Journal of Intelligent and Connected Vehicles
2024, **7**(2): 97–107
https://doi.org/10.26599/JICV.2023.9210027

Research Article

Localization and mapping algorithm based on Lidar-IMU-Camera fusion

Yibing Zhao[1,✉], Yuhe Liang[1], Zhenqiang Ma[1], Lie Guo[1], Hexin Zhang[2]

[1]*School of Mechanical Engineering, Dalian University of Technology, Dalian 116024, China*
[2]*Department of the Built Environment, Technology University of Eindhoven, Eindhoven 5600MB, the Netherlands*

Received: June 16, 2023; Revised: August 18, 2023; Accepted: December 18, 2023

ABSTRACT: Positioning and mapping technology is a difficult and hot topic in autonomous driving environment sensing systems. In a complex traffic environment, the signal of the Global Navigation Satellite System (GNSS) will be blocked, leading to inaccurate vehicle positioning. To ensure the security of automatic electric campus vehicles, this study is based on the Lightweight and Ground-Optimized Lidar Odometry and Mapping on Variable Terrain (LEGO-LOAM) algorithm with a monocular vision system added. An algorithm framework based on Lidar-IMU-Camera (Lidar means light detection and ranging) fusion was proposed. A lightweight monocular vision odometer model was used, and the LEGO-LOAM system was employed to initialize monocular vision. The visual odometer information was taken as the initial value of the laser odometer. At the back-end opti9mization phase error state, the Kalman filtering fusion algorithm was employed to fuse the visual odometer and LEGO-LOAM system for positioning. The visual word bag model was applied to perform loopback detection. Taking the test results into account, the laser radar loopback detection was further optimized, reducing the accumulated positioning error. The real car experiment results showed that our algorithm could improve the mapping quality and positioning accuracy in the campus environment. The Lidar-IMU-Camera algorithm framework was verified on the Hong Kong city dataset UrbanNav. Compared with the LEGO-LOAM algorithm, the results show that the proposed algorithm can effectively reduce map drift, improve map resolution, and output more accurate driving trajectory information.

KEYWORDS: Lightweight and Ground-Optimized Lidar Odometry and Mapping on Variable Terrain (LEGO-LOAM), monocular vision system, error state Kalman filter, odometer

1 Introduction

In complex traffic environments, GNSS signals can be blocked, resulting in inaccurate vehicle positioning. To ensure the safety of unmanned vehicles, simultaneous localization and mapping (SLAM) can use vehicle sensors to determine the vehicle's attitude and surrounding environment. On the basis of the SLAM algorithm framework, this study focuses on the mapping and positioning problem of unmanned vehicles with multisensor fusion. On the basis of the Lightweight and Ground-Optimized Lidar Odometry and Mapping on Variable Terrain (LEGO-LOAM) algorithm, a Lidar-IMU-Camera fusion algorithm framework is proposed.

Lidar-based mapping and localization systems were first applied to robotics research. Lidar acquires the distance between itself and objects in the surrounding environment by emitting and reflecting laser light. A point cloud is generated to describe the spatial geometry of the environment. The Rao-Blackwellization Particle Filter (RBPF) algorithm was developed by Murphy et al. (2001), combining the particle filter technique and the Rao-Blackwellization algorithm. This method created a milestone in the history of mapping and positioning systems based on particle filters. The fast simultaneous localization and mapping (FastSLAM) series of algorithms employed RBPF and Kalman filter algorithms to obtain robot attitude information. Map features were estimated using the Kalman filter algorithm. The map and pose information were updated via back-end optimization. The serious particle degradation problem was solved by the Gaussian distribution hypothesis of reference particles (Montemerlo and Thrun, 2003a; Montemerlo et al., 2003b). Grisetti et al. (2007) proposed the GMapping algorithm by improving the RBPF particle filter algorithm, decomposing the mapping and positioning process into two independent modules, integrating the observation results of odometers into a map, adopting an adaptive resampling strategy to reduce the number of reuses, and ensuring sustainable and reasonable particle diversity. Furthermore, the risk of particle degradation decreased, and the GMapping algorithm was widely used in two-dimensional (2D) raster maps. Kohlbrecher et al. (2011) proposed HectorSLAM. They combined the matching method of inertial measurement units (IMUs) and lidar to ensure that the matching results were more robust. To eliminate gradient ascent, multiresolution maps were adopted to avoid local optimization. Therefore, it could adapt better to a variety of different environments. Zhang and Singh (2014) proposed the Lidar Odometry and Mapping (LOAM) algorithm. In this algorithm, edge features and surface features were proposed to replace whole-frame laser point cloud matching. Coarse matching with high frequency and fine registration with low frequency were proposed. The algorithm without loopback detection was suitable for multithread rotating lidar and solid-state lidar. Similarly, Shan and Englot (2018)

✉ Corresponding author.
E-mail: zhaoyibing@dlut.edu.cn

proposed the LEGO-LOAM algorithm on the basis of the LOAM algorithm, which added a loopback detection module. When the defined time threshold was exceeded, the Euclidean distance was used to assess whether the position and attitude of two frames were within the predetermined threshold. If the answer was yes, loopback was detected; otherwise, no loopback was detected. This greatly reduced the pose error in most environments. The back-end optimization module adopted the GTSAM (Kaess et al., 2012) to obtain the pose information. Cartographer (Hess et al., 2016), an algorithm proposed by Google, realized mapping and positioning on 2D and three-dimensional (3D) Lidar. They innovatively proposed a fast loopback detection method called SPA, which matched the established mini map with the point cloud in real time and finally eliminated the accumulated errors. To satisfy the demand of a variety of mappings, Shan et al. (2020) proposed LIO-SAM, a mapping and positioning method based on tightly coupled inertial lidar. This method integrated multiple relative and absolute measurements by using the factor graph optimization method. It could complete robot pose estimation and high-precision maps. In addition, to improve real-time performance, position and pose information was optimized by marginalizing old laser frames rather than by matching laser frames to the global map. This method was adapted to various large outdoor environments and obtained a better positioning effect and clear 3D point cloud map. Chen et al. (2022) proposed a weighted NDT-LOAM algorithm. They employed weighted cells of the normal distribution transform (NDT) based on the surface properties of point clouds and distance values and then constructed new residual functions with weight values. This method successfully reduced the cumulative error of the lidar odds and mapping in real-time (LOAM) algorithm. Park et al. (2020) considered the intensity information measured by lidar to improve the LOAM algorithm. This method could handle the ambiguity issue generated by the planar structure of buildings more successfully by considering the lidar intensity data in the residual function of feature matching. In addition to point-line matching and point-surface matching in the LOAM algorithm, intensity information matching was added, which effectively helped to increase the accuracy of the LOAM algorithm.

Since self-built mapping and positioning systems were proposed, research on visual-based mapping and positioning systems has become more extensive due to their relatively low price and abundant environmental information (Klein and Murray, 2007). It includes various visual sensors, such as monocular vision, binocular vision, and fisheye cameras. As early as 2007, Andrew et al. (2007) proposed MonoSLAM, a monocular vision-based system for positioning and mapping. The algorithm estimated the real-time camera pose by tracking sparse feature points generated by front-end odometry. The back-end optimization module employed the extended Kalman filter approach. It can enhance the front-end's pose data and create a sparse map from the feature points of key frames. Klein and Murray (2007) adopted the multithreading method for the first time to construct a mapping and positioning system named parallel tracking and mapping (PTAM). This algorithm employs a multithreading method to simultaneously track feature points, estimate position, and pose information and construct maps. In the back-end optimization part, the pose information was estimated for the first time by using the nonlinear optimization method. Taihú et al. (2017) integrated the binocular camera sensor S-PTAM based on the PTAM. According to different image processing methods, visual-based mapping and positioning algorithms consist of two parts: direct approaches and feature point approaches. The direct approach was based on gray information in the image, adopted optical flow tracking for matching, and estimated camera motion information and pixel reprojection by minimizing photometric errors (Khattak, 2017). Engel et al. (2014) proposed large-scale direct monocular SLAM (LSD-SLAM). This method is a typical visual-based map building and positioning system. Assuming constant luminosity, interframe camera motion information could be obtained using the direct method. A semidense environment map was constructed, which provided more abundant environmental information than a sparse map. However, the algorithm adopted the feature point method to match the loopback detection part. Mur-Artal et al. (2015) proposed the ORB-SLAM system, a typical visual odometer model based on feature points. There were fewer newly added feature mapping points; however, they had high quality, matching efficiency, and good robustness. Loopback detection was added to eliminate accumulated errors and improve positioning accuracy. ORB-SLAM2 (Mur-Artal and Tardós, 2017) was an improvement over ORB-SLAM in terms of speeding up the dictionary method calculation. Since it applied a quadtree structure instead of an array to store feature points, the selection speed of feature points improved. Many mini maps have been added to ORB-SLAM3 (Campos et al., 2021). When the tracking thread was lost, ORB-SLAM3 would query and match the previous mini map. If the match was successful, the tracking thread continued to work. If it failed to open a mini map again, there would be no output of any order. Instead, the map was reconstructed with the current pose for positioning. In loopback detection, these mini maps were also used for matching. If coincidence points were found, they were matched. This improved the speed and accuracy of loopback detection. Several vision sensors, including pinhole cameras and fisheye cameras, were supported by ORB-SLAM3. A visual inertial navigation odometer (Xu et al., 2018) was built on a tightly coupled nonlinear optimization system. It integrates the visual feature tracking of the direct method with the preintegral IMU data. The factor graph optimization method was used to compute the camera's attitude and trajectory information and construct a sparse point cloud map simultaneously. Twosections were included in visual inertial odometry. The first part was based on sparse visual feature tracking to acquire depth information. The second part was based on tightly coupled IMU preintegral information and dense visual tracking information to obtain camera trajectory and pose information. The direct method requires a small amount of calculations and has high real-time performance. However, it is difficult to meet the assumption that the grayscale is unchanged. In this case, the feature method could be widely used. The feature descriptor described the visual features more stably and robustly since it is not easily affected by light.

