}(\hat{\mathbf{z}}_n^{l_k}, \boldsymbol{x}_k)$, and $\mathbf{r}_0$ are residuals for IMU, visual measurements, LiDAR edge features and planar features, and state prior, respectively. $\|\cdot\|_{\mathbf{P}}$ is the Mahalanobis norm, and ρ is the Huber norm.

We utilize a sliding window to store image keyframes. The features in $\mathcal{C}$ are observed in the current image and the sliding window. For the the ℓ-th feature that is first observed in the i-th image and then seen again in the k-th image, the residual is defined as

$$\begin{aligned}\mathbf{r}_{\mathcal{C}}(\mathcal{X}) =& \hat{\bar{\mathcal{P}}}_\ell^{c_k} - \frac{\mathcal{P}_\ell^{c_k}}{\|\mathcal{P}_\ell^{c_k}\|} \\ \mathcal{P}_\ell^{c_k} =& \mathbf{R}_b^c(\gamma_{b_k}^{b_{k-\Delta k}}\mathbf{R}_w^{b_k}(\mathbf{R}_{b_i}^w(\gamma_{b_i}^{b_{i-\Delta i}})^{-1}(\mathbf{R}_c^b\frac{1}{\lambda_\ell}\hat{\bar{\mathcal{P}}}_\ell^{c_i} \\ & + \boldsymbol{t}_c^b - \boldsymbol{\alpha}_{b_i}^{b_{i-\Delta i}}) + \boldsymbol{t}_{b_i}^w - \boldsymbol{t}_{b_k}^w) + \boldsymbol{\alpha}_{b_k}^{b_{k-\Delta k}} - \boldsymbol{t}_c^b)\end{aligned} \tag{17}$$

where $\hat{\overline{\mathcal{P}}}_\ell^{c_i} = {\pi_c}^{-1}([\hat{u}_\ell^{c_i}\ \hat{v}_\ell^{c_i}\ 1]^T)$ is the first observation of the ℓ-th in the i-th image, and $\hat{\overline{\mathcal{P}}}_\ell^{c_k} = {\pi_c}^{-1}([\hat{u}_\ell^{c_k}\ \hat{v}_\ell^{c_k}\ 1]^T)$ is the observation of the same feature in the k-th image. π_c represents the camera projection function. $(\mathbf{R}_{b_i}^w, \boldsymbol{t}_{b_i}^w)$ is the output of the tightly coupled LIVO at frame i, and $(\mathbf{R}_{b_k}^w, \boldsymbol{t}_{b_k}^w)$ is the pose to be estimated at frame k.

VI. SENSOR DEGRADATION

In extreme underground environments, LiDAR and cameras will degenerate differently. In this section, we analyze the degenerate cases of LiDAR and cameras and illustrate our solutions to improve the system's robustness.

A. Self-similar Structure

1) State Definition: We define the system state $\boldsymbol{x}_k$ as a linear combination of degenerate degrees of freedom (DOF) $\boldsymbol{x}_k^d$ and non-degenerate DOF $\boldsymbol{x}_k^u$:

$$\begin{aligned}\boldsymbol{x}_k &\triangleq \boldsymbol{x}_k^d + \boldsymbol{x}_k^u = \left[\boldsymbol{t}_{b_k}^w, \mathbf{v}_{b_k}^w, \mathbf{q}_{b_k}^w, \mathbf{b}_a, \mathbf{b}_g, s_1, s_2, \cdots, s_n\right] \\ \boldsymbol{x}_k^u &= \boldsymbol{x}_k - \boldsymbol{x}_k^d = \left[\boldsymbol{t}_{b_k}^{w,u}, \mathbf{v}_{b_k}^w, \mathbf{q}_{b_k}^{w,u}, \mathbf{b}_a, \mathbf{b}_g, s_1, s_2, \cdots, s_n\right]\end{aligned} \tag{18}$$

where $\boldsymbol{x}_k^d$ and $\boldsymbol{x}_k^u$ are two independent variables. Define the degenerate state variable $\boldsymbol{S}_k$ as follows:

$$\boldsymbol{S}_k = [s_{t_x}^k, s_{t_y}^k, s_{t_z}^k, \mathbf{0}_{1\times3}, s_{q_w}^k, s_{q_x}^k, s_{q_y}^k, s_{q_z}^k, \mathbf{0}_{1\times3}, \mathbf{0}_{1\times3}, \mathbf{0}_{1\times n}] \tag{19}$$

