Robotica (2026), **44**, pp. 846–864
doi:10.1017/S0263574726103397

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

A robust simultaneous localization and mapping method for agricultural environments: integrating spatial modeling and noisy estimation

Ziyang Wang[1], Wenfu Tong[1], Haibo Zhou[1], Mingjun Li[2], Xiaozhuo Wang[3] and Biyao Cheng[3]

[1]State Key Laboratory of Precision Manufacturing for Extreme Service Performance, College of Mechanical and Electrical Engineering, Central South University, Changsha, China
[2]Guangdong Institute of Modern Agricultural Equipment, Guangzhou, China
[3]Xi'an Institute of Electromechanical Information Technology, Xi'an, China
Corresponding author: Haibo Zhou; Email: zhouhaibo@csu.edu.cn

Received: 1 December 2025; **Revised:** 19 February 2026; **Accepted:** 23 March 2026; **First published online:** 22 April 2026

Keywords: agricultural environments; simultaneous localization and mapping; spatial feature modeling; time-varying noise estimation

Abstract

To address the challenges posed by highly repetitive structures, unstructured terrain, and pronounced sensor disturbances in agricultural environments, this paper proposes a simultaneous localization and mapping framework that integrates explicit spatial feature modeling with time-varying noise estimation. First, a curved-voxel-based intensity–geometry joint probability model is constructed to transform conventional point features into spatial features capable of effectively capturing local structural patterns. These features are then aligned by minimizing the joint probability distance using the normal distributions transform, thereby enhancing feature discriminability and registration stability in highly repetitive scenes. Second, a maximum a posteriori-based recursive noise estimator is designed. By employing a dual sub-filter architecture, the proposed estimator enables joint modeling and online optimization of the noise parameters associated with LiDAR odometry and the inertial measurement unit, thereby improving the system's adaptability to terrain-induced perturbations and sensor uncertainties. Experimental results demonstrate that the proposed method achieves a mean relative translation error of 1.42% and a mean relative rotation error of 1.53°/100 m in agricultural scenarios. In addition, the average error in row spacing estimation derived from the reconstructed map is constrained within 0.071 m. Compared with baseline methods, the proposed system exhibits significant advantages in both pose estimation accuracy and agronomic structural perception capability.

1. Introduction

With the rapid advance of intelligent and automated agriculture, high-resolution mapping and high-precision navigation have become the core enablers for the stable operation of smart agricultural machinery in complex field scenarios [1]. Accurate terrain reconstruction and spatial modeling of crop distribution are essential for precision farming management, automated harvesting, and optimal resource allocation [2]. Nevertheless, conventional remote sensing exhibits pronounced limitations in near-ground perception. Its spatiotemporal resolution and real-time capability fail to meet the demands of fine-grained agricultural tasks. Specifically, in densely vegetated farmland, existing remote-sensing data rarely deliver the centimeter-level environmental details required for field operations [3]. Addressing this bottleneck, simultaneous localization and mapping (SLAM) demonstrates distinct advantages. By concurrently estimating ego-pose and constructing an environmental map, SLAM provides agricultural robotic systems with high-accuracy positional references and environmental representations, thereby establishing a reliable foundation for precision spraying, autonomous harvesting, and other fine-scale field operations [4, 5].

© The Author(s), 2026. Published by Cambridge University Press.

https://doi.org/10.1017/S0263574726103397 Published online by Cambridge University Press

Based on sensor modality, SLAM systems can be classified into vision-based V-SLAM- and LiDAR-based L-SLAM. V-SLAM estimates pose by processing RGB image sequences and is characterized by low hardware cost and high integration, yet its robustness is degraded by variable illumination in typical agricultural scenes [6]. Conversely, L-SLAM exploits high-precision 3D point clouds acquired by LiDAR for matching and optimization, maintaining consistent accuracy under diverse illumination and weather conditions [7], and has thus become the preferred solution for complex agricultural environments.

Accurate pose estimation and consistent map construction constitute the core challenges of SLAM, and the overall performance of L-SLAM heavily relies on the efficiency and robustness of point cloud registration algorithms [8]. Among existing approaches, feature-based registration methods (e.g., LOAM [9], LeGO-LOAM [10], and their variants [11–13]) have become mainstream by extracting key geometric features such as edge and planar points from point clouds, achieving a favorable trade-off between computational efficiency and estimation accuracy. Nevertheless, these methods are highly sensitive to the richness and discriminability of environmental features. In scenarios with sparse or highly repetitive structures, they are prone to mismatches, leading to degraded pose estimation or even state divergence [14, 15]. To address this, research has primarily progressed in two directions: (1) enhancing feature distinctiveness by incorporating auxiliary information such as intensity [13] and curvature [16] and (2) implementing multi-sensor fusion strategies, particularly tightly coupling LiDAR with inertial measurement unit (IMU) [17], such as LIO-SAM [18] and FAST-LIO2 [19]. This integration results in LiDAR–inertial odometry (LIO), which leverages high-frequency short-term motion constraints from IMU to improve system robustness. These approaches have achieved notable progress in general scenarios and partially alleviated issues related to feature degradation and motion discontinuity.

However, in complex agricultural environments, these solutions still face significant challenges. First, the prevalence of highly repetitive spatial structures, such as orderly arranged fruit trees and uniform furrows, hinders the establishment of stable and discriminative constraints, even with feature enhancement. Consequently, existing studies often resort to feature extraction strategies tailored to specific crop types, which, while effective in localized scenarios, exhibit limited cross-scene adaptability given the substantial diversity in crop species and planting patterns [20]. Second, during operations on loose soil or uneven terrain, agricultural robots frequently encounter jolts and slippage, inducing significant and non-stationary changes in sensor noise characteristics. Notably, most existing LiDAR–IMU fusion methods assume fixed noise parameters, a static modeling approach ill-suited to the dynamically varying noise profiles encountered in agricultural tasks. This mismatch in fusion weights can undermine the stability and long-term accuracy of state estimation [21].

Therefore, this paper proposes an L-SLAM framework tailored to complex agricultural robotic operations, with contributions in feature modeling and noise estimation. First, a curved-voxel (CV) [22] representation that fuses intensity and geometric structure is introduced to enhance registration stability in environments with repetitive structures. Second, a time-varying recursive noise estimator is devised to jointly model the noise characteristics of the LiDAR odometry (LO) and the IMU, thereby improving the system's adaptability to terrain disturbances and dynamic uncertainties. The main contributions of this paper are as follows:

(1) A robust L-SLAM architecture is proposed to counteract feature degradation caused by repetitive structures and dynamic sensor noise in agriculture, significantly improving mapping accuracy and operational stability in typical farmland and orchard scenarios.

(2) An intensity–geometry joint probability model based on CV is constructed. Point features are transformed into spatial features, and the distribution-to-distribution normal-distributions transform (D2D-NDT) algorithm [14] is employed to minimize the joint probability distance, thereby enhancing matching accuracy in feature-degraded environments.

(3) A maximum a posteriori (MAP)-based time-varying recursive noise estimator is designed. Employing a dual-filter architecture [23] enables adaptive modeling and joint recursive

https://doi.org/10.1017/S0263574726103397 Published online by Cambridge University Press

optimization of LO and IMU noise parameters, strengthening pose estimation accuracy and robustness under complex agricultural terrain.

2. Related work

2.1. Feature enhancement methods in L-SLAM

Traditional feature-based point cloud registration schemes, typified by LOAM [9], typically estimate inter-frame motion by extracting and matching key geometric features such as edge and planar points. The efficacy of these methods is highly dependent on the richness and discriminability of environmental features [24]. To enhance the distinctiveness of geometric features in sparse or structurally repetitive scenarios, researchers have incorporated auxiliary information such as LiDAR intensity and curvature, while also refining feature description strategies [25]. For instance, Wang et al. [13] proposed intensity-SLAM, which integrates intensity information into the optimization process to improve feature discrimination. Guo et al. [14] supplemented and reinforced sparse geometric features by fusing intensity data, thereby enhancing overall feature quality. Zhang et al. [26] further introduced an adaptive geometry–intensity weighting strategy to jointly model multimodal information during feature extraction. Choi et al. [16] employed local quadric surface modeling and optimized the registration process using a symmetric error function, bolstering system robustness in weak-feature environments.

In agricultural robotics applications, researchers have further tailored feature extraction and matching strategies by leveraging structural priors specific to farming environments. For example, Kim et al. [20] developed P-AgSLAM, a dual-LiDAR framework for cornfields, which extracts geometric features of corn stalks as stable landmarks to achieve reliable localization in row-crop settings. Xia et al. [27] targeted trellised orchard scenarios, utilizing the trellis structure as a stable matching target in conjunction with the NDT for point cloud registration. Aguiar et al. [28] achieved high-precision localization in vineyards by integrating point features with half-plane features. Teng et al. [29] proposed an adaptive mapping module that employs a consistency filtering mechanism to retain only stable points for matching, achieving centimeter-level accuracy in heavily occluded environments.

Although these methods have improved the adaptability of L-SLAM in agricultural settings to some extent, their feature enhancement strategies remain largely confined to point-level or localized geometric refinements. In environments characterized by large-scale repetitive structures, such as orderly crop rows and uniform furrows, point clouds still exhibit high similarity at the macro-structural level, making it difficult to establish stable and globally consistent geometric constraints [30]. Moreover, approaches reliant on specific crop morphologies or man-made structures, while effective in targeted scenarios, exhibit strong dependencies on crop type and environmental layout, limiting their generalizability across diverse agricultural contexts. Departing from these methods, this paper adopts a spatial structure modeling perspective by representing local point clouds using CV [22], thereby providing a more stable and generalizable feature representation for repetitive structural environments.

