Title	Three-dimensionalized feature-based LiDAR-visual odometry for online mapping of unpaved road surfaces
Author(s)	Lee, Junwoon; Kurisu, Masamitsu; Kuriyama, Kazuya
Citation	Journal of Field Robotics. 2024, 41(5), p. 1452-1468
Version Type	VoR
URL	https://hdl.handle.net/11094/95750
rights	This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.
Note	

Received: 9 April 2023 | Revised: 15 January 2024 | Accepted: 20 March 2024

DOI: 10.1002/rob.22334

RESEARCH ARTICLE

WILEY

Three-dimensionalized feature-based LiDAR-visual odometry for online mapping of unpaved road surfaces

Junwoon Lee | **Masamitsu Kurisu** | **Kazuya Kuriyama**

Komatsu MIRAI Construction Equipment Cooperative Research Center, Graduate School of Engineering, Osaka University, Osaka, Japan

Correspondence
Masamitsu Kurisu, Komatsu MIRAI Construction Equipment Cooperative Research Center, Graduate School of Engineering, Osaka University, Osaka 565-0871, Japan.
Email: kurisu@jrl.eng.osaka-u.ac.jp

Abstract

Automated maintenance and motion planning for unpaved roads are research areas of great interest in the field robotics. Constructing such systems necessitates the development of surface maps for unpaved roads. However, the lack of distinctive features on unpaved roads degrades the performance of light detection and ranging (LiDAR)-based mapping. To address this problem, this paper proposes three-dimensionalized feature-based LiDAR-visual odometry (TFB odometry) for the online mapping of unpaved road surfaces. TFB odometry introduces a novel interpolation concept to directly estimate the three-dimensional coordinates of the image features using LiDAR. Furthermore, LiDAR intensity-weighted motion estimation is proposed to effectively mitigate the effects of dust, which significantly impact the performance of LiDAR. Finally, TFB odometry includes pose graph optimization to efficiently fuse global navigation satellite system data and poses estimated from motion estimation. Through field experiments on unpaved roads, TFB odometry demonstrated successful online full mapping and outperformed other simultaneous localization and mapping methods. Additionally, it demonstrated remarkable performance in accurately mapping road surface anomalies, even in dusty regions.

KEYWORDS
LiDAR, online mapping, road surface anomaly, road surface mapping, sensor fusion, SLAM, unpaved road

1 | INTRODUCTION

Unpaved roads are considered as those that lack surface treatment, such as asphalt or cement (Saeed et al., 2020). These can be found in different settings from rural areas and grasslands to mountains, deserts, mines, underground tunnels, and even on other planets, like, Mars. Unpaved roads are more vulnerable to damage than that of paved roads because of their lack of surface treatment, making them more prone to ruts, potholes, and erosion. These anomalies pose risks to the safety of vehicles and their passengers, and exert a negative impact on fuel efficiency. Hence, governments and companies spend significant resources on the periodic maintenance of unpaved roads (Burningham & Stankevich, 2005; Saeed et al., 2020).

Mapping unpaved roads can constitute a highly effective approach for optimizing maintenance schedules and minimizing costs. Through the implementation of an online mapping system, a dynamic three-dimensional (3D) map of unpaved roads can be continuously generated, enabling quantitative assessment of road conditions and identification of anomalies. This information can aid the precise determination of maintenance intervals. Moreover, the real-time generation of a comprehensive 3D map empowers vehicles to navigate around obstacles on the road with enhanced situational awareness. Therefore, this paper presents a novel online mapping system for unpaved roads, leveraging sