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

LMC-VIO: Lane Model-Constrained Monocular Inertial Visual SLAM for High-Precision Localization in Highway Scenes

Maosheng Yan[1,2], Man Luo[3,4], Yuan Guo[3*], Bijun Li[1,2], Jian Zhou[1,2]

***Abstract*—Continuous stability, as one of the core modules of the autopilot system, is particularly important for its performance. However, as the vehicle speed increases, the system positioning error may be amplified, consequently introducing deviations in the positioning consistency of the system. The inherent high speeds and motion constraints in highway environments introduce new challenges for feature matching, particularly in vision-based vehicle localization, where initialization and scale estimation biases are further expanded. Lane markings, characterized by their simple and uniform structures and high distinctiveness from the surrounding environment, serve as effective features for matching-based localization in autonomous driving. This paper introduces a high-precision and robust vehicle localization method based on lane model constraints. Initially, leveraging lane model parameters from prior maps, we track and model lane line detections across consecutive frames to enhance the completeness of lane representation. The tracking results, combined with prior map data on lane widths, are employed to optimize scale parameters. Subsequently, real-time detected lanes are matched with prior maps through point-map association to constrain the vehicle's heading angle. Finally, map matching results are integrated into existing visual local odometry methods to perform real-time localization optimization, thereby improving localization performance. Experimental evaluations conducted on a self-collected highway dataset demonstrate that the incorporation of lane models significantly enhances system localization accuracy.**

***Index Terms*—Highway scenes, Lane Model-based localization, Visual inertial odometry , Robustness improvement.**

I. Introduction

THE automotive industry is currently undergoing a transformation from traditional vehicles to intelligent ones. Many vehicles are now equipped with high-precision localization devices, cameras, Inertial Measurement Units(IMUs), LiDAR(Light Detection and Ranging), and various other sensors. These advanced sensing systems provide the technological foundation for intelligent driving functionalities such as perception, decision-making, and path planning, significantly enhancing driving safety [1]. SLAM plays a crucial role in the field of autonomous driving by enabling real-time environmental mapping and vehicle localization through the integration of cameras and other sensors, thereby facilitating high-precision autonomous navigation in complex and dynamic environments. However, the practical performance of vision-based SLAM technologies in intricate urban road scenarios still faces certain limitations. Urban road environments often exhibit relatively homogeneous structures, with high redundancy in environmental features such as buildings and road signs. This redundancy can lead to confusion during the feature extraction and matching processes of Visual SLAM systems, adversely affecting the system's localization accuracy and stability [2]. Most existing visual-inertial system methods predominantly focus on indoor, urban, or suburban environments, with limited research dedicated to vehicles operating on highways. Although visual-inertial SLAM algorithms have achieved significant advancements on several public datasets, their performance in actual urban road environments often falls short, failing to meet the stringent requirements for high-precision localization and mapping. Therefore, optimizing SLAM technology specifically for urban road scenarios to enhance its stability and accuracy in real-world applications remains a critical challenge in current autonomous driving research. Addressing this challenge is essential for the reliable deployment of autonomous vehicles in diverse and dynamic urban settings, ensuring both safety and efficiency in their operations.

In recent years, map models [3]–[5] have continuously evolved and have been widely applied. In the realm of autonomous driving, high-precision maps enable vehicles to achieve accurate localization [6] and stable vehicle control in complex traffic environments by providing precise lane-level navigation information, offering a thorough basis for environmental perception for autonomous vehicles [7]; the accuracy of this information can typically reach centimeter-level. However, in highway scenarios, the high speed can easily lead to image blurring, which in turn affects the accuracy of front-end feature extraction and matching; on the other hand, in road scenarios with repetitive features, visual inertial systems based on feature point tracking are prone to failure. Furthermore, the high cost of high-precision maps presents numerous challenges for large-scale applications.

To address the aforementioned issues, this paper proposes a lightweight method capable of achieving high-precision vehicle positioning on highways. The proposed method requires only images collected by visual sensors and measurements from the IMU unit to achieve accurate positioning on a prior model. By fully utilizing the lane line information provided by SD map [8] and combining it with real-time sensor data, this method effectively enhances vehicle positioning accuracy in highway environments without the need for additional equipment.

[1]Luojia Laboratory Hubei, 129 Luoyu Rd, Wuhan, 430079, China
[2]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430072, China
[3]SCHOOL OF AUTOMOTIVE ENGINEERING, Wuhan University of Technology, Wuhan 430070, China
[4]Dongfeng USharing Technology Co., Ltd, Wuhan 430056, China
Corresponding Author: yuang@whut.edu.cn

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

The high-precision robust positioning method based on lane model constraints (LMC-VIO) proposed in this paper aims to improve the positioning accuracy and stability of autonomous vehicles in highway environments. The method uses anchor-based lane detection to detect and track lane line information and combines inverse perspective transformation of camera imaging model for cross-domain matching. The known lane width and shape information provided by the SD map are used to constrain the camera pose. In summary, the main contributions of this paper are as follows.

- This paper introduces a cost-effective localization approach specifically designed for highway scenarios, enabling standard vehicles to achieve high localization accuracy without the need for expensive equipment.
- We propose a novel map matching method that leverages real-time detected lane models to perform point-map association with known maps, thereby constraining vehicle motion and enhancing localization reliability.
- The proposed map matching results are incorporated into existing visual local odometry methods to facilitate real-time localization optimization. Experimental evaluations conducted on a self-collected highway dataset demonstrate that our method significantly improves localization accuracy compared to existing open-source approaches.