2.2. Noise modeling method based on LiDAR-IMU fusion in L-SLAM

Although feature enhancement improves accuracy to a certain extent, LiDAR alone still suffers from inherent limitations in temporal resolution and short-term motion constraints. To address this, contemporary L-SLAM systems widely adopt LiDAR–IMU multi-sensor fusion strategies, leveraging high-frequency inertial measurements from the IMU to provide short-term motion priors for pose estimation, thereby enhancing system continuity and robustness [31]. Relevant approaches primarily fall into two categories: factor graph optimization and Kalman filtering. Factor graph-based methods, exemplified by LIO-SAM [18], flexibly model LiDAR and IMU observations as constraint factors and jointly optimize them within a sliding window or global framework, offering advantages in modeling flexibility and global consistency [32]. These methods have been extensively applied in agricultural scenarios. For instance, Hong et al. [21] proposed a hierarchically coupled LiDAR–IMU system that achieves high-precision localization and mapping in orchard environments through the synergistic design of factor

https://doi.org/10.1017/S0263574726103397 Published online by Cambridge University Press

graph optimization and a geometric observer. Rincon et al. [33] incorporated trunk observation factors into the factor graph, reducing localization error to less than 0.18 m in apple and pear orchard datasets. Kim et al. [34] further improved the stability of localization and mapping results in agricultural settings by introducing a dynamic weight allocation mechanism based on covariance analysis into the factor graph. Gong et al. [35] developed a confidence region optimization algorithm to tightly couple LiDAR, IMU, and ultrasonic data within a factor graph framework, achieving high-precision localization in indoor plant factories.

In contrast, Kalman filtering-based methods typically employ an error-state model to realize tight coupling of LiDAR and IMU data through recursive updates, offering high computational efficiency and real-time performance. Representative methods include FAST-LIO2 [19] and LIO-EKF [36]. This category of methods has also seen widespread application in agricultural robotics. For example, Huang et al. [37] integrated ground constraints and loop closure detection into the FAST-LIO2 framework to achieve robust pose estimation in complex greenhouse environments. Yu et al. [38] utilized an unscented Kalman filter to fuse LiDAR and IMU observations, improving localization accuracy and system robustness in typical agricultural scenarios such as farm roads. Hoang et al. [39] proposed an odometry solution for autonomous agricultural machinery by first estimating IMU attitude precisely using an error-state Kalman filter and then integrating it into a multi-sensor fusion pipeline.

However, regardless of whether a factor graph or Kalman filtering approach is adopted, existing studies commonly treat sensor noise covariances as static or empirically determined constants [7]. In agricultural operational environments characterized by significant terrain undulations and complex wheel–terrain interactions, sensor noise often exhibits pronounced time-varying and nonstationary characteristics. Conventional static noise modeling approaches fail to accurately capture these variations, which can lead to mismatched fusion weights and degraded estimation performance. Departing from the static assumptions regarding noise statistics in existing methods, this study introduces an online recursive noise estimation mechanism to adaptively update the noise model, specifically addressing the time-varying noise properties prevalent in agricultural settings.

3. Methodology

3.1. System overview

This paper proposes an L-SLAM system tailored for agricultural environments, with the overall architecture illustrated in Figure 1. The system primarily takes LiDAR and IMU data as inputs and synergistically integrates a spatial feature enhancement module with a noise-state joint optimization module to output high-precision pose estimates and mapping results. For spatial feature modeling, the system constructs an intensity-geometry joint probabilistic model using CV [22], transforming point features into spatial features. Subsequently, the D2D-NDT [14] is employed to minimize the distribution distance between the current frame and the local map, yielding LO results. In the joint noise-state optimization stage, a time-varying recursive noise estimator dynamically models the noise parameters of both the LO and IMU. This is coupled with a dual-filter architecture to jointly optimize the system state and noise parameters, thereby enhancing the overall accuracy and robustness of the system.

3.2. Spatial feature modeling for repetitive structure

To address the feature matching challenges arising from highly repetitive structures in agricultural environments, an intensity–geometry joint probability model based on CV [22] is proposed. This model enhances point cloud discriminability within structures exhibiting similar geometric shapes and spatial distributions, suppressing mismatches and improving the accuracy and robustness of LO registration.

To accurately characterize local geometric properties of farmland scenes, the system first partitions the point cloud space using a volume of interest (VOI). The physical domain of a VOI is defined as

$$V_t = \{p_k | p_k \in P_t, \rho_k < L_{\max}, h_{\min} < z_k < h_{\max}\} \tag{1}$$

https://doi.org/10.1017/S0263574726103397 Published online by Cambridge University Press

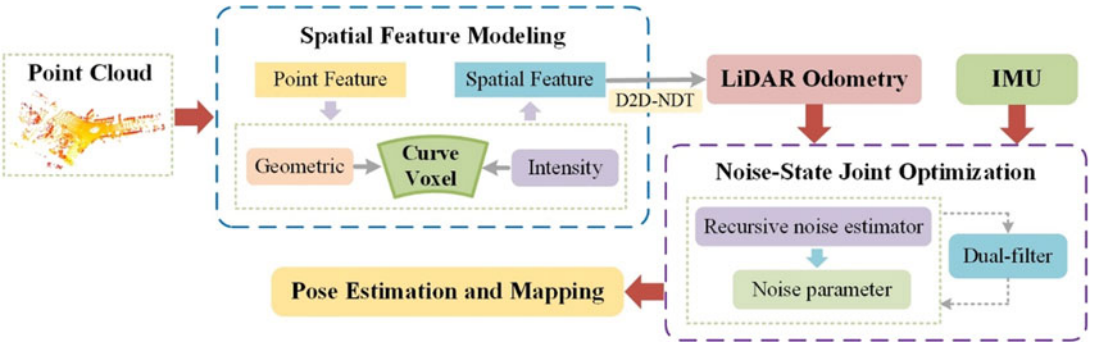


Figure 1. *System overview.*

where p_k represents points within the VOI of the current frame P_t, and $\rho_k = \sqrt{x_k^2 + y_k^2}$, $L_{\max}$, $h_{\min}$ and $h_{\max}$ are constants defining the VOI's radial and elevation boundaries.

Next, intensity information is introduced as a complementary dimension to eliminate matching ambiguities arising from highly similar geometric structures. The LiDAR intensity value I is determined by

$$I = \frac{P_r}{P_e} = \eta \frac{\rho \cos \alpha}{R^2} \tag{2}$$

where P_e denotes the laser transmission power, P_r represents the received signal energy, η is a device-specific constant, ρ is the material reflectivity, R is the distance between LiDAR and object surface, and α is the incidence angle between the laser beam and surface normal.

However, in agricultural scenes, severe incident angle variations on foliage and furrows, render raw intensities unreliable. We address this limitation by employing a geometry-based intensity correction method [13]. For any given point $p \in R^3$, we use its neighboring points p_1 and p_2 to compute the local normal vector, subsequently deriving the incidence angle for intensity correction. This process enables a more accurate representation of crop or soil surface properties.

$$n = \frac{(p - p_1) \times (p - p_2)}{|p - p_1| \cdot |p - p_2|} \tag{3}$$

$$\cos \alpha = \frac{p^T \cdot n}{|p|} \tag{4}$$

To accommodate the within-row point density variation (dense near, sparse far), we employ a distance threshold instead of a point-count threshold [40]. We construct a viewpoint-invariant evaluation function for both roughness and intensity distribution, which scales the geometric relationship between a point p_i and the points p_{i-n} and p_{i+n} adjacent to it on either side. This function is defined as

$$\sigma_i^c = \frac{1}{N_i} \sum_{n=1}^{N_i} \frac{\delta d \, ||(p_{i+n} - p_i) - (p_{i-n} - p_i)||}{\min \left(||(p_{i+n} - p_i)|| \, , ||(p_{i-n} - p_i)|| \right)} \tag{5}$$

$$\sigma_i^I = \frac{1}{N_i} \sum_{n=1}^{N_i} \frac{\delta d \, ||(\eta_{i+n} - \eta_i) - (\eta_{i-n} - \eta_i)||}{\min \left(||(\eta_{i+n} - \eta_i)|| \, , ||(\eta_{i-n} - \eta_i)|| \right)} \tag{6}$$

where η_i represents the corrected intensity threshold. Furthermore, N_i denotes the number of neighboring points on each side, which is calculated as

https://doi.org/10.1017/S0263574726103397 Published online by Cambridge University Press

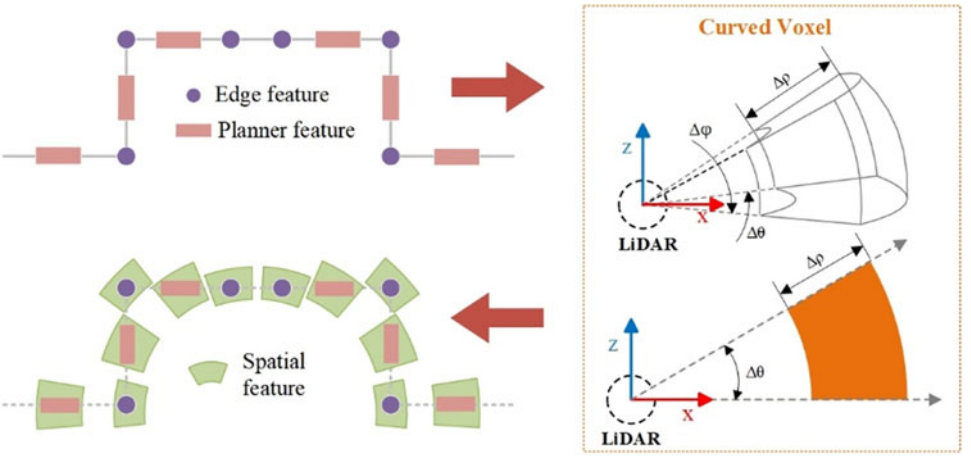


Figure 2. *Schematic of spatial feature modeling.*

$$N_i^L = \min_{n>0}(n)\ s.t.\ ||(p_{i-n} - p_i)|| \geq \delta d \tag{7}$$

$$N_i^R = \min_{n>0}(n)\ s.t.\ ||(p_{i+n} - p_i)|| \geq \delta d \tag{8}$$

$$N_i = \max\left(N_i^L, N_i^R\right) \tag{9}$$

Subsequently, we adopt a weighted fusion method for feature extraction, where ω^c and ω^I represent the weights of geometric and intensity distributions, respectively. During each scan, edge and plane features are selected from regions with larger and smaller σ_i:

$$\sigma_i = \omega^c \cdot \sigma_i^c + \omega^I \cdot \sigma_i^I \tag{10}$$

As shown in Figure 2, the core procedure for transforming point features into spatial features constructs an intensity–geometry joint probability model within CV. This process begins by mapping the point cloud space V_t to a spherical coordinate system. Subsequently, the point cloud is partitioned into CVs based on distance, angular resolution, and sparsity.