where $s_{t_x}^k, s_{t_y}^k, s_{t_z}^k, s_{q_w}^k, s_{q_x}^k, s_{q_y}^k, s_{q_z}^k$ are all binary numbers. For the degenerate DOF, the correspondence degenerate state variable s^k is 1, otherwise it is 0. Hence, the degenerate DOF $\boldsymbol{x}_k^d$ can be obtained by the following formula:

$$\boldsymbol{x}_k^d = \boldsymbol{x}_k \circ \boldsymbol{S}_k \tag{20}$$

where $\circ$ represents the Hadamard Product (i.e., the element-wise product) between two vectors. The pose transform $\boldsymbol{x}_{b_k}^{b_{k-1}}$ between b_k frame and b_{k-1} frame is defined as $\boldsymbol{x}_{b_k}^{b_{k-1}} = [\boldsymbol{t}_{b_k}^{b_{k-1}}, \mathbf{0}_{1\times3}, \mathbf{q}_{b_k}^{b_{k-1}}, \mathbf{0}_{1\times3}, \mathbf{0}_{1\times3}, \mathbf{0}_{1\times n}]$, which is obtained by $\boldsymbol{x}_k$ and $\boldsymbol{x}_{k-1}$ as

$$\mathbf{q}_{b_k}^{b_{k-1}} = (\mathbf{q}_{b_{k-1}}^w)^{-1} \otimes \mathbf{q}_{b_k}^w,\ \boldsymbol{t}_{b_k}^{b_{k-1}} = \mathbf{R}_w^{b_{k-1}}(\boldsymbol{t}_{b_k}^w - \boldsymbol{t}_{b_{k-1}}^w) \tag{21}$$

2) Analysis: For a point i in the k-th LiDAR measurements, its coordinate is denoted as $\mathcal{X}_{k,i}^{l_k}$. The point i is projected to the l_{k-1} frame as follows:

$$\mathcal{X}_{k,i}^{l_{k-1}} = \mathbf{R}_b^l(\mathbf{R}_w^{b_{k-1}}(\mathbf{R}_{b_k}^w \mathcal{X}_{k,i}^{b_k} + \boldsymbol{t}_{b_k}^w - \boldsymbol{t}_{b_{k-1}}^w) - \boldsymbol{t}_l^b) \tag{22}$$

where $\mathcal{X}_{k,i}^{b_k} = \mathbf{R}_l^b \mathcal{X}_{k,i}^{l_k} + \boldsymbol{t}_l^b$. $(\mathbf{R}_{b_{k-1}}^w, \boldsymbol{t}_{b_{k-1}}^w)$ is the output of LIVER at b_{k-1} frame, and $(\mathbf{R}_{b_k}^w, \boldsymbol{t}_{b_k}^w)$ is the pose to be estimated at b_k frame. Furthermore, the residuals for LiDAR features represent point to line distance and point to plane distance, which are expressed as follows:

$$\begin{aligned}\mathbf{r}_{\mathcal{H}}\left(\hat{\mathbf{z}}_m^{l_k}, \boldsymbol{x}_k\right) &= \left|(\mathcal{X}_{k,i}^{l_{k-1}} - \mathcal{X}_{k-1,j}^{l_{k-1}}) \cdot \frac{\mathcal{X}_{jl} \times \mathcal{X}_{jm}}{\|\mathcal{X}_{jl} \times \mathcal{X}_{jm}\|}\right| \\ \mathbf{r}_{\mathcal{E}}\left(\hat{\mathbf{z}}_m^{l_k}, \boldsymbol{x}_k\right) &= \frac{\left|\left(\mathcal{X}_{k,i}^{l_{k-1}} - \mathcal{X}_{k-1,j}^{l_{k-1}}\right) \times \left(\mathcal{X}_{k,i}^{l_{k-1}} - \mathcal{X}_{k-1,l}^{l_{k-1}}\right)\right|}{|\mathcal{X}_{jl}|} \\ \mathcal{X}_{jl} &= \mathcal{X}_{k-1,j}^{l_{k-1}} - \mathcal{X}_{k-1,l}^{l_{k-1}}, \mathcal{X}_{jm} = \mathcal{X}_{k-1,j}^{l_{k-1}} - \mathcal{X}_{k-1,m}^{l_{k-1}}\end{aligned} \tag{23}$$