The remainder of this paper is organized as follows. Section 2 reviews the related work in the field; Section 3 details the proposed high-precision localization method based on lane model constraints; Section 4 presents the experimental results; and Section 5 concludes the study.

II. Related Work

A. Visual SLAM

Visual SLAM primarily estimates motion by continuously capturing environmental texture information through cameras, thereby achieving self-localization and map construction. Research in the field of visual SLAM is diverse, mainly categorized into direct matching, feature-based methods, and hybrid approaches. Direct matching is based on the photometric invariance assumption, estimating inter-frame pose transformations by minimizing inter-frame photometric errors, where LSD-SLAM [9] constructs a semi-dense map using image gradient information; however, this method is sensitive to camera intrinsic parameters and is prone to pose loss during rapid camera movements. DSO [10], based on accurate sparse direct structure, utilizes a sliding window to achieve robust pose estimation, enhancing stability in low-texture matching. however, this method is only suitable for small scenes and has poor robustness. Based on the classic ORB (Oriented Brief) [11] features in computer vision, Mur-Artal et al [12] proposed a real-time visual SLAM method that innovatively uses three threads for localization and sparse mapping: the tracking thread, local optimization thread, and loop thread; however, this method incurs significant CPU performance overhead and only supports monocular cameras. To improve this method, they subsequently proposed ORB-SLAM2 [13], which not only supports monocular cameras but also stereo and RGB-D cameras, achieving higher accuracy than ICP or photometric errors. Building on previous work, Campos et al [14] proposed a visual-inertial SLAM system that utilizes IMU information to achieve robust localization and map construction in high-speed motion scenarios. SVO [15] is one of the representatives of hybrid methods, achieving efficient pose estimation and map construction in small ranges by extracting FAST features and all pixels with non-zero inter-frame gradients, however, the omission of back-end optimization and loop closure leads to cumulative errors and difficulties in re-localization after loss of localization.

The visual SLAM system realizes the estimation of the pose through the observation of the environment texture, so the problem of low robustness of the localization of weak feature scenes caused by its own motion or feature degradation still needs to be solved. At present, some scholars have studied the scene with insufficient saliency of visual features, and used edge line features to solve the problem of unstable matching caused by insignificant point features in the scene. However, on the one hand, in some feature degradation scenarios, it may be difficult to extract edge lines that can effectively constrain the pose. On the other hand, due to the decrease in the number of features caused by the change of light, it is difficult to solve the problem by relying on the edge line.

B. Map-Matching-Based Localization

Map-based localization is a technique that determines a vehicle's position by precisely matching vehicle sensor data, such as cameras, LiDAR, and Inertial Measurement Units (IMUs), with map data. Ma et al. [16] constructed a sparse high-precision map containing only traffic sign locations and key lane line features, constraining the lateral position by matching lane detection results with lane models in the map and constraining the longitudinal position by matching traffic sign information. Schönberger et al. [17] combined the scene's 3D semantic information with geometric information, utilizing high-level abstract features during localization by constructing a local 3D semantic map, extracting descriptors, and performing 3D-3D matching between the query map and the database map to determine the camera pose. Li et al. [18] uses environment semantic landmarks for position recognition and loop closure detection. Guo et al. [19] proposes a lane-level localization method based on the vehicle's lateral displacement and motion estimation model, aimed at calculating the vehicle's lateral displacement within a lane and its lateral movement across multiple lanes. The association of semantic features is more intuitive and clear compared to intensity or geometric features, which helps reduce potential association errors during the map matching process. However, the limitation of this method lies in its reliance on pre-labeled high-precision maps for semantic features. When map annotations are incomplete, the matching performance is adversely affected. Additionally, the extraction and association of semantic features require complex algorithms and models, increasing computational overhead and imposing high demands on real-time performance.

To address these challenges and enable high-precision localization for ordinary vehicles using standard maps, this study

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

proposes a cost-effective lane model-based localization method (LMC-VIO) leveraging prior SD maps. Vehicles equipped with basic cameras and IMU use GNSS only for initialization, while lane lines are extracted from camera images. Prior lane width and shape information from SD map constrain the camera pose, with these constraints integrated into an open-source visual-inertial framework to enhance localization accuracy. This approach improves lane feature recognition and camera-based localization, enabling more precise vehicle positioning on the road.

III. METHODOLOGY

A. Overview

The structure of the proposed LMC-VIO system is illustrated in Figure 1. Specifically, the proposed framework comprises three sub-modules: prior map-assisted parameter calibration, point-to-map matching, and vision-based optimized localization. Firstly, the system employs a neural network to detect, extract, and associate lane line information from images, correlating it with lane lines in the map model. Subsequently, the identified lane lines are modeled through a line fitting module, and camera scale parameter calibration is performed based on the prior lane line information provided in the map and the camera imaging model. Following this, lane lines from consecutive frames are modeled, and the vehicle's heading is constrained using lateral constraints provided by the map. Finally, by integrating the constraints from the first two steps as factors into the existing monocular visual-inertial odometry system, the real-time localization accuracy is enhanced. This integration effectively improves the system's ability to accurately determine the vehicle's position on the road.