The polar coordinate system $P(\rho, \theta, \phi)$ of a CV comprises radial distance, azimuth angle, and polar angle. Taking the m, n, o voxel as an example, it is defined as:

$$\begin{aligned} CV_{m,n,o} = \{P(\rho,\theta,\phi) = |&\Delta\rho \cdot (m-1) \leq \rho < \Delta\rho \cdot m, \\ &\Delta\theta \cdot (n-1) \leq \theta < \Delta\theta \cdot n, \\ &\Delta\phi \cdot (o-1) \leq \phi < \Delta\phi \cdot o\} \end{aligned} \tag{11}$$

where $\Delta\rho, \Delta\theta$ and $\Delta\phi$ represent the incremental units adjusted based on point cloud sparsity and distance.

It is important to identify the $CV_{m,n,o|i}$ containing the feature point p_i in order to establish matching constraints for local structural information. For brevity, indices i and j hereafter are used to denote $m, n, o|i$ and $m, n, o|j$, respectively.

The point cloud within a CV is assumed to follow a Gaussian distribution $N(\mu_i, \sum_i)$, where the probability density parameters are computed as

$$\mu_i = \frac{1}{|CV_i|} \sum_{p_i \in CV_i} PI_i \tag{12}$$

$$\sum_i = \frac{1}{|CV_i| - 1} \sum_{p_i \in CV_i} (PI_i - \mu_i)(PI_i - \mu_i)^T \tag{13}$$

https://doi.org/10.1017/S0263574726103397 Published online by Cambridge University Press

where μ_i is the mean vector, $\sum_i$ is the covariance matrix, and $PI_i = (\rho_i, \theta_i, \phi_i, I_i)$ encodes the joint statistical properties of spatial geometry and intensity, elevating a single point feature to a spatial descriptor that represents local structure.

The matching of local structural information is achieved using the D2D-NDT [14] method to minimize the joint probability distribution distance between the current frame and the local map. The optimization objective is given below:

$$f_{d2d} = \sum_{CV_i} \sum_{CV_j} -d_1 \exp\left(-\frac{d_2}{2} \mu_{ij}^T \left(R^T \sum_i R + \sum_j\right)^{-1} \mu_{ij}\right) \tag{14}$$

where $\mu_{ij} = R\mu_i + t - \mu_j$ and d_1, d_2 are regularization factors.

Iterative optimization is used to align the local structural information, and the resulting LO serves as input for the subsequent state estimation layer.

3.3. Noise–state joint optimization via recursive estimation

To counteract the pronounced time-varying and non-stationary noise characteristics of multi-source sensors caused by severe terrain vibrations and wheel slip during operation in unstructured farmland, a time-varying recursive noise estimator is devised. Its objective is to estimate the noise statistics of the LO and the IMU under dynamic motion conditions in real time. Coupled with this estimator, a dual-filter architecture jointly propagates and refines both noise parameters and system states through recursive inference, thereby markedly improving pose-estimation accuracy and robustness in agricultural environments.

The system model consists of state-observation equations. The state vector is defined as follows:

$$x_k = \begin{bmatrix} p_k & \varphi_k & v_k & w_k \end{bmatrix}^T \tag{15}$$

where p_k and φ_k denote position and orientation, respectively, and v_k and w_k represent linear and angular velocities, respectively.

The state transition and observation equations are

$$\hat{x}_k = F_k \hat{x}_{k-1} + B_k u_k + w_k \tag{16}$$

$$z_k = H_k \hat{x}_k + v_k \tag{17}$$

where F_k is the state transition matrix, B_k is the control input matrix, u_k is the control input, and w_k and v_k denote the process and observation noise, respectively.

Position and orientation updates are derived from velocities as follows:

$$\begin{bmatrix} p \\ \varphi \end{bmatrix} = \begin{bmatrix} p + R(\varphi)\, v \Delta t \\ \varphi + J(\varphi)\, w \Delta t \end{bmatrix} \tag{18}$$

where $R(\varphi) \in SO(3)$ is the rotation matrix, $J(\varphi)$ is the Jacobian mapping relating angular velocity to Euler derivatives, and Δt is the sampling interval.

The joint observation model $Z_k = [\, Z_L \; Z_I \,]^T$ combines LO and IMU. Velocity is used instead of pose as the observation metric, based on the EKF-LOAM [41] principles.

The linear velocity v_L and angular velocity w_L are obtained from the LO as:

$$Z_L = \begin{bmatrix} v_L & w_L \end{bmatrix}^T \tag{19}$$

$$\begin{bmatrix} v_L \\ w_L \end{bmatrix} = \begin{bmatrix} R(\varphi_L) & o_{3\times3} \\ o_{3\times3} & J(\varphi_L) \end{bmatrix} \begin{bmatrix} \Delta p_L \\ \Delta \varphi_L \end{bmatrix} \frac{1}{\Delta t_L} \tag{20}$$

where Δp and $\Delta \varphi$ represent the position and direction changes in LO, respectively.

The angular velocity w_I is measured from the IMU as

$$Z_I = w_I \tag{21}$$

https://doi.org/10.1017/S0263574726103397 Published online by Cambridge University Press

Traditional fusion schemes assume constant noise covariances, but in complex agricultural scenes, the noise characteristics of LO and IMU exhibit significant heterogeneity. If fixed parameters are directly used, it will cause fusion weight mismatch and state estimation divergence problems. To address this, an MAP time-varying recursive noise estimator is formulated. The associated noise parameters are defined as:

$$w_k = \frac{1}{k}\sum_{j=1}^{k}\left[\hat{x}_{j,k} - F_{j-1}\hat{x}_{j-1,k}\right] \tag{22}$$

$$v_k = \frac{1}{k}\sum_{j=1}^{k}\left[Z_j - H_j\hat{x}_{j,k}\right] \tag{23}$$

$$Q_k = Cov\left[w_k, w_j\right] = \frac{1}{k}\sum_{j=1}^{k}\left[\hat{x}_{j,k} - F_{j-1}\hat{x}_{j-1,k} - w\right]\left[\hat{x}_{j,k} - F_{j-1}\hat{x}_{j-1,k} - w\right]^T \tag{24}$$

$$R_k = Cov\left[v_k, v_j\right] = \frac{1}{k}\sum_{j=1}^{k}\left[z_j - H_j\hat{x}_{j,k} - r\right]\left[z_j - H_j\hat{x}_{j,k} - v\right]^T \tag{25}$$

where w_k and v_k are the process and observation noise, respectively, and Q_k and R_k are the process and observation noise covariance, respectively.

However, the noise parameters contain complex multi-step smoothing terms, such as $\hat{x}_{j-1,k}$ and $\hat{x}_{j,k}$, which make it difficult to achieve recursive computation. To address this issue, we adopt a single-step smoothing approximation of $\hat{x}_{j-1,j} \approx \hat{x}_{j-1,k}, \hat{x}_{j,j} \approx \hat{x}_{j,k}$, thereby obtaining

$$w_k = \frac{1}{k}\sum_{j=1}^{k}\left[\hat{x}_{j,j} - F_{j-1}\hat{x}_{j-1,j}\right] \tag{26}$$

$$Q_k = \frac{1}{k}\sum_{j=1}^{k}\left[\hat{x}_{j,j} - F_{j-1}\hat{x}_{j-1,j} - w\right]\left[\hat{x}_{j,j} - F_{j-1}\hat{x}_{j-1,j} - w\right]^T \tag{27}$$

$$v_k = \frac{1}{k}\sum_{j=1}^{k}\left[Z_j - H_j\hat{x}_{j,j}\right] \tag{28}$$

$$R_k = \frac{1}{k}\sum_{j=1}^{k}\left[z_j - H_j\hat{x}_{j,j} - v\right]\left[z_j - H_j\hat{x}_{j,j} - v\right]^T \tag{29}$$

The noise estimator is effectively integrated into the system by using the dual-filter structure proposed by Gao et al. [23]. This structure consists of two parts: noise estimation propagation and state estimation correction. The former obtains the single-step smoothed value $\hat{x}_{k-1,k}$ of the propagating noise estimates based on the $k-1$ step estimate $\hat{x}_{k-1,k-1}$ and the k step innovation $\overline{\varepsilon}_k$ as follows:

$$\overline{x}_{k,k-1} = F_{k-1}\hat{x}_{k-1,k-1} + w_{k-1} \tag{30}$$

$$\overline{\varepsilon}_k = z_k - H_k\overline{x}_{k,k-1} - v_k \tag{31}$$

$$P_{k,k-1} = F_{k-1}P_{k-1}F_{k-1}^T + Q_{k-1} \tag{32}$$

$$\overline{K}_k = P_{k-1}F_{k-1}^T H_k^T\left(H_k P_{k,k-1}H_k^T + R_k\right)^{-1} \tag{33}$$

$$\hat{x}_{k-1,k} = \hat{x}_{k-1,k-1} + \overline{K}_k\overline{\varepsilon}_k \tag{34}$$

The latter corrects the state estimate $\hat{x}_{k-1,k}$ using the single-step smoothed value $\hat{x}_{k,k}$ as

$$\hat{x}_{k,k-1} = F_{k-1}\hat{x}_{k-1,k} + w_{k-1} \tag{35}$$

$$\varepsilon_k = z_k - H_k\hat{x}_{k,k-1} - v_k \tag{36}$$

$$K_k = P_{k,k-1}H_k^T\left(H_k P_{k,k-1}H_k^T + R_k\right)^{-1} \tag{37}$$

$$\hat{x}_{k,k} = \hat{x}_{k,k-1} + K_k\varepsilon_k \tag{38}$$

$$P_k = (I - K_k H_k)\,P_{k,k-1} \tag{39}$$

https://doi.org/10.1017/S0263574726103397 Published online by Cambridge University Press

The noise estimator can be redefined according to the dual-filter structure, expressed in (40) and (41) as follows:

$$\hat{x}_{j,j} - F_{j-1}\hat{x}_{j-1,j} = K_j\varepsilon_j + w_{j-1} \tag{40}$$

$$Z_j - H_j\hat{x}_{j,j} = Z_j - H_j\left(\hat{x}_{j,j-1} + K_j\varepsilon_j\right) = \left(I - H_jK_j\right)\varepsilon_j + v_j \tag{41}$$

Furthermore, an exponentially weighted fading mechanism combined with a dynamic-window strategy is introduced to accurately capture the time-varying nature of sensor noise. Specifically, the exponential weighting term is defined as

$$\alpha^k = \frac{\beta}{N_k + 1} \tag{42}$$

where α^k is the exponential weighting term, β is a constant term, and N_k is the window size. The term α^k is adjusted according to N_k: when N_k is small, α^k is larger, giving higher weight to recent data; when N_k is large, α^k is smaller, which allows more historical data to be considered.