Multisensor information fusion refers to the synchronous optimization of multiple sensor information or data through filtering or optimization algorithms to obtain the required decision and estimation information (Wang et al., 2020). Similarly, in mapping and positioning systems, multisensor information (such as lidar, IMU, and vision information) can complement each other to obtain highly robust, high-precision map and trajectory information in complex traffic environments. Feng et al. (2022) proposed a multimodal key scene searching method for autonomous vehicle trials. Yuan et al. (2022) reported the safety and responsibility of autonomous vehicles by analyzing Bayesian random parameters with an ordered probit model. Dong et al. (2022) proposed the creation and evaluation of image converters for comprehensible autonomous vehicle systems. Ding et al. (2022) proposed enhanced eco-driving strategies for connected electric vehicles based on reinforcement learning. The fusion of multisensor information for mapping and positioning overcomes the shortcomings of a single sensor. Mourikis and Roumeliotis (2007) proposed the multistate constrained Kalman filter (MSCKF) algorithm, which innovatively integrates visual sensors and inertial navigation based on filtering and applies a Kalman filter under multistate constraints. Moreover, this algorithm can derive the geometric constraints of cameras at multiple locations even if the positions of 3D road sign features are not stored in the Kalman filter's state vector. This algorithm integrates visual information and IMU information. Compared with other purely visual mapping and positioning algorithms, it can run normally under the conditions of fast motion and feature degradation with good robustness. The highly accurate MSCKF algorithm was suitable for estimating the pose and posture in most scenarios and was well applied in embedded platforms. Leutenegger et al. (2015) proposed a new algorithm named Open Key-frame-based Visual-Inertial SLAM (OKVIS). To calculate camera posture information, this algorithm developed an optimization technique based on a sliding window. The front-end odometer of this device employed the BRISK algorithm. The preintegral IMU result was used as the front-end odometer's initial value, and the image's features were retrieved and matched using the initial value. The reprojection error of visual information and the residual error of the IMU were taken as optimization variables, and the pose information was optimized by nonlinear optimization iteration. However, the algorithm lacked a loopback detection module. Zhang and Singh (2015) proposed the visual-Lidar odds and mapping (VLOAM) algorithm. By fusing a visual odometer and a laser odometer, good position, pose information, and maps can be output under fast-moving and poor illumination conditions. Qin et al. (2018) proposed an algorithm named Vins-Mono. This approach was based on the integration of an IMU with monocular vision. The Kanade–Lucas–Tomasi (KLT) approach was employed to track the visual features. Zuo et al. (2019) proposed a binocular vision inertial navigation model with efficient multistate constraints. They generated a navigation map by applying a 3D laser point cloud map with boundary errors. The multistate constrained Kalman filter algorithm reforms the drift problem of vision-based state estimation. Wisth et al. (2021) proposed an efficient odometer model integrating vision, Lidar, and IMU sensors. This approach prevents traditional interframe point cloud matching from obtaining only suboptimal pose information. Pose estimation and laser point cloud construction were completed by optimizing the residual sensor information. Wei et al. (2021) proposed a mapping and positioning system integrating vision and Lidar, Direct Visual Lidar Odometry, and Mapping (DV-LOAM) (Wang et al. 2021). It is composed of three modules. The visual odometer was used to track visual features in the first module to obtain camera pose information, and an improved sliding window model was employed to further optimize the pose information. To suppress the effects of dynamic objects, the second module uses a laser odometer to further optimize the position and pose information between visual key frames. The third module combined laser loopback detection with visual loopback detection based on the word bag model to optimize the global pose and eliminate the accumulated pose errors.

In this study, an algorithm framework based on the fusion of Lidar, IMU, and monocular vision sensors is developed. This framework aims to further enhance the mapping and positioning accuracy of the algorithm in complicated situations. To obtain more precise pose information, the error state Kalman filter is employed to tightly connect the data from the LEGO-LOAM odometer and the data from the visual odometer. A lightweight monocular vision odometer model is adopted, and monocular vision is initialized in combination with the LEGO-LOAM system. The visual odometer information is taken as the laser odometer's initial value, and the visual word bag model is used for loopback detection. The initial value is determined by the detection result, and the pose information is further optimized via lidar loopback detection. In this case, the overall cumulative position error could be further reduced. The parallel operation approach is employed. If the monocular vision system or LEGO-LOAM system fails, another subsystem can continue to work to improve the robustness of sensor degradation.

2 Materials and methods

This study proposes an algorithm framework based on Lidar-IMU-camera fusion. In Fig. 1, the algorithm framework is displayed. To reduce the amount of calculation, the algorithm uses parallel operations. The monocular vision system or LEGO-LOAM system is detected. In the case of failure, another subsystem will continue to work, improving the robustness of sensor degradation.

2.1 Feature depth correlation

To better initialize the visual odometry, the Lidar-IMU system is

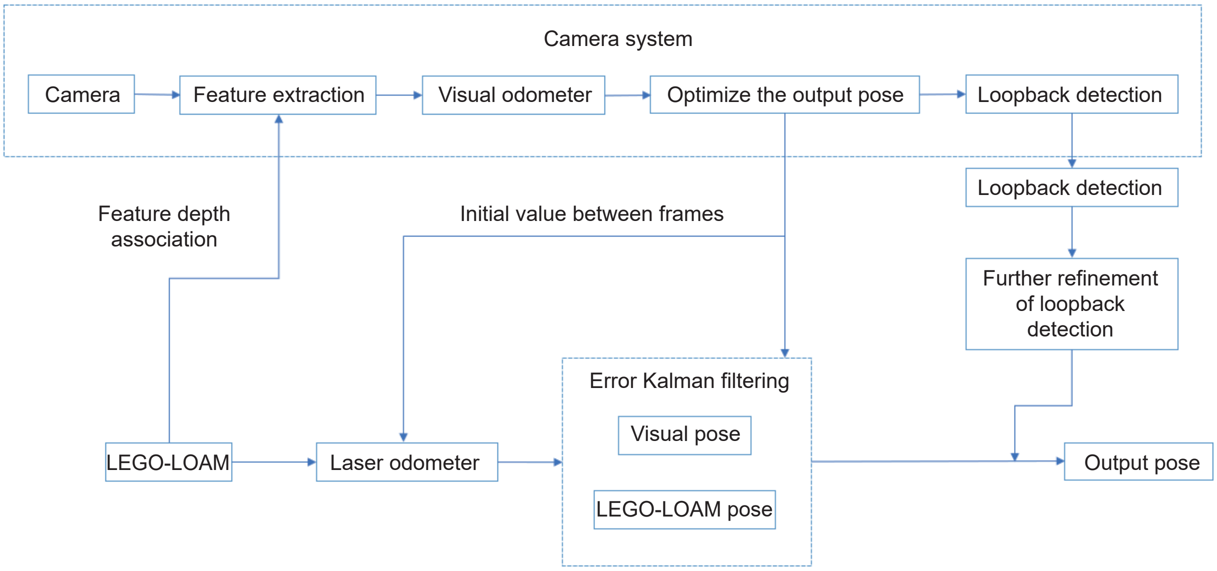

Fig. 1 Algorithm framework based on Lidar-IMU-Camera fusion.

https://doi.org/10.26599/JICV.2023.9210027

employed to estimate the initial state value for monocular visual odometry and align the time stamps between the laser frame and the visual frame through interpolation calculations. Since the laser points are sparse, multiple frames of point clouds can be combined into one dense point cloud map. The visual features and lidar points are projected onto the unit circle whose center coincides on the place where the camera's coordinate system began. The K-D Tree (KDT, K-Dimension Tree) algorithm is applied to find the former three of the nearest three-dimensional laser points to each visual point. Then, the depth of the visual point is derived using the average depth distance of these three laser points. If the average depth exceeds 2, the visual feature association fails. This approach aims to prevent the detection of occlusion points. In Fig. 2, O_c is the origin of the camera coordinate system. The laser points are blue, and the visual points are green.

