Degradation-Aware LiDAR-Thermal-Inertial SLAM

Yu Wang, Yufeng Liu, Lingxu Chen, Haoyao Chen, *Senior Member, IEEE*, and Shiwu Zhang, *Member, IEEE*

***Abstract*—During robotic disaster relief missions, state estimation still faces significant challenges, especially when GNSS is denied or sensor perception undergoes degradation. In this letter, we introduce a degradation-aware LiDAR-Thermal-Inertial SLAM, DaLiTI, that leverages the complementary nature of multi-modal information to achieve robust and precise state estimation in perceptually challenging environments. The system utilizes an iterated error state Kalman filter (IESKF) to loosely integrate LiDAR, thermal infrared camera, and IMU measurements. We propose an adaptive fusion mechanism that dynamically weights and fuses LiDAR and thermal measurements based on real-time modal quality to prevent failure information from propagating throughout the system. Experimental results demonstrate that, compared with state-of-the-art methods, DaLiTI maintains competitive performance in conventional environments and exhibits superior robustness and accuracy in degraded scenarios such as fire scenes or chemical plants with gas leaks.**

***Index Terms*—Adaptive fusion, LiDAR-thermal-inertial SLAM, perceptual degradation.**

I. Introduction

OVER the decades, various successful SLAM frameworks have emerged for real-time state estimation, primarily vision-based or LiDAR-based. However, a single-modality system faces environmental constraints, with performance heavily impacted by the use case conditions [1]. Vision-based methods require stable and distinct features and may encounter data association errors in environments with poor illumination and low texture [2]. LiDAR-based solutions rely on the environment for sufficient geometric constraints, which often fail in structure-less or self-similar environments such as long tunnels or open fields [3]. Recent advances in multi-modal fusion, notably LiDAR-inertial-visual SLAM, have demonstrated significant potential [4]. However, visual clutter such as dust, fog, and smoke can scatter light, disrupting measurements from cameras and LiDARs, thus reducing the reliability of the state estimator.

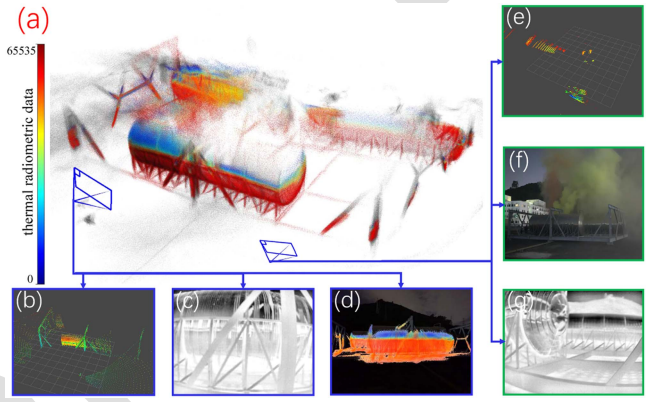


Fig. 1. DaLiTI performance in a chemical plant leakage scenario. (a) geometric-thermal map built by DaLiTI. (b)(c) LiDAR and thermal measurements in an area without smoke. (d) Projection of geometric-thermal map onto RGB image in an area without smoke. (e)(f)(g) LiDAR, visual camera, and thermal measurements in an area with smoke.

Thermal infrared cameras have attracted considerable attention in recent years, operating in the long-wave infrared spectrum and capturing thermal-radiometric information. In contrast to visual cameras, thermal cameras exhibit evident advantages in hazardous disaster areas, as they remain unaffected by dark and smoke. However, thermal images suffer from low contrast, high noise, and lack of color and detail, making data association challenging [5]. Inspired by the significant change in thermal radiation at the edge of objects, our previous work, named ETIO [6], utilizes the edge information to establish reliable correspondences and further performs state estimation by combining IMU pre-integration with reprojection errors of edge features. While ETIO robustly handles complex illumination and smoke conditions for state estimation at all daytime, its monocular-inertial system lacks reliable depth observation, leading to lower accuracy than LiDAR-based methods in non-degraded environments, thus limiting its application scope.

We introduce Degradation-aware LiDAR-Thermal-Inertial SLAM (DaLiTI-SLAM) to overcome the limitations above. Following a loosely coupled manner, DaLiTI utilizes the IMU for forward state propagation and parallel processes LiDAR and thermal measurements based on IESKF as shown in Fig. 1, considering the complementarity of perception between different modal sensors, a multi-modal adaptive information fusion scheme is designed for the real-time update of the state, guided by the quality evaluation of perception. This scheme can effectively prevent the propagation of degraded information throughout the system, thereby enhancing the robustness and resilience of state estimation. The main contributions of this letter are threefold:

Received 29 January 2025; accepted 8 June 2025. Date of publication; date of current version. This article was recommended for publication by Associate Editor B. Englot and Editor J. Civera upon evaluation of the reviewers' comments. This work was supported in part by the National Natural Science Foundation of China under Grant U21A20119 and Grant U1713206 and in part by Shenzhen Science and Innovation Committee Funds under Grant RCJC20231211090050082, Grant SZXJP20230703093206015, and Grant JCYJ20241202123714019. *(Yu Wang and Yufeng Liu contributed equally to this work.) (Corresponding author: Haoyao Chen.)*

Yu Wang and Shiwu Zhang are with the Department of Precision Machinery and Precision Instrumentation, University of Science and Technology of China, Hefei 230027, China (e-mail: wangyuustc@ustc.edu.cn; swzhang@ustc.edu.cn).

Yufeng Liu is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: raymond.lau.lyf@gmail.com).

Lingxu Chen and Haoyao Chen are with the School of Mechanical Engineering and Automation, Harbin Institute of Technology, Shenzhen 518055, China (e-mail: rabbitno9@outlook.com; hychen5@hit.edu.cn).

Our implementation is available at https://github.com/HITSZ-NRSL/DaLiTI.

This article has supplementary downloadable material available at https://doi.org/10.1109/LRA.2025.3581127, provided by the authors.

Digital Object Identifier 10.1109/LRA.2025.3581127

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

1) We introduce DaLiTI, a novel degradation-aware LiDAR-Thermal-Inertial SLAM framework, which integrates complementary information loosely for reliable and real-time estimation in degraded environments.
2) An efficient perception quality evaluation method is designed for degradation detection, whose quantization results are further utilized as weights for multi-modal information adaptive fusion in IESKF-based state update.
3) Extensive experiments demonstrate that DaLiTI, compared with state-of-the-art methods, achieves competitive performance in conventional environments while maintaining robust and accurate state estimation during perceptual degradation. To contribute to the community, the datasets and code are made publicly available.

II. Related Works

In this section, we first review LiDAR-visual-inertial SLAM frameworks from the data fusion perspective. Subsequently, we discuss thermal odometry, closely related to our work.

A. Lidar-Visual-Inertial Slam

Prior LiDAR-visual-inertial odometry (LVIO) research can be categorized into loosely and tightly coupled approaches based on their multi-modal fusion schemes. Loosely coupled methods involve separate subsystems for state estimation, fused for refinement. Examples include V-LOAM [7], which uses VIO to predict relative transformations refined by LiDAR odometry, and Comp-SLAM [8], which utilizes VIO and TIO estimates to provide priors for scan matching and selective LiDAR pose propagation in challenging environments. Super Odometry [9] employs a factor graph with an IMU-centric paradigm, where visual-inertial and LiDAR-inertial odometry offer priors to constrain IMU biases. Unlike the work above, DAMS-LIO [1], MIMOSA [4], and Switch-SLAM [10] adjust their fusion architectures based on modal quality analysis. If a modality-specific estimate is detected as unreliable, its updates are temporarily detached from the fusion framework until the degraded estimation subsystem locally recovers.