The window size N_k is adjusted based on ${\sigma_k}^2$, which is the variance of the innovation sequence within the window. This process is represented by the following equations:

$$\sigma_k^2 = \frac{1}{N_k}\sum_{i=k-N_k+1}^{k}\left(\varepsilon_i - \overline{\varepsilon}_k\right)^2 \tag{43}$$

$$N_k = N_{k-1} + \gamma\left({\sigma_k}^2 - \overline{\sigma}^2\right) \tag{44}$$

where ${\sigma_k}^2$ reflects the noise fluctuation in the current window. Furthermore, ε_i represents the i-th innovation value, $\overline{\varepsilon}_k$ is the mean of the innovation sequence within the window, $\overline{\sigma}^2$ is the historical mean variance, and γ is a tuning factor.

In the presence of severe noise fluctuations, the window size is increased to improve the smoothness of the estimation results. When the noise stabilizes, the window size is reduced to ensure the system's real-time responsiveness.

Thus, the final noise estimator is defined as

$$w_k = w_{k-1} + \alpha^k\left[K_j\varepsilon_j + w_{j-1}\right] \tag{45}$$

$$Q_k = \left(1 - \alpha^k\right)Q_{k-1} + \alpha^k\left[K_j\varepsilon_j\varepsilon_j^TK_j^T\right] \tag{46}$$

$$v_k = v_{k-1} + \alpha^k\left[\left(I - H_jK_j\right)\varepsilon_j + v_j\right] \tag{47}$$

$$R_k = \left(1 - \alpha^k\right)R_{k-1} + \alpha^k\left[\left(I - H_jK_j\right)\varepsilon_j\varepsilon_j^T\left(I - H_jK_j\right)^T\right] \tag{48}$$

4. Experiments

4.1. Experimental setup

To evaluate the proposed method, experiments are conducted along two dimensions: generalizability and agricultural adaptability. General performance is assessed on public datasets (KITTI [42], NCLT [43]), while task-oriented functionality is validated in both simulated agricultural environments (Gazebo) and real-world farmland. Evaluation metrics include relative translational error (RTE) and relative rotational error (RRE) as quantitative baselines [44], and qualitative analysis is carried out in combination with the generated point cloud maps. RTE and RRE, established metrics in the SLAM domain, quantify the discrepancies between the estimated relative pose transformations and the ground truth over a time interval Δ. For a given time i, the relative pose transformation error E_i is defined as:

$$E_i = \left(Q_i^{-1}Q_{i+\Delta}\right)^{-1}\left(P_i^{-1}P_{i+\Delta}\right) \tag{49}$$

where Q denotes the ground truth pose sequence and P the estimated pose sequence.

https://doi.org/10.1017/S0263574726103397 Published online by Cambridge University Press

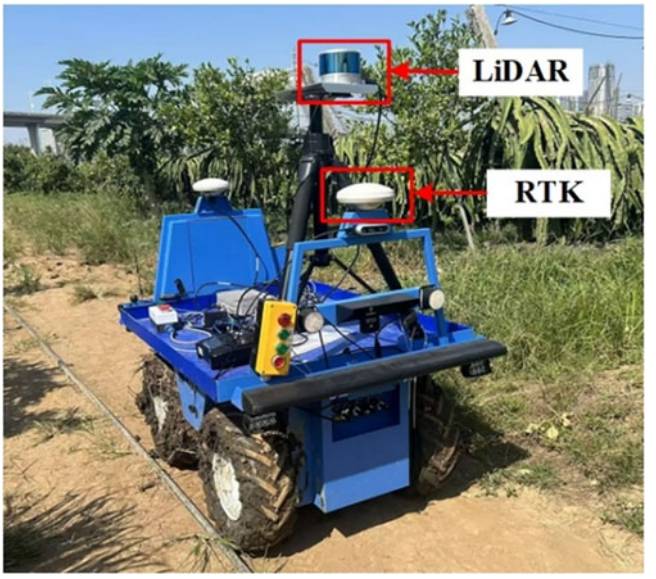


Figure 3. *Experimental platform.*

Subsequently, RTE and RRE are calculated as:

$$RTE_i = ||trans\ (E_i)|| \tag{50}$$

$$RRE_i = \angle\ (rot\ (E_i)) \tag{51}$$

where $trans(E_i)$ extracts the translational vector from E_i, and $\angle(rot(E_i))$ converts the rotational matrix extracted from E_i into an angular representation.

In agricultural scenes, row spacing is adopted as the core agronomic structural indicator. Row spacing, a key spatial parameter in agricultural remote sensing and agronomic mapping, directly reflects the consistency and precision of inter-row crop distribution. By quantifying the structural fidelity of the reconstructed map, this metric evaluates SLAM's potential for remote sensing detail supplementation. Computation proceeds as follows: after orthographic projection of the reconstructed cloud, crop centers are extracted and trunk centerlines are fitted, and the Euclidean distance between adjacent plants is calculated and compared with ground truth.

In terms of hardware configuration, the public dataset and Gazebo experiments are executed on a notebook with an Intel Core i7-12700H CPU and 32 GB RAM under Ubuntu 20.04, with ROS Noetic. Real-field experiments employ an agricultural transport robot equipped with an LS-C16 LiDAR, and centimeter-level ground truth is obtained via RTK (as shown in Figure 3).

4.2. Generalization validation on public datasets

To evaluate the generalization performance of the proposed method under different environmental conditions, this paper conducts experimental validation using the widely recognized KITTI [42] and NCLT [43] datasets in the SLAM community. The KITTI dataset, a classic benchmark in autonomous driving, was collected from real urban environments and features rich structured roads and buildings, alongside geometrically sparse regions such as highways. It is well-suited for assessing system robustness under feature-degraded conditions. The NCLT dataset, a long-term large-scale collection from a campus environment, exhibits significant temporal spans and seasonal variations, thereby effectively testing the consistency and stability of SLAM systems during prolonged operation. The selection of these two representative public datasets for comparative experiments aims to systematically verify the generalization capability of the proposed method on standard benchmark datasets, demonstrating that its core design

https://doi.org/10.1017/S0263574726103397 Published online by Cambridge University Press

Table I. *Relative pose errors on the KITTI dataset.*

Seq.	A-LOAM	Intensity-SLAM	F-LOAM	Ours
01	2.69/0.66	6.22/0.07	3.21/1.14	1.12/0.64
03	1.57/0.69	2.34/0.77	1.04/0.90	0.92/0.65
06	0.79/0.31	1.82/0.78	1.63/1.52	0.69/0.22
10	1.67/0.63	2.25/1.18	1.18/1.29	0.75/0.50
Avg.	1.68/0.57	3.16/0.70	1.77/1.21	0.87/0.50

Note that the performance is presented in the "A/B" format, where "A" and "B" represent the RTE (%) and °/100 m, respectively.

is not an overfitting optimization tailored to specific environments. Moreover, the geometric monotony and feature sparsity issues present in these datasets share intrinsic similarities with the feature degradation caused by orderly crop rows in agricultural environments, thereby providing empirical support for its applicability in farming scenarios.

The evaluation first focuses on the KITTI dataset. Two representative scenario types are selected: (1) feature-sparse highway sequences (01, 10) with large scale topology and (2) complex urban sequences (03, 06) containing wide streets and multiple corner obstacles. Since the KITTI dataset used in this experiment only provides LiDAR data, the spatial feature modeling module of the proposed method is exclusively enabled to validate its effectiveness in pure LO tasks. Three representative state-of-the-art pure LO methods are selected as baselines for comparison. A-LOAM [9], an efficient implementation of the LOAM framework, is employed as a baseline method due to its stable performance, which relies on matching extracted edge and planar points. F-LOAM [11] adopts a pure scan-to-map framework, directly registering each point cloud frame to a local map to achieve higher computational efficiency while maintaining accuracy. Intensity-SLAM [13] leverages both geometric features and point cloud intensity information to enhance robustness in complex environments.

Table I shows the pose estimation results obtained by different methods. The results demonstrate that by constructing a joint probabilistic model based on CV, which transforms traditional point features into spatial features, the localization accuracy of the system is significantly improved. Across the four sequences, the average RTE and RRE are reduced to 0.87% and 0.50°/100 m, respectively. This represents a 48.2% improvement in translation accuracy and a 14% improvement in rotation accuracy compared to the second-best method, A-LOAM, thereby further validating the effectiveness of the proposed method in enhancing LO precision. F-LOAM achieves translation accuracy comparable to A-LOAM, but its rotation error is considerably higher than that of other methods. This indicates that while the scan-to-map framework can improve computational efficiency, it exhibits limited capability in estimating high-frequency rotational motions, making it susceptible to error accumulation. In contrast, Intensity-SLAM shows a larger translation error, as its performance heavily relies on the consistency of environmental intensity features. In outdoor scenarios affected by varying illumination, the reliability of intensity features diminishes, consequently degrading localization accuracy.

