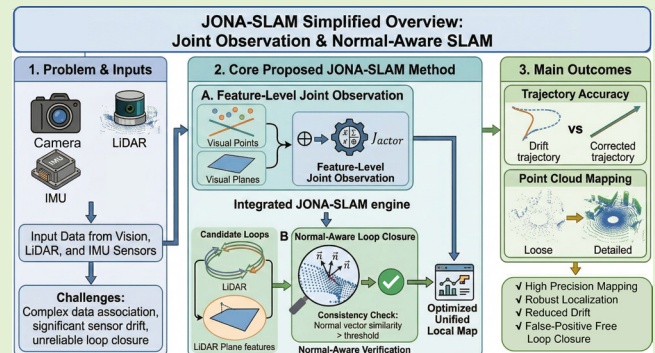


JONA-SLAM: Tightly Coupled SLAM With Joint Observation and Normal-Aware Loop Closure

Qiuwei Jian^{ID}, Qinglu Ma, Xinyu Wang, and Saleem Ullah^{ID}

Abstract—In the field of autonomous navigation, accurate localization and mapping are of paramount importance, especially for autonomous vehicles operating in global navigation satellite system (GNSS)-challenged or denied environments. This article presents joint observation and normal-aware (JONA)-simultaneous localization and mapping (SLAM), a tightly coupled light detection and ranging (LiDAR)-visual-inertial SLAM system designed to enhance state estimation robustness for autonomous driving platforms in complex scenarios such as urban canyons and tunnels. Building upon our previous research on point-line feature extraction, the proposed system strengthens spatial constraints by constructing a LiDAR-visual joint observation factor. This factor enables the tight coupling of LiDAR-derived geometric information with visual texture features at the feature level, effectively addressing the observation degradation common in GNSS-denied areas. Furthermore, a normal-aware hybrid loop closure detection method is proposed, incorporating point cloud normal consistency verification to significantly improve loop identification reliability and suppress false positives in challenging semistructured scenes. Evaluation results on 22 public sequences and a self-collected campus dataset demonstrate that JONA-SLAM achieves superior or competitive performance compared to state-of-the-art methods. Across 23 experimental test cases, the proposed method reduces the trajectory error [root mean square error (RMSE)] by an average of 35.02%. While maintaining real-time performance, the system demonstrates superior environmental adaptability and global consistency, providing a high-precision and robust localization solution for autonomous driving and versatile mobile platforms in degraded environments.

Index Terms—Joint observation factor, light detection and ranging (LiDAR)-visual-inertial fusion, normal-aware loop closure, tightly coupled simultaneous localization and mapping (SLAM).



I. INTRODUCTION

OVER the past two decades, simultaneous localization and mapping (SLAM) has become a fundamental technology for autonomous systems and intelligent vehicles [1], [2]. While early single-sensor setups are cost-effective, they face

inherent limitations: cameras are sensitive to illumination and textureless scenes, while light detection and ranging (LiDAR) is susceptible to sparse geometry and motion distortion. Particularly for autonomous driving in complex metropolitan areas, single-sensor approaches often fail to meet the stringent reliability demands in global navigation satellite system (GNSS)-denied or challenged environments, such as urban canyons and tunnels, where satellite signals are frequently blocked by high-rise buildings or underground structures [3].

While tightly coupled multisensor fusion has gained significant traction due to its superior performance in jointly optimizing raw measurements [4], [5], existing frameworks still struggle with several critical bottlenecks when transitioning between open roads and these constrained areas. Specifically, many systems lack sufficient feature-level integration to anchor visual features with LiDAR geometry in low-texture areas, while the structural information offered by line features remains underdeveloped [6]. Furthermore, loop-closure detection in large-scale urban settings remains susceptible to false matches due to repetitive environmental structures, compromising long-term global consistency [7], [8].

Received 9 February 2026; revised 26 February 2026; accepted 26 February 2026. Date of publication 12 March 2026; date of current version 16 April 2026. This work was supported in part by the Science and Technology Innovation Key Research and Development Program of Chongqing under Grant CSTB2022TIAD-STX0003, in part by Chongqing Natural Science Foundation Project under Grant CSTB2023NSCQ-MSX0551, and in part by the Research and Innovation Program for Graduate Students in Chongqing under Grant CYS25539. The associate editor coordinating the review of this article and approving it for publication was Dr. Zengke Li. (Corresponding author: Qiuwei Jian.)

Qiuwei Jian is with the School of Mechatronics and Vehicle Engineering, Chongqing Jiaotong University, Chongqing 400074, China (e-mail: jqw@mails.cqjtu.edu.cn).

Qinglu Ma and Xinyu Wang are with Chongqing Jiaotong University, Chongqing 400074, China (e-mail: qlm@cqjtu.edu.cn; xyw@mails.cqjtu.edu.cn).

Saleem Ullah is with the Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Khan, Punjab 64200, Pakistan (e-mail: saleem.ullah@kfuait.edu.pk).

Digital Object Identifier 10.1109/JSEN.2026.3669355

To address these challenges, this article presents JONA-SLAM (Joint Observation and Normal-Aware Loop Closure), a robust tightly coupled LiDAR-visual-inertial framework tailored for autonomous navigation in degraded environments. Expanding on our previous point-line extraction work [9], this system introduces a LiDAR-Visual Joint Observation Factor for seamless feature-level fusion by treating LiDAR geometry as dynamic observations. Furthermore, a normal-aware hybrid loop closure module is proposed, combining visual appearance with surface normal verification to ensure global consistency even in scenes with high geometric ambiguity. Our primary contributions are as follows.

- 1) We propose a unified observation model achieving tight feature-level fusion by integrating LiDAR spatial constraints with visual features, specifically enhancing robustness in scenes with sparse visual or geometric textures.
- 2) We introduce a robust loop detection mechanism leveraging LiDAR normal-consistency verification to eliminate false positives in complex environments.
- 3) We implement joint observation and normal-aware (JONA)-SLAM and validate its effectiveness through extensive evaluations on various datasets. The results demonstrate that our framework consistently outperforms state-of-the-art methods, such as LiDAR-visual-inertial odometry via smoothing and mapping (LVI-SAM), by providing superior localization accuracy and robustness.

The remainder of this article is organized as follows. Section II reviews related work. Section III introduces the JONA-SLAM. Section IV presents experimental results. Conclusions and future work are summarized in Section V.