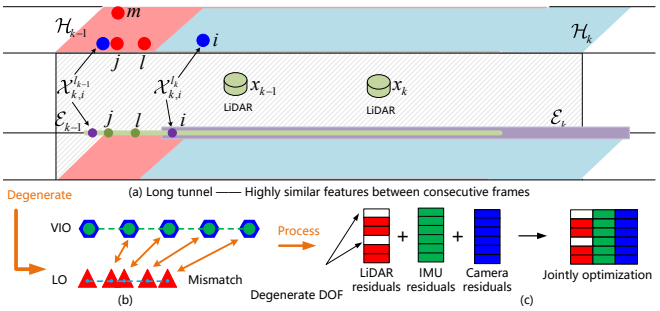

Fig. 5. Illustration of LiDAR degeneracy in the long tunnel of underground environments. (a) The long tunnel is the presence of self-similar structures, in which the LiDAR odometry system is degenerate. (b) Hence, there is a mismatch between the LiDAR odometry and visual-inertial odometry. (c) The LiDAR measurement residuals participating in the tightly coupled optimization do not contain degenerate DOF $\boldsymbol{x}_k^d$.

where $\mathcal{X}_{k-1,j}^{l_{k-1}}$, $\mathcal{X}_{k-1,l}^{l_{k-1}}$, and $\mathcal{X}_{k-1,m}^{l_{k-1}}$ are the coordinates of points j, l, and m, respectively. As shown in Fig. 5(a), for planar residual, i is a point in planar points sets $\mathcal{H}_k$ at l_k frame. j, l, m are all planar points in $\mathcal{H}_{k-1}$ at l_{k-1} frame. For edge residual, i is a point in edge points sets $\mathcal{E}_k$ at l_k frame. j and l are edge points in $\mathcal{E}_{k-1}$ at l_{k-1} frame. For LiDAR slam, $(\mathbf{R}_{b_k}^w, \boldsymbol{t}_{b_k}^w)$ is obtained by solving the following nonlinear optimization problem:

$$\min_{\boldsymbol{x}_k}\left\{\sum_{m\in\mathcal{E}}\left\|\mathbf{r}_{\mathcal{E}}\left(\hat{\mathbf{z}}_m^{l_k}, \boldsymbol{x}_k\right)\right\|_{\mathbf{P}_m^{l_k}}^2 + \sum_{n\in\mathcal{H}}\left\|\mathbf{r}_{\mathcal{H}}\left(\hat{\mathbf{z}}_n^{l_k}, \boldsymbol{x}_k\right)\right\|_{\mathbf{P}_n^{l_k}}^2\right\} \tag{24}$$

However, as shown in Fig. 5(a), when a robot moves in a long tunnel, the projected point i and the points j, l, m are on the same long plane, which leads to $\mathbf{r}_{\mathcal{H}}(\hat{\mathbf{z}}_n^{l_k}, \boldsymbol{x}_k)$, $\mathbf{r}_{\mathcal{E}}(\hat{\mathbf{z}}_m^{l_k}, \boldsymbol{x}_k) \rightarrow \mathbf{0}$. Hence, for the optimization problem of (24), it is impossible to obtain the optimal solution for the degenerate DOF $\boldsymbol{x}_k^{d*}$, which leads to the unstable estimation of degenerate DOF as shown in Fig. 5(b). The constraints of the degenerate DOF do not exist. Finally, the residuals for LiDAR features with respect to $\boldsymbol{x}_k^d$ destroy the optimization results of (16). Hence, we introduce visual-IMU information to solve this problem.