Mapping results are shown in Figure 4(a–d). The proposed method produces complete and continuous point cloud maps under all tested conditions, accurately reconstructing spatial structure. Even in geometrically sparse regions (e.g., flat highway sections), the system maintains stable mapping quality without discernible drift. Figure 4(e–h) presents the trajectory estimation outcomes of different methods. The trajectory obtained through spatial feature registration aligns most closely with the ground truth, exhibiting substantially lower errors compared to the baseline methods. In contrast, the trajectories generated by other methods display varying degrees of drift in feature-sparse regions. Particularly, Intensity-SLAM suffers from severe drift in Sequence 01, leading to a complete system failure.

Subsequently, the complete system (integrating both LiDAR and IMU data) is validated on the NCLT dataset. Four representative LIO methods are selected as baselines for comparison: LIO-SAM, a classic factor graph-based optimization method; FAST-LIO2 [19] and LIO-EKF [36], which are representative methods under the Kalman filtering framework; and HC-LIO [21], a method specifically designed

https://doi.org/10.1017/S0263574726103397 Published online by Cambridge University Press

Table II. *Relative pose errors on the NCLT dataset.*

Seq.	LIO-SAM	FAST-LIO2	LIO-EKF	HC-LIO	Ours
20,120,429 (N1)	0.74/1.96	0.21/1.83	0.28/1.73	0.45/1.50	0.19/1.31
20,120,615 (N2)	0.56/0.94	0.58/0.36	0.55/0.46	0.71/0.90	0.64/0.55
20,130,110 (N3)	0.53/1.26	0.31/0.92	0.34/0.98	0.62/1.13	0.21/0.68
Avg.	0.61/1.39	0.37/1.04	0.39/1.06	0.59/1.18	0.35/0.85

Note that the performance is presented in the "A/B" format, where "A" and "B" represent the RTE (%) and °/100 m, respectively.

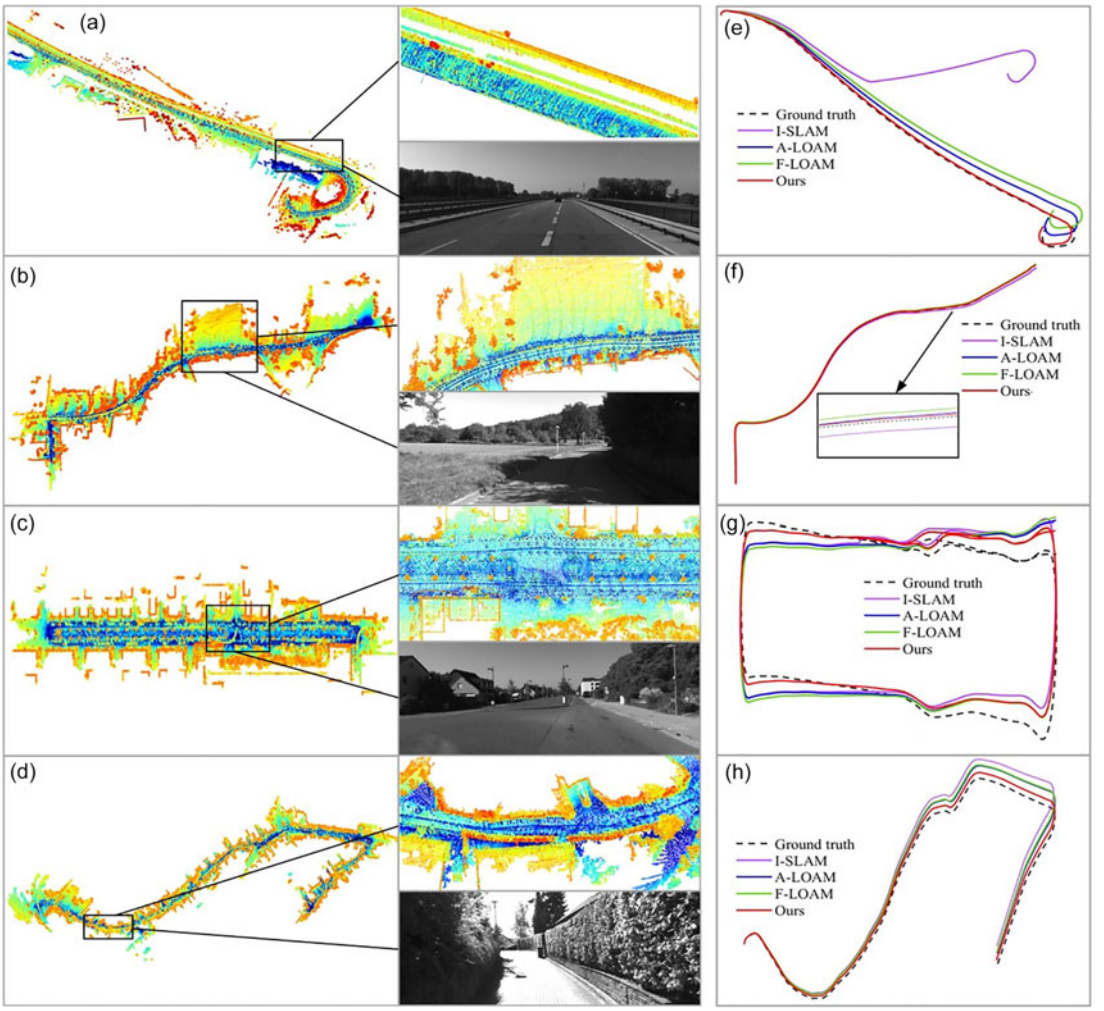


Figure 4. *Mapping and trajectory results on the KITTI. (a–d) Mapping results on Seq. 01, 03, 06, and 10, (e–h) corresponding trajectory results.*

for agricultural operational environments. It should be noted that pure LO methods are excluded from this comparison, as they cannot operate effectively on this dataset. The pose estimation results of these methods are summarized in Table II. The proposed complete system achieves the best performance across most sequences, with average RTE and RRE controlled at 0.35% and 0.85°/100 m, respectively. LIO-EKF and FAST-LIO2 exhibit comparable overall performance, demonstrating high accuracy and stability in translation estimation, though their rotation estimation accuracy is relatively limited.

https://doi.org/10.1017/S0263574726103397 Published online by Cambridge University Press

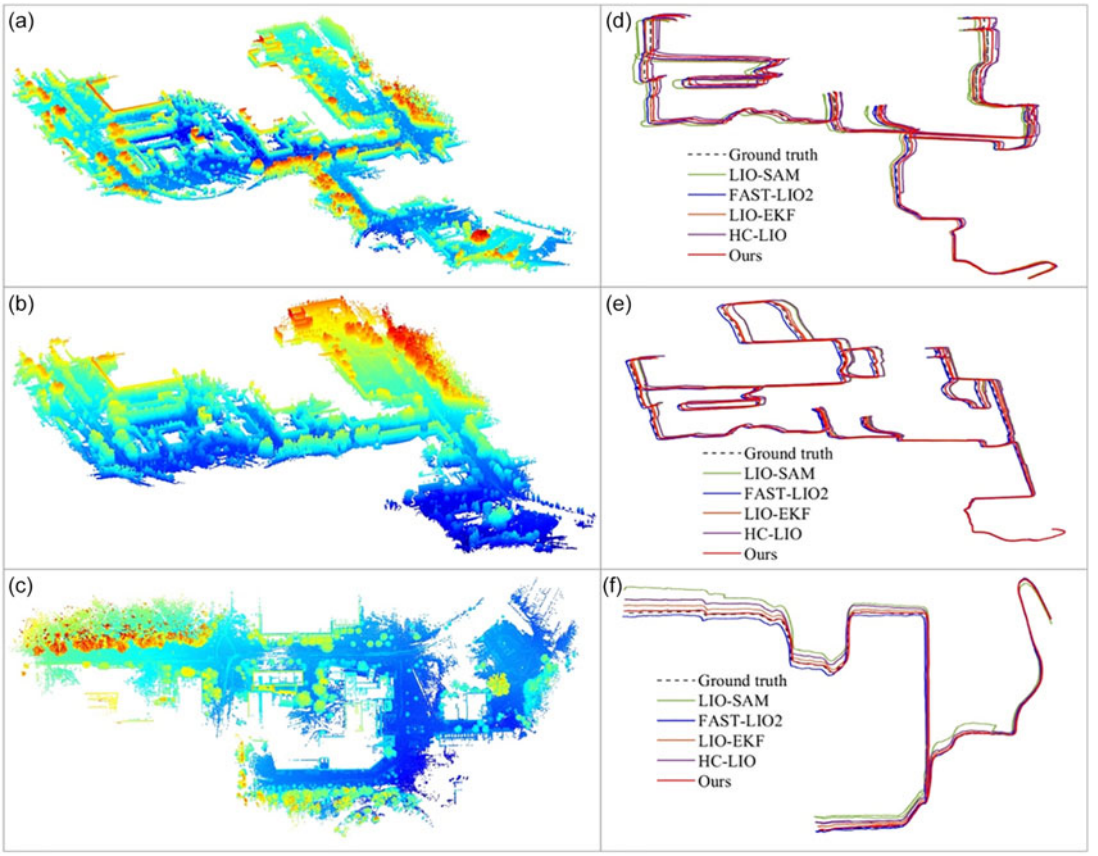


Figure 5. *Mapping and trajectory results on the NCLT. (a)–(c) Mapping results on Seq. N1, N2, and N3, (d)–(f) corresponding trajectory results.*

In contrast, leveraging the noise-state joint optimization mechanism, the proposed method significantly enhances the system's ability to constrain rotational motion, resulting in more stable and improved pose estimation accuracy. HC-LIO outperforms LIO-SAM, but both methods lag notably behind the filter-based approaches. This outcome indicates that in large-scale, long-term operational scenarios like NCLT, factor graph-based methods face greater challenges in terms of real-time performance and consistency. Moreover, while HC-LIO is tailored for agricultural environments, its cross-domain adaptability in general scenarios remains limited, further highlighting the advantages of the proposed method in terms of cross-scene applicability.