II. RELATED WORK

In recent years, multisensor fusion SLAM has attracted increasing attention from researchers due to its potential for higher precision and robustness [10]. Current fusion methods generally fall into two categories: loosely coupled and tightly coupled [11]. While loosely coupled methods offer advantages in modularity and computational simplicity, tightly coupled methods typically achieve superior accuracy and robustness by jointly optimizing raw sensor measurements.

Among tightly coupled frameworks, Kalman filter (KF)-based methods are widely used due to their high computational efficiency [12]. For instance, improvements to the unscented KF (UKF) via Schmidt orthogonalization and square-root decomposition enhance SLAM performance [13]. Fast-LIO employs an iteratively extended KF (EKF) to tightly fuse LiDAR features with IMU data for robust navigation, also introducing a state-dependent Kalman gain for optimization [14]. Its successor, Fast-LIO2, enables direct raw point registration and real-time map updates without feature extraction [15]. However, recursive filtering-based methods often face challenges in complex autonomous driving scenarios. They are prone to linearization errors in highly non-linear environments and lack a native mechanism for incorporating global constraints, such as loop closures, to ensure long-term consistency [16], [17].

To address these gaps, optimization-based frameworks, particularly factor graph optimization, have become a dominant paradigm. Tightly coupled optimization methods achieve superior performance by integrating diverse constraints. For example, RaI-SLAM [18] combines automotive radar with IMU for accurate localization without GNSS, while Switch-SLAM [19] employs a switching structure between LiDAR and visual odometry to maintain robustness. Focusing on multimodal integration, LVI-SAM [20] incorporates LiDAR, visual, and inertial odometry within a factor graph, allowing its subsystems to operate independently or cooperatively. In a different vein, R3LIVE++ [21] emphasizes dense map construction by maintaining separate geometric and photometric mapping threads. Beyond these sensors, the fusion of LiDAR, wheel odometry, and red-green-blue-depth (RGB-D) cameras has yielded improvements in robustness [22]. A notable instance is LiDAR-aided multi-view (LAMV)-SLAM, which tightly couples different sensor data to estimate and refine errors, enabling real-time dense mapping without graphics processing unit (GPU) acceleration [23]. Alternatively, another novel approach involves running sensors independently in the front-end and in $\mathbf{p}_f(u_f, v_f)$ incorporating all measurements into back-end optimization [24]. Unlike recursive filters, these factor graph-based systems allow for iterative relinearization and the seamless integration of global loop closure factors, ensuring a globally consistent trajectory.

Despite advancements, existing systems struggle with noise-sensitive matching and unreliable loop closure. We present JONA-SLAM, a tightly coupled LiDAR-visual-inertial framework designed for robust localization in complex autonomous driving. By integrating LiDAR-assisted depth recovery with a unified dual-modal observation model, JONA-SLAM maintains state estimation in textureless or geometrically ambiguous scenes. Additionally, our hybrid loop closure, fusing point cloud normals with visual features, suppresses false matches to ensure global consistency.

III. METHODOLOGY

This article presents JONA-SLAM, a tightly coupled LiDAR-visual-inertial SLAM system. While the front-end utilizes the point-line feature extraction from our previous work, this research significantly extends the framework by introducing a LiDAR-visual joint observation factor for deep feature-level fusion and a normal-aware hybrid loop closure mechanism for global consistency. The integrated pipeline of the proposed system is illustrated in Fig. 1.

A. LiDAR-Assisted Depth Recovery and Tightly Coupled Observation Model

To achieve deep fusion between the visual and LiDAR subsystems, this system constructs a LiDAR-assisted visual feature depth recovery and tightly coupled observation framework. The framework first utilizes the precise geometric information from LiDAR point clouds to provide a metric scale for visual features. These multimodal observations are then uniformly encoded into “LiDAR-Visual Joint Observation Factors” and fed into a sliding window optimizer, enabling tightly coupled at the state estimation level.

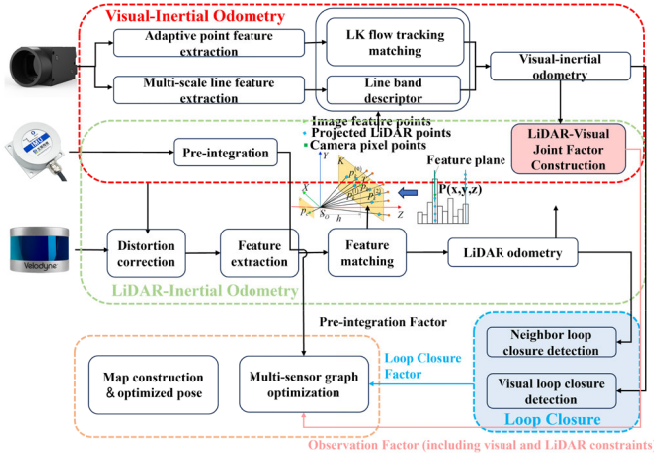


Fig. 1. JONA-SLAM framework diagram. The red and blue parts represent the novel contributions of this article.

For the visual point feature, we first project the LiDAR point cloud onto the image plane to generate a sparse depth map $D_{space}(u, v)$. We then search for the K nearest projected neighbors of $\mathbf{p}_f$ in D_{space} and calculate the standard deviation σ_z of their depths. To further ensure the precision of the metric scale and filter out geometric uncertainties, a local planarity criterion based on eigenvalue analysis is introduced. The planarity factor η is defined via the eigenvalues $\lambda_1, \lambda_2, \lambda_3$ of the neighborhood covariance matrix

$$\eta > \frac{\lambda_2 - \lambda_3}{\lambda_1}. \quad (1)$$

Only when $\eta > \tau_{planar}$ and $\sigma_z < \tau_{flat}$, the median depth is taken as d_f . These thresholds are established based on universal physical-geometric constraints: τ_{planar} is set to 0.8 to ensure sufficient planarity of the local neighborhood, while τ_{flat} is set to 0.05 m to strictly limit geometric fluctuations. Rather than being tuned for specific sequences, these parameters are kept constant across all test scenarios. Since they are derived from fundamental local geometric properties, they offer strong generalization across various outdoor driving environments and effectively mitigate depth misassociation at object boundaries. Otherwise, if the local geometry exhibits high curvature, the candidate points are back-projected to fit a local plane using random sample consensus (RANSAC), and the median depth of points conforming to the plane model is selected. For visual point features that cannot be directly associated, depth interpolation is performed based on successfully associated surrounding points using Delaunay triangulation. For the line feature L_{img} , we first uniformly sample along the line segment and map its projection rays into the LiDAR coordinate system to form a search volume, from which local point cloud $\mathcal{P}_L$ is collected. A supporting plane is then fit using RANSAC

$$\prod : \mathbf{n}^T \mathbf{X} + d = 0. \quad (2)$$

The projection rays of the line segment's start and end pixels are then intersected with plane $\prod$ to obtain the 3-D endpoints $\mathbf{P}_s^L, \mathbf{P}_e^L$ in the LiDAR coordinate system. Finally, the endpoints are transformed into the world coordinate system to derive the