When the vision system or the Lidar system detects that the number of feature points is lower than the threshold, the system is considered to be invalid. At this point, the pose residual of its failed system must be set to zero until the initialization is successful. The other system with no failure will continue to calculate the current pose information and build a point cloud map based on feature tracking. Fig. 3 shows the image data and laser point cloud data information based on our algorithm. Fig. 3a shows the visual image data, and Fig. 3b shows the laser point cloud data. In Fig. 3a, the green points indicate that they have been associated successfully, while the red points indicate that they failed to be successfully associated.

In this study, the error state Kalman filter fusion algorithm integrates the residual error of the Lidar-IMU system with the reprojection error based on the visual odometer. For the sake of simplified calculation and deduction, the symbols $\boxplus$ and $\boxminus$ represent the projection and the magnitude relationship between matrix vectors, respectively. M represents the $3D$ rotation space, the dimension of M is n, R refers to the rotation matrix and r is the rotation vector. The meanings of the symbols are shown in Eqs. (1) and (2):

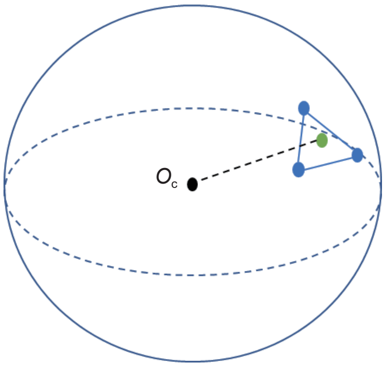

Fig. 2 Feature depth association based on a spherical coordinate system.

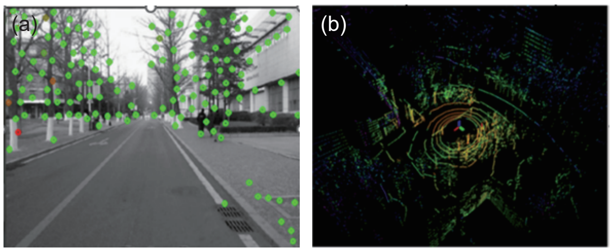

Fig. 3 Frame of image data and laser point cloud data information obtained via our algorithm. (a) Visual image data; (b) laser point cloud data.

$$\boxplus : M \times R^n \to M, \boxminus : M \times M \to R^n \tag{1}$$

$$\begin{cases} \begin{bmatrix} R \\ a_1 \end{bmatrix} \boxplus \begin{bmatrix} r \\ a_2 \end{bmatrix} = \begin{bmatrix} R \cdot \exp(r) \\ a_1 + a_2 \end{bmatrix} \\ \begin{bmatrix} R_1 \\ a_1 \end{bmatrix} \boxplus \begin{bmatrix} R_2 \\ a_2 \end{bmatrix} = \begin{bmatrix} \log\left(R_2^T R_1\right) \\ a_1 + a_2 \end{bmatrix} \end{cases} \tag{2}$$

As shown in Eqs. (1) and (2), $\exp(\cdot)$ and $\log(\cdot)$ represent the transformation relationships between the rotation matrix and rotation vector, respectively.

2.2 Inertial measurement unit

Let x_i be the state of the IMU survey, as shown in Eq. (3):

$$\boldsymbol{x}_i = \begin{bmatrix} {}^{G}R_{I_i}^{T} & {}^{G}\boldsymbol{p}_{I_i}^{T} & {}^{I}R_{C_i}^{T} & {}^{I}\boldsymbol{p}_{C_i}^{T} & {}^{G}\boldsymbol{v}_{i}^{T} & \boldsymbol{b}_{g_i}^{T} & \boldsymbol{b}_{a_i}^{T} \end{bmatrix}^{T} \tag{3}$$

where ${}^{G}R_{I_i}^{T}$ represents the rotation of the IMU to the global coordinate system transformation matrix, ${}^{G}\boldsymbol{p}_{I_i}^{T}$ denotes the IMU position in the global coordinate system, ${}^{I}R_{C_i}^{T}$ is the rotation transformation matrix from the IMU to the camera, ${}^{I}\boldsymbol{p}_{C_i}^{T}$ is the position of the camera in the IMU coordinate system, ${}^{G}\boldsymbol{v}_{i}^{T}$ represents the linear speed of the IMU in the global coordinate system, $\boldsymbol{b}_{g_i}^{T}$ denotes the IMU vertical acceleration deviation, and $\boldsymbol{b}_{a_i}^{T}$ is the acceleration deviation.

The IMU preintegration model is adopted to obtain the error state at $t + 1$, as shown in Eq. (4):

$$x_{t+1} = x_i \boxplus (\Delta t f(x_i, u_i, w_i)) \tag{4}$$

As shown in Eqs. (1)–(4), $\boxplus$ has the same meaning as in Eq. (1), Δt is the time difference between t and $t + 1$, x_i represents the observed data, u_i denotes the input data, and w_i indicates the bias data.

The state estimator is shown in Eq. (5):

$$\check{x}_{i+1} = \check{x}_i \boxplus (\Delta t f(\check{x}_i, u_i, 0)) \tag{5}$$

where $\check{x}_{i+1}$ and $\check{x}_i$ are the estimated data. $\boxplus$ and Δt are as indicated in Eq. (4).

2.3 Lidar measurements

Taking $\check{x}_{k+1}$as the current estimated data of x_{k+1} and converting the j-th feature point to a global coordinate system, the measurement residual is shown in Eq. (6):

$$r_l\left(\check{x}_{k+1}, {}^{L}p_j\right) = u_j^T({}^{G}p_J - q_j) \tag{6}$$

where ${}^{L}p_j$ represents the reprojection points, and ${}^{G}p_j$ denotes the position in the global coordinate system. q_j indicates the centroid, and u_j^T indicates the normal value.

Let n_j be the measurement noise of point ${}^{L}p_j$, and by compensating for the noise from${}^{L}p_j$, the resulting true point position is shown in Eq. (7):

$${}^{L}p_j = {}^{L}p_j^{gt} + n_j, n_j\ \mathcal{N}(0, \sum n_j) \tag{7}$$

where ${}^{L}p_j^{gt}$ is the location of the real point and ${}^{L}p_j$ is the location of the measurement point.

The zero residual caused by the true point position and the real state x_{k+1} is shown in Eq. (8):

$$0 = r_l\left(x_{k+1}, {}^{L}p_j^{gt}\right) = r_l\left(\check{x}_{k+1}, {}^{L}p_j\right) + H_j^l \delta \check{x}_{k+1} + \alpha_j \tag{8}$$

It forms a posterior distribution of $\delta \check{x}_{k+1}$. The parameterization of x_{k+1} in error $\delta \check{x}_{k+1}$ is shown in Eqs. (9) and (10). It is defined byα_j referring to the bias amount. $\boldsymbol{H}_j^l$ is the Jacobian matrix of the state residuals. ${}^{G}\check{R}_{I_{k+1}}$represents the rotation of the IMU to the global coordinate system transformation matrix. ${}^{I}R_L$ denotes the transformation matrix between the IMU and lidar. F_{p_j} and $F_{p_j}^T$ indicate the optimization function and its own transposition, respectively.

$$H_j^l = \frac{\partial r_l\left(\check{x}_{k+1} \boxplus \delta \check{x}_{k+1}, {}^{L}p_j\right)}{\partial \delta \check{x}_{k+1}} |_{\delta \check{x}_{k+1}} = 0 \tag{9}$$

$$\sum\nolimits_{\alpha_j} = F_{p_j} \sum\nolimits_{n_j} F_{p_j}^T F_{p_j} = \left(\frac{\partial r_l\left(\check{x}_{k+1}, {}^{L}p_j\right)}{\partial^L p_j} \right) = {}^{G}\check{R}_{I_{k+1}} {}^{I}R_L \tag{10}$$