Figure 5 shows the mapping and trajectory results on the NCLT dataset. From Figure 5(a–c), it can be seen that the proposed full system reconstructs large-scale scene details, including street layouts and vegetation distributions. The trajectory results in Figure 5(d–f) show that the proposed method consistently achieves high trajectory accuracy across all test sequences, with estimated trajectories aligning most closely with the ground truth. In contrast, the baseline methods exhibit more pronounced trajectory errors in certain sequences.

To quantify the joint optimization effect of the proposed time-varying recursive noise estimator and the dual-filter architecture, a comparative experiment was conducted on sequence N3 against a conventional EKF baseline. Error statistics are shown in Figure 6. The joint optimization confines absolute position error in the XY plane to 0.89 m and keeps heading error within 1.02°, both significantly outperforming the EKF. The error traces reveal pronounced fluctuations for the EKF in local regions,

https://doi.org/10.1017/S0263574726103397 Published online by Cambridge University Press

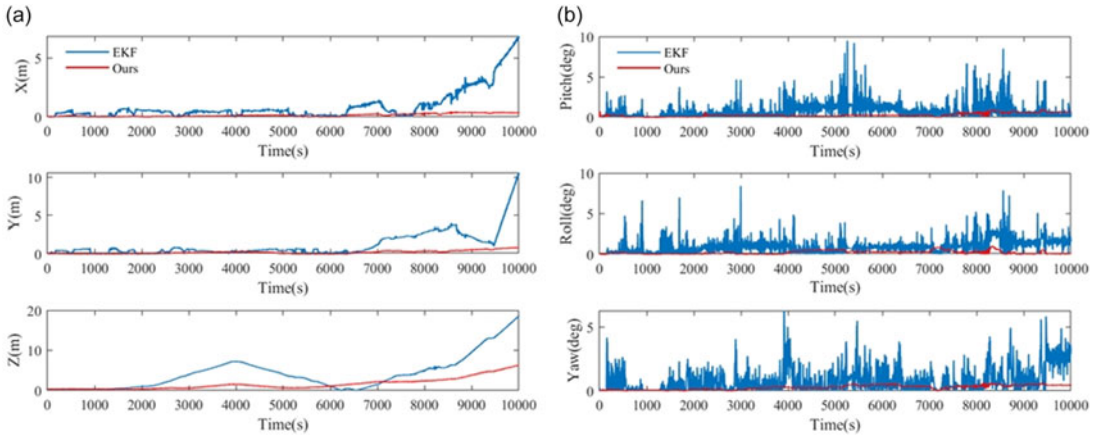


Figure 6. *Error statistics. (a) Translation error, (b) rotation error.*

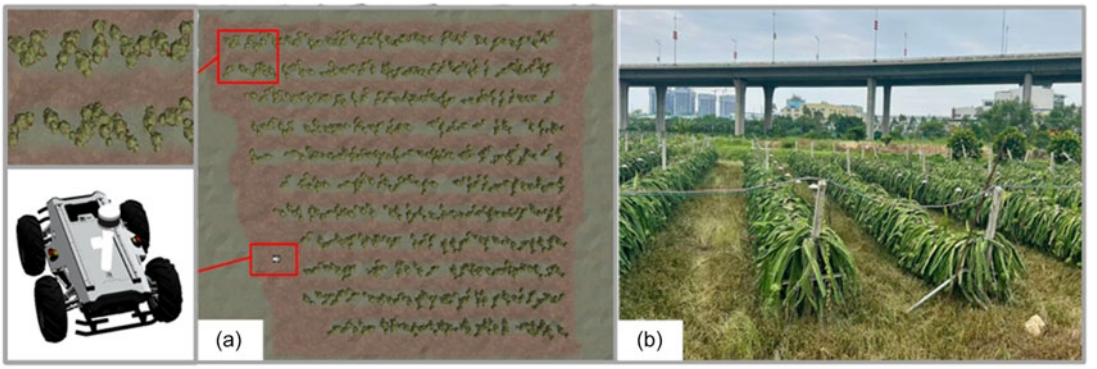


Figure 7. *Simulated and real agricultural scenes. (a) Simulated agricultural scene, (b) real agricultural scene.*

whereas the adaptive adjustment of noise parameters by the recursive estimator effectively suppresses these fluctuations, yielding superior stability and reliability in pose estimation.

Collectively, the public dataset results demonstrate that the proposed spatial feature modeling and joint noise–state optimization framework is fully transferable across scenarios. Although originally designed to address structural degradation and nonstationary noise in agriculture, the technical framework confers universal advantages in complex environments.

4.3. Agricultural scene validation

To comprehensively evaluate performance in agricultural contexts, experiments were conducted in both simulated and real-world scenes. As shown in Figure 7, both environments contain regularly spaced, highly repetitive crop rows and feature complex, undulating terrain, providing stringent tests of scene adaptability.

The proposed system is compared with three representative methods: A-LOAM [9], a pure LO method; FAST-LIO2 [19], a representative LIO method (both of which performed well in the generalizability validation); and HC-LIO [21], a method specifically designed for agricultural scenarios. The pose estimation results are summarized in Table III. The results show that the proposed system achieves average RTE and RRE values of 1.42% and 1.53°/100 m, respectively, significantly outperforming all baseline methods. In the simulation environment, A-LOAM and FAST-LIO2 exhibit comparable accuracy, indicating that the IMU noise is in an ideal state under these conditions, and the primary factor

https://doi.org/10.1017/S0263574726103397 Published online by Cambridge University Press

Table III. *Relative pose errors on the agricultural scenes.*

Seq.	A-LOAM	FAST-LIO2	HC-LIO	Ours
Simulated	3.01/2.33	2.84/2.08	1.92/1.75	1.68/1.37
Real	6.84/5.93	3.37/2.69	2.15/2.02	1.15/1.69
Avg.	4.93/4.13	3.11/2.39	2.04/1.89	1.42/1.53

Note that the performance is presented in the "A/B" format, where "A" and "B" represent the RTE (%) and °/100 m, respectively.

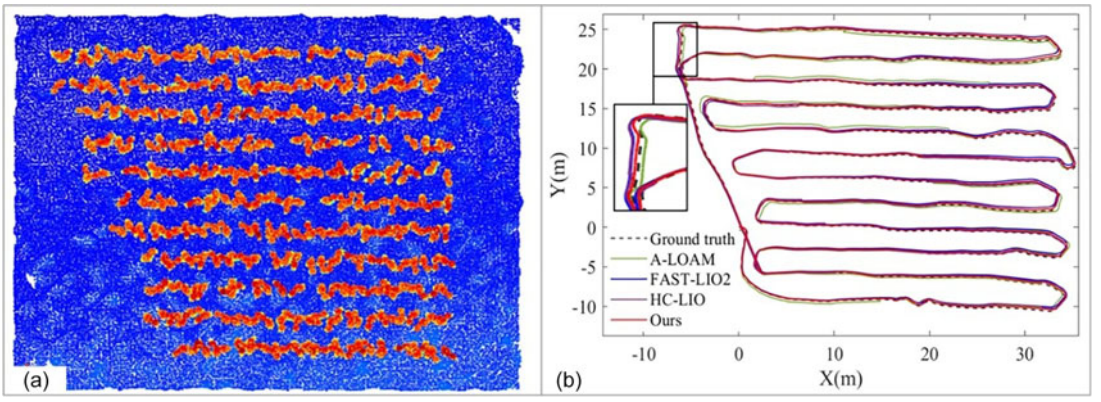


Figure 8. *Mapping and trajectory results on the simulated agricultural scenes. (a) Mapping result, (b) trajectory result.*

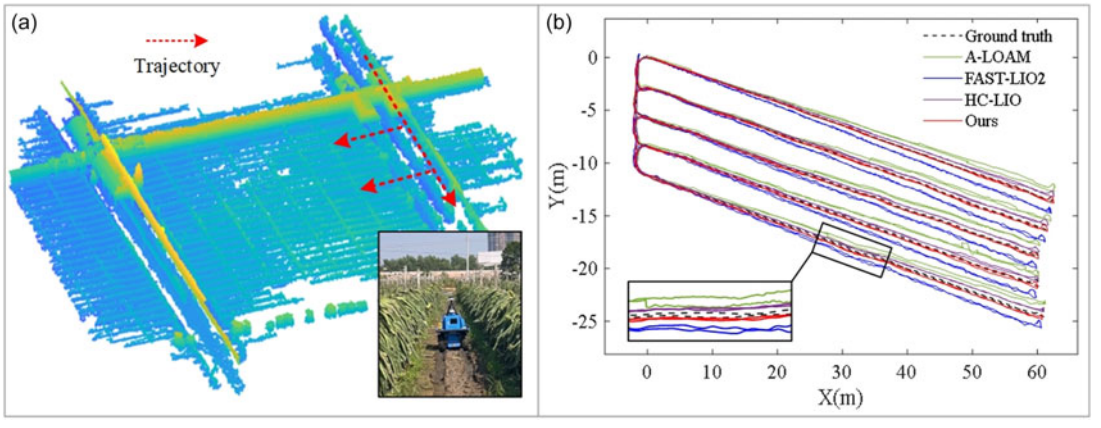


Figure 9. *Mapping and trajectory results on the real agricultural scenes. (a) Mapping result, (b) trajectory result.*

limiting system accuracy is the feature extraction and registration performance of the LO. Under these conditions, both the proposed method and HC-LIO achieve better localization accuracy than A-LOAM and FAST-LIO2 by improving their feature extraction and matching strategies. However, the spatial feature modeling approach proposed more effectively characterizes the geometric structures of regularly arranged crops in agricultural environments, resulting in superior localization accuracy compared to HC-LIO. In real agricultural scenarios, the accuracy of A-LOAM decreases significantly. In contrast, all LIO methods that fuse IMU data outperform A-LOAM, demonstrating that a single sensor is insufficient to meet high-precision localization requirements in complex real-world environments. Specifically, by integrating spatial feature modeling with noise-state joint optimization, the proposed method achieves a 43.6% improvement in RTE and a 10.6% improvement in RRE over the second-best method, HC-LIO,

https://doi.org/10.1017/S0263574726103397 Published online by Cambridge University Press