In contrast, tightly coupled methods directly fuse raw sensor measurements for state estimation through joint optimization. Based on [11] and [12], Shan et al.. introduced LVI-SAM [13], which jointly optimizes constraints from LiDAR, vision, IMU pre-integration, and loop closure using a factor graph. Meanwhile, R2LIVE [14] estimates the state by minimizing the reprojection error of LiDAR and visual features within the IESKF framework. Unlike the feature-based methods mentioned above, FAST-LIVO [15] and R3LIVE [16] directly use raw LiDAR points and image pixels for data association, achieving higher computational efficiency. More recently, Z. Yuan et al.. [17] proposed a sweep reconstruction method to align LiDAR sweeps with image timestamps, allowing the LIO module to determine states at all imaging moments accurately.

While tightly coupled methods enhance accuracy and robustness by fully utilizing sensor measurements, their strict joint optimization limits flexibility and increases error propagation vulnerability. In contrast, loosely coupled methods offer resilience, plug-and-play functionality, and flexible sensor switching, making them suitable for complex tasks like the DARPA Subterranean Challenge [18] and planetary exploration. Maintaining robust state estimation is crucial, given the inevitable perceptual degradation in disaster areas. Thus, we adopt a loosely coupled strategy for multi-modal fusion.

B. Thermal Odometry

Raw thermal images generally have a pixel depth of 14 or 16 bits. Early thermal odometry works mainly adhered to visual odometry frameworks, necessitating rescaling these images to 8 bits for algorithmic compatibility. Vidas et al.. [19] introduced the first thermal odometry, utilizing GFTT features tracked via optical flow on rescaled images. However, rescaling may result in detail loss and abrupt appearance changes, causing association errors. Khattak et al.. circumvented this by directly utilizing full radiometric data for motion estimation, establishing image correspondences without rescaling [2]. Compared to rescaled thermal images, full radiometric thermal images remain more consistent even after a high-heat object suddenly appears in the FOV. As an alternative, our previously proposed ETIO [6] utilized edge information for data association instead of the raw thermal image, which effectively suppresses the influence of noise and low-contrast areas. Tightly coupling thermal odometry and LiDAR odometry, EIL-SLAM [20] generates a dense depth map for recovering the thermal feature scale. Furthermore, integrating infrared and LiDAR data for place recognition reduces accumulated drift and enhances global consistency.

With deep learning advancements, neural networks have been integrated into thermal odometry. TP-TIO [21] is the first tightly-coupled deep thermal-inertial odometry, using CNNs for feature extraction from thermal images and IMU-assisted tracking. DeepTIO [22], introduced by Saputra et al.., is a pose regression framework using an end-to-end neural network. Their subsequent work [23] presents a comprehensive thermal-inertial SLAM, incorporating neural abstraction, graph-based optimization, and neural loop closure detection using global descriptors. Jiang et al. [24] combine conventional and learning-based methods, proposing a real-time system with SVD-based image enhancement for feature detection and a lightweight optical flow network for tracking. While promising, these learning-based approaches require GPUs for acceleration and face challenges in transferability, limiting their application in disaster rescue. Given this, our study adopts a conventional framework for real-time state estimation using only CPUs tailored for rescue robots.

III. Methodology

An overview of the proposed DaLiTI, which primarily comprises three major modules: forward propagation, multi-sensor measurement processing, and degradation-aware adaptive fusion state estimation, is illustrated in Fig. 2. The system first predicts the state vector and covariance using forward propagation based on a discrete-time model. Subsequently, LiDAR and thermal measurements are processed, and their quality is evaluated independently via a degradation quantizer. Finally, an adaptive fusion state estimator based on an IESKF is utilized for state update by combining the prior distribution from IMU, the posterior distribution derived from LiDAR and thermal measurements, and the normalized degradation factor.

We assume the IMU, LiDAR, and the monocular thermal infrared camera are rigidly mounted with known extrinsic parameters. Specifically, the transformation from the LiDAR frame ${}^{L}(\cdot)$ to the IMU frame ${}^{I}(\cdot)$ is denoted as ${}^{I}\mathbf{T}_{L} = ({}^{I}\mathbf{R}_{L}, {}^{I}\mathbf{P}_{L})$,

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

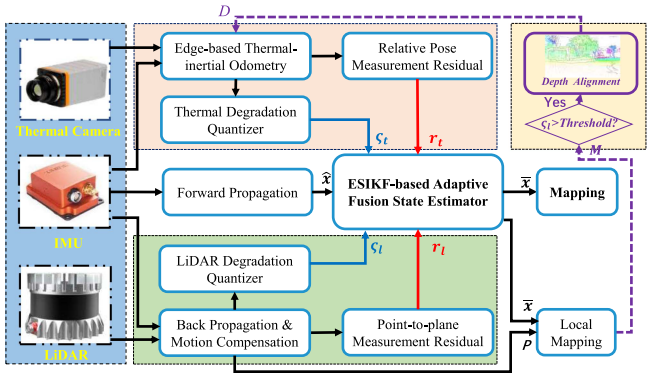


Fig. 2. Overview of the DaLiTI system.

and the transformation from the camera frame ${}^{C}(\cdot)$ to the IMU frame ${}^{I}(\cdot)$ is denoted as ${}^{I}\mathbf{T}_{C} = ({}^{I}\mathbf{R}_{C}, {}^{I}\mathbf{P}_{C})$.

A. Forward Propagation Based on Discrete IMU Model

For simplicity, we consider the IMU frame as the body frame and designate the first body frame as the global frame ${}^{G}(\cdot)$. The continuous kinematic model is expressed as:

$$\begin{aligned} {}^{G}\dot{\mathbf{p}}_{I} &= {}^{G}\mathbf{v}_{I}, \quad \dot{\mathbf{b}}_{\mathbf{a}} = \mathbf{n}_{\mathbf{b}_{\mathbf{a}}}, \quad \dot{\mathbf{b}}_{\mathbf{g}} = \mathbf{n}_{\mathbf{b}_{\mathbf{g}}} \\ {}^{G}\dot{\mathbf{v}}_{I} &= {}^{G}\mathbf{R}_{I}(\mathbf{a}_{m} - \mathbf{b}_{\mathbf{a}} - \mathbf{n}_{\mathbf{a}}) + {}^{G}\mathbf{g} \\ {}^{G}\dot{\mathbf{R}}_{I} &= {}^{G}\mathbf{R}_{I}\lfloor \omega_{m} - \mathbf{b}_{\mathbf{g}} - \mathbf{n}_{\mathbf{g}} \rfloor_{\wedge} \end{aligned} \tag{1}$$

where ${}^{G}\mathbf{p}_{I}$, ${}^{G}\mathbf{v}_{I}$, and ${}^{G}\mathbf{R}_{I}$ represent the IMU's position, velocity, and attitude in the global frame, respectively; ${}^{G}g$ denotes the gravity vector; the raw accelerometer and gyroscope measurements from the IMU are given by $\mathbf{a}_{m}$ and ω_{m}, respectively, with measurement noises $\mathbf{n}_{\mathbf{a}}$ and $\mathbf{n}_{\mathbf{g}}$; the IMU biases, modeled as random walks driven by Gaussian noises $\mathbf{n}_{\mathbf{b}_{\mathbf{a}}}$ and $\mathbf{n}_{\mathbf{b}_{\mathbf{g}}}$, are denoted by $\mathbf{b}_{\mathbf{a}}$ and $\mathbf{b}_{\mathbf{g}}$; the skew-symmetric matrix of a vector, denoted by $\lfloor \cdot \rfloor_{\wedge}$, maps the cross-product operation.