2.4 Residual model of the visual odometer

The visual odometer information is integrated, and the PNP reprojection error is adopted as the residual term of the error state Kalman filter fusion. For a certain feature point s, its pixel coordinate is $P_s = [u_s, v_s]^T$, and its corresponding point in 3D space is P_s^G. Then, the residual of this point is shown in Eq. (11):

$$r_s\left(X_{k+1}, P_s, P_s^G\right) = P_s - \pi\left(P_s^G\right) \tag{11}$$

where $\pi(\cdot)$ function in Eq. (11) represents the camera model, that is, $\pi\left(P_s^G\right) = KTP_s^G$. Here, K is the gain factor in Kalman filtering, and T denotes the transformation matrix. Both P_s and P_s^G have certain noise, as shown in Eqs. (12) and (13), respectively:

$$P_s^G = P_s^{Ggt} + np_s, np_s\ N(0, \sum\nolimits_{np_s}) \tag{12}$$

$$P_s = P_s^{gt} + np_s, np_s\ N(0, \sum\nolimits_{np_s}) \tag{13}$$

where $P_s^{Ggt}$and P_s^{gt} are true values and np_s is a term with a Gaussian distribution that represents the noise.

For the residual item for the Taylor expansion in the first order, the error of the real value is 0, where $\beta_s \sim N(0, \sum_{\beta_s})$ and represents the bias amount, as shown in Eqs. (14)–(17):

$$0 = r_s\left(X_{k+1}, P_s, P_s^G\right) \approx r_s\left(\check{X}_{k+1}, P_s, P_s^G\right) + H_s \delta \check{X}_{k+1} + \beta_s \tag{14}$$

$$H_s = \frac{\Phi r_s\left(\check{X}_{k+1} \boxplus \delta \check{X}_{k+1}, P_s, P_s^G\right)}{\Phi \delta \check{X}_{k+1}} |_{\delta \check{X}_{k+1}=0} \tag{15}$$

$$\sum\nolimits_{\beta_s} = \sum\nolimits_{np_s} + F_{P_s} \sum\nolimits_{np_s} F_{P_s}^T \tag{16}$$

$$F_{P_s} = \frac{\Phi r_s\left(\check{X}_{k+1}, P_s, P_s^G\right)}{\Phi P_s^G} \tag{17}$$

where F_{p_s} and $F_{p_s}^T$ indicate the optimization function and its own transposition, respectively.

2.5 Sensor information fusion and the error state Kalman filter update

Combined with prior IMU information, the Lidar odometer and visual odometer are posterior distribution information, and the maximum posterior estimation of the residual term is shown in Eq. (18):

$$\begin{aligned} \min\nolimits_{\delta \check{X}_{k+1}} \Bigg[& \left(\| \check{X}_{k+1} \boxminus \widehat{X_{k+1}} + H \delta \check{X}_{k+1} \|^2_{\sum_{\delta \widehat{X_{k+1}}}} \right) \\ & + \left(\sum\nolimits_{j=1}^{m} \| r_l\left(\check{X}_{k+1}, P_j^L\right) + H_j^L \delta \check{X}_{k+1} \|^2_{\sum_{\alpha_j}} \right) \\ & + \left(\sum\nolimits_{s=1}^{t} \| r_c\left(\check{X}_{k+1}, P_s, P_s^G\right) + H_s \delta \check{X}_{k+1} \|^2_{\sum_{\beta_s}} \right) \Bigg] \end{aligned} \tag{18}$$

As described in Eq. (18), ⊟ has the same meaning as that in Eq. (1), and H is the Kalman filter observation matrix. The IMU motion error is the first term, followed by the laser odometer error and the visual odometer error.

Then, the Kalman gain K can be calculated using Eq. (19):

$$K = \left(H^T R^{-1} H + P^{-1}\right)^{-1} H^T R^{-1} \tag{19}$$

where R denotes the rotation matrix, and P indicates the state estimation covariance matrix. The status estimation is updated as shown in Eq. (20):

$$\check{X}_{k+1} = \check{X}_{k+1} \boxplus \left(-KZ_{k+1} - (I - KH)(\check{H})^{-1} \left(\check{X}_{k+1} \boxminus \widehat{X_{k+1}} \right) \right) \tag{20}$$

Then, I continues iterating, and let $\widehat{X_{k+1}} = \check{X}_{k+1}, \sum^{\wedge}_{\sum_{\delta \widehat{X_{k+1}}}} = (I - KH) \sum^{\wedge}_{\sum_{\delta X_{k+1}}}$. The symbols ⊞ and ⊟ are as indicated in Eq. (1). They explain the projection and the magnitude relationship between matrix vectors.

3 Comparative test and analysis of results

The computer platform used in this study was Lenovo Xiaoxin V3000, 2015 edition, with an I7–4800U CPU and an integrated graphics card. The algorithm in this study was tested in a real vehicle. The distribution of equipment and sensors is shown in Fig. 4. Velodyne_16–line Lidar, USB_IMU, and KONKA cameras were used in the experiment. Equation (21) shows the external parameters between the IMU and camera.

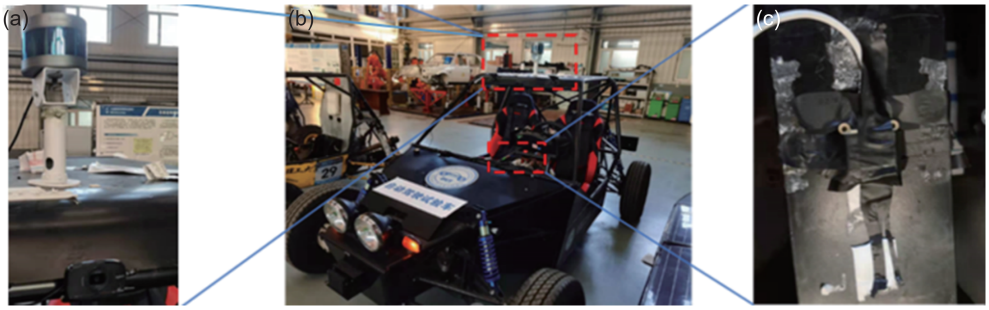

Fig. 4 Unmanned vehicles and sensor equipment.

$$\begin{bmatrix} 999.34608718980233 & -1.5715484428488590 & -3.2564114721728155 \\ 3.2359037356803094 & -13.131917124154624 & 91.139003669937865 \\ -1.6133527815482926 & -92.179026615676858 & -12.614792047622947 \end{bmatrix} \tag{21}$$

When there is no GNSS, GPS or other high-precision positioning sensors available to obtain real path information, this study determines the odometer error by determining whether the vehicle returns to sthe origin. It should be performed without loopback detection. Fig. 5 shows a comparison of the closed-loop effect between the LEGO-LOAM system and the Lidar-IMU-Camera system. Fig. 5a shows the trajectory and closed-loop effect based on the Lidar-IMU-Camera system, Fig. 5b shows the trajectory and closed-loop effect based on the LEGO-LOAM system, and the three-axis coordinate system is the endpoint coordinate system. It can be seen from the details of Figs. 5a and 5b that the closed-loop effect of the Lidar-IMU-Camera system is slightly better than that of the latter.

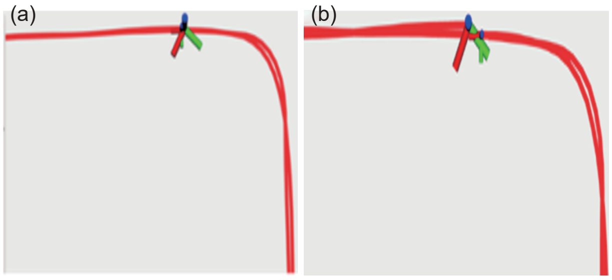

Fig. 5 Trajectory and closed-loop effect based on the LEGO-LOAM system and Lidar-IMU-Camera system. (a) Trajectory and closed-loop effect based on Lidar-IMU-Camera system; (b) trajectory and closed-loop effect based on LEGO-LOAM system.