3) Solution: The solution is Visual-IMU-assisted LiDAR degeneracy handling method. Specifically, when $|\mathbf{r}_{\mathcal{H}}(\hat{\mathbf{z}}_n^{l_k}, \boldsymbol{x}_k)| + |\mathbf{r}_{\mathcal{E}}(\hat{\mathbf{z}}_m^{l_k}, \boldsymbol{x}_k)|$ is less than the threshold $\mathbf{r}_{th}$, we consider that there are self-similar areas in the current environment, which may lead to LiDAR degeneracy. Hence, visual IMU information is introduced to determine whether LiDAR is degenerate, the LiDAR odometry (LO) $\boldsymbol{x}_k^L$ and visual-inertial odometry (VIO) $\boldsymbol{x}_k^{VI}$ are solved in two threads. The pose transforms $\boldsymbol{x}_{b_k}^{b_{k-1},L}$ and $\boldsymbol{x}_{b_k}^{b_{k-1},VI}$ of LO and VIO are obtained by (21). We identify the DOF whose difference $|\boldsymbol{x}_{b_k}^{b_{k-1},L}(i) - \boldsymbol{x}_{b_k}^{b_{k-1},VI}(i)|$ between LO and VIO exceeds the threshold $\mathbf{x}_{th}$ as the degenerate DOF, the corresponding degenerate state variable $\boldsymbol{S}_k(i)$ is set to 1, which is the mismatch between LO and VIO as shown in Fig. 5(b). Hence, the degenerate state variable $\boldsymbol{S}_k$ is determined. Finally, $\boldsymbol{x}_k^d$ and $\boldsymbol{x}_k^u$ can be derived by (18)-(20).

To eliminate the effect of LiDAR degeneracy, we project

point i to l_{k-1} frame only by non-degenerate DOF $\boldsymbol{x}_{k,u}$:

$$\mathcal{X}_{k,i}^{l_{k-1}} = \mathbf{R}_b^l(\mathbf{R}_{w,u}^{b_{k-1}}(\mathbf{R}_{b_k}^{w,u}\mathcal{X}_{k,i}^{b_k} + \boldsymbol{t}_{b_k}^{w,u} - \boldsymbol{t}_{b_{k-1}}^{w,u}) - \boldsymbol{t}_l^b) \quad (25)$$

Substituting (25) into (23), we convert the residuals (23) for LiDAR features into terms independent of the degenerate DOF $\boldsymbol{x}_k^d$. Therefore, the residuals for LiDAR features participating in the tightly coupled optimization (16) will not contain degenerate DOF $\boldsymbol{x}_k^d$, as shown in Fig. 5(c). Finally, $\boldsymbol{x}_k^d$ is estimated through visual measurements and IMU residuals.

B. Poor Lighting Conditions

Li *et al.* [25] propose a novel and lightweight deep learning method, which achieves 1000/11 (GPU/CPU) FPS high-frequency image processing and lighting enhancement under a wide range of lighting conditions. Visual SLAM generally uses grayscale images as input to improve efficiency. Hence, to make it more efficient, we modify the input and output of the deep neural network as grayscale images. Finally, the deep neural network is successfully introduced into our system. After image enhancement of the input images as shown on the right of Fig. 6, the impact of poor lighting conditions on the algorithm is ultimately eliminated.

VII. EXPERIMENTAL RESULTS

We evaluate the proposed LIVER through public datasets and real-world experiments. In the first experiment, we compare the proposed algorithm with the various methods on public datasets. The performance of the method is further shown through numerical analysis in detail. Then, a series of experiments is conducted in the real world to evaluate the performance of the algorithm, including dark indoor long corridors and underground mines. In both public datasets and real-world experiments, we compare the proposed LIVER with LVI-SAM [9] (LiDAR+Monocular+IMU), LIO-SAM [27] (LiDAR+IMU), FAST-LIO2 [28] (LiDAR+IMU), and VINS-Mono [24] (Monocular+IMU). The experiments in this section are conducted on a laptop with an Intel i7-11700K CPU and NVIDIA GeForce GTX 3070 GPU in Ubuntu Linux.

A. Dataset Comparison

We evaluate the proposed LIVER using M2DGR datasets [26]. The datasets are collected on a GAEA ground mobile robot, which contains a Velodyne VLP-32C LiDAR, a Realsense d435i camera, a Handsfree A9 IMU, and a Ublox M8T GPS (for ground truth).

Most sequences are in poor lighting conditions, and the results are recorded in Table I. In sequences of poor lighting conditions, after running for a while, the visual-inertial module of LVI-SAM fails, and the degenerate LVI-SAM turns into a LINS. Since LIVER adopts the image enhancement algorithm, visual information can normally participate in tight coupling optimization in these sequences. The root-mean-square error (RMSE) of all sequences in M2DGR datasets is shown in Table I, which is evaluated by the absolute trajectory error (ATE)[2]. LIVER achieves the lowest average RMSE in most sequences with respect to the GPS measurements, which are treated as ground truth. The above results verify that the proposed LIVER is more accurate and robust.