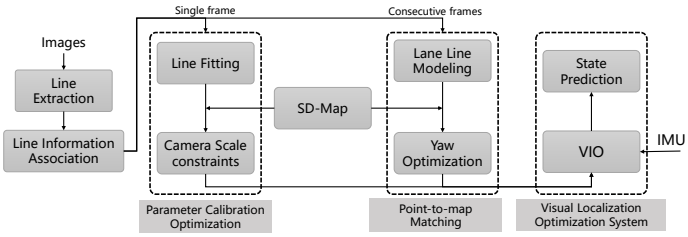


Fig. 1: A block diagram illustrating the full pipeline of the LMC-VIO system

B. Prior Map-Assisted Parameter Calibration

In camera-based vehicle localization, accurate scale estimation is a critical challenge in highway scenarios, as uncertainties can lead to significant pose errors. To address this issue, we propose a camera-based parameterization method that leverages lane information from a prior map to effectively constrain the system's pose and thereby estimate the camera's scale parameter.

This study employs the CLRNet [20] network to detect lane lines from the collected image sequences. For lane line association and the completion of partially occluded lane lines in images, we utilize a model-based completion strategy described in prior work [19]. Briefly, lane lines are represented using anchor-based geometric primitives and tracked across frames via a Kalman filter and Hungarian matching. RANSAC fitting, vanishing point constraints, and motion consistency are used to suppress noise and infer missing segments, thereby enhancing the continuity and robustness of lane line representation.

The coordinate systems involved in this study include the image coordinate system, the camera coordinate system, and the road coordinate system, as illustrated in the figure below:

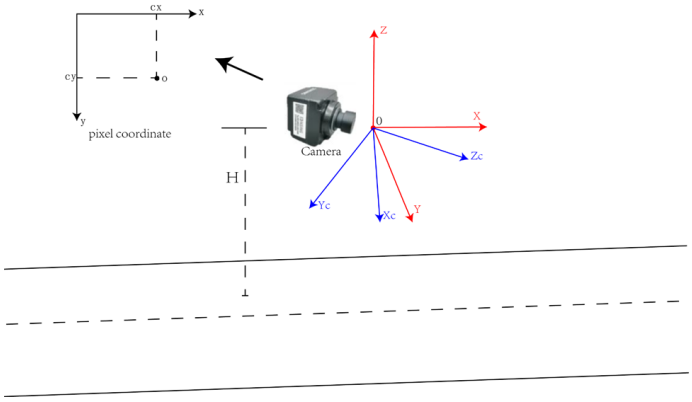


Fig. 2: Definition of coordinate systems

Here, $o-xy$ represents the image coordinate system, $O-X_cY_cZ_c$ represents the camera coordinate system, $O-XYZ$ represents the road coordinate system, and H denotes the height of the camera. The origin of the camera coordinate system coincides with the origin of the road coordinate system. For any given point $\boldsymbol{P}=(X,Y,Z)$ in the lane space, the relationship between the road coordinates and image coordinates can be described by the pinhole camera model:

$$\lambda\begin{pmatrix}x\\y\\1\end{pmatrix}=\boldsymbol{K}\boldsymbol{R}_{\text{pitch}}\,\boldsymbol{R}_{\text{yaw}}\,\boldsymbol{R}_{\text{roll}}\begin{pmatrix}X\\Y\\Z\end{pmatrix} \tag{1}$$

where $\boldsymbol{R}_{\text{pitch}}$, $\boldsymbol{R}_{\text{yaw}}$, $\boldsymbol{R}_{\text{roll}}$ are the rotation matrices corresponding to the three-axis angles pitch, yaw, and roll. K is the internal reference matrix of the camera. λ is the scale parameter; and the road is assumed to be flat, i.e., Eq. $Z=-H$. The rotation matrix in this paper is in the form of roll-yaw-pitch.

We utilize Inverse Perspective Mapping to project image-space lane lines into a top-down view for matching against a map. By detecting the lane lines in an image, we can determine the vanishing point from their intersection. According to [21], this vanishing point can be utilized to calculate the initial values of the camera's pitch and yaw angles. The relevant relationships are shown as follows:

$$\theta_p=\arctan\left(\frac{c_y-v_y}{f_y}\right) \tag{2}$$

$$\theta_y=\arctan\left(\frac{c_x-v_x}{f_x\cdot\cos\left(\theta_p\right)}\right) \tag{3}$$

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

Where c_x, c_y are the center of projection of the camera coordinates and f_x, f_y are the focal lengths of the camera. θ_p is the camera's pitch angle; θ_y is the camera's yaw angle. v_x, v_y are the pixel coordinates of the vanishing point point. A detailed flowchart is illustrated in Figure 3. For single-frame interaction, we extract lane lines in the image and apply Inverse Perspective Mapping to transform them into the ground plane. We use the pixel positions of corresponding points on adjacent lane lines to compute the projected lane width as a function of the camera's roll angle. We can compute the roll angle by constructing a constraint relationship based on the actual lane width or the parallelism condition. Accordingly, we formulate an objective function to minimize the error between the computed and the expected physical lane width, resulting in a residual that directly constrains the roll parameter.

$$\theta_r^* = \underset{\theta_r}{\operatorname{argmin}} \sum_{i=1}^{N-1} \left(d_i\left(\theta_r\right) - d_i\right)^2 \tag{4}$$

$d_i (i = 0, 1, 2 \ldots)$ denotes the distances from the vehicle to the adjacent lane lines on the corresponding road segments in the SD map. The camera pose can be estimated from the lane model and is denoted as R_{lane}, which can be used as a constraint factor.