Fig. 6 is the trajectory matching effect based on the LEGO-LOAM system and the satellite map, where Fig. 6a is the trajectory information based on the LEGO-LOAM system, and Fig. 6b is the trajectory matching result related to the satellite map. It can be seen from the dotted box that the trajectory based on the LEGO-LOAM system is not satisfactory n the road center, and there is a slight deviation. Fig. 7 shows the trajectory matching effect based on the Lidar-IMU-camera system and the satellite map, Fig. 7a shows the trajectory information from the Lidar-IMU-camera system, and Fig. 7b shows the trajectory matching result related to the satellite map. It can be seen in the white dotted box that the trajectory based on the Lidar-IMU-Camera system can well coincide with the real road, and it is in the center of the road without excessive deviation. Therefore, the Lidar-IMU-Camera system can greatly improve the pose accuracy.

This study collected data from the Liu Changchun gymnasium of Dalian University of Technology. Fig. 8 shows the 3D map of the Lidar-IMU Camera system for the gymnasium, and Fig. 9

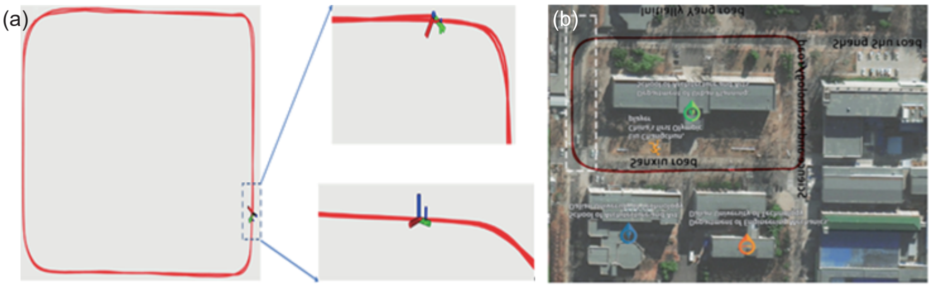

Fig. 6 Trajectory matching effect based on the LEGO-LOAM system and satellite map. (a) Trajectory information based on the LEGO-LOAM system; (b) trajectory matching result related to the satellite map.

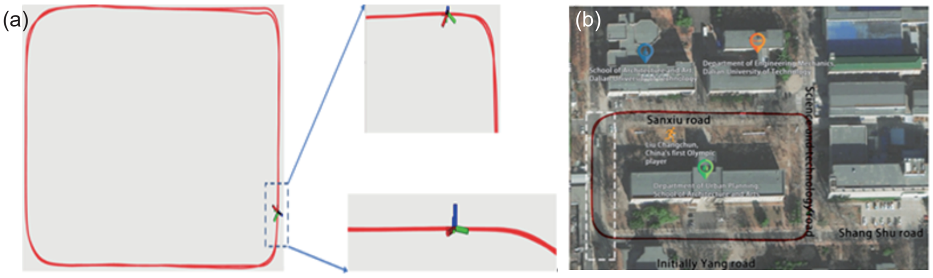
(a)
(b)
School of Architecture and Art Dalian University of Technology
Department of Engineering Mechanics, Dalian University of Technology
Sanxiu road
Liu Changchun, China's first Olympic player
Department of Urban Planning School of Architecture and Arts
Science and technology road
Shang Shu road
Initially Yang road

Fig. 7 Trajectory matching effect based on the Lidar-IMU-Camera system and satellite map. (a) Trajectory information from the Lidar-IMU-Camera system; (b) trajectory matching result related to the satellite map.

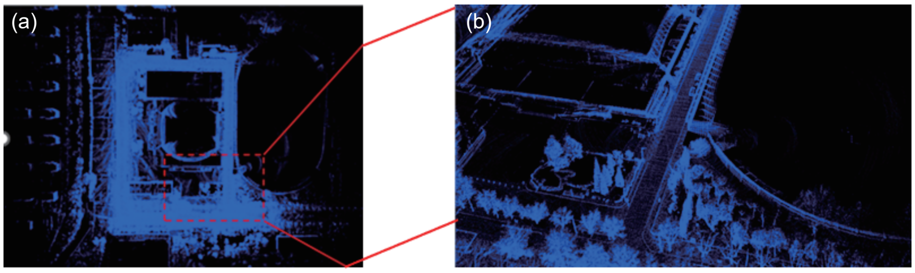

Fig. 8 3D point cloud map of DLUT gymnasium based on the Lidar-IMU Camera system.

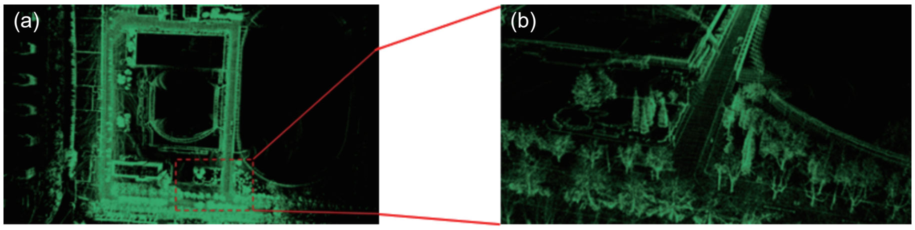

Fig. 9 3D point cloud map of DLUT gymnasium based on the Lidar-IMU system.

shows the 3D map of the LEGO-LOAM system. The red dotted boxes in Figs. 8 and 9 are the detailed displays. Comparing the overall map and the details of these two systems, it can be seen that the overall mapping effect is slightly different. However, the Lidar-IMU-Camera system has more details and is clearer and less noisy than the LEGO-LOAM system.

Fig. 10 shows the trajectory and closed-loop effect of the DLUT gymnasium based on the Lidar-IMU-Camera system. Fig. 11 shows the trajectory and closed-loop effect of the DLUT gymnasium based on the LEGO-LOAM system. The blue dotted frame in Fig. 10 shows the closed-loop details based on the Lidar-IMU-Camera system. Here, the starting point coordinates are in the large three-axis coordinate system, while the ending point coordinates are in the tiny three-axis coordinate system. Similarly, the blue dotted frame in Fig. 11 shows the closed-loop details based on the LEGO-LOAM system. The large three-axis coordinate system represents the starting point coordinates, and the small three-axis coordinate system represents the endpoint coordinates. It is also determined whether the closed-loop effect is satisfactory by comparing the distances between the two three-axis coordinate systems, as shown in the blue dotted boxes in Figs. 10 and 11. The starting point and the end point of Fig. 10 almost coincide, and the starting point and the end point of Fig. 11 have large distance differences. In summary, by comparing the mapping and trajectory information based on the two systems, the Lidar-IMU-Camera system outperforms the LEGO-LOAM system.

Fig. 12 denotes satellite map of the Dalian University of Technology gymnasium. Fig. 13 shows the matching effect between the LEGO-LOAM system map and the real satellite map. Fig. 13a is the voxelized 3D point cloud map of the DLUT gymnasium based on the LEGO-LOAM system. Fig. 13b is the matching result of the voxelized map and the satellite map. The white dashed box in Fig. 13b shows that there is a deviation between the map and the real building, and there is no good overlap. This proves that there is a certain drift in the LEGO-LOAM system map.

Fig. 14 shows the map matching effect based on the Lidar-IMU-Camera system and the real satellite map. Fig. 14a is the 3D point cloud map of the Dalian University of Technology gymnasium based on the Lidar-IMU-Camera system. Fig. 14b shows the matching effect between the voxelized map and the satellite map. It can be seen in the white dashed box in Fig. 14b that the map and the real building can overlap well, and compared with the results of the LEGO-LOAM system, the deviation is small. Therefore, it is further proven that the Lidar-IMU-Camera system greatly improves the effect of mapping.

Fig. 15 shows the trajectory matching effect based on the LEGO-LOAM system and the satellite map, and Fig. 15a shows the trajectory information based on the LEGO-LOAM system for the Dalian University of Technology gymnasium. Fig. 15b shows the trajectory matching result related to the satellite map. Compared to the real road, the trajectory based on the LEGO-LOAM system obviously deviates from the road center. Fig. 16 shows the

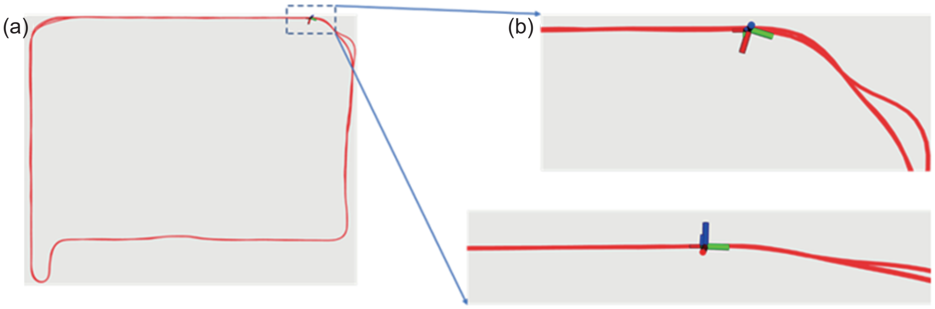

Fig. 10 Trajectory information and closed-loop effect of DLUT gymnasium based on the Lidar-IMU-Camera system.