We define $\boxplus$ and $\boxminus$ as the plus and minus of a state vector on a manifold, enabling state update in the tangent space. Then, we discretize the continuous kinematic model at the IMU sampling period Δt with a zero-order holder, yielding the state transition model for the ith IMU measurement as:

$$\mathbf{x}_{i+1} = \mathbf{x}_{i} \boxplus (\Delta t \cdot \mathbf{f}(\mathbf{x}_{i}, \mathbf{u}_{i}, \mathbf{w}_{i})). \tag{2}$$

The system state $\mathbf{x}$, input $\mathbf{u}$, process noise $\mathbf{w}$ and the state transition function $\mathbf{f}$ are defined as:

$$\begin{aligned} \mathbf{x} &\triangleq [{}^{G}\mathbf{R}_{I}, {}^{G}\mathbf{p}_{I}, {}^{G}\mathbf{v}_{I}, \mathbf{b}_{\mathbf{g}}, \mathbf{b}_{\mathbf{a}}, {}^{G}\mathbf{g}]^{\top} \\ \mathbf{u} &\triangleq [\omega_{m}^{\top}, \mathbf{a}_{m}^{\top}], \quad \mathbf{w} \triangleq [\mathbf{n}_{\mathbf{g}}^{\top}, \mathbf{n}_{\mathbf{a}}^{\top}, \mathbf{n}_{\mathbf{b}_{\mathbf{g}}}^{\top}, \mathbf{n}_{\mathbf{b}_{\mathbf{a}}}^{\top}]^{\top} \end{aligned}$$

$$\mathbf{f}(\mathbf{x}, \mathbf{u}, \mathbf{w}) = \begin{bmatrix} \omega_{m} - \mathbf{b}_{\mathbf{g}} - \mathbf{n}_{\mathbf{g}} \\ {}^{G}\mathbf{v}_{I} + ({}^{G}\mathbf{R}_{I}(\mathbf{a}_{m} - \mathbf{b}_{\mathbf{a}} - \mathbf{n}_{\mathbf{a}}) + {}^{G}\mathbf{g})\Delta t/2 \\ {}^{G}\mathbf{R}_{I}(\mathbf{a}_{m} - \mathbf{b}_{\mathbf{a}} - \mathbf{n}_{\mathbf{a}}) + {}^{G}\mathbf{g} \\ \mathbf{n}_{\mathbf{b}_{\mathbf{g}}} \\ \mathbf{n}_{\mathbf{b}_{\mathbf{a}}} \end{bmatrix} \tag{3}$$

By setting the process noise $\mathbf{w}_{i}$ in (2) to zero, the forward propagation to predict the $\hat{\mathbf{x}}_{i+1}$ can be defined as:

$$\hat{\mathbf{x}}_{i+1} = \hat{\mathbf{x}}_{i} \boxplus (\Delta t \cdot \mathbf{f}(\mathbf{x}_{i}, \mathbf{u}_{i}, 0)). \tag{4}$$

To facilitate the linearization of attitude variables, we introduce the error state $\tilde{\mathbf{x}}$ into (4) as:

$$\begin{aligned} \tilde{\mathbf{x}}_{i+1} &= \mathbf{x}_{i+1} \boxminus \hat{\mathbf{x}}_{i+1} \\ &= (\mathbf{x}_{i} \boxplus (\Delta t \cdot \mathbf{f}(\mathbf{x}_{i}, \mathbf{u}_{i}, \mathbf{w}_{i}))) \\ &\quad \boxminus (\hat{\mathbf{x}}_{i} \boxplus (\Delta t \cdot \mathbf{f}(\hat{\mathbf{x}}_{i}, \mathbf{u}_{i}, 0))) \\ &\approx \mathbf{F}_{\tilde{\mathbf{x}}}\tilde{\mathbf{x}}_{i} + \mathbf{F}_{\mathbf{w}}\mathbf{w}_{i} \end{aligned} \tag{5}$$

with covariance $\hat{\mathbf{P}}_{i+1}$ propagated as:

$$\hat{\mathbf{P}}_{i+1} = \mathbf{F}_{\tilde{\mathbf{x}}}\hat{\mathbf{P}}_{i}\mathbf{F}_{\tilde{\mathbf{x}}}^{\top} + \mathbf{F}_{\mathbf{w}}\mathbf{Q}\mathbf{F}_{\mathbf{w}}^{\top}, \tag{6}$$

where $\mathbf{F}_{\tilde{\mathbf{x}}}$ and $\mathbf{F}_{\mathbf{w}}$ are approximate Jacobian matrix, and $\mathbf{Q}$ is the covariance matrix of $\mathbf{w}$.

B. Processing of LiDAR Measurements

Upon receiving a LiDAR scan at time t_{k}, we utilize the backward propagation method proposed in [25] to compensate for motion distortion, which estimates the LiDAR pose of each point in the scan concerning the pose at the scan end time based on IMU measurements, and subsequently projects all points onto the scan end time as $\{{}^{L}p_{j}\}$. Based on the point-to-plane distance, we formulate the LiDAR measurement observation function as follows:

$$\begin{aligned} \mathbf{0} &= \boldsymbol{h}_{l}\left(\mathbf{x}_{k}, {}^{L}\mathbf{p}_{j} + {}^{L}\mathbf{n}_{j}\right) \\ &= {}^{G}\mathbf{u}_{j}^{\top}\left({}^{G}\mathbf{R}_{I_{k}}\left({}^{I}\mathbf{R}_{L}\left({}^{L}\mathbf{p}_{j} + {}^{L}\mathbf{n}_{j}\right) + {}^{I}\mathbf{p}_{L}\right) + {}^{G}\mathbf{p}_{I_{k}} - {}^{G}\mathbf{q}_{j}\right) \end{aligned} \tag{7}$$

where ${}^{L}\mathbf{n}_{j}$ represents the measurement noise of ${}^{L}\mathbf{p}_{j}$; ${}^{G}\mathbf{u}_{j}^{\top}$ is the normal vector of the neighboring plane, and ${}^{G}\mathbf{q}_{j}$ is corresponding point of the plane in the map for ${}^{L}\mathbf{p}_{j}$, respectively.

C. Processing of Thermal Measurements

Once a thermal infrared image is captured, it is essential to address the low contrast and noise in thermal infrared images for accurate data association before detailing the precise formulation of the thermal measurement model. Given that the edges of objects exhibit a notable response in thermal radiation, we follow our previous work [6] that utilizes edge features to construct robust image alignment. We first converted the raw images into binarized edge images using the Difference of Gaussians (DoG) method to effectively filter out noise and low-contrast areas. To deal with the lack of photometric information in edge images, we introduce the distance transform to improve the traditional KLT tracker:

$$\arg\min_{d} \sum_{x=u_{x}-w_{x}}^{u_{x}+w_{x}} \sum_{y=u_{y}-w_{y}}^{u_{y}+w_{y}} \left[D_{r}(x,y) - D_{c}\left(x + d_{x}, y + d_{y}\right)\right]^{2}, \tag{8}$$

where D_{r} and D_{c} represent the distance fields corresponding to the edge images I_{e}^{r} and I_{e}^{c}, respectively. $d = [d_{x}, d_{y}]$ denotes the 2D position displacement on I_{e}^{c} for a edge feature $u = [u_{x}, u_{y}]$ on I_{e}^{r}. An Adaptive Distance Transform-aided KLT (ADT-KLT) tracker proposed by [6] is utilized, which switches between DT-KLT and traditional KLT schemes based on edge stability, enhancing feature tracking robustness. The detail of the ADT-KLT tracker can refer to [6].