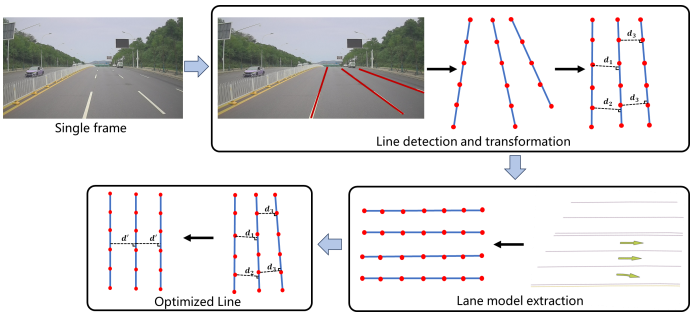


Fig. 3: Pipeline of Parameter Calibration Optimization

Then, the camera's scale factor can be computed based on the actual lane width.

$$\lambda = f(H, \mathbf{R}, \mathbf{K}, \mathbf{p}, \mathbf{d_i}) \tag{5}$$

C. Point-to-Map Matching

Based on the results of the camera parameter constraint optimization model described in the previous section, we initialize the camera pose. To ensure high consistency between the lane model in the image and the actual road environment, we adopt a lateral constraint method based on the prior SD map. The workflow of this module is shown in Figure 4. Firstly, the detected lane line sequences are transformed into a bird's-eye view via inverse perspective mapping, and lane lines across consecutive frames are modeled. Subsequently, the extracted lane points are matched with the map to further refine the vehicle pose estimation.

After calculating the perspective-transformed points based on the camera projection model, let the series of point coordinates on the lane line be represented as $\{(x_i, y_i)\}$. Common lane line fitting methods include parabolic curves and cubic B-spline curves. Given that highways are mostly straight or gently curved, linear models offer a good balance between accuracy and efficiency, this paper adopts a linear fitting approach to model the lane lines. Firstly, this can further constrain the camera's roll angle and scale using the aforementioned formulation. Secondly, let the normal vector of the fitted line be denoted as $\mathbf{V}_{\text{heading}} = (n_x, n_y)$, and the normal vector of the lane model in the SD map be denoted as $\mathbf{V}'_{\text{lane}} = (n'_x, n'_y)$. Using the angle between the fitted line's normal vector n and the high-definition map's lane model normal vector $\mathbf{n}'$, the loss function is defined as follows:

$$\theta_{\text{yaw}}^* = \arg\min_{\theta_{\text{yaw}}} \left\{ \arccos\left(\frac{\mathbf{v}_{\text{heading}} \cdot \mathbf{v}_{\text{lane}}}{\|\mathbf{v}_{\text{heading}}\| \cdot \|\mathbf{v}_{\text{lane}}\|} \right) \right\} \tag{6}$$

where $\|\mathbf{v}_{\text{heading}}\|$ and $\|\mathbf{v}_{\text{lane}}\|$ are the magnitudes of the respective normal vectors. The camera pose can be estimated from the lane model and is denoted as q_{lane}, while the VINS-Fusion system provides its own pose estimate, denoted as q_{wc}. A residual can be constructed between them as: $\Delta q = q_{wc} \otimes q_{\text{lane}}^{-1}$.

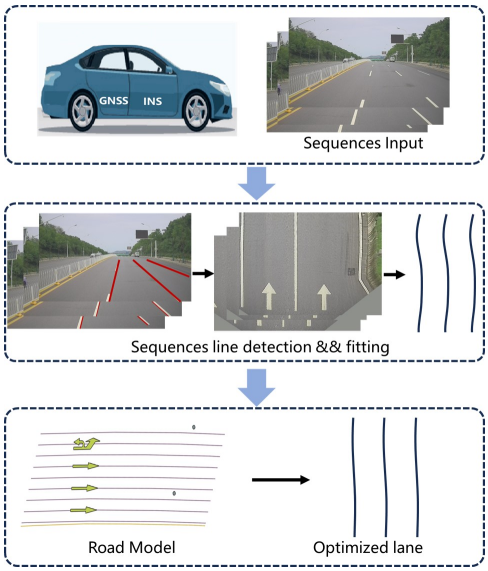


Fig. 4: Pipeline of Point-to-map matching

D. Visual Localization Optimization System

The field of visual-inertial localization boasts numerous excellent open-source frameworks. The algorithm proposed in this paper is based on the VINS-Fusion [22] framework, wherein the system achieves high-precision localization by integrating inputs from a monocular camera, an Inertial Measurement Unit (IMU). Notably, GPS data is utilized solely for the initialization of position and orientation.

Building upon the aforementioned foundation, we comprehensively consider multiple constraints and employ an optimization algorithm to jointly optimize the existing data. The

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

nonlinear optimization model used in this study is presented as follows:

$$f(\boldsymbol{X}) = \arg\min_{\mathbf{X}} \left\{ \sum \left(\left\| \mathbf{r}_M \left(\boldsymbol{M}_{k-1}^{k}, \mathbf{X} \right) \right\|^2 + \right. \right. \\ \left. \left. \left\| \mathbf{r}_R \left(\boldsymbol{R}_k, \boldsymbol{X} \right) \right\|^2 + \left\| \mathbf{r}_P \left(\boldsymbol{P}_k, \boldsymbol{X} \right) \right\|^2 + \left\| \mathbf{r}_I \left(\boldsymbol{I}_{k-1}^{k}, \mathbf{X} \right) \right\|^2 \right\} \tag{7}$$

In the equation, X represents the position of the vehicle, $\mathbf{r}_M \left(\boldsymbol{M}_{i-1}^{i}, \mathbf{X} \right)$ denotes the pose transformation relationship calculated from matching features between the $(\mathbf{i}-\mathbf{1})^{\text{th}}$ and $\mathbf{i}^{\text{th}}$ frames, $\mathbf{r}_R \left(\boldsymbol{R}_k, \boldsymbol{X} \right)$ represents the result based on loop closure detection, and $\mathbf{r}_P \left(\boldsymbol{P}_k, \boldsymbol{X} \right)$ represents the result of camera pose optimization based on all parameter constraints, $\mathbf{r}_I \left(\boldsymbol{I}_{k-1}^{k}, \mathbf{X} \right)$ represents the IMU constraint between adjacent keyframes. $\| \cdot \|^2$ denotes the Euclidean distance.