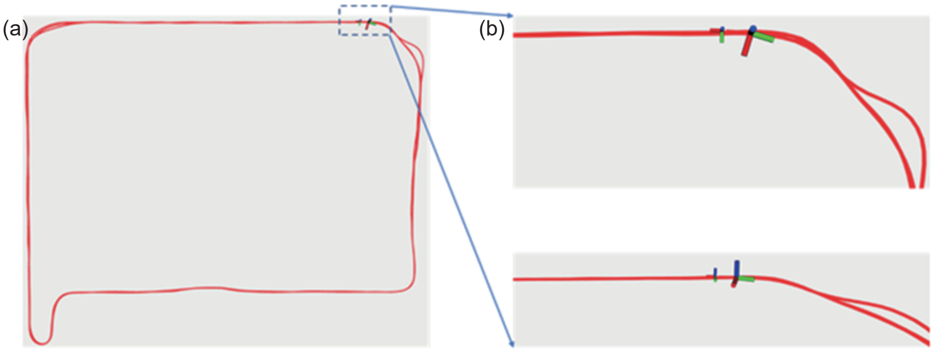

Fig. 11 Trajectory information and closed-loop effect of DLUT gymnasium based on the LEGO-LOAM system.

https://doi.org/10.26599/JICV.2023.9210027

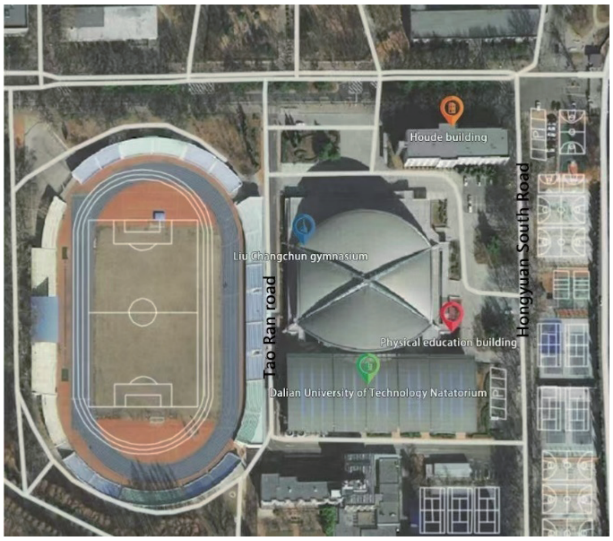
Houde building
Liu Changchun gymnasium
Hongyuan South Road
Tao Ran road
Physical education building
Dalian University of Technology Natatorium

Fig. 12 Satellite map of the Dalian University of Technology gymnasium.

trajectory of the Lidar-IMU-Camera system and the trajectory of the satellite map. Fig. 16a shows the trajectory information based on the Lidar-IMU-Camera system for the Dalian University of Technology Gymnasium. Fig. 16b shows the matching results with satellite maps. It can also be seen in the white dotted box in Fig. 16b that the trajectory based on the Lidar-IMU-Camera system coincides well with the real road. It is in the center of the road without much deviation. Therefore, the Lidar-IMU-Camera system can greatly improve the pose accuracy. Fig. 17 shows the trajectory comparison between the Lidar-IMU-Camera and the LEGO-LOAM systems. The blue line is the trajectory of the Lidar-IMU-Camera algorithm, and the red line is the LEGO-LOAM algorithm trajectory. It can be observed that the trajectory based on our algorithm in this paper is significantly closer to the road center.

In this study, the Hong Kong city dataset UrbanNav is used to verify the LEGO-LOAM system and the Lidar-IMU-Camera fusion algorithm. The EVO software is used to compare the trajectory obtained by our algorithm with the real trajectory. Fig. 18 shows the error comparison information from the LEGO-LOAM system and the Lidar-IMU-Camera system. Fig. 18a shows the error information from the LEGO-LOAM system. The maximum error of the LEGO-LOAM system is 1.51 m, and the minimum error is 0.129 m. Fig. 18b shows the error information from the Lidar-IMU-Camera system. The maximum error is 0.213 m, and the minimum error is 0.002 m. In summary, by analyzing and comparing the error information from these two algorithms, the Lidar-IMU-Camera fusion algorithm reduces the maximum error from 1.51 to 0.213 m and the minimum error from 0.129 to 0.002 m. Therefore, the fusion algorithm based on Lidar-IMU-Camera greatly improves the trajectory accuracy of the LEGO-LOAM system.

In terms of the mapping effect, as shown in Fig. 19, the red frame shows the construction details based on the LEGO-LOAM system, which is related to the largest trajectory error. When the unmanned vehicle passed through the road once more, the point cloud maps did not coincide well, and obvious drift occurred.

The point cloud map constructed based on the Lidar-IMU-Camera fusion algorithm is shown in Fig. 20. Compared with Fig. 19, the map is much clearer. It has fewer noise points and drift points. The red frame in Fig. 20 shows the details related to the

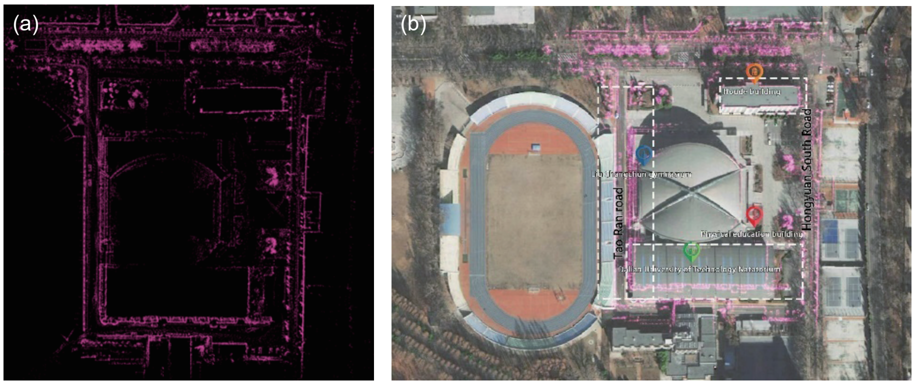

Fig. 13 Matching effect between the LEGO-LOAM system map and the real satellite map. (a) Voxelized 3D point cloud map from the LEGO-LOAM system; (b) matching result of the voxelized map and the satellite map.

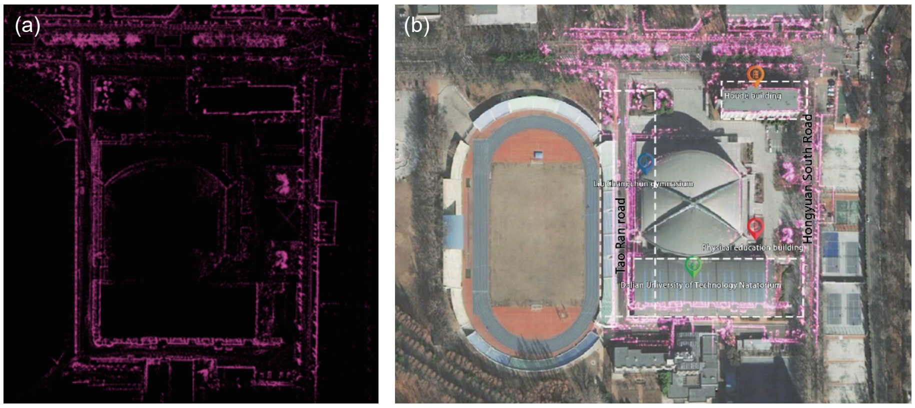

Fig. 14 Matching effect based on the Lidar-IMU-Camera system and the real satellite map. (a) 3D point cloud map based on the Lidar-IMU-Camera system; (b) matching effect between the voxelized map and the satellite map.