There are two ways to initialize the depth of edge features. The corresponding depth value is directly assigned if edge features are successfully associated with depth through LiDAR measurements in Section III-D. Otherwise, the feature depth will be initialized through triangulation based on multiple observations. Assuming ${}^{c_i}p_l$ and ${}^{c_j}p_l$ are the lth edge feature tracking in the ith and jth frames by ADT-KLT tracker, the reprojection distance can be defined as:

$$\begin{aligned}
\boldsymbol{r}_c(\mathbf{x}_i) &= \left[b_1\ b_2\right]^\top \cdot \left({}^{c_j}\hat{\bar{\mathcal{P}}}_l - \frac{{}^{c_j}\mathcal{P}_l}{\|{}^{c_j}\mathcal{P}_l\|}\right) \\
{}^{c_j}\hat{\bar{\mathcal{P}}}_l &= \pi^{-1}\left({}^{c_j}p_l\right) \\
{}^{c_j}\mathcal{P}_l &= {}^{C}\mathbf{R}_I\left({}^{I_j}\mathbf{R}_G\left({}^{G}\mathbf{R}_{I_j}\left({}^{I}\mathbf{R}_C\,\frac{1}{\lambda_l}\,\pi^{-1}\left({}^{c_i}\mathbf{p}_l\right)+{}^{C}\mathbf{p}_I\right)\right.\right. \\
&\quad \left.\left.+{}^{G}\mathbf{P}_{I_i}-{}^{G}\mathbf{P}_{I_j}\right)-{}^{C}\mathbf{P}_I\right), \qquad (9)
\end{aligned}$$

where λ_l is the inverse depth of $l\ th$ edge feature, and π^{-1} is the back projection function; ${}^{c_j}\mathcal{P}_l$ and ${}^{c_j}\hat{\bar{\mathcal{P}}}_l$ are the projected vector on the unit sphere of ${}^{c_i}p_l$ and ${}^{c_j}p_l$, respectively; b_1 and b_2 are two arbitrarily selected orthogonal bases that span the tangent plane of ${}^{c_j}\hat{\bar{\mathcal{P}}}_l$.

By aligning (9) with IMU pre-integration residual, we can obtain the IMU state when the thermal infrared image is captured. Furthermore, we obtained the variances of the TIO to serve as inputs for the update process of IESKF. Readers can refer to [6] for the details.

D. Cross-Modality Data Association

1) LiDAR to Thermal Feature Depth Association: Scale ambiguity is an inherent characteristic of monocular systems. To aid visual odometry in recovering the scale of state estimation, previous works [7], [13] project local dense depth maps onto the image plane by integrating multiple frames of point clouds and then use the nearest points around visual feature points for depth registration. However, in this study, feature points are typically located at image edges, with the nearest points distributed in the foreground and background. Consequently, using the average depth of these neighboring points as the feature depth results in significant depth estimation errors.

To address this issue, we use an iterative thresholding approach to filter out background points for cross-modality data association (CMA). Initially, the k-nearest neighbors were searched on the projection map. If their depth variance exceeds a threshold, the feature point is identified as being on an object's edge, and back-facing points are excluded. An illustration of the proposed depth estimation is shown in Fig. 3. Considering that the depth measurement is inaccurate when LiDAR perception is degraded, LiDAR measurement is used for depth only when $\boldsymbol{\zeta}_l$ is higher than the threshold $\boldsymbol{\zeta}_{th}$.

2) LiDAR to Thermal System Initialization: The quality of initialization is vital in optimization-based thermal inertial estimation frameworks. Adequate initialization requires sufficient excitation of various degrees of freedom. However, ground robots often struggle to provide excitation along the z-axis, *roll*-axis, and *pitch*-axis. To enhance the initialization accuracy of thermal odometry, we directly interpolate the estimated state obtained by LiDAR-IMU and replace the keyframe state of TIO only during the initialization phase of the thermal inertial estimation, thereby obtaining a better estimated prior.

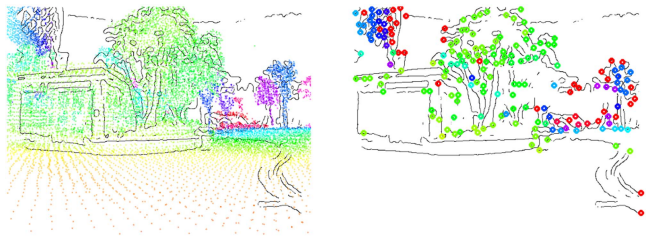


Fig. 3. Result of depth alignment. Left: LiDAR projection on edge thermal image, color represents depth. Right: Tracked feature points with depth. Deep red represents no depth alignment.

E. State Update Based on Multi-Modal Measurements

1) LiDAR Frame Residual Calculation: For LiDAR frame, we approximate the measurement model (7) by its first order approximation at $\hat{\mathbf{x}}_k$:

$$\mathbf{r}_l\left(\hat{\mathbf{x}}_k, {}^{L}\mathbf{p}_j\right) = \mathbf{0} - \boldsymbol{h}_l\left(\mathbf{x}_k, {}^{L}\mathbf{p}_j\right) \simeq \mathbf{H}^l\tilde{\mathbf{x}}_\mathbf{k} + \mathbf{v}_l, \qquad (10)$$

where $\mathbf{v}_l \in \mathcal{N}(\mathbf{0}, \mathbf{R}_l)$ is the Gaussian measurement noise. $\mathbf{H}^l$ is given as

$$\mathbf{H}^l = {}^{G}\mathbf{u}_j^\top\begin{bmatrix}\mathbf{H}^{l1}\ \mathbf{I}_3\ \mathbf{0}_{3\times12} & \mathbf{H}^{l2} & {}^{G}\hat{\mathbf{R}}_{I_k} & \mathbf{0}_{3\times6}\end{bmatrix} \qquad (11)$$

$$\mathbf{H}^{l1} = {}^{G}\hat{\mathbf{R}}_{I_k}\lfloor{}^{L}\mathbf{p}_j + {}^{I}\hat{\mathbf{p}}_l\times\rfloor,\ \ \mathbf{H}^{l2} = -{}^{G}\hat{\mathbf{R}}_{I_k}{}^{I}\hat{\mathbf{R}}_l\lfloor{}^{l}\mathbf{p}_j\times\rfloor. \qquad (12)$$

2) Thermal Residual Calculation: Through Section III-C, we can directly obtain the thermal rate odometry $\mathbf{x}^t$ in the IMU frame. Thus, the observation function can be obtained:

$$\begin{aligned}
\mathbf{x}^t &= \left[{}^{G}\mathbf{R}_I, {}^{G}\mathbf{p}_I\right]^\top \\
\mathbf{r}_t &= \mathbf{0} - \mathbf{h}_t\left(\mathbf{x}_k, \mathbf{x}_i^t\right) \simeq \mathbf{H}^t\tilde{\mathbf{x}}_k + \mathbf{v}_t, \qquad (13)
\end{aligned}$$

where $\mathbf{v}_t \in \mathcal{N}(\mathbf{0}, \mathbf{R}_t)$ is the Gaussian noise and $\mathbf{H}^t$ is:

$$\mathbf{H}^t = \begin{bmatrix}\mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\end{bmatrix}. \qquad (14)$$