3-D line segment L_{3D} . Depth consistency across all sampling points along the line segment is checked to filter out abnormal segments that span depth-discontinuous boundaries. Through the mechanism described above, we obtain a collection of 3-D visual features with an accurate metric scale

$$\mathcal{F}_{LV} = \{\mathbf{P}_i^{3D}, \mathbf{P}_j^{3D}\}. \quad (3)$$

To deeply integrate the recovered depth information into state estimation, we construct a LiDAR-Visual Joint Observation Factor. This factor treats the LiDAR-derived depth not as a fixed value, but as a soft constraint with uncertainty, respecting both visual reprojection and LiDAR geometric consistency. For a 3-D point feature, its joint observation error in the k th frame is defined as follows:

$$\mathbf{e}_{LV,k,i} = \begin{bmatrix} \mathbf{e}_{k,i}^{vis} \\ \mathbf{e}_{k,i}^{lidar} \end{bmatrix}. \quad (4)$$

The visual reprojection error term ensures that the projection of the feature onto the image plane is consistent with the observation

$$\mathbf{e}_{k,i}^{vis} = \mathbf{z}_{k,i}^{vis} - \pi(\mathbf{T}_{WC,k} \cdot \mathbf{P}_i^{3D}) \quad (5)$$

where $\mathbf{z}_{k,i}^{vis}$ is the pixel observation, $\pi(\cdot)$ is the camera projection function, and $\mathbf{T}_{WC,k}$ is the camera pose of the k th frame. To ensure that the 3-D position of the feature is consistent with the geometric structure of the local point cloud, the point-to-plane distance is adopted for point features

$$\mathbf{e}_{k,i}^{lidar} = \mathbf{n}^T (\mathbf{T}_{WL,k} \cdot \mathbf{P}_i^{3D} - \mathbf{q}) \quad (6)$$

where $\mathbf{n}$ and $\mathbf{q}$ are the plane parameters fit from the LiDAR point cloud in the vicinity of feature point $\mathbf{P}_i^{3D}$, and $\mathbf{T}_{WL,k}$ is the LiDAR pose of the k th frame. For line features, the error term is constructed in a similar manner, using the point-to-line (or point-to-plane) distance by sampling points along the line segment. This metric is independent of the specific positions of endpoints, thereby ensuring robustness even when endpoints vanish or shift due to viewpoint changes and occlusions.

The state estimation update $\delta\mathcal{X}$ is derived by minimizing the joint energy function via the Gauss-Newton method. To demonstrate the substantive coupling between LiDAR geometry and visual manifolds, the Joint Jacobian matrix $\mathbf{J}_{JO}$ is explicitly derived as

$$\mathbf{J}_{LV} = \begin{bmatrix} \mathbf{J}_{vis} \\ \mathbf{J}_{geo} \end{bmatrix} = \begin{bmatrix} \frac{\partial \mathbf{r}_{reproj}}{\partial \delta \xi} \\ \mathbf{n}^T [-[\mathbf{R}_k \mathbf{P}_w + \mathbf{t}_k]_{\times} \quad \mathbf{n}^T] \end{bmatrix} \quad (7)$$

where $[\cdot]_{\times}$ denotes the skew-symmetric matrix. This formulation allows the LiDAR geometric gradient $\mathbf{J}_{vis}$ to directly refine the camera pose manifold. Such integration ensures system observability even in texture-less environments where the visual Jacobian $\mathbf{J}_{geo}$ becomes rank-deficient. The covariance matrix Σ_{LV} of this joint factor is a block-diagonal matrix, containing the visual observation uncertainty Σ_{vis} and the LiDAR geometric fitting uncertainty Σ_{geo} , respectively.

Algorithm 1 Enhanced Loop Closure Detection with Normal Verification

Input: Current frame features F_t , historical keyframes $K_{1:t-1}$, time threshold T_{th} , spatial threshold D_{th} , normal threshold Γ_{th}
Output: Loop closure constraints $\mathcal{C}$, updated keyframe database

```

function DETECT_LOOPS ( $F_t, K_{1:t-1}$ )
  ( $P_t, \mathbf{n}_t$ )  $\leftarrow$  EXTRACT_LIDAR_FEATURES ( $F_t$ )
   $V_t \leftarrow$  EXTRACT_VISUAL_FEATURES ( $F_t$ )
   $K_v \leftarrow$  SEARCH_VISUAL_KEYFRAMES( $V_t, K_{1:t-1}$ )
  if  $K_v \neq \emptyset$  then
    if LK_TRACKING ( $V_t, K_v$ ) and  $|t - t_{K_v}| > T_{th}$ 
      and  $\|\mathbf{p}_t - \mathbf{p}_{K_v}\| < D_{th}$  then
         $\psi \leftarrow$  COMPUTE_NORMAL_SIMILARITY ( $\mathbf{n}_t, \mathbf{n}_{K_v}$ )
        if  $\psi > \Gamma_{th}$  then
           $\mathbf{e}_{joint} \leftarrow$  CONSTRUCT_JOINT_FACTOR(
             $V_t, P_t, K_v, \mathbf{n}_t, \mathbf{n}_{K_v}$ )
           $\mathcal{C} \leftarrow \mathcal{C} \cup \{\mathbf{e}_{joint}\}$ 
        end if
      end if
    end if
  return  $\mathcal{C}$ 
end function
  
```

B. Keyframe Selection and Loop Closure Detection

When the system state change relative to the previous keyframe exceeds a threshold and feature counts meet requirements, the frame is selected as a keyframe. We tightly couple and store the LiDAR point cloud, visual features, and odometry poses. To enhance reliability, this article proposes a hierarchical multimodal fusion-based loop closure detection method. Crucially, instead of an independent parallel search, visual candidates directly trigger a LiDAR-based geometric cross-validation. By incorporating LiDAR normal information and a structural similarity score under unified spatiotemporal constraints, the system strengthens geometric constraints. This approach effectively reduces single-modal mismatches and improves global consistency in complex scenarios. The optimized process is detailed in Algorithm 1.