IV. EXPERIMENT

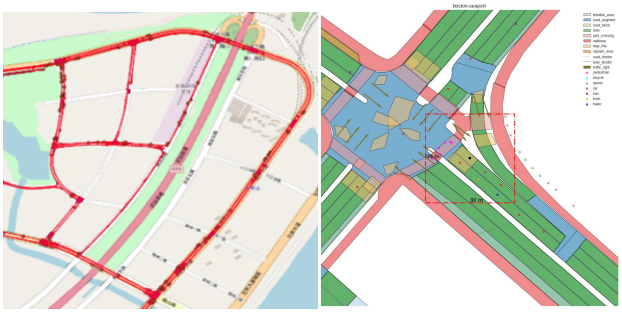
A. Experiments on Self-Collected Dataset

1) System Setup and Datasets: To evaluate the accuracy of the proposed LMC-VIO method, experiments were conducted on a test vehicle using self-collected data. The experiments utilized GNSS data collected by the high-precision on-board integrated navigation positioning system as the groundtruth to validate the experimental results. This equipment is capable of synchronously acquiring image sequences and position information during vehicle operation.The short-range camera used operates at a frequency of 10 Hz, with a horizontal field of view greater than 100° and a vertical field of view greater than 50°. The resolution of the captured images is 1920×1080. The type of high-precision on-board integrated navigation system used is the INS570D, with the IMU and GNSS operating at frequencies of 100 Hz and 10 Hz, respectively.

The experimental equipment was mounted on an autonomous vehicle. The dataset for this study was acquired on approximately 10 km of highway in Wuhan City, Hubei Province, encompassing multiple road segments. The prior SD map of the experimental area is illustrated in Figure 5, where the left image depicts the visualization results of integrating the road network into OpenStreetMap, and the right image presents Map visualisation in Nuscenes. The SD map utilized in our work has an absolute accuracy of approximately 20 cm, providing reliable geometric priors for lane-based constraints.

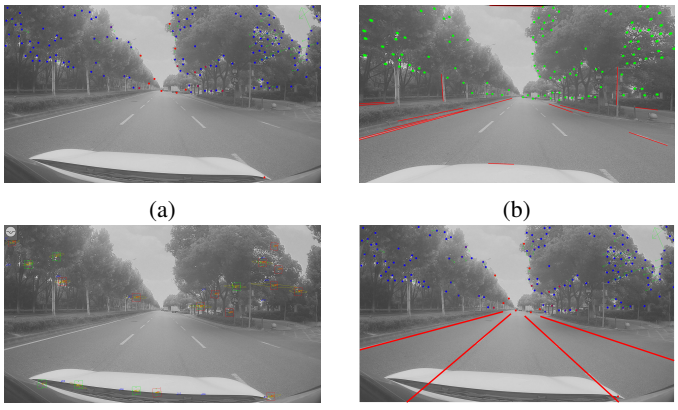
2) Evaluation for VIO: For the monocular visual-inertial localization system on highways, this study conducts experimental evaluations focusing on localization accuracy and heading angle analysis. To assess the performance of the proposed lane-level localization method, experiments are conducted using data from multiple different road segments for comparative analysis of localization accuracy.

To evaluate the generality of the proposed framework, we conducted comparative analyses of the accuracy of different monocular visual-inertial localization systems, including PL-VINS, VINS-Fusion, ROVIO, and OpenVINS. Experiments were carried out on four distinct trajectories. As presented in Table I , our proposed method consistently demonstrated superior performance across self-collected datasets. On highways, the high vehicle speeds result in motion blur, which increases matching errors between adjacent frames. Additionally, some visual-inertial systems require IMU data to estimate the scale factor during initialization. However, in challenging environments such as highways, the IMU does not provide sufficient rotational data to accurately estimate the scale. In our system, we first utilize GNSS data to initialize the camera's orientation and, in combination with a prior SD map, initialize the camera's scale parameters. Furthermore, during the operational phase, map matching is employed to constrain the camera's orientation.



(a) (b)

Fig. 5: Visualization of Road Network Results (a) Visualization of the Road Network Structure Integrated into OpenStreetMap (b) Map visualisation in Nuscenes



(a) (b) (c) (d)

Fig. 6: Scenario Features Extracted Using Different Algorithms on Highways (a) Vins-Fusion (b) PLVins (c) Rovio (d) LMC-VIO

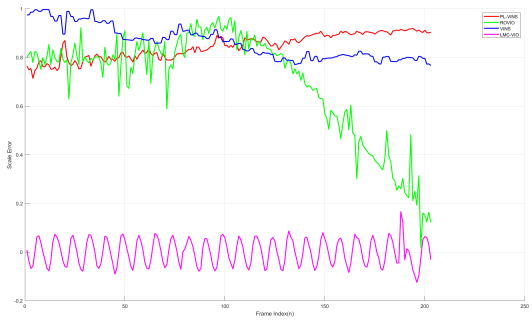
As shown in Figure 6, feature distributions in highway environments are sparse, with most features concentrated in roadside vegetation and few on the road surface. In VINS-Fusion and ROVIO, point features are mainly extracted from trees, while PL-VINS adds line features but suffers instability at high speeds, leading to mismatches. To overcome these limitations, our method enhances VINS-Fusion with robust lane detection, enabling stable lane marking extraction to improve visual-inertial odometry.