3) LiDAR and Thermal Degradation Quantizer: To adaptively fuse the LiDAR and thermal measurements, we introduce a degradation quantizer to evaluate the quality of the two sensors. Considering that the robustness of the system depends on the number of features in data associations, we define a universal degradation factor based on information density as:

$$\boldsymbol{\zeta}(\alpha) = \frac{2}{1+e^{-\alpha}} - 1 \qquad (15)$$

where α is the sensor information density. For LiDAR and thermal degradation factor $\boldsymbol{\zeta}^l$ and $\boldsymbol{\zeta}^t$, we define α_l and α_t as:

$$\alpha_l = \frac{N}{\beta N_{la}},\ \ \alpha_t = \frac{M}{M_{ta}}, \qquad (16)$$

in which N is the number of effective plane features of LiDAR and N_{la} is technically total points; M is the number of edge feature points and M_{ta} is the total pixels in the thermal image. N_{la} and M_{ta} are fixed and can be obtained from the sensor datasheet. N and M can be obtained for the data association per frame. β is a scale parameter representing the original confidence between the LiDAR and thermal measurements, typically set to

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

be less than 1 since LiDAR can directly measure environmental depth information.

4) Multi-Modal Information Adaptive Fusion for State Update: As DaLiTI does not depend on the order of sensor acquisitions, we assume the latest kth frame of observation corresponds to either a thermal or LiDAR state. These states are fully passive and based on real data. Specifically, for the latest factor $\boldsymbol{\zeta}_k$ and its Jacobian $\mathbf{H}_k$. If the latest sensor frame is the ith thermal frame, then $\boldsymbol{\zeta}_k = \boldsymbol{\zeta}_i^t$ and $\mathbf{H}_k = \mathbf{H}_i^t$, else it would be $\boldsymbol{\zeta}_k = \boldsymbol{\zeta}_j^l$ and $\mathbf{H}_k = \mathbf{H}_j^l$ for the latest jth LiDAR frame. Regardless of whether the latest frame is thermal or LiDAR, the kth error state satisfies the following prior distribution:

$$\begin{aligned}\mathbf{x}_k \boxminus \hat{\mathbf{x}}_k = &\left(\hat{\mathbf{x}}_k^\kappa \boxplus \tilde{\mathbf{x}}_k^\kappa\right) \boxminus \hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^\kappa \boxminus \hat{\mathbf{x}}_k + \mathbf{J}^\kappa \hat{\mathbf{x}}_k^\kappa \\ &\backsim \mathcal{N}\left(0, \hat{\mathbf{P}}_k\right),\end{aligned} \tag{17}$$

where $\mathbf{J}^\kappa$ represents the partial derivative of $\left(\hat{\mathbf{x}}_k^\kappa \boxplus \tilde{\mathbf{x}}_k^\kappa\right) \boxminus \hat{\mathbf{x}}_k$ with respect to $\tilde{\mathbf{x}}_k^\kappa$. Then, the Kalman Gain calculation and the update process are applied:

$$\begin{aligned}\mathbf{K} &= \mathbf{P}_{\text{pred}}\mathbf{H}^\top \left(\mathbf{H}\mathbf{P}_{\text{pred}}\mathbf{H}^\top + \mathbf{V}\right)^{-1} \\ \hat{\mathbf{x}}_k^{\kappa+1} &= \hat{\mathbf{x}}_k^\kappa \boxplus \left(-\mathbf{K}\mathbf{z}_k^\kappa - (\mathbf{I} - \mathbf{K}\mathbf{H})\left(\mathbf{J}^\kappa\right)^{-1}\left(\hat{\mathbf{x}}_k^\kappa \boxminus \hat{\mathbf{x}}_k\right)\right).\end{aligned} \tag{18}$$

The above process is deployed iteratively until convergence. Once converged, the optimal state and covariance estimates are computed using a weighted combination of the converged iteration results and the predicted state as:

$$\begin{aligned}\overline{\mathbf{x}}_k &= \tilde{\boldsymbol{\zeta}} \cdot \hat{\mathbf{x}}_k^{\kappa+1} \boxplus (1 - \tilde{\boldsymbol{\zeta}}) \cdot \hat{\mathbf{x}}_k \\ \overline{\mathbf{P}}_k &= \tilde{\boldsymbol{\zeta}} \cdot (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P} + (1 - \tilde{\boldsymbol{\zeta}})\mathbf{P}_{\text{pred}},\end{aligned} \tag{19}$$

where $\hat{\mathbf{x}}_k$ represents predicted state in forward propagation. In cases where the current thermal or LiDAR measurement degrades, the IMU-predicted state is partially considered for state updating rather than merely serving as a prior update as in traditional methods. $\tilde{\boldsymbol{\zeta}}$ is the normalized degradation factor. Specifically, if the latest frame is thermal, $\tilde{\boldsymbol{\zeta}} = \tilde{\boldsymbol{\zeta}}_k^t$; otherwise, $\tilde{\boldsymbol{\zeta}} = \tilde{\boldsymbol{\zeta}}_k^l$. Both $\tilde{\boldsymbol{\zeta}}_k^l$ and $\tilde{\boldsymbol{\zeta}}_k^t$ are normalized by:

$$\tilde{\boldsymbol{\zeta}}_k^l = \tilde{\boldsymbol{\zeta}}_j^l = \frac{\boldsymbol{\zeta}_j^l}{\boldsymbol{\zeta}_j^l + \boldsymbol{\zeta}_i^t}, \quad \tilde{\boldsymbol{\zeta}}_k^t = \tilde{\boldsymbol{\zeta}}_i^t = \frac{\boldsymbol{\zeta}_i^t}{\boldsymbol{\zeta}_j^l + \boldsymbol{\zeta}_i^t}. \tag{20}$$

Normalization is crucial as it balances the importance of LiDAR and thermal measurements. In LiDAR-friendly environments (typically non-smoky with structural features), $\boldsymbol{\zeta}_j^l$ tends to be much higher than $\boldsymbol{\zeta}_i^t$, causing $\tilde{\boldsymbol{\zeta}}_j^l$ to approach 1 and $\tilde{\boldsymbol{\zeta}}_i^t$ to near 0, thus LiDAR dominates. In smoky conditions that affect LiDAR but not thermal, $\boldsymbol{\zeta}_j^l$ drops sharply while $\boldsymbol{\zeta}_i^t$ remains stable, shifting $\tilde{\boldsymbol{\zeta}}_j^l$ towards 0 and $\tilde{\boldsymbol{\zeta}}_i^t$ towards 1, allowing thermal to dominate. This smooth dominance switch is facilitated by quantization and normalization.

IV. Experiments

The proposed DaLiTI was verified with extensive experiments both indoors and outdoors. As shown in Fig. 4, a sensor box integrated with LiDAR (Ouster OS1-64), thermal infrared camera (Xenics Gobi+ 640), IMU (Xsens MTi-G-710) and RGB-D camera (Realsense D435i). A wheeled robot equiped with a multi-sensors device was utilized for real-world experiments. Notably, all experiments were conducted on a laptop powered by an Intel i7-11800H processor without a GPU. For the 64 beams LiDAR with 1024 horizontal resolution, the N_{la} is set to 64×1024, and for the thermal camera with 640×480 pixels, the M_{ta} is set to 640×480. In addition, The parameter β of the quantizer is set to 0.1, which can be verified as a suitable parameter by sensitivity analysis in Fig. 11.