As shown in Algorithm 1, the visual loop closure module performs dual verification through LK optical flow tracking and temporal consistency checks on candidate image matches. Only keyframes satisfying these criteria are accepted as valid loop candidates. Crucially, instead of an independent parallel search, these visual candidates directly trigger a LiDAR-based geometric cross-validation. Specifically, the LiDAR consistency is quantified by a weighted similarity score ψ , which evaluates both the orientation alignment and the structural distribution of local point clusters

$$\psi = \alpha \cdot \exp\left(-\frac{(1 - \mathbf{n}_{curr}^\top \mathbf{n}_{hist})^2}{2\sigma_n^2}\right) + (1 - \alpha) \cdot \frac{\text{tr}(\mathbf{S}_{curr} \mathbf{S}_{hist})}{\|\mathbf{S}_{curr}\|_F \|\mathbf{S}_{hist}\|_F} \quad (8)$$

where $\mathbf{n}$ is the average surface normal, and $\mathbf{S}$ represents the scatter matrix characterizing the local geometric topology.

$\alpha \in [0, 1]$ is a balancing factor that regulates the relative importance between normal orientation consistency and geometric structural similarity. In our experiments, α is set to 0.5 and the acceptance threshold is set to $\Gamma_{th} = 0.85$. These values are established based on the geometric rigidity of static scenes to ensure that only loop candidates with highly consistent surface orientations are accepted. To demonstrate robustness, these parameters remain constant across all evaluations without per-sequence tuning. Since this gatekeeping mechanism is derived from fundamental geometric properties rather than dataset-specific features, it offers strong generalization and effectively prevents false-positive loops from corrupting the factor graph. By incorporating these unified spatiotemporal constraints, the dual-modality detection framework effectively suppresses false positives in perceptually ambiguous environments, significantly improving both recognition accuracy and global consistency.

C. Factor Graph Optimization

Factor graph optimization facilitates efficient state estimation in SLAM by leveraging sparsity and incremental smoothing. This framework ensures global trajectory consistency while maintaining real-time computational performance. In this article, the factor graph is constructed using the IMU preintegration factor, the LiDAR-Visual Joint Observation Factor, and the loop closure constraint factor. These factors are integrated into a tightly coupled optimization framework, resulting in the optimized global pose estimation and a globally consistent point cloud map.

As described previously, the IMU preintegration factor is obtained by integrating the IMU measurements between two consecutive keyframes. The LiDAR-Visual Joint Observation Factor is derived by tightly coupling LiDAR-derived spatial constraints with visual point-line features, effectively bridging multimodal observations within the sliding window optimizer. The loop closure factor is created by matching the current keyframe with candidate loop closure keyframes. These factors are incorporated into the factor graph and optimized to yield the global pose and point cloud map. The objective function for global optimization is defined as

$$\left\{ \begin{array}{l} \arg \min_{\mathcal{X}} \left\{ \sum_{(i,j) \in I_{pre}} \|\mathbf{e}_{ij}\|_{\Sigma_{pre}}^2 + \sum_{(i,j) \in I_{joint}} \|\mathbf{e}_{ij}\|_{\Sigma_{joint}}^2 \right. \\ \quad \left. + \sum_{(i,j) \in I_{loop}} \|\mathbf{e}_{ij}\|_{\Sigma_{loop}}^2 \right\} \\ \mathbf{e}_{ij} = \ln \left(T_{ij} (T_i^w)^{-1} T_j^w \right) \end{array} \right. \quad (9)$$

where $\mathcal{X}$ represents the set of all state variables to be optimized, including the poses, velocities, and IMU biases of the keyframes. Unlike conventional methods that treat sensors independently, our joint observation factor ($\mathbf{e}_{ij}, (i, j) \in Joint$) unifies visual re-projection errors and LiDAR geometric residuals into a single cost term. To balance the scale differences

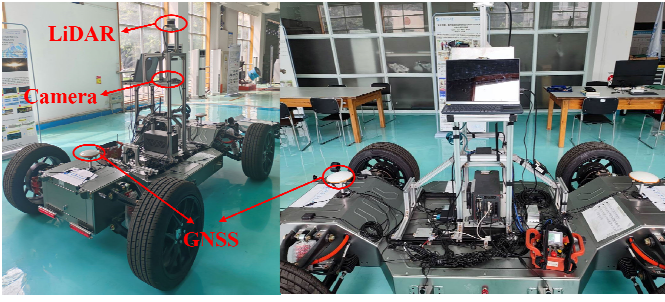


Fig. 2. Experimental equipment.

and noise characteristics between these modalities, the residuals are weighted by the inverse of their covariance matrices. Specifically, the joint covariance Σ_{LV} is designed to be adaptive: while the visual covariance Σ_{vis} is fixed based on reprojection noise, the LiDAR geometric covariance Σ_{geo} is dynamically adjusted according to the depth uncertainty σ_z and planarity η derived in Section III-A. This mechanism assigns higher confidence to features with precise planar structures while accurately modeling the uncertainty of features with higher depth variance. Through global optimization, the updated poses and IMU biases are obtained. The refined IMU bias is subsequently fed back to the preintegration module to calibrate future measurements, ensuring the temporal consistency of the factor graph. After n keyframes, the system resets the factor graph optimizer to bound computational complexity and maintain real-time performance.

IV. EVALUATION

In this section, the performance of the proposed JONA-SLAM is evaluated to validate the tightly coupled framework. The experimental platform is the Apollo D-KIT lite, where GNSS provides ground truth for positioning accuracy evaluation. We utilize KITTI, M2DGR, and a self-collected campus dataset for testing. KITTI Sequence 00 is used to evaluate long-term drift suppression in large-scale residential areas, while Sequence 08 is selected to test the robustness of the normal-aware loop closure module under challenging opposite-direction travel. The M2DGR dataset, which contains features such as severe lighting variations, extremely low illumination, and geometrically degraded hallways, provides an excellent benchmark to evaluate the performance of JONA-SLAM in extreme scenarios. Additionally, the self-collected campus dataset covers a 500-m loop representing a complex unstructured environment with frequent pedestrians and multiple speed bumps. These bumps induce significant platform jitter and point cloud distortion, providing a rigorous test for the LiDAR-visual joint optimization and the JOHL module in maintaining global consistency under erratic motion. Equipment details are shown in Fig. 2.