In highway scenarios, the vehicle's high speed can cause significant scale errors in Visual-Inertial Odometry. Moreover, insufficient rotational motion during initialization can hinder accurate scale estimation, adversely affecting later algorithm

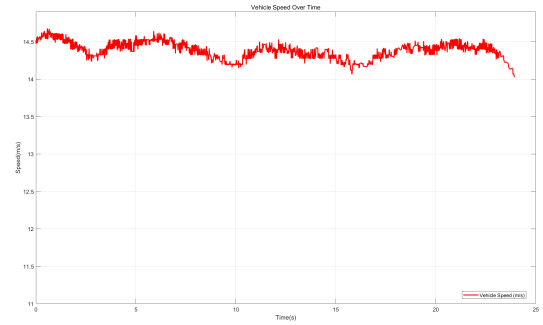
IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

TABLE I: Comparison of Different Methods (RMSE in meters)

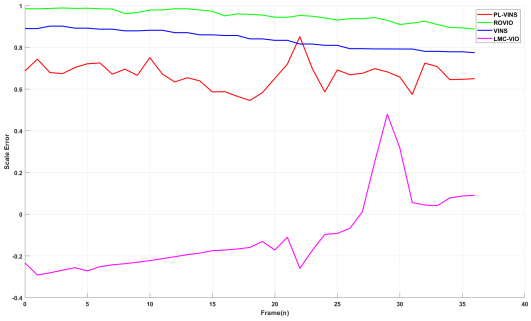
Seq.	Length (m)	Avg. Vel. (m/s)	VINS-Fusion	PL-VINS	ROVIO	LMC-VIO
01	344.91	14.55	30.34	8.00	62.90	1.44
02	356.04	15.09	18.96	6.11	44.31	1.48
03	525.89	19.85	25.90	8.65	35.47	1.78
04	348.74	14.38	17.55	14.32	89.39	1.50



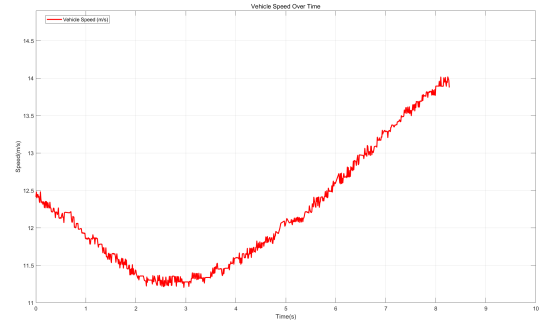
(a) Scale Error under Nearly Constant Speed



(b) Velocity under Nearly Constant Speed



(c) Scale Error during Acceleration and Deceleration



(d) Velocity under Acceleration and Deceleration

Fig. 7: Analysis results for different sequences.

performance. Figure 7 illustrates the scale error results of various algorithms (Vins-Fusion, PLVins, ROVIO, and LMC-VIO) on representative datasets. Figures 7(a) and 7(c) show the scale errors on two datasets, while Figures 7(b) and 7(d) display the corresponding vehicle speeds. In this paper, we employ the ratio method for analysis: for the trajectory estimated by the VIO system, the traveled distance d_1 within a certain time period is calculated, and the corresponding traveled distance d_2 in the groundtruth is computed. The ratio is determined using the following formula:

$$\text{ratio} = (d_2 - d_1)/d_2$$

When the ratio approaches 0, it indicates a more accurate scale estimation; conversely, when the result approaches 1, it signifies a larger scale estimation error. As shown in the results, the scale estimation errors of Vins-Fusion and PLVins exhibit significant fluctuations throughout the process, with consistently large errors. ROVIO shows considerable variation in scale estimation during the process but continuously optimizes the scale in subsequent stages. In contrast, our algorithm demonstrates strong robustness in scale estimation under high-speed scenarios.

As illustrated in the results in Figure 8, (a) and (b) show the differences with the groundtruth in experiments conducted on two datasets. Existing visual-inertial algorithms generally exhibit poor scale estimation performance in high-speed scenarios, with the estimated trajectory length often significantly smaller than the true trajectory length. However, in our method, both the trajectory length and yaw angle optimization achieved good results.