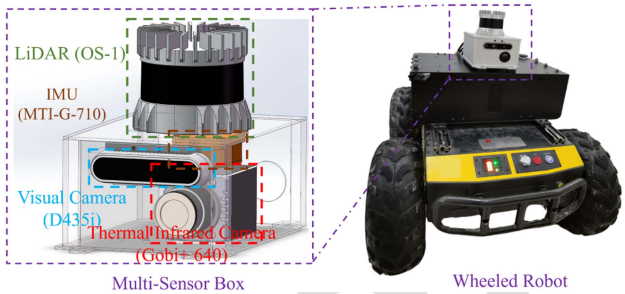


Fig. 4. Robot platform for real-world experiments.

TABLE I
Comparison of RMSE ATE [m] in Conventional Environments

Dataset	indoor-1 (43m)	indoor-2 (73m)	campus-1 (481m)	campus-2 (1027m)	street (750m)	Avg
FAST-LIO2	0.026	0.040	0.643	**0.870**	2.611	0.838
LIO-SAM	0.023	**0.027**	0.776	1.041	2.524	0.878
LVI-SAM	0.026	0.028	0.757	0.987	2.439	0.847
R2LIVE	0.043	0.049	1.251	2.513	2.950	1.361
R3LIVE	0.032	0.061	1.510	1.843	2.744	1.238
FAST-LIVO	0.041	0.039	0.985	1.104	2.577	0.949
DaLiTI	**0.022**	0.032	**0.621**	0.934	**2.438**	**0.809**

Since there are no open-source LiDAR-Thermal-Inertial SLAM frameworks, we can only compare DaLiTI with existing LVIO frameworks. We directly serve the 8-bit mono grayscale mapped thermal image as the LVIO frameworks' visual input. We compared our approach with state-of-the-art LVIO solutions, i.e. LVI-SAM [13], FAST-LIVO [15], R2LIVE [14], R3LIVE [16], and fundamental LIO frameworks like LIO-SAM [11] and FAST-LIO2 [25]. We used their default parameters and evaluated results without loop-closure correction. Accuracy was assessed using the RMSE of Absolute Trajectory Error (ATE), with ground truth provided by the motion capture system for indoor tests and GNSS for outdoor tests.

A. Experiments in Conventional Environments

Before deploying it into degraded environments, we validated DaLiTI in typical settings like indoors, campuses and streets to ensure its feasibility in general applications. Indoor-1 was conducted in general light conditions, while indoor-2 campuses and streets were conducted at night with low illumination. Table I outlines the state estimation errors. The results show that DaLiTI outperformed other open-source solutions in 3 out of 5 sequences and has the lowest average error. Notably, despite using thermal observations, existing LVIO solutions did not significantly improve upon LIO and sometimes had higher errors. The abundant structural features provide sufficient geometric constraints for LiDAR. Integrating low-quality thermal

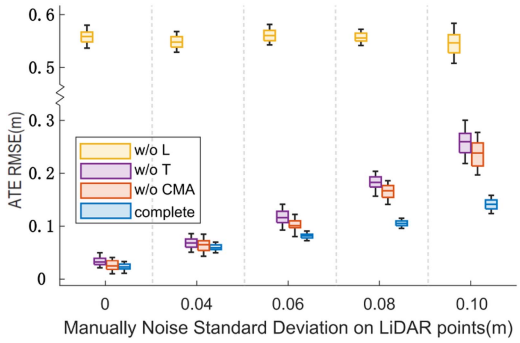
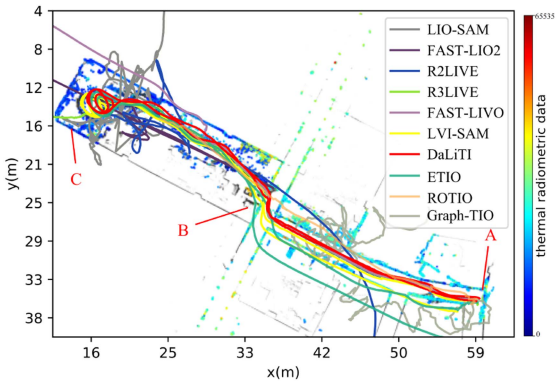
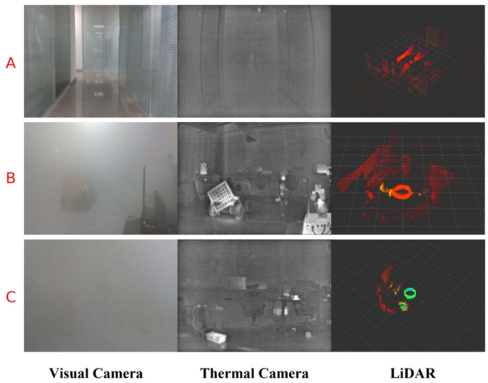


Fig. 5. Impact of manual noise injection on state estimation accuracy.



(a) Trajectories comparison in the indoor degraded environment.



(b) Perception performance at location A, B, and C.

Fig. 6. Comparison of trajectories and perception performance in the indoor degraded environment. (a) Trajectories comparison in the indoor degraded environment. (b) Perception performance at location A, B, and C.

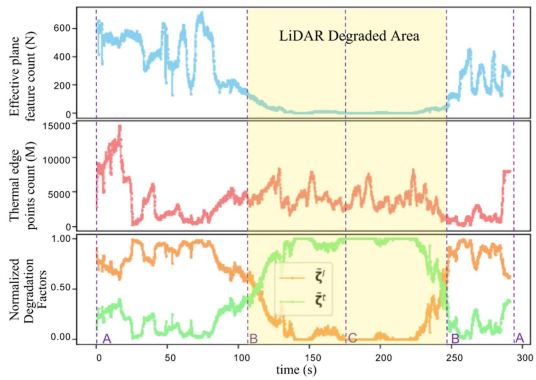


Fig. 7. Degradation analysis in indoor fire scenario.

measurements with high-precision LiDAR in LVIO did not enhance accuracy but degraded performance. In contrast, DaLiTI achieves competitive performance by real-time quality assessment of both modalities and adaptive fusion of measurements. In scenarios rich in structural features, it primarily relies on LiDAR for state updates, supplementing with thermal measurements in structural feature-poor areas to ensure accurate and continuous performance.

To further validate our multi-modal adaptive fusion scheme's robustness, even in the presence of unreliable LiDAR measurements, we manually introduced Gaussian noise $\mathcal{N}(0, \sigma^2)$ to the raw LiDAR point cloud during state estimation. Additionally, we conducted an ablation analysis to evaluate the impact of excluding LiDAR measurement, thermal measurement, or CMA module—on DaLiTI performance. Besides the complete DaLiTI, three variants are investigated:

- DaLiTI (w/o T): System without thermal measurements.
- DaLiTI (w/o L): System without LiDAR measurements.
- DaLiTI (w/o CMA): System without using CMA module.

Taking the indoor-1 sequence as an example, Fig. 5 compares the performance with different LiDAR noise levels (σ, ranging from 0 to 0.1 meters, with a step size of 0.02 meters). For each σ, the variants were executed 10 times for accuracy analysis. The box plot shows that with increased noise variance, the ATE's mean and covariance of the methods that use LiDAR measurements increase. In contrast, introducing thermal measurements and the CMA module can achieve higher accuracy performance and lower covariance bound, proving its effectiveness in enhancing state estimation robustness.