Fig. 6. *Left:* Dark indoor long corridors - challenges include illumination changes and straight long corridors. *Middle, Right:* Underground mine - challenges include poor lighting conditions, straight long tunnels, self-similar areas, and low texture. The *Middle* is the original gray image, and the *Right* is the enhanced image.

TABLE I
RMSE OF ATE IN M2DGR DATASETS IN METERS.

Sequence	VINS-Mono	FAST-LIO2	LIO-SAM	LVI-SAM	LIVER
gate_01	1.426	0.384	0.146	0.148	**0.142**
gate_02	fail	0.475	0.327	0.327	**0.310**
gate_03	7.381	0.333	**0.112**	**0.112**	0.117
street_01	fail	0.530	0.843	0.851	**0.257**
street_02	fail	2.858	3.844	3.844	**2.835**
street_03	7.470	0.380	0.140	0.138	**0.136**
street_04	fail	3.866	1.038	1.039	**0.591**
street_05	fail	**0.379**	0.405	0.412	0.444
street_06	fail	0.591	**0.420**	0.423	0.453
street_08	4.762	0.387	0.188	0.191	**0.166**

The best results are marked in bold. *fail* means the RMSE of ATE is larger than 10 meters. Since we find that the ground truth of sequence street_07 is unreliable, no results for this sequence are given.

B. Real-World Experiments

As shown in Fig. 6, the data of the real-world experiments is collected on an unmanned ground vehicle SCOUT 2.0, which contains a Velodyne VLP-16 LiDAR, a Realsense d455 camera, an Xsens MTi-300 IMU, and an Intel NUC PC.

1) Dark indoor long corridors experiments: The long indoor corridor is in the Laboratory building of the Nankai University campus. The datasets are recorded at 10 o'clock in the evening. There are no lights in some areas of the corridor and the lighting conditions are very poor, as shown on the left of Fig. 6. The scenario is similar to the underground mine, which is the presence of long corridors.

We start and end the data-gathering process at the same position to validate end-to-end translation and rotation errors, as provided in Table II. The trajectories of all methods are shown in Fig. 8. The lowest end-to-end translation and rotation errors are achieved by LIVER. Note that the loop

[2]https://github.com/MichaelGrupp/evo

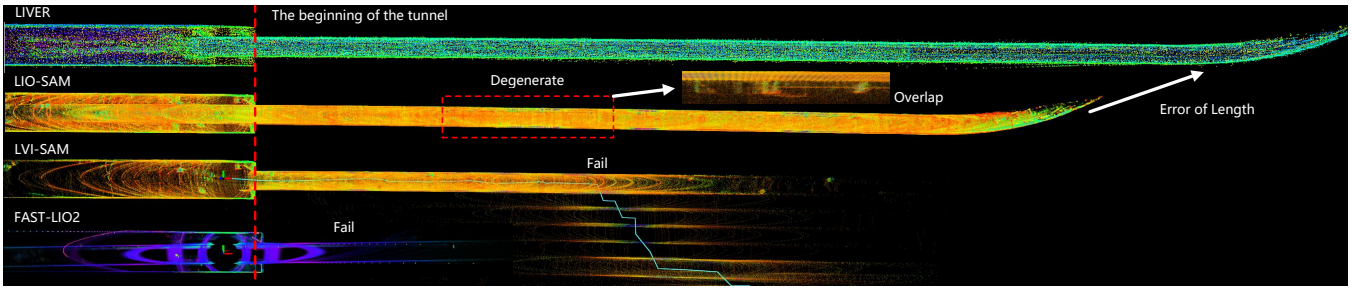

Fig. 7. Point cloud map of LIVER, LIO-SAM, LVI-SAM, and FAST-LIO2 from underground mine experiment. The map of VINS-Mono is not shown due to its failure to generate meaningful results. The map of LIVER is the most regular and the most accurate in length.

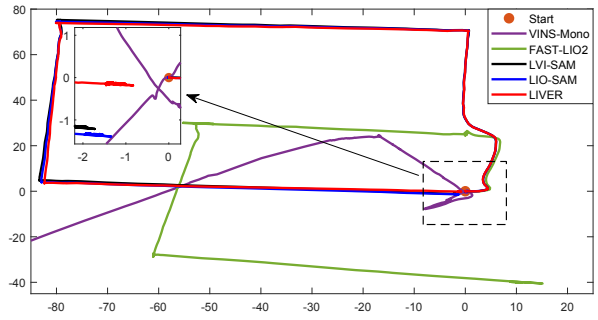

Fig. 8. Trajectories of LIVER, LIO-SAM, LVI-SAM, FAST-LIO2, and VINS-Mono in the dark indoor long corridors experiment.