A. Experiments on the KITTI Dataset

To evaluate the generalization capability of JONA-SLAM, we first benchmarked it against the classic visual-LiDAR-inertial framework, LVI-SAM, using the KITTI dataset. As a widely recognized benchmark, KITTI effectively simulates

TABLE I
TRAJECTORY ERROR COMPARISON (RMSE) (UNIT: m)

Seq	LVI-SAM		JONA-SLAM		
	w/o	w/	w/o	w/o Normal	w/
00	7.36	3.69	3.39	1.95	1.71
05	4.48	3.47	3.22	1.35	1.08
07	1.18	0.81	0.52	0.38	0.23
08	7.41	4.56	3.10	1.98	1.68

TABLE II
QUANTITATIVE EVALUATION: F1-SCORE COMPARISON, ABLATION STUDY, AND AVERAGE COMPUTATIONAL TIME (ms)

Seq	00	05	07	08	Time
DBoW2	0.842	0.761	0.611	0.568	12.5
ScanContext	0.932	0.896	0.665	0.621	22.4
ISC	0.908	0.924	0.638	0.678	18.2
V-only(DBoW2 + LK)	0.875	0.875	0.875	0.875	15.2
L-only(SC+ Ψ)	0.943	0.915	0.692	0.656	24.8
Simple Fusion (Parallel)	0.935	0.902	0.680	0.635	37.6
JONA-SLAM (w/o Normal)	0.951	0.938	0.725	0.668	16.2
JONA-SLAM	0.968	0.961	0.885	0.711	17.8

complex real-world driving environments for evaluating trajectory root mean square error (RMSE). To further evaluate the impact of different loop closure strategies on trajectory accuracy, we extended the JONA-SLAM results into three configurations: 1) w/o Loop (odometry mode); 2) w/o Normal (loop closure enabled but without normal-consistency verification); and 3) full JONA-SLAM. This multistage comparison isolates the specific accuracy gains provided by the normal-aware constraints beyond basic sensor fusion and standard loop closure. Comparative results are detailed in Table I.

As shown in Table I, JONA-SLAM achieves superior performance across nearly all sequences, primarily due to the proposed LiDAR-visual joint factor that maintains high observability in feature-degraded environments. The incremental precision gains across the three configurations highlight the decisive impact of our loop closure strategy: while the JONA-SLAM (w/o Normal) variant significantly reduces RMSE compared to the w/o Loop baseline by eliminating long-term drift, the full JONA-SLAM further refines the global pose. This is particularly evident in Sequences 07 and 08, where the normal-aware mechanism acts as a stringent gatekeeper to prevent false-positive loops from corrupting the backend factor graph, thereby ensuring superior global consistency and suppressing trajectory distortions. Furthermore, to evaluate global consistency, we compared our normal-aware hybrid loop closure strategy against DBoW2 [25], Scan Context [26], and ISC [27] using the maximum F_1 -score (the harmonic mean of precision and recall) as a comprehensive metric to reflect the peak balance between detection accuracy and coverage. The quantitative evaluation also includes an extended multistage ablation study. Specifically, JONA-SLAM is compared against visual-only (V-only), LiDAR-only (L-only), a loosely coupled simple fusion baseline, and the JONA-SLAM (w/o Normal) variant. Detailed comparative results and ablation metrics are summarized in Table II.

As summarized in Table II, JONA-SLAM achieves a superior average F_1 -score of 0.881, effectively eliminating trajectory drift and reducing global RMSE in Table I. The performance gap between JONA-SLAM and the w/o Normal variant (0.885 versus 0.725 in Sequence 07) directly highlights the necessity of geometric topology verification. In Sequence 07, characterized by repetitive roadside layouts similar to tunnels or corridors, JONA-SLAM (0.885) significantly outperforms DBoW2 (0.611), Scan Context (0.665), and the w/o Normal baseline (0.725). Even in the most challenging Sequence 08 featuring reverse visits, JONA-SLAM maintains a high F_1 -score of 0.711, whereas DBoW2 drops to 0.568. The comparative results demonstrate that our normal-aware scoring mechanism acts as a stringent gatekeeper, utilizing LiDAR geometric surface consistency to eliminate false positives caused by appearance-based perceptual aliasing in repetitive structures. Furthermore, JONA-SLAM ensures real-time feasibility in large-scale scenarios (e.g., Sequence 00) with an average latency of 17.8 ms, nearly twice as fast as the 37.6 ms of Simple Fusion. Notably, the addition of the normal-consistency check introduces negligible computational overhead, increasing the latency by only 1.6 ms compared to the w/o Normal version (16.2 ms). This efficiency stems from our vision-triggered logic, which avoids redundant global searching in dense point clouds by performing LiDAR-based validation only when potential visual candidates are identified.

B. Experiments on the M2DGR Dataset

To further evaluate the robustness and generalization capability of JONA-SLAM in complex environments, we conducted comparative experiments on the M2DGR dataset. Unlike the smooth profiles of KITTI, M2DGR features aggressive motion, high-dynamic transitions, and complete darkness. Our method is benchmarked against state-of-the-art (SOTA) frameworks, including R3LIVE++ (a radiance-map-based direct method) and dast LiDAR-inertial-visual odometry 2 (FAST-LIVO2) [28] (an iterated extended kalman filter (IEKF)-based efficient fusion baseline). The visualized trajectory comparisons, generated via the python package for the evaluation of odometry and SLAM (EVO) tool, are presented in Fig. 3 to provide an intuitive performance assessment.

As illustrated by the trajectory plots and magnified details in Fig. 1, JONA-SLAM exhibits superior resilience against the extreme challenges of the M2DGR dataset. In the door_01/02 sequences, abrupt indoor-outdoor lighting transitions cause severe visual feature degradation; while LVI-SAM and FAST-LIVO2 suffer from significant trajectory jitter at transition points, JONA-SLAM maintains continuity by reinforcing visual data with LiDAR-derived geometric constraints. In the low-light gate_02 loop, where appearance-based methods fail due to texture scarcity, our normal-based matching strategy ensures precise loop alignment, eliminating the ghosting and misalignment seen in LVI-SAM. While R3LIVE++ excels in stable outdoor environments, JONA-SLAM demonstrates superior global consistency during high-dynamic movements and severe illumination shifts. Quantitative absolute pose error (APE) metrics are summarized in Table III.