Table IV. *Row-spacing estimation results on the agricultural scenes.*

Seq.	Truth (m)	A-LOAM (m)	FAST-LIO2 (m)	HC-LIO (m)	Ours (m)
Simulated	3.673	3.855	3.782	3.745	3.731
Real	4.164	4.583	4.397	4.305	4.247
Avg-error	–	0.301	0.168	0.103	0.071

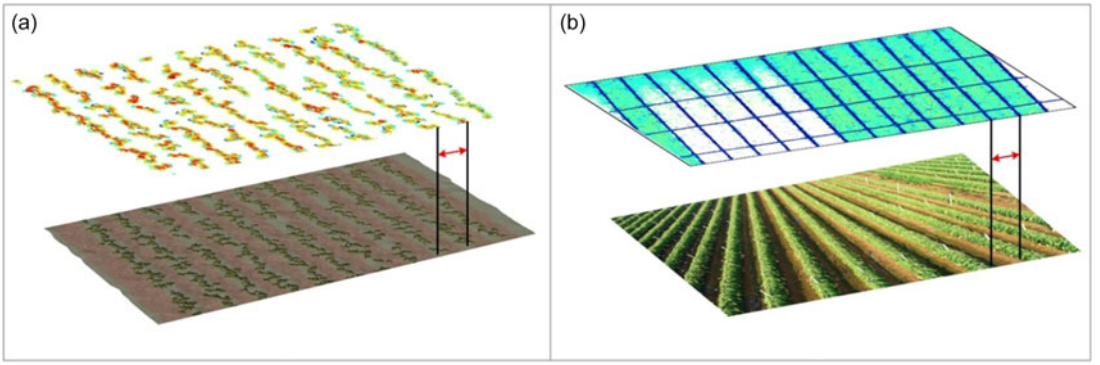


Figure 10. *Row-spacing estimation in agricultural scenes.*

highlighting its significant advantages. As a SLAM method specifically tailored for agricultural environments, HC-LIO outperforms the general-purpose FAST-LIO2 in this scenario through its hierarchical coupling design, which further underscores the necessity of developing targeted algorithms to address the unique challenges of agricultural settings.

Figures 8 and 9 show the mapping and trajectory results for the simulated and real scenes, respectively. The proposed method constructs maps that preserve geometric fidelity and spatial continuity, accurately reproducing agricultural scene structures and providing reliable data for subsequent precision operations. Regarding trajectory results, in the simulation environment, A-LOAM and FAST-LIO2 exhibit multiple minor drifts in structurally repetitive areas. Although HC-LIO shows significantly better trajectory stability than these methods, it still presents slight deviations at certain turning points. In the real-world scenario, the trajectory drift of A-LOAM is considerably more severe. While both FAST-LIO2 and HC-LIO maintain a certain level of trajectory accuracy, their trajectories are noticeably affected by the rugged terrain, exhibiting pronounced "zigzag" fluctuations. HC-LIO shows smaller fluctuation amplitudes, reflecting the advantages of its scenario-specific design. In contrast, the proposed method demonstrates superior stability in both scenarios, with its trajectory closely matching the ground truth, thereby fully validating its strong adaptability in complex agricultural environments.

Further, row spacing is extracted from the generated maps (shown in Figure 10) as an agronomic metric to verify the proposed system's value for near-ground remote sensing supplementation. The estimation results of row spacing are shown in Table IV. The mean row spacing values extracted by the proposed method are 3.731 and 4.247 m, respectively, with estimation errors controlled within 0.071 m, which is the closest to the measured ground truth. In contrast, the estimates obtained by other methods exhibit varying levels of deviation. Figure 11 further shows the distribution statistics of row spacing estimation errors for each method. The proposed method achieves the lowest median error and the narrowest interquartile range, indicating that its estimation results are not only highly accurate overall but also exhibit low fluctuation and excellent stability. HC-LIO shows a lower median error and a narrower distribution range compared to A-LOAM and FAST-LIO2, which is consistent with its superior localization accuracy and confirms the effectiveness of its design tailored for agricultural environments.

https://doi.org/10.1017/S0263574726103397 Published online by Cambridge University Press

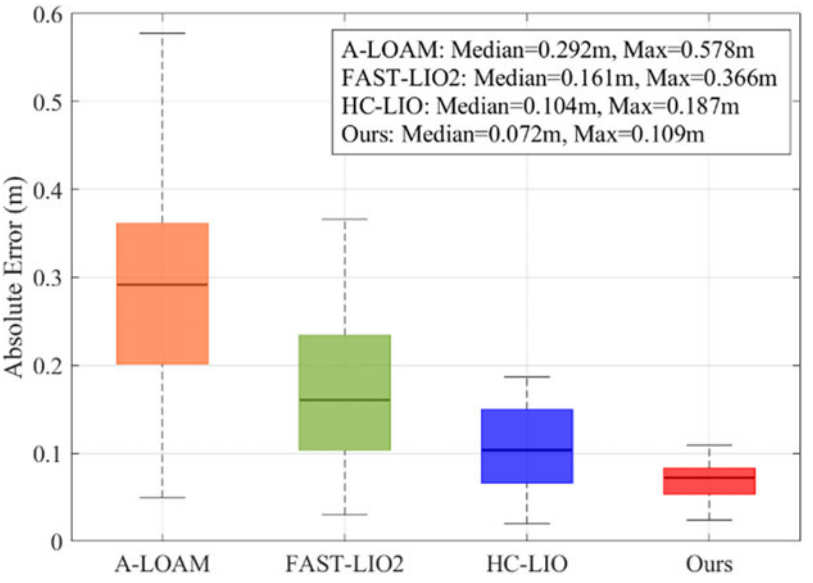


Figure 11. *Statistical distribution of row-spacing estimation error.*

These results demonstrate that general-purpose methods such as A-LOAM and FAST-LIO2 struggle to reconstruct agronomic structures in a stable manner. Although HC-LIO effectively enhances performance in agricultural scenarios, its accuracy and stability still fall short of the proposed method. The proposed system not only provides high-precision localization and mapping capabilities to support precision agriculture operations but also, through stable and accurate map reconstruction, delivers high-resolution, quantifiable supplementary information on the near-ground 3D structure of crop structures, which are often difficult for conventional remote sensing systems to resolve.

5. Conclusion

This paper proposes a robust L-SLAM system that integrates spatial feature modeling with noise-state joint optimization to address the repetitive structures, severe terrain disturbances, and non-stationary sensor noise prevalent in agricultural environments. To improve registration stability in structurally redundant scenes, the system employs a CV-based joint probability model that fuses intensity and geometric information. For pose estimation under dynamic and uncertain conditions, a time-varying recursive noise modeling scheme is introduced within an MAP optimization framework, further enhanced by a dual-filter architecture that jointly refines states and noise parameters. Experimental validation on public datasets demonstrates the generalizability of the proposed method. Comparative experiments in both simulated and real-world agricultural environments further confirm the system's significant advantages in localization accuracy, mapping quality, and the ability to reconstruct agronomic structures.

However, the proposed method has certain limitations. Its superior performance largely depends on the presence of discernible macro-geometric structures in the environment; in extremely sparse and unstructured scenarios, the system's constraint capability may diminish. Moreover, the complex models introduced to achieve high precision and strong adaptability impose higher computational demands. Future work will focus on developing lightweight models and acceleration algorithms to enhance real-time performance on embedded platforms, and to explore enhancement strategies for extremely unstructured environments, thereby advancing the large-scale application of SLAM systems in precision agriculture.

https://doi.org/10.1017/S0263574726103397 Published online by Cambridge University Press

Author contribution. Ziyang Wang: Writing original draft, Validation, Investigation, Methodology. Wenfu Tong: Visualization, Investigation, Supervision. Haibo Zhou: Writing review & editing, Methodology. Mingjun Li: Supervision, Investigation. Xiaozhuo Wang: Writing review & editing. Biyao Cheng: Writing review & editing, Methodology.

Financial support. This work was supported in part by National Natural Science Foundation of China (No. U24B2051), Changsha Major Science and Technology Open Competition Project (kq2503005) and Natural Science Foundation of Hunan Province (No. 2023JJ30694).

Competing interests. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical approval. This article does not involve human or animal participation or data; therefore, ethics approval is not applicable.

Declaration of generative AI and AI-assisted technologies in the writing process. During the preparation of this work, the authors used ChatGPT in order to translate and polish this article. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