TABLE II

END-TO-END TRANSLATION AND ROTATION ERRORS

Error type	VINS-Mono	FAST-LIO2	LIO-SAM	LVI-SAM	LIVER
Translation (m)	fail	41.894	3.147	2.949	**2.683**
Rotation (degree)	fail	7.580	4.227	7.493	**3.784**

The best results are marked in bold.

closure detection is disabled for all methods in this test to validate the system in a pure odometry mode.

2) Underground mine experiment: We test the robustness of the proposed algorithm in the underground tunnel shown in the middle and right of Fig. 6, which is challenging due to the presence of self-similar tunnels and extremely dark environments. The lack of illumination is compensated by handheld light sources, which can only provide poor lighting conditions. In this experiment, the robot first moves along the direction of the tunnel to its endpoint and then returns to the starting point. Note that the robot *keeps moving forward*, without moving backward, in this experiment.

In this experiment, we set $\mathbf{r}_{th} = 100$, $\mathbf{x}_{th} = 0.05$. Fig. 7 shows the mapping results of four methods with LiDAR. There is apparent LiDAR degeneracy in the map of LIO-SAM, which is much shorter than the actual tunnel length and is severely disorderly. On the other hand, LVI-SAM fails to position and map. FAST-LIO2 is completely degenerate and its mapping result is invalid. The map built by LIVER is the most regular and the most accurate in length.

The position results of all methods are shown in Fig. 9. The estimated position of LIVER is that the robot keeps moving forward. The marked parts in the estimation of x

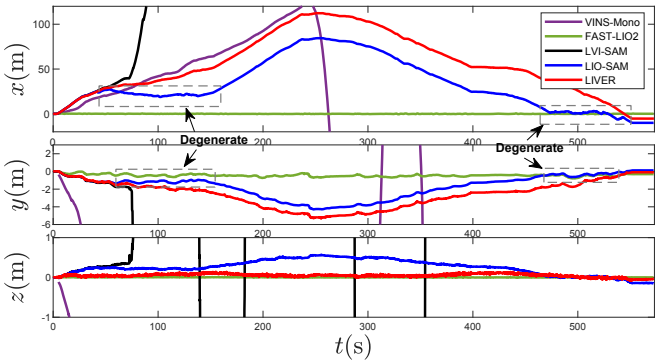

Fig. 9. Position results of all methods in underground mine experiment. *Top, Mid:* The results of LIO-SAM are degenerate in the marked part.

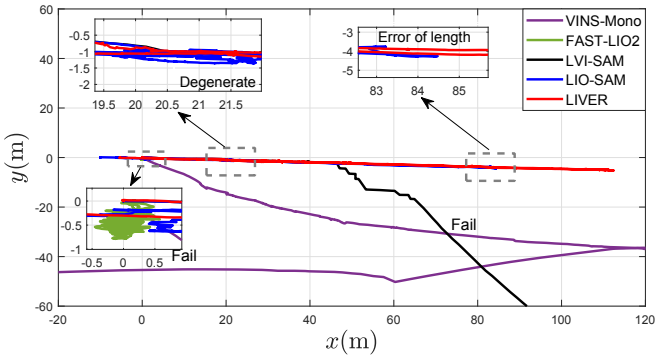

Fig. 10. Trajectories of all methods in the underground mine experiment. The trajectory of LIO-SAM is much shorter than LIVER because of degeneracy. LVI-SAM and FAST-LIO2 fail to obtain effective pose estimation.

and y DOF show that LIO-SAM is degenerate, the estimated pose moves back and forth. The trajectories of all methods are provided in Fig. 10. The trajectory length of LIVER is the most consistent with the actual length. However, the trajectory length of LIO-SAM is much shorter than the actual length because of degeneracy. LVI-SAM and VINS-Mono receive low-quality images without processing, hence VINS-Mono and the subsystem VIS of LVI-SAM fail to complete positioning. Moreover, LVI-SAM cannot handle the degeneracy of LiDAR. Hence, LVI-SAM fails to obtain effective pose estimation. Similarly, FAST-LIO2 obtains poor results without considering LiDAR degeneracy.