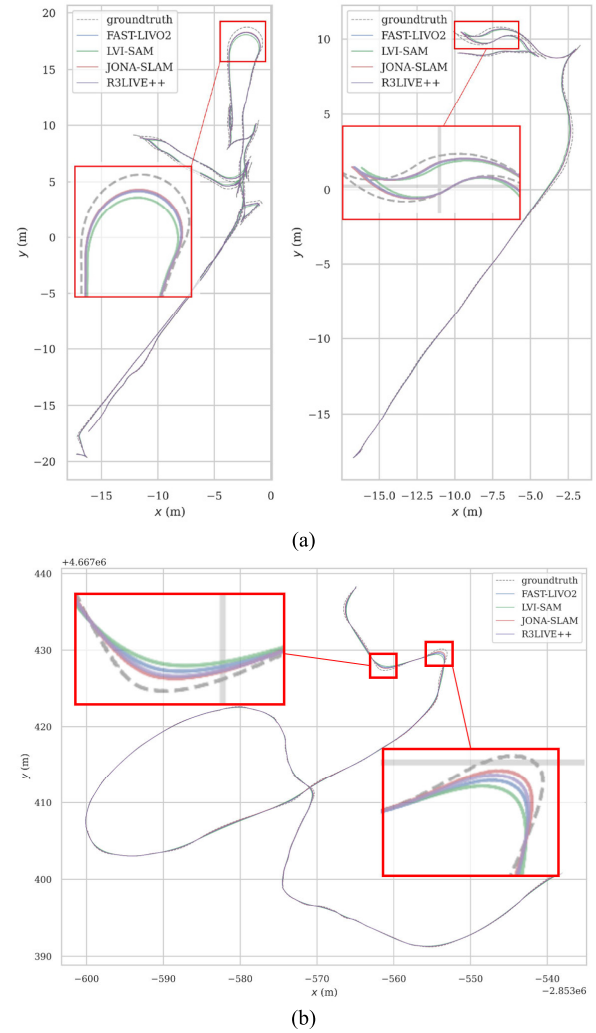


Fig. 3. Trajectory comparisons. (a) door_01 and door_02 and (b) gate_02.

TABLE III
RMSE OF APE IN M2DGR DATASETS IN METERS

Sequence	LVI-SAM	FAST-LIVO2	R3LIVE++	JONA-SLAM
door_01	0.265	0.215	0.201	0.178
door_02	0.198	0.165	0.158	0.141
gate_01	0.152	0.132	0.125	0.106
gate_02	0.357	0.332	0.315	0.292
gate_03	0.154	0.102	0.098	0.083
hall_01	0.252	0.218	0.208	0.198
hall_02	0.282	0.258	0.241	0.204
hall_03	0.351	0.282	0.245	0.196
hall_04	0.855	0.812	0.792	0.725
hall_05	1.113	0.942	0.842	0.754
room_01	0.141	0.118	0.102	0.079
room_02	0.132	0.098	0.102	0.113
room_03	0.162	0.148	0.134	0.124
street_02	3.46	3.12	2.65	3.112
street_03	0.138	0.122	0.112	0.099
street_04	0.932	0.415	0.285	0.315
street_05	0.345	0.305	0.295	0.298
street_06	0.384	0.352	0.355	0.342
Mean	0.531	0.452	0.406	0.398

Quantitative metrics in Table III show JONA-SLAM achieving a mean RMSE of 0.398 m, highlighting its superior adaptability. In loop-heavy sequences like the gate series

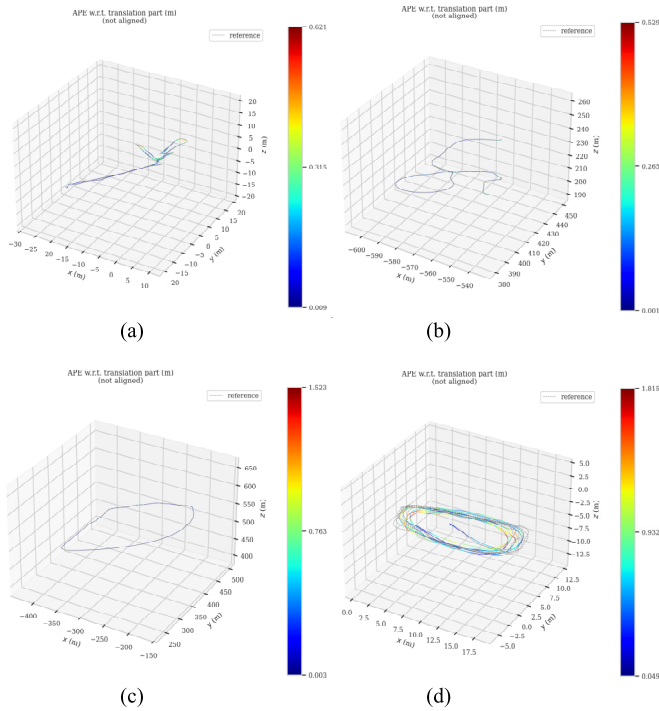


Fig. 4. JONA-SLAM APE error maps of JONA-SLAM. (a) door_01, (b) roomdark_02, (c) gate_04, and (d) hall_05.

and hall_03, our method outperforms others by leveraging stable environmental normal structures. Even when visual texture fails due to illumination shifts, normal-constrained loop closure ensures precise alignment and eliminates cumulative drift. In high-motion scenarios such as hall_04 and hall_05, where motion blur disrupts direct and feature-based methods, our LiDAR-reinforced approach yields results superior to R3LIVE++. While R3LIVE++ remains highly precise in open settings like street_02, JONA-SLAM proves more resilient against dynamic interference and visual degradation. To validate the normal constraints, we selected four challenging sequences: hall_05 (geometric degeneracy), roomdark_02 (visual failure), door_01 (abrupt lighting), and street_04 (large-scale loop). This multidimensional ablation study confirms the impact of normal vectors on overall system robustness. The APE error maps are illustrated in Fig. 4.

As shown in Fig. 4, JONA-SLAM maintains high spatial consistency across four representative degraded sequences despite long-distance maneuvers and complex environmental transitions. The error distribution remains uniform, with the vast majority of the trajectories staying within the dark blue regions, representing low error magnitudes. Particularly during the straight-line acceleration phases in hall_05 and the aggressive turns in roomdark_02, the trajectories exhibit no significant red or yellow spikes in error. This qualitatively confirms the system's capability to maintain robust pose estimation even under severe visual texture loss or geometric degeneracy. Detailed quantitative improvements in accuracy before and after the introduction of normal constraints are summarized in Table IV.