[1] L. Liu, F. Yang, X. Liu, Y. Du, X. Li, G. Li, D. Chen, Z. Zhu and Z. Song, "A review of the current status and common key technologies for agricultural field robots," *Comput. Electron. Agric.* **227**, 109630 (2024).
[2] S. Yang, Y. Xing, T. Xing, H. Deng and Z. Xi, "Multisensors fusion SLAM-aided forest plot mapping with backpack dual-LiDAR system," *IEEE J. Sel. Top. Appl.* **17**, 16051–16070 (2024).
[3] S. Shao, G. Han, H. Jia, X. Shi, T. Wang, C. Song and C. Hu, "Multirobot collaborative SLAM based on novel descriptor with LiDAR remote sensing," *IEEE J. Sel. Top. Appl.* **17**, 19317–19327 (2024).
[4] H. Liu, G. Xu, B. Liu, Y. Li, S. Yang, J. Tang, K. Pan and Y. Xing, "A real time LiDAR-Visual-Inertial object level semantic SLAM for forest environments," *ISPRS J. Photogram. Remote Sens.* **219**, 71–90 (2025).
[5] J. Qu, Z. Qiu, L. Li, K. Guo and D. Li, "Map construction and positioning method for LiDAR SLAM-based navigation of an agricultural field inspection robot," *Agronomy* **14**(10), 2365 (2024).
[6] G. Li, L. Yu and S. Fei, "A deep-learning real-time visual SLAM system based on multi-task feature extraction network and self-supervised feature points," *Measurement* **168**, 108403 (2021).
[7] Z. Wang, X. Tang, H. Zhou, L. Xiao and J. a. Duan, "3D LiDAR SLAM for natural guided navigation of automated guided vehicles in unstructured scenes," *Meas. Sci. Technol.* **36**(2), 026317 (2025).
[8] I. Vizzo, T. Guadagnino, B. Mersch, L. Wiesmann, J. Behley and C. Stachniss, "KISS-ICP: In defense of point-to-point ICP – Simple, accurate, and robust registration if done the right way," *IEEE Robot. Autom. Lett.* **8**(2), 1029–1036 (2023).
[9] J. Zhang and S. Singh, "LOAM: Lidar odometry and mapping in real-time," *Robot.: Sci. Syst.* **2**(9), 1–9 (2014).
[10] T. Shan and B. Englot, "Lego-Loam: Lightweight and Ground-Optimized Lidar Odometry and Mapping on Variable Terrain," **In:** *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2018) pp. 4758–4765.
[11] H. Wang, C. Wang, C.-L. Chen and L. Xie, "F-LOAM: Fast LiDAR Odometry and Mapping," **In:** *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2021) pp. 4390–4396.
[12] P. Zhou, X. Guo, X. Pei and C. Chen, "T-LOAM: Truncated least squares LiDAR-only odometry and mapping in real time," *IEEE Trans. Geosci. Remote.* **60**, 1–13 (2022).
[13] H. Wang, C. Wang and L. Xie, "Intensity-SLAM: Intensity assisted localization and mapping for large scale environment," *IEEE Robot. Autom. Lett.* **6**(2), 1715–1721 (2021).
[14] H. Guo, J. Zhu and Y. Chen, "E-LOAM: LiDAR odometry and mapping with expanded local structural information," *IEEE Trans. Intell. Veh.* **8**(2), 1911–1921 (2022).
[15] S. A. Melvin, H. Wang and A. Yazdani, "Global localization for mobile robots in symmetrical indoor environments: A review, practical challenges, and experimental validation," *Robotica* **44**(1), 343–385 (2026).
[16] S. Choi, H.-W. Chae, Y. Jeung, S. Kim, K. Cho and T.-W. Kim, "Fast and versatile feature-based LiDAR odometry via efficient local quadratic surface approximation," *IEEE Robot. Autom. Lett.* **8**(2), 640–647 (2023).
[17] C. Shen, W. Lin, S. Sun, W. Ouyang, B. Shi and L. Sun, "DV-LIO: LiDAR-inertial Odometry based on dynamic merging and smoothing voxel," *Robotica* **43**(4), 1456–1468 (2025).
[18] 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:** *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2020) pp. 5135–5142.
[19] W. Xu, Y. Cai, D. He, J. Lin and F. Zhang, "FAST-LIO2: Fast direct LiDAR-inertial odometry," *IEEE Trans. Robot.* **38**(4), 2053–2073 (2022).
[20] K. Kim, A. Deb and D. J. Cappelleri, "P-AgSLAM: In-row and under-canopy SLAM for agricultural monitoring in cornfields," *IEEE Robot. Autom. Lett.* **9**(6), 4982–4989 (2024).

https://doi.org/10.1017/S0263574726103397 Published online by Cambridge University Press

[21] Y. Hong, R. Ma, C. Li, C. Shao, J. Huang, Y. Zeng and Y. Chen, "Three-dimensional localization and mapping of multiagricultural scenes via hierarchically-coupled LiDAR-inertial odometry," *Comput. Electron. Agric.* **227**, 109487 (2024).
[22] S. Park, S. Wang, H. Lim and U. Kang, "Curved-Voxel Clustering for Accurate Segmentation of 3D LiDAR Point Clouds with Real-Time Performance," **In:** *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2019) pp. 6459–6464.
[23] W. Gao, J. Li, G. Zhou and Q. Li, "Adaptive Kalman filtering with recursive noise estimator for integrated SINS/DVL systems," *J. Navigat.* **68**(1), 142–161 (2014).
[24] Z. Wang, H. Zhou and J. A. Duan, "DALO-SLAM: LiDAR-only SLAM based on dynamic removal and adaptive feature enhancement in GNSS-denial environment," *Meas. Sci. Technol.* **36**(6), 066316 (2025).
[25] Z. Xu, J. Li, S. Wei, L. Tang and H. Chen, "A multi-metric LiDAR Simultaneous Localization and Mapping system based on intensity feature assistance and two-step decoupling estimation," *Robotica*, **43**(8), 1–20 (2025).
[26] X. Zhang, H. Zhang, C. Qian, B. Li and Y. Cao, "A LiDAR-intensity SLAM and loop closure detection method using an intensity cylindrical-projection shape context descriptor," *Int. J. Appl. Earth Obs. Geoinf.* **122**, 103419 (2023).
[27] Y. Xia, X. Lei, J. Pan, L. Chen, Z. Zhang and X. Lyu, "Research on orchard navigation method based on fusion of 3D SLAM and point cloud positioning," *Front. Plant Sci.* **14**, 1207742 (2023).
[28] A. S. Aguiar, F. N. D. Santos, H. Sobreira, J. Boaventura-Cunha and A. J. Sousa, "Localization and mapping on agriculture based on point-feature extraction and semiplanes segmentation from 3D LiDAR data," *Front. Robot. AI* **9**, 832165 (2022).
[29] P. Teng, Y. Zhang, T. Yamane, M. Kogoshi, T. Yoshida, T. Ota and J. Nakagawa, "Accuracy evaluation and branch detection method of 3D modeling using backpack 3D lidar SLAM and UAV-SfM for peach trees during the pruning period in winter," *Remote Sens.-BASEL* **15**(2), 408 (2023).
[30] Y. Xiong, X. Zhang, W. Gao, Y. Wang, J. Liu, Q. Qu, S. Guo, Y. Shen and J. Li, "GF-SLAM: A novel hybrid localization method incorporating global and arc features," *IEEE Trans. Autom. Sci. Eng.* **22**, 6653–6663 (2025).
[31] Z. Wang, X. Liu, L. Yang and F. Gao, "SW-LIO: A sliding window based tightly coupled liDAR-inertial odometry," *IEEE Robot. Autom. Lett.* **8**(10), 6675–6682 (2023).
[32] R. Li, X. Zhang, S. Zhang, J. Yuan, H. Liu and S. Wu, "BA-LIOM: Tightly coupled laser-inertial odometry and mapping with bundle adjustment," *Robotica* **42**(3), 684–700 (2024).
[33] D. Rapado-Rincon and G. Kootstra, "Tree-SLAM: Semantic object SLAM for efficient mapping of individual trees in orchards," *Smart Agric. Technol.* **12**, 101439 (2025).
[34] H. G. Kim, H. M. Lee and S. H. Lee, "A new covariance intersection based integrated SLAM framework for 3D outdoor agricultural applications," *Electron Lett* **60**(9), e13206 (2024).
[35] L. Gong, B. Gao, Y. Sun, W. Zhang, G. Lin, Z. Zhang, Y. Li and C. Liu, "preciseSLAM: Robust, real-time, LiDAR–Inertial–Ultrasonic tightly-coupled SLAM with ultraprecise positioning for plant factories," *IEEE Trans. Ind. Inform.* **20**(6), 8818–8827 (2024).
[36] Y. Wu, T. Guadagnino, L. Wiesmann, L. Klingbeil, C. Stachniss and H. Kuhlmann, "LIO-EKF: High Frequency LiDAR-Inertial Odometry using Extended Kalman Filters," **In:** *2024 IEEE International Conference on Robotics and Automation (ICRA)* (2024) pp. 13741–13747.
[37] Z. Huang, N. Yang, R. Cao, Z. Li, Y. He and X. Feng, "Autonomous navigation system in various greenhouse scenarios based on improved FAST-LIO2," *Comput. Electron. Agric.* **234**, 110279 (2025).
[38] Y. Yu, Z. Li, B. Dai, J. Pan and L. Xu, "High-precision mapping and real-time localization for agricultural machinery sheds and farm access roads environments," *Agriculture-LONDON* **15**(21), 2248 (2025).
[39] Q. H. Hoang and G.-W. Kim, "IMU augment tightly coupled lidar-visual-inertial odometry for agricultural environments," *IEEE Robot. Autom. Lett.* **9**(10), 8483–8490 (2024).
[40] D. Chang, R. Zhang, S. Huang, M. Hu, R. Ding and X. Qin, "WiCRF: Weighted bimodal constrained liDAR odometry and mapping with robust features," *IEEE Robot. Autom. Lett.* **8**(3), 1423–1430 (2023).
[41] G. P. C. Junior, A. M. C. Rezende, V. R. F. Miranda, R. Fernandes, H. Azpurua, A. A. Neto, G. Pessin and G. M. Freitas, "EKF-LOAM: An adaptive fusion of liDAR SLAM with wheel odometry and inertial data for confined spaces with few geometric features," *IEEE Trans. Autom. Sci. Eng.* **19**(3), 1458–1471 (2022).
[42] A. Geiger, P. Lenz and R. Urtasun, "Are We Ready for Autonomous Driving? The Kitti Vision Benchmark Suite," **In:** *2012 IEEE Conference on Computer Vision and Pattern Recognition* (2012) pp. 3354–3361.
[43] N. Carlevaris-Bianco, A. K. Ushani and R. M. Eustice, "University of Michigan North Campus long-term vision and lidar dataset," *Int. J. Robot. Res.* **35**(9), 1023–1035 (2016).
[44] J. Sturm, N. Engelhard, F. Endres, W. Burgard and D. Cremers, "A Benchmark for the Evaluation of RGB-D SLAM Systems," **In:** *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems* (2012) pp. 573–580.

Cite this article: Z. Wang, W. Tong, H. Zhou, M. Li, X. Wang and B. Cheng (2026). "A robust simultaneous localization and mapping method for agricultural environments: integrating spatial modeling and noisy estimation", Robotica **44**, 846–864. https://doi.org/10.1017/S0263574726103397

https://doi.org/10.1017/S0263574726103397 Published online by Cambridge University Press