TABLE III

RATES OF INLIER VISUAL FEATURES - AVERAGE [%] (VARIANCE [%])

Sequence	gate_01	gate_02	gate_03*	street_01	street_02*	corridor
LVI-SAM	95.52(29.7)	93.56(60.5)	**97.69**(**4.4**)	86.13(245.3)	**95.63**(**8.2**)	95.77(110.5)
LIVER	**95.54**(**15.0**)	**95.22**(**14.6**)	96.72(6.0)	**91.70**(**20.9**)	94.30(13.5)	**96.40**(**12.7**)
Sequence	street_03	street_04	street_05	street_06	street_08	mine
LVI-SAM	76.24(403.7)	89.50(122.6)	90.45(127.3)	**94.40**(43.4)	80.94(374.7)	91.49(82.9)
LIVER	**90.49**(**24.6**)	**92.64**(**9.1**)	**91.96**(**10.1**)	94.06(**8.4**)	**91.14**(**23.1**)	**95.84**(**16.1**)

The best result are marked in bold, indicating more stable feature extraction and tracking. * are sequences in good lighting conditions, *corridor* denotes the dark indoor long corridors experiments, *mine* represents underground mine experiment.

C. Analysis

The key benefit of using the lighting enhancement deep network is increased inlier visual features. Table III shows the rates of inlier visual features for LVI-SAM and LIVER. The VIS of LVI-SAM is adopted from VINS-Mono, hence the results of VINS-Mono are not given. On most poor lighting conditions datasets, LIVER with illumination enhancement can extract more inlier visual features. Hence, optical flow tracking is more stable.

In the underground mine experiment, the angle between the x-axis of the robot and the direction of the tunnel is $-2.67°$, as shown in Fig. 10. Hence, the degenerate direction of LIO-SAM is a vector consisting of x and y DOF. As shown in Fig. 9, the estimation results of LIO-SAM in the y-axis are degenerate.

VIII. CONCLUSIONS

In this paper, we propose a tightly coupled LiDAR-inertial-visual state estimator with high robustness for underground environments. Our approach presents novel solutions to LIV fusion initialization and robust sensor degeneracy handling. These features allow for robust estimation in extreme underground environments. We show comparable performance to state-of-the-art open-source LIVO systems in typical conditions and superior performance in extreme conditions, such as underground mines. In future research, we are interested in multi-robot mapping in underground environments and deep learning-based visual feature detection in poor lighting conditions. Moreover, we will study threshold adaptive adjustment strategies for the degeneracy handling approach to improve the versatility of the algorithm.

REFERENCES

[1] K. Ebadi et al., "Present and Future of SLAM in Extreme Environments: The DARPA SubT Challenge," *IEEE Trans. Robot.*, vol. 40, pp. 936–959, 2024.

[2] Z. Song et al., "IR-VIO: Illumination-Robust Visual-Inertial Odometry Based on Adaptive Weighting Algorithm With Two-Layer Confidence Maximization," *IEEE/ASME Trans. Mechatron.*, vol. 28, no. 4, pp. 1920-1929, 2023.

[3] *DARPA*. "DARPA Subterranean Challenge." Accessed: Mar. 8, 2022. [Online]. Available: https://www.subtchallenge.com

[4] D. Wisth, M. Camurri, S. Das and M. Fallon, "Unified Multi-Modal Landmark Tracking for Tightly Coupled Lidar-Visual-Inertial Odometry," *IEEE Robot. Automat. Lett.*, vol. 6, no. 2, pp. 1004–1011, 2021.

[5] X. Zuo et al., "LIC-Fusion: LiDAR-Inertial-Camera Odometry," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2019, pp. 5848–5854.

[6] X. Zuo et al., "LIC-Fusion 2.0: LiDAR-Inertial-Camera Odometry with Sliding-Window Plane-Feature Tracking," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2020, pp. 5112–5119.

[7] C. Zheng et al., "FAST-LIVO: Fast and Tightly-coupled Sparse-Direct LiDAR-Inertial-Visual Odometry," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2022, pp. 4003–4009.

[8] J. Lin et al., "R3LIVE: A Robust, Real-time, RGB-colored, LiDAR-Inertial-Visual tightly-coupled state Estimation and mapping package," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2022, pp. 10672–10678.