B. Experiments in Degraded Environments

This section evaluated DaLiTI's applicability in challenging environments by constructing two disaster scenarios: an indoor fire and an outdoor chemical plant explosion. We degraded the environment using artificial smoke from fire drill canisters. Considering the perception degradation of LiDAR, we additional introduced three TIO frameworks for comparison: ROTIO [2], ETIO [6] and Graph-TIO [23].

1) Indoor Fire Scenario: As shown in Fig. 6(a), during testing, the wheeled robot started at the corridor entrance, traversed a textureless corridor, entered a dark, smoke-filled warehouse, and returned via the same corridor. The total length of the trajectory is about 110 m. Points A, B, and C mark the starting position, warehouse entrance, and thickest smoke area. Fig. 6(b) demonstrates that the performance of the visual camera and LiDAR significantly declined as the robot ventured deeper into the warehouse. In contrast, the thermal camera was less affected by smoke and illumination, but the image contrast in the corridor with thermal radiation distribution is lower. Since motion capture system cannot provide ground truth in this environment, we evaluated trajectory accuracy using accumulated drift as the robot returned to its starting point. Fig. 6(a) shows that DaLiTI outperformed other methods with the least drift (<0.1 meters), validating its superiority and reliability in degraded environments. Fig. 7 records the number of effective LiDAR plane features N, thermal edges M, and normalized degradation factor $\tilde{\zeta}_k^l, \tilde{\zeta}_k^t$. When the robot moves in the corridor, DaLiTI can extract many plane features from the LiDAR sweep, while the

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

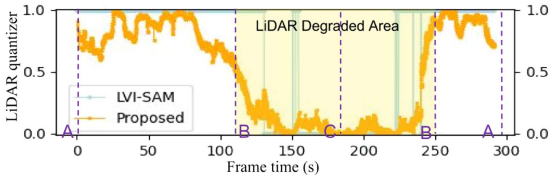


Fig. 8. Comparison of the degradation evaluation results.

edge features extracted from the thermal infrared images are very few due to the low contrast. However, LiDAR degraded due to smoke, which reduced plane features as the robot entered the warehouse. At the same time, thermal camera captured richer edge features from thermal radiation sources unaffected by smoke and darkness. The proposed degradation quantizer rapidly captures changes in the sensor's perception quality, supporting multi-modal adaptive fusion during state update. By leveraging the perception complementarity, DaLiTI achieves accurate and robust state estimation.

In contrast, all other comparison methods, except ETIO and LVI-SAM, failed due to significant trajectory errors in the warehouse. Although the thermal infrared camera is not affected by smoke, its image quality is poor, especially in the corridor area, so the cumulative drift of ETIO is large. LVI-SAM's success can be attributed to utilizing the minimum eigenvalue of the Hessian matrix in the optimization process, called degeneracy factor in [3], for failure detection. When the degeneracy factor falls below a specified threshold, LiDAR odometry is reinitialized, and the corresponding constraints are excluded from the factor graph. Consequently, the state prevents degraded LiDAR measurements in the warehouse from being updated until LiDAR perception recovers. However, the Hessian matrix of the observation model compresses the feature constraint information into a square matrix of a fixed dimension for efficient degeneracy detection, which inevitably leads to information loss, and the eigenvalues are susceptible to measurement noise interference. Additionally, the binary judgment of degeneracy may lead to insufficient utilization of the constraints' information. Instead, we adopt the effective feature proportion for degeneracy evaluation, which can more directly reflect the number of effective constraints.

Fig. 8 compares LiDAR measurement weight for DaLiTI and LVI-SAM. In the smoke-filled warehouse, many LiDAR observations are still considered normal by LVI-SAM, even in the smoky areas, inevitably introducing unreliable observations into the factor graph optimization. In contrast, DaLiTI can swiftly respond to environmental degradation and block LiDAR observations from participating in system state updates, thereby achieving lower accumulated drift than LVI-SAM.

2) Outdoor Chemical Plant Leakage Scenario: To evaluate our system in an outdoor LiDAR degraded and visually challenging environment, we conducted a nighttime experiment where a robot moved around a smoke-obscured water tank (Fig. 1(f)), simulating a reconnaissance mission at a leaking chemical plant. The total length of the trajectory is 60 m. Fig. 9 presents the comparative trajectories, and Table II summarizes the estimation errors. DaLiTI performed best, with its geometric-thermal radiation mapping (Fig. 1(a) and (d)) showcasing well global consistency. Meanwhile, LVI-SAM also delivered comparable performance benefitting from the degradation detection and disposal mechanisms, while other LiDAR-based methods

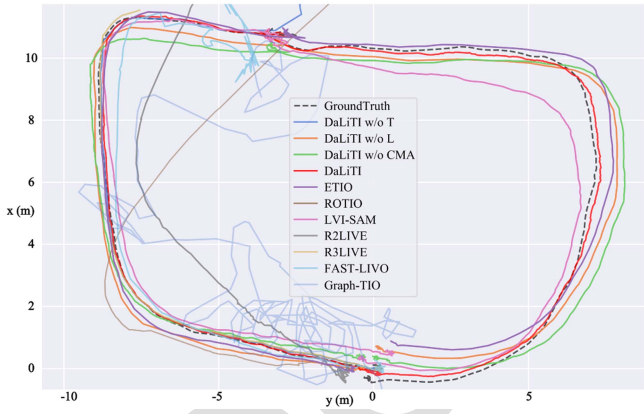


Fig. 9. Trajectories comparison in outdoor degraded environment.

TABLE II
COMPARISON OF RMSE ATE [M] IN OUTDOOR CHEMICAL PLANT LEAKAGE SCENARIO

ETIO	LVI-SAM	DaLiTI			
		w/o T	w/o L	w/o CMA	complete
1.283	1.767	-	1.085	0.938	**0.306**

The blank (-) represents failure.

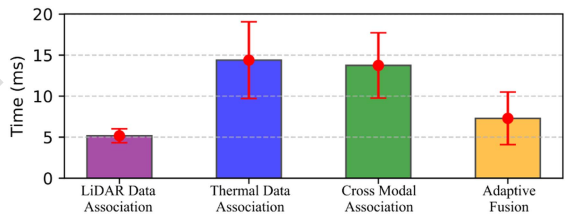


Fig. 10. Time consumption statistics.

struggled with sudden degradation, resulting in significant trajectory deviations.

Table II shows that DaLiTI (w/o T) degrades to a LiDAR-inertial system without thermal measurements, losing smoke interference resistance and failing trajectory estimation. Conversely, DaLiTI (w/o L) still maintains trajectory estimation by leveraging the consistency of the thermal camera through thermal-inertial fusion, even outperforming LVI-SAM. This phenomenon also occur in ETIO, which suggests that in degraded environments, traditional multi-modal fusion may not always enhance state estimation. For DaLiTI (w/o CMA), thermal and LiDAR modalities function independently, leveraging LiDAR measurements for state updates in areas without smoke, resulting in slightly higher accuracy than DaLiTI (w/o L). However, the absence of CMA hinders the enhancement of thermal depth estimation aided by LiDAR. Consequently, the complete DaLiTI, leveraging modal complementarity, achieves the best results, highlighting the importance of an effective fusion scheme in multi-modal SLAM.