TABLE IV
ABLATION ANALYSIS ON NORMAL CONSTRAINTS

Sequence	Features	JONA-SLAM(w/o Normal)	JONA-SLAM
door_01	indoor/outdoor	0.211	0.178
roomdark_02	indoor/dark	0.208	0.158
street_04	dark/look back	0.352	0.315
hall_05	Random walk	0.938	0.754

TABLE V
APE EVALUATION IN REAL-WORLD SCENARIOS IN METERS

Error	LVI-SAM	FAST-LIVO2	R3LIVE++	JONA-SLAM
Max	5.673	4.892	4.281	4.309
Mean	2.823	2.541	2.215	2.275
Med	3.781	3.120	2.248	2.357
Min	0.296	0.185	0.082	0.104
Rmse	4.027	3.425	2.418	2.404

Based on the ablation data in Table IV, JONA-SLAM achieves a substantial reduction in ATE RMSE across all four representative degraded sequences following the integration of normal constraints. Correlated with the APE 3-D error maps, the performance gains are particularly prominent in the hall_05 sequence—a long corridor environment characterized by significant geometric degeneracy. In this scenario, the error decreases from 0.938 to 0.754 m. This 24.23% precision improvement provides strong evidence that normal constraints offer a definitive directional prior for the optimizer, effectively suppressing the axial sliding often encountered by LiDAR-only systems in feature-sparse environments. In sequences such as roomdark_02 and door_01, which involve visual failure or abrupt exposure shifts, the reduction in error reflects the robustness of normal information as a high-order geometric prior. When visual residuals fail to provide reliable constraints due to environmental degradation, normal constraints ensure that the factor graph optimization maintains the correct spatial topological dimensions, preventing trajectory drift. Furthermore, in the street_04 sequence, these constraints enhance pose stability over long distances by refining the descriptive accuracy of the global environmental structure.

C. Experiments on the Self-Collected Dataset

To further assess the generalization capability of JONA-SLAM, we conducted evaluations on a self-collected dataset. The APE and relative pose error (RPE) values for different algorithms were output using the EVO tools. The specific error values are shown in Table V.

The quantitative evaluation on the self-collected dataset, summarized in Table V, demonstrates that JONA-SLAM achieves high-precision performance comparable to the state-of-the-art R3LIVE++ framework. While R3LIVE++ exhibits a marginal advantage in localized metrics like the minimum and mean errors, JONA-SLAM yields a superior APE RMSE of 2.404 m. This improvement indicates that our method provides better global robustness against the abrupt disturbances and structural irregularities typical of real-world campus environments. Compared to baseline frameworks such as LVI-SAM and FAST-LIVO2, JONA-SLAM achieves a

substantial precision gain of 40.30% relative to LVI-SAM, highlighting the practical value of the proposed normal-aware geometric constraints in suppressing pose drift during challenging maneuvers.

V. CONCLUSION AND FUTURE WORK

This article presents JONA-SLAM, a tightly coupled LiDAR-visual-inertial SLAM system designed to enhance localization accuracy and robustness in complex and GNSS-challenged environments. By shifting the focus from basic feature extraction to tight multisensor fusion, the system introduces a joint observation factor that effectively fuses LiDAR-derived spatial constraints with visual point-line features. Furthermore, a normal-aware hybrid loop closure mechanism is developed, leveraging geometric consistency to significantly suppress false matches and ensure global trajectory precision. Quantitative evaluation across 22 public sequences from KITTI and M2DGR datasets, along with the self-collected campus dataset, demonstrates that JONA-SLAM achieves high-precision state estimation. In the real-world campus scenario, JONA-SLAM achieves an RMSE of 2.40 m, outperforming LVI-SAM at 4.03 m and R3LIVE++ at 2.42 m. Overall, based on 23 experimental test cases, the proposed method yields an average RMSE reduction of 35.02% relative to the LVI-SAM framework. These results, maintained at a precision of two decimal places, provide objective and statistically verified evidence for the performance of the system. In real-world campus scenarios, the system demonstrated superior stability under abrupt disturbances and structural irregularities, providing a robust localization foundation for autonomous navigation. Future work will focus on optimizing computational efficiency and exploring more advanced feature enhancement techniques to further improve adaptability under extreme perceptual degradation.