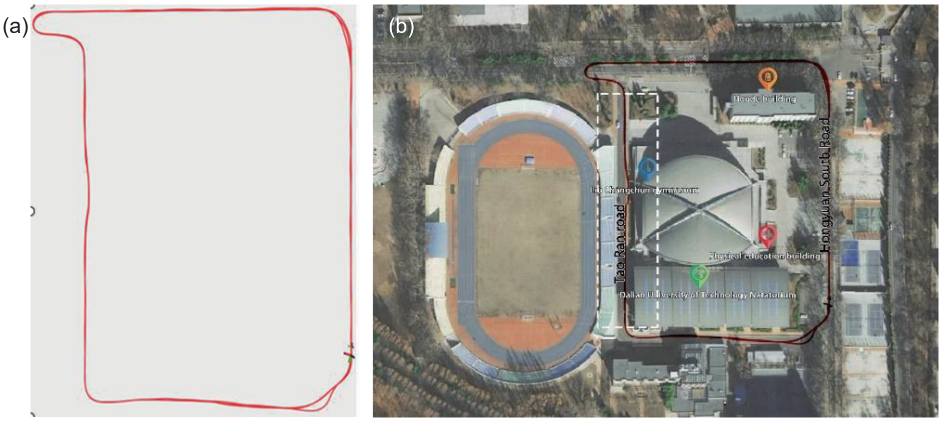

Fig. 15 Matching effect of trajectory based on the Lidar-IMU system and satellite map. (a) Trajectory information based on the LEGO-LOAM system; (b) trajectory matching result related to the satellite map.

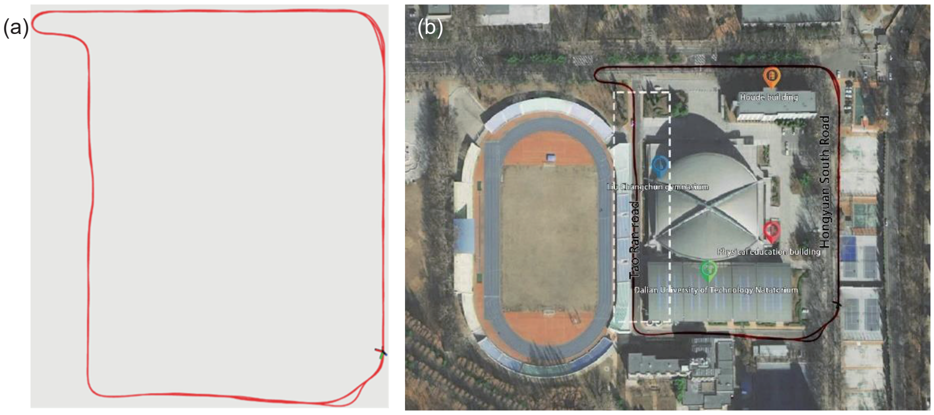

Fig. 16 Matching effect of trajectory based on the Lidar-IMU-Camera system and satellite map. (a) Trajectory information based on the Lidar-IMU-Camera system; (b) trajectory matching result related to the satellite map.

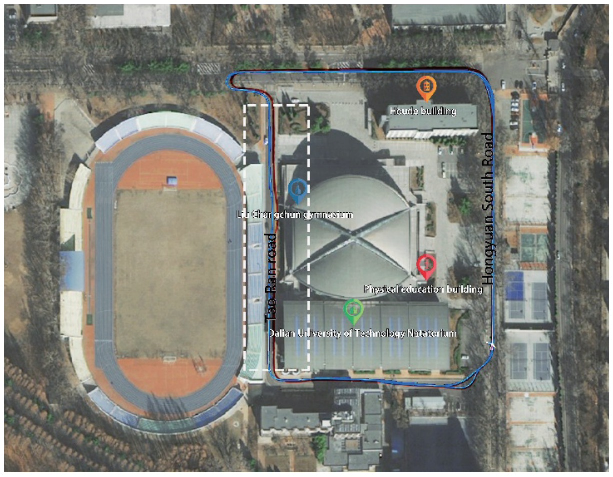
Houde building
Liu Changchun gymnasium
Tiao-Ran road
Hongyuan South Road
Physical education building
Dalian University of Technology Natatorium

Fig. 17 Trajectory comparison between the Lidar-IMU-Camera and the LEGO-LOAM systems.

corresponding part of Fig. 19. When the unmanned vehicle returns to this section, the map coincides well and does not produce excessive drift. Therefore, the point cloud map of the Lidar-IMU-Camera system is much clearer and more accurate than that of the LEGO-LOAM system.

4 Conclusions

This study proposes an algorithm based on Lidar-IMU-Camera fusion, and more accurate positions are obtained by using the error state Kalman filter. Furthermore, the visual word bag model is employed, and the initial value is the detection outcome. The pose information is further optimized by Lidar closed-loop detection. The algorithm is verified with the LEGO-LOAM system in an open-source dataset and a real vehicle. The demonstration results are satisfactory. In the open-source UrbanNav dataset and the Hong Kong city dataset, our algorithm offers a much higher mapping accuracy than does the LEGO-LOAM system. The performance of the algorithm is demonstrated to be superior to that of current traditional open-source algorithms. In the real vehicle test, the mapping and trajectory information based on the Lidar-IMU-Camera system and the LEGO-LOAM system are compared in two scenarios. It can be further concluded that the effect of the Lidar-IMU-Camera system is better than that of the LEGO-LOAM system.

Replication and data sharing

The program code within this research can be made accessible upon request via email to the corresponding author.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant Nos. 51975088 and 51975089).

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

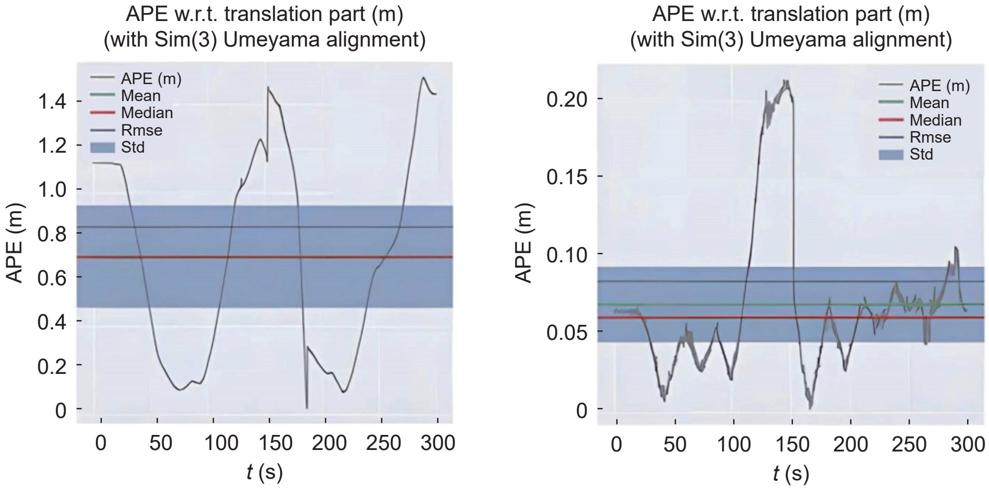

Fig. 18 Comparison of error information between the LEGO-LOAM system and Lidar-IMU-Camera system. (a) Error information from the LEGO-LOAM system; (b) error information from the Lidar-IMU-Camera system.

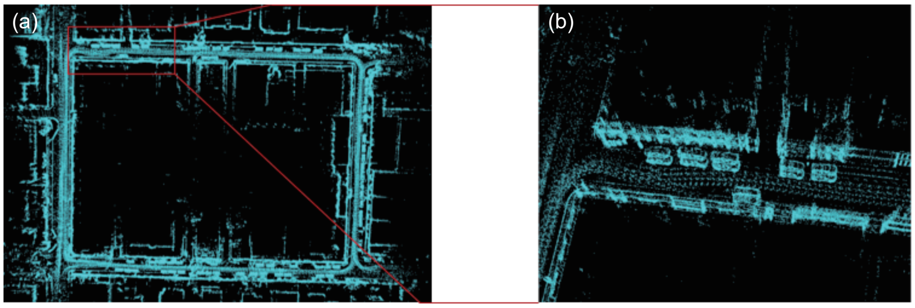

Fig. 19 Mapping effect based on the LEGO-LOAM system in the Hong Kong urban dataset UrbanNav.

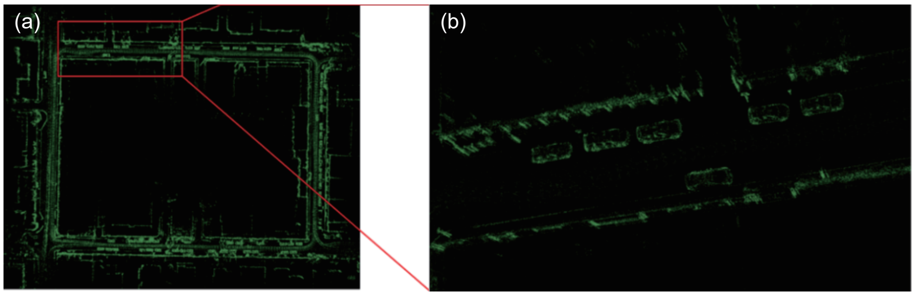

Fig. 20 Mapping effect based on the Lidar-IMU-Camera system in the Hong Kong Urban dataset UrbanNav.