B. Experiments on NuScenes Dataset

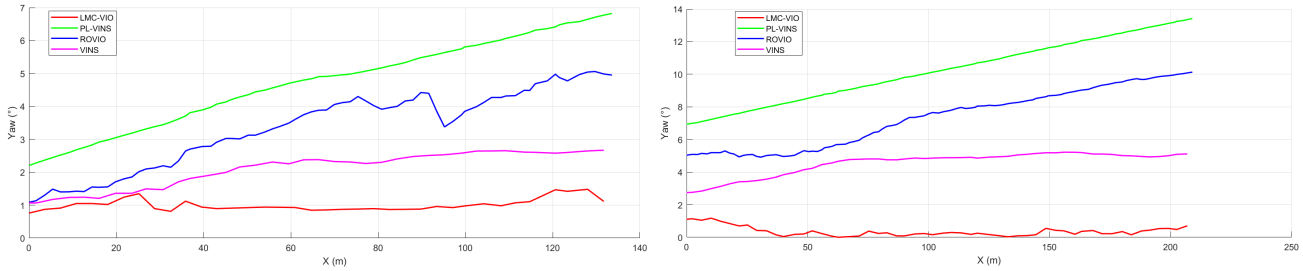
1) Datasets Introduction: This work is also evaluated on the NuScenes dataset [23]. It consists of 1000 driving scenes, each 20 seconds long, collected in Boston and Singapore. The dataset is particularly notable for its comprehensive, 360-degree sensor suite, which includes 6 cameras, 1 LiDAR, 5 RADAR sensors, as well as IMU, GPS, and detailed map data. The scale and sensor diversity of NuScenes make it a robust platform for validating autonomous driving algorithms.

2) Evaluation for VIO: Firstly, we analyze the absolute trajectory accuracy of different algorithms on several representative trajectories. As shown in Table II, the results demonstrate that our proposed method consistently maintains high accuracy. Next, we evaluate the heading angle

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

TABLE II: APE errors of different algorithms on the NuScenes dataset (unit: m)

Scene	Length (m)	Avg. Vel. (m/s)	VINS-Fusion	PL-VINS	ROVIO	LMC-VIO
Scene-09	185.12	9.64	48.58	42.49	38.42	1.29
Scene-23	151.16	7.79	13.93	12.08	192.32	1.22
Scene-24	90.43	4.68	10.08	11.33	27.04	1.13



(a) (b)

Fig. 8: Results of the yaw analysis

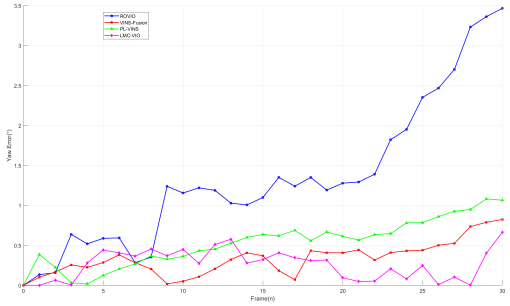


Fig. 9: Yaw error analysis

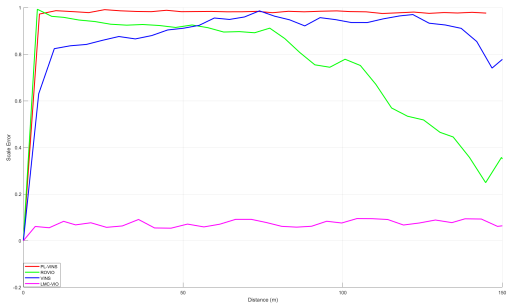


Fig. 10: Scale error

error of the algorithm within selected scenes. Specifically, we compute the absolute difference between the estimated heading angle and the ground-truth value. The results are presented in Figure 9, where the horizontal axis represents the frame index and the vertical axis shows the absolute heading angle error. We observe that in certain scenes within the NuScenes dataset, missing lane markings lead to increased yaw estimation errors. Finally, we estimate the scale error on representative sequences. As shown in Figure 10, our method consistently maintains a reliable and accurate scale estimation across different scenarios. This demonstrates the effectiveness of the proposed lane-based constraints in enhancing scale consistency.

C. System Real-time Analysis

To verify real-time performance, we conducted evaluations on multiple trajectory sequences. Table III presents the real-time performance analysis of our system, showing the incremental computational cost of each component. As the lane detection and association modules are integrated with the VINS-Fusion baseline, the processing time per frame moderately increases. Crucially, the final LMC-VIO system maintains a frame rate suitable for real-time operation in autonomous driving applications.

TABLE III: System Real-time analysis

Method	Processing Time (ms)/Frame	FPS
VINS-Fusion (Baseline)	45.0 ± 3.0	22.2
+ Lane Detection (CLRNet)	57.5 ± 3.8	17.4
+ Lane Association (DeepSORT)	62.0 ± 4.2	16.1
LMC-VIO (Complete System)	65.0 ± 4.5	15.4

D. Analysis

Our system is built on the VINS-Fusion framework to achieve high-precision localization in highway environments. However, practical deployment faces challenges due to the unique feature distribution in such settings. Highways contain sparse and less distinctive landmarks, with most extracted features coming from roadside shrubs and dynamic objects such as vehicles and pedestrians. These moving objects complicate feature matching and can cause large localization errors. During initialization, high vehicle speeds prevent sufficient rotational motion, unlike handheld devices, leading to inadequate rotational transformation information. Moreover, high speeds exacerbate error accumulation in IMU pre-integration, reducing initialization accuracy. Poor initialization then propagates through subsequent operations, diminishing overall system stability and reliability.

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

During the process of matching with the prior SD map, the lane lines detected in the images are accurately aligned with those in the map to constrain the camera's scale parameters and orientation. Incorporating these constraints into the localization system's optimization process significantly improves overall accuracy. Specifically, the lane line matching process leverages structured road information from existing maps to provide critical geometric constraints for visual-inertial fusion, effectively reducing IMU drift and correcting visual sensor failures caused by lighting changes or dynamic occlusions.