To quantitatively evaluate the real-time performance of DaLiTI, we also recorded the time consumption of all the sequences for analysis. The time consumption statistics with a mean and a standard deviation per round for each component are shown in Fig. 10. Parallel computation and multi-threading technologies are adopted for different modules. The state update frequency is determined by the highest execution time in the three threads.

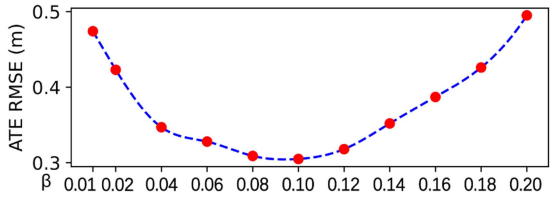


Fig. 11. Comparison of different β in the degradation quantizer.

The maximum output frequency of state estimation is about 30 Hz greater than the image input (25 Hz), so DaLiTI can work in real time without GPU acceleration.

We also conducted a sensitivity analysis on β in the degradation quantizer ((15)). As shown in Fig. 11, with the increase of β, the error of DaLiTI decreases first and then increases. This is because β can balance the modal preferences in adaptive fusion. Excessive β over-weights thermal measurements, impairing LiDAR measurements utilization in conventional conditions, while insufficient β over-relies on LiDAR, weakening degradation resistance. By setting suitable β, the proposed adaptive fusion can make a trade-off between robustness and accuracy.

V. Conclusion and Future Work

In this study, we focus on leveraging the complementarity of multi-modal information to address the challenge of state estimation in degraded environments and introduce DaLiTI, a novel LiDAR-Thermal-Inertial SLAM framework. Building upon an IESKF, DaLiTI uses IMU for forward propagation and LiDAR/thermal measurements for state updating in a loosely coupled manner. To prevent degraded information from propagating throughout the system, we proposed a degradation-aware multi-modal information adaptive fusion scheme, which dynamically weights and fuses LiDAR and thermal measurements based on real-time modal quality analysis. Compared with state-of-the-art methods, extensive experiments demonstrate that DaLiTI achieves competitive performance in conventional environments and maintains robust and accurate state estimation during perceptual degradation. In future work, we will focus on the theoretical model of perception degradation to carry out more precise degradation detection.

References

[1] F. Han, H. Zheng, W. Huang, R. Xiong, Y. Wang, and Y. Jiao, "DAMS-LIO: A degeneration-aware and modular sensor-fusion LiDAR-inertial odometry," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2023, pp. 2745–2751.

[2] S. Khattak, C. Papachristos, and K. Alexis, "Keyframe-based thermal–inertial odometry," *J. Field Robot.*, vol. 37, no. 4, pp. 552–579, 2020.

[3] J. Zhang, M. Kaess, and S. Singh, "On degeneracy of optimization-based state estimation problems," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2016, pp. 809–816.

[4] N. Khedekar, M. Kulkarni, and K. Alexis, "MIMOSA: A multi-modal SLAM framework for resilient autonomy against sensor degradation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2022, pp. 7153–7159.

[5] H. Gil, M. -H. Jeon, and A. Kim, "Fieldscale: Locality-aware field-based adaptive rescaling for thermal infrared image," *IEEE Robot. Automat. Lett.*, vol. 9, no. 7, pp. 6424–6431, Jul. 2024.

[6] Y. Wang, H. Chen, Y. Liu, and S. Zhang, "Edge-based monocular thermal-inertial odometry in visually degraded environments," *IEEE Robot. Automat. Lett.*, vol. 8, no. 4, pp. 2078–2085, Apr. 2023.

[7] J. Zhang and S. Singh, "Visual-LiDAR odometry and mapping: Low-drift, robust, and fast," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2015, pp. 2174–2181.

[8] S. Khattak, H. Nguyen, F. Mascarich, T. Dang, and K. Alexis, "Complementary multi–modal sensor fusion for resilient robot pose estimation in subterranean environments," in *Proc. Int. Conf. Unmanned Aircr. Syst.*, 2020, pp. 1024–1029.

[9] S. Zhao, H. Zhang, P. Wang, L. Nogueira, and S. Scherer, "Super odometry: IMU-centric LiDAR-visual-inertial estimator for challenging environments," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2021, pp. 8729–8736.

[10] J. Lee et al., "Switch-SLAM: Switching-based LiDAR-inertial-visual SLAM for degenerate environments," *IEEE Robot. Automat. Lett.*, vol. 9, no. 8, pp. 7270–7277, Aug. 2024.

[11] T. Shan, B. Englot, D. Meyers, W. Wang, C. Ratti, and D. Rus, "LIO-SAM: Tightly-coupled LiDAR inertial odometry via smoothing and mapping," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2020, pp. 5135–5142.

[12] 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, Aug. 2018.

[13] 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.*, 2021, pp. 5692–5698.

[14] J. Lin, C. Zheng, W. Xu, and F. Zhang, "R ^{2} LIVE: A robust, real-time, LiDAR-inertial-visual tightly-coupled state estimator and mapping," *IEEE Robot. Automat. Lett.*, vol. 6, no. 4, pp. 7469–7476, Oct. 2021.

[15] C. Zheng, Q. Zhu, W. Xu, X. Liu, Q. Guo, and F. Zhang, "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.

[16] J. Lin and F. Zhang, "R ^{3} LIVE: 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.

[17] Z. Yuan, J. Deng, R. Ming, F. Lang, and X. Yang, "SR-LIVO: LiDAR-inertial-visual odometry and mapping with sweep reconstruction," *IEEE Robot. Automat. Lett.*, vol. 9, no. 6, pp. 5110–5117, Jun. 2024.

[18] 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.

[19] S. Vidas and S. Sridharan, "Hand-held monocular SLAM in thermal-infrared," in *Proc. IEEE 12th Int. Conf. Control Automat. Robot. Vis.*, 2012, pp. 859–864.

[20] W. Chen, Y. Wang, H. Chen, and Y. Liu, "EIL-SLAM: Depth-enhanced edge-based infrared-LiDAR SLAM," *J. Field Robot.*, vol. 39, no. 2, pp. 117–130, 2022.

[21] S. Zhao, P. Wang, H. Zhang, Z. Fang, and S. Scherer, "TP-TIO: A robust thermal-inertial odometry with deep thermalpoint," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2020, pp. 4505–4512.

[22] M. R. U. Saputra et al., "DeepTIO: A deep thermal-inertial odometry with visual hallucination," *IEEE Robot. Automat. Lett.*, vol. 5, no. 2, pp. 1672–1679, Apr. 2020.

[23] M. R. U. Saputra, C. X. Lu, P. P. B. de Gusmao, B. Wang, A. Markham, and N. Trigoni, "Graph-based thermal–inertial SLAM with probabilistic neural networks," *IEEE Trans. Robot.*, vol. 38, no. 3, pp. 1875–1893, Jun. 2022.

[24] J. Jiang, X. Chen, W. Dai, Z. Gao, and Y. Zhang, "Thermal-inertial SLAM for the environments with challenging illumination," *IEEE Robot. Automat. Lett.*, vol. 7, no. 4, pp. 8767–8774, Oct. 2022.

[25] W. Xu, Y. Cai, D. He, J. Lin, and F. Zhang, "FAST-LIO2: Fast direct LiDAR-inertial odometry," *IEEE Trans. Robot.*, vol. 38, no. 4, pp. 2053–2073, Aug. 2022.