References

Andrew J., Ian D., Nicholas D., Olivier S., 2007. MonoSLAM: Real-Time Single Camera SLAM. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, 1052–1067.

Campos, C., Elvira, R., Rodríguez, J. J. G., Montiel, J. M. M., Tardós, J. D., 2021. ORB-SLAM3: An accurate open-source library for visual, visual-inertial, and multimap SLAM. IEEE Trans Robot, 37, 1874–1890.

Chen, S., Ma, H., Jiang, C., Zhou, B., Xue, W., Xiao, Z., et al., 2022. NDT-LOAM: Real-time lidar odometry and mapping with weighted NDT and an LFA. IEEE Sens J, 22, 3660–3671.

Ding, H., Li, W., Xu, N., Zhang, J., 2022. An enhanced eco-driving strategy based on reinforcement learning for connected electric vehicles: Cooperative velocity and lane-changing control. J Intell Connect Veh, 5, 316–332.

Dong, J., Chen, S., Miralinaghi, M., Chen, T., Labi, S., 2022. Development and testing of an image transformer for explainable autonomous driving systems. J Intell Connect Veh, 5, 235–249.

Engel, J., Schöps, T., Cremers, D., 2014. LSD-SLAM: Large-scale direct monocular SLAM. In: European Conference on Computer Vision, 834–849.

Feng, T., Liu, L., Xing, X., Chen, J., 2022. Multimodal critical-scenarios search method for test of autonomous vehicles. J Intell Connect Veh, 5, 167–176.

Grisetti, G., Stachniss, C., Burgard, W., 2007. Improved techniques for grid mapping with rao-blackwellized particle filters. IEEE Trans Robot, 23, 34–46.

Hess, W., Kohler, D., Rapp, H., Andor, D., 2016. Real-time loop closure in 2D LIDAR SLAM. In: 2016 IEEE International Conference on Robotics and Automation (ICRA), 1271–1278.

Kaess, M., Johannsson, H., Roberts, R., Ila, V., Leonard, J. J., Dellaert, F., 2012. iSAM2: Incremental smoothing and mapping using the bayes tree. Int J Robot Res, 31, 216–235.

Khattak, S., 2017. Multi-modal landmark detection and tracking for odometry estimation in degraded visual environments. M.D. Dissertation. NV, USA: University of Nevada.

Klein, G., Murray, D., 2007. Parallel tracking and mapping for small AR workspaces. In: 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, 225–234.

Kohlbrecher, S., von Stryk, O., Meyer, J., Klingauf, U., 2011. A flexible and scalable SLAM system with full 3D motion estimation. In: IEEE International Symposium on Safety, Security, and Rescue Robotics, 155–160.

Leutenegger, S., Lynen, S., Bosse, M., Siegwart, R., Furgale, P., 2015. Keyframe-based visual-inertial odometry using nonlinear optimization. Int J Robot Res, 34, 314–334.

Montemerlo, M., Thrun, S., 2003a. Simultaneous localization and mapping with unknown data association using FastSLAM. In: 2003 IEEE International Conference on Robotics and Automation, 1985–1991.

Montemerlo, M., Thrun, S., Koller, D., Wegbreit, B., 2003b. FastSLAM 2. 0: An improved particle filtering algorithm for simultaneous localization and mapping that provably converges. In: Proceedings of the 18th international joint conference on Artificial intelligence, 1151–1156.

Mourikis, A. I., Roumeliotis, S. I., 2007. A multi-state constraint Kalman filter for vision-aided inertial navigation. In: Proceedings 2007 IEEE International Conference on Robotics and Automation, 3565–3572.

Mur-Artal, R., Montiel, J. M. M., Tardós, J. D., 2015. ORB-SLAM: A versatile and accurate monocular SLAM system. IEEE Trans Robot, 31, 1147–1163.

Mur-Artal, R., Tardós, J. D., 2017. ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras. IEEE Trans. Robot, 33, 1255–1262.

Murphy, K., Russell, S., 2001. Rao-blackwellised particle filtering for dynamic Bayesian networks. In: Sequential Monte Carlo Methods in Practice, 499–515.

Park, Y. S., Jang, H., Kim, A., 2020. I-LOAM: Intensity enhanced LiDAR odometry and mapping. In: 2020 17th International Conference on Ubiquitous Robots (UR), 455–458.

Taihú, P., Thomas, F., Gastón, C., Pablo De, C., Javier, C., Julio Jacobo, B., 2017. S-PTAM: Stereo parallel tracking and mapping. Robot Auton Syst, 93, 27–42.

Qin, T., Li, P., Shen, S., 2018. VINS-mono: A robust and versatile monocular visual-inertial state estimator. IEEE Trans Robot, 34, 1004–1020.

Shan, T., Englot, B., 2018. 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), 4758–4765.

Shan, T., Englot, B., Meyers, D., Wang, W., Ratti, C., Rus, D., 2020. Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 5135–5142.

Wang, M., He, L., Yu, L., Chao, S., 2020. Mobile robot localization algorithm based on multi-sensor information fusion. J Meas Sci Instrum, 11, 152–160.

Wei, W., Jun, L., Chenjie, W., Bin, L., Cheng, Z., 2021. DV-LOAM: Direct visual LiDAR odometry and mapping. Remote Sens, 13, 3340.

Wisth, D., Camurri, M., Das, S., Fallon, M., 2021. Unified multi-modal landmark tracking for tightly coupled lidar-visual-inertial odometry. IEEE Robot Autom Lett, 6, 1004–1011.

Xu, W., Choi, D., Wang, G., 2018. Direct visual-inertial odometry with semi-dense mapping. Comput Electr Eng, 67, 761–775.

Yuan, Q., Xu, X., Wang, T., Chen, Y., 2022. Investigating safety and liability of autonomous vehicles: Bayesian random parameter ordered probit model analysis. J Intell Connect Veh, 5, 199–205.

Zhang, J., Singh, S., 2014. LOAM: lidar odometry and mapping in real-time. In: Robotics: Science and Systems Conference, 1–9.

Zhang, J., Singh, S., 2015. Visual-lidar odometry and mapping: Low-drift, robust, and fast. In: 2015 IEEE International Conference on Robotics and Automation (ICRA), 2174–2181.

Zuo, X., Geneva, P., Yang, Y., Ye, W., Liu, Y., Huang, G., 2019. Visual-inertial localization with prior LiDAR map constraints. IEEE Robot Autom Lett, 4, 3394–3401.

Yibing Zhao received the B.S. and M.S. degrees in agriculture bioenvironmental and energy engineering from Jilin Agricultural University, China, in 2002 and 2005, respectively, and the Ph.D. degree in vehicle application engineering from Jilin University, China, in 2008. She is currently an Associate Professor in the School of Mechanical Engineering, Dalian University of Technology, Dalian. Her research interests include field obstacle detection and tracking and collision avoidance control of unmanned vehicles, intelligent vehicle environment awareness and autonomous navigation.

Yuhe Liang received the B.S. degree in vehicle engineering from Hebei University of Technology, Tianjin, China, in 2021. He is currently a M.S. student at the School of Mechanical Engineering, Dalian University of Technology, China. His research interests include multisensor fusion localization and semantic SLAM.

Zhenqiang Ma received the B.S. degree in vehicle engineering from Shandong University of Technology, China, in 2019, and the M.S. degree in vehicle engineering from Dalian University of Technology, China, in 2022. He is currently a Ph.D. student at the Department of Information Science, Wuhan University, China. His research interests include multisensor fusion localization and semantic SLAM.

Lie Guo received the B.S., M.S., and Ph.D. degrees in vehicle engineering from Jilin University, China, in 2000,, 2003, and 2007, respectively. He is currently an associate professor in the School of Mechanical Engineering, Dalian University of Technology, China. His research interests include intelligent human-computer interaction and collaborative control, intelligent networked vehicles, intelligent vehicles, and automotive safety assisted driving

Hexin Zhang is a Ph.D. candidate of Technology University of Eindhoven, majoring in urban planning & transportation. Her research interests include travel behavior, emerging transportation, and discrete choice model.

https://doi.org/10.26599/JICV.2023.9210027