In complex scenarios such as rainy conditions, faded markings, or lane occlusions caused by traffic congestion, when a lane line is missing, the occluded segment can be reconstructed using the lane completion model. In areas where all lane markings disappear, the system applies a buffering strategy, using the constraints from the last valid lane model as the initialization prior. While this approach helps maintain continuity, it may introduce some instability in scale estimation. However, our method has a critical limitation, it relies heavily on the accuracy of the prior map. When the local accuracy of lane lines in the map is poor, our scale and camera pose estimation also degrade, and in some cases, this can even affect the accuracy of the underlying VIO system.

V. CONCLUSIONS

Due to the suboptimal performance of existing visual-inertial systems in highway environments, this paper proposes a localization system LMC-VIO based on vehicle-end perception and SD map matching. The system leverages lane prior information provided by existing maps and constrains the detected lane lines from image perception by matching them with map data in terms of lane width and shape, ultimately imposing constraints on the camera's scale and heading angle. Road segments on a highway in Wuhan were selected as test scenarios.

ACKNOWLEDGMENT

This work was supported in part by Luojia Laboratory, Hubei, in part by the National Natural Science Foundation of China (Grant No. 42201480), in part by the Hubei Provincial Natural Science Foundation of China (Grant No. 2024AFB778), and in part by the Key Research and Development Program of Hubei Province (Grant No. 2024BAB078).

REFERENCES

[1] S. Zheng, J. Wang, C. Rizos, W. Ding, and A. El-Mowafy, "Simultaneous localization and mapping (slam) for autonomous driving: concept and analysis," *Remote Sensing*, vol. 15, no. 4, p. 1156, 2023.

[2] I. A. Kazerouni, L. Fitzgerald, G. Dooly, and D. Toal, "A survey of state-of-the-art on visual slam," *Expert Systems with Applications*, vol. 205, p. 117734, 2022.

[3] Y. Guo, J. Zhou, Q. Dong, B. Li, J. Xiao, and Z. Li, "Refined high-definition map model for roadside rest area," *Transportation Research Part A: Policy and Practice*, vol. 195, p. 104463, 2025.

[4] J. Xiao, S. Wang, J. Zhou, Z. Tian, H. Zhang, and Y.-F. Wang, "Mim: High-definition maps incorporated multi-view 3d object detection," *IEEE Transactions on Intelligent Transportation Systems*, 2024.

[5] J. Zhou, Y. Guo, Y. Bian, Y. Huang, and B. Li, "Lane information extraction for high definition maps using crowdsourced data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 7, pp. 7780–7790, 2022.

[6] H. Wang, C. Xue, Y. Zhou, F. Wen, and H. Zhang, "Visual semantic localization based on hd map for autonomous vehicles in urban scenarios," in *2021 IEEE International conference on robotics and automation (ICRA)*. IEEE, 2021, pp. 11 255–11 261.

[7] G. Elghazaly, R. Frank, S. Harvey, and S. Safko, "High-definition maps: Comprehensive survey, challenges and future perspectives," *IEEE Open Journal of Intelligent Transportation Systems*, 2023.

[8] J. Li, P. Jia, J. Chen, J. Liu, and L. He, "Local map construction methods with sd map: A novel survey," *arXiv e-prints*, pp. arXiv–2409, 2024.

[9] J. Engel, T. Schöps, and D. Cremers, "Lsd-slam: Large-scale direct monocular slam," in *European conference on computer vision*. Springer, 2014, pp. 834–849.

[10] J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 3, pp. 611–625, 2017.

[11] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf," in *2011 International conference on computer vision*. Ieee, 2011, pp. 2564–2571.

[12] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "Orb-slam: a versatile and accurate monocular slam system," *IEEE transactions on robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[13] R. Mur-Artal and J. D. Tardós, "Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras," *IEEE transactions on robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.

[14] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardós, "Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1874–1890, 2021.

[15] C. Forster, M. Pizzoli, and D. Scaramuzza, "Svo: Fast semi-direct monocular visual odometry," in *2014 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2014, pp. 15–22.

[16] W.-C. Ma, I. Tartavull, I. A. Bârsan, S. Wang, M. Bai, G. Mattyus, N. Homayounfar, S. K. Lakshmikanth, A. Pokrovsky, and R. Urtasun, "Exploiting sparse semantic hd maps for self-driving vehicle localization," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 5304–5311.

[17] J. L. Schönberger, M. Pollefeys, A. Geiger, and T. Sattler, "Semantic visual localization," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 6896–6906.

[18] C. Li, B. Zhou, and Q. Li, "Semanticcslam: Using environment landmarks for cooperative simultaneous localization and mapping," *IEEE Internet of Things Journal*, 2024.

[19] Y. Guo, J. Zhou, Q. Dong, Y. Bian, Z. Li, and J. Xiao, "A lane-level localization method via the lateral displacement estimation model on expressway," *Expert Systems with Applications*, vol. 243, p. 122848, 2024.

[20] T. Zheng, Y. Huang, Y. Liu, W. Tang, Z. Yang, D. Cai, and X. He, "Clrnet: Cross layer refinement network for lane detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 898–907.

[21] Y. Huang, J. Zhou, B. Li, J. Xiao, and Y. Cao, "Roll-sensitive online camera orientation determination on the structured road," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 43, pp. 687–693, 2022.

[22] T. Qin, S. Cao, J. Pan, and S. Shen, "A general optimization-based framework for global pose estimation with multiple sensors," *arXiv preprint arXiv:1901.03642*, 2019.

[23] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuscenes: A multimodal dataset for autonomous driving," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11 621–11 631.