[9] T. Shan, B. Englot, C. Ratti and D. Rus, "LVI-SAM: Tightly-coupled Lidar-Visual-Inertial Odometry via Smoothing and Mapping," in *Proc. IEEE Int. Conf. Robot. Automat.*, pp. 5692–5698, 2021.

[10] X. Lang et al., "Coco-LIC: Continuous-Time Tightly-Coupled LiDAR-Inertial-Camera Odometry Using Non-Uniform B-Spline," *IEEE Robot. Automat. Lett.*, vol. 8, no. 11, pp. 7074–7081, 2023.

[11] K. Ebadi et al., "LAMP: Large-Scale Autonomous Mapping and Positioning for Exploration of Perceptually-Degraded Subterranean Environments," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 80–86.

[12] S. Khattak et al., "Complementary MultiModal Sensor Fusion for Resilient Robot Pose Estimation in Subterranean Environments," in *Proc. Int. Conf. Unman. Aircra. Syst.*, 2020, pp. 1024–1029.

[13] Y. Chang et al., "LAMP 2.0: A Robust Multi-Robot SLAM System for Operation in Challenging Large-Scale Underground Environments," *IEEE Robot. Automat. Lett.*, vol. 7, no. 4, pp. 9175–9182, 2022.

[14] J. Zhang and S. Singh, "Laser-visual-inertial Odometry and Mapping with High Robustness and Low Drift," *J. Field Robot.*, vol. 35, no. 8, pp. 1242–1264, 2018.

[15] A. I. Mourikis and S. I. Roumeliotis, "A Multi-State Constraint Kalman Filter for Vision-aided Inertial Navigation," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2007, pp. 3565–3572.

[16] J. Zhang et al., "On degeneracy of optimization-based state estimation problems," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2016, pp. 809–816.

[17] T. Tuna et al., "X-icp: Localizability-aware lidar registration for robust localization in extreme environments," *IEEE Trans. Robot.*, vol. 40, pp. 452–471, 2024.
T. Tuna et al., "X-ICP: Localizability-Aware LiDAR Registration for Robust Localization in Extreme Environments," IEEE Trans. Robot., vol. 40, pp. 452–471, Nov. 2024.

[18] J. Nubert et al., "Learning-based Localizability Estimation for Robust LiDAR Localization," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2022, pp. 17–24.

[19] F. Han et al., "DAMS-LIO: A Degeneration-Aware and Modular Sensor-Fusion LiDAR-inertial Odometry," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2023, pp. 2745-2751.

[20] J. Zhang and S. Singh, "LOAM: Lidar odometry and mapping in real-time," in *Proc. Robotics: Sci. Syst. Conf.*, 2014, pp. 1–9.

[21] D. Galvez-López and J. D. Tardos, "Bags of Binary Words for Fast Place Recognition in Image Sequences," *IEEE Trans. Robot.*, vol. 28, no. 5, pp. 1188–1197, 2012.

[22] G. Kim, S. Choi and A. Kim, "Scan Context++: Structural Place Recognition Robust to Rotation and Lateral Variations in Urban Environments," *IEEE Trans. Robot.*, vol. 38, no. 3, pp. 1856–1874, 2022.

[23] M. Kaess et al., "iSAM2: Incremental smoothing and mapping using the Bayes tree," *Int. J. Robot. Res.*, vol. 31, no. 2, pp. 216–235, 2012.

[24] T. Qin, P. Li and S. Shen, "VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator," *IEEE Trans. Robot.*, vol. 34, no. 4, pp. 1004–1020, 2018.

[25] C. Li, C. Guo and C. C. Loy, "Learning to Enhance Low-Light Image via Zero-Reference Deep Curve Estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 8, pp. 4225–4238, 2022.

[26] J. Yin, A. Li, T. Li, W. Yu and D. Zou, "M2DGR: A Multi-Sensor and Multi-Scenario SLAM Dataset for Ground Robots," *IEEE Robot. Automat. Lett.*, vol. 7, no. 2, pp. 2266–2273, 2022.

[27] T. Shan et al., "LIO-SAM: Tightly-coupled Lidar Inertial Odometry via Smoothing and Mapping," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2020, pp. 5135–5142.

[28] W. Xuet al., "FAST-LIO2: Fast Direct LiDAR-Inertial Odometry," *IEEE Trans. Robot.*, vol. 38, no. 4, pp. 2053–2073, 2022.