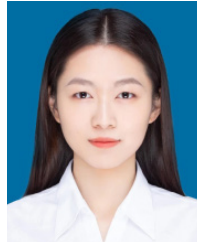
REFERENCES

- [1] T. Ran, L. Yuan, J. Zhang, D. Tang, and L. He, "RS-SLAM: A robust semantic SLAM in dynamic environments based on RGB-D sensor," *IEEE Sensors J.*, vol. 21, no. 18, pp. 20657–20664, Sep. 2021.
- [2] Z. Lin, Q. Zhang, Z. Tian, P. Yu, and J. Lan, "DPL-SLAM: Enhancing dynamic point-line SLAM through dense semantic methods," *IEEE Sensors J.*, vol. 24, no. 9, pp. 14596–14607, May 2024.
- [3] B. Zhou, C. Li, S. Chen, D. Xie, M. Yu, and Q. Li, "ASL-SLAM: A LiDAR SLAM with activity semantics-based loop closure," *IEEE Sensors J.*, vol. 23, no. 12, pp. 13499–13510, Jun. 2023.
- [4] J. Zhao, X. Zhang, K. He, Z. Qi, and Y. Yang, "Tightly coupled visual-inertial SLAM with dual-modal confidence perception," *IEEE Sensors J.*, vol. 25, no. 23, pp. 43138–43149, Dec. 2025.
- [5] J. Fu, H. Zhu, K. Zhang, T. Ma, J. Liu, and Ye Li, "Online exploratory coverage path planning of incremental SLAM for autonomous vehicles," *IEEE Trans. Ind. Informat.*, vol. 22, no. 3, pp. 1861–1870, Mar. 2025, doi: [10.1109/TII.2025.3632027](https://doi.org/10.1109/TII.2025.3632027).
- [6] Z. Ying and H. Li, "IMM-SLAMMOT: Tightly-coupled SLAM and IMM-based multi-object tracking," *IEEE Trans. Intell. Vehicles*, vol. 9, no. 2, pp. 3964–3974, Feb. 2024.
- [7] J. Zheng, K. Zhou, and J. Li, "Visual-inertial-wheel SLAM with high-accuracy localization measurement for wheeled robots on complex terrain," *Measurement*, vol. 243, Feb. 2025, Art. no. 116356.
- [8] K. Xu, Y. Hao, S. Yuan, C. Wang, and L. Xie, "AirSLAM: An efficient and illumination-robust point-line visual SLAM system," *IEEE Trans. Robot.*, vol. 41, pp. 1673–1692, 2025.
- [9] Q. Jian, Q. Ma, X. Wang, and S. Ullah, "A high-precision and fast autonomous vehicle SLAM based on multiscale adaptive feature extraction," *IEEE Sensors J.*, vol. 26, no. 2, pp. 3030–3044, Jan. 2026.
- [10] H. Liu, S. Pan, X. Wang, W. Gao, B. Yu, and H. Zhao, "Long-term localization method integrated with voxel mapping LiDAR odometry and adaptive updating map in diverse environment," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1–14, 2025.
- [11] S. Song, Q. Zhu, and H. Hu, "RTS-SLAM: Robust, tightly coupled, and semantic-enhanced LiDAR-inertial SLAM," *IEEE Sensors J.*, vol. 25, no. 20, pp. 38613–38626, Oct. 2025.
- [12] Y. Song, Z. Zhang, J. Wu, Y. Wang, L. Zhao, and S. Huang, "A right invariant extended Kalman filter for object based SLAM," *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 1316–1323, Apr. 2022.
- [13] M. Tang, Z. Chen, and F. Yin, "Robot tracking in SLAM with Masreliez-Martin unscented Kalman filter," *Int. J. Control, Autom. Syst.*, vol. 18, no. 9, pp. 2315–2325, Sep. 2020.
- [14] W. Xu and F. Zhang, "FAST-LIO: A fast, robust LiDAR-inertial odometry package by tightly-coupled iterated Kalman filter," *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 3317–3324, Apr. 2021.
- [15] 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.
- [16] T. Du, D. Shan, and P. Guo, "PFLIO-SAM: Tightly coupled polarization camera/optical flow/LiDAR/IMU odometry via smoothing and mapping," *IEEE Sensors J.*, vol. 25, no. 19, pp. 36243–36253, Oct. 2025.
- [17] C. Zhang et al., "LIWOM-GD: Enhanced LiDAR-inertial-wheel odometry and mapping by fusion with ground constraint and dynamic points elimination," *IEEE Sensors J.*, vol. 24, no. 19, pp. 30287–30303, Oct. 2024.
- [18] D. C. Herraes, M. Zeller, D. Wang, J. Behley, M. Heide, and C. Stachniss, "RaI-SLAM: Radar-inertial SLAM for autonomous vehicles," *IEEE Robot. Autom. Lett.*, vol. 10, no. 6, pp. 5257–5264, Jun. 2025.
- [19] J. Lee et al., "Switch-SLAM: Switching-based LiDAR-inertial-visual SLAM for degenerate environments," *IEEE Robot. Autom. Lett.*, vol. 9, no. 8, pp. 7270–7277, Aug. 2024.
- [20] 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. Autom. (ICRA)*, May 2021, pp. 5692–5698.
- [21] J. Lin and F. Zhang, "R3LIVE++: A robust, real-time, radiance reconstruction package with a tightly-coupled LiDAR-inertial-visual state estimator," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 12, pp. 11168–11185, Dec. 2024.
- [22] J. Yin, D. Luo, F. Yan, and Y. Zhuang, "A novel LiDAR-assisted monocular visual SLAM framework for mobile robots in outdoor environments," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.
- [23] J. Huang, S. Wen, W. Liang, and W. Guan, "VWR-SLAM: Tightly coupled SLAM system based on visible light positioning landmark, wheel odometer, and RGB-D camera," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–12, 2023.
- [24] C.-C. Chou and C.-F. Chou, "Efficient and accurate tightly-coupled visual-LiDAR SLAM," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 14509–14523, Sep. 2022.
- [25] J. Chen, P. Hou, Z. Cao, T. Xu, and J. Zhao, "An improved monocular SLAM system by combining geometric features with semantic probability," *IEEE Sensors J.*, vol. 25, no. 14, pp. 27111–27125, Jul. 2025.
- [26] G. Kim and A. Kim, "Scan context: Egocentric spatial descriptor for place recognition within 3D point cloud map," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Madrid, Spain, Oct. 2018, pp. 4802–4809.
- [27] K. Fischer, M. Simon, S. Milz, and P. Mader, "MagneticPillars++: Efficient LiDAR odometry via deep frame-to-keyframe point cloud registration," *Appl. Artif. Intell.*, vol. 39, no. 1, Dec. 2025, doi: [10.1080/08839514.2025.2472105](https://doi.org/10.1080/08839514.2025.2472105).
- [28] C. Zheng et al., "FAST-LIVO2: Fast, direct LiDAR-inertial-visual odometry," *IEEE Trans. Robot.*, pp. 326–346, 2024.



Qiuwei Jian received the bachelor's degree from Chongqing Technology and Business University, Chongqing, China, in 2022. He is currently pursuing the master's degree with Chongqing Jiaotong University, Chongqing.

His research interest includes intelligent transportation.



Xinyu Wang received the bachelor's degree from Chongqing Jiaotong University, Chongqing, China, in 2021, where she is currently pursuing the Ph.D. degree.

Her current research interests include intelligent transportation and safety, and autonomous driving technology.



Qinglu Ma received the Ph.D. degree in computer technology and science from Chongqing University, Chongqing, China, in 2012.

He is currently a Professor with Chongqing Jiaotong University, Chongqing. His current research interests include intelligent transportation and safety, and autonomous driving technology.



Saleem Ullah received the B.Sc. degree in computer science from The Islamia University of Bahawalpur, Bahawalpur, Pakistan, in 2003, and the M.I.T degree in computer science from Bahauddin Zakariya University, Multan, Pakistan, in 2005.

From 2006 to 2009, he worked as a Network/IT Administrator at different companies. Currently, he has been working as an Associate Professor with the Khwaja Fareed University of Engineering and Information Technology, Rahim

Yar Khan, Pakistan, since February 2016. He has almost 14 years of industry experience in the field of IT. He is an Active Researcher in the field of ad hoc networks, Internet of Things (IoT), congestion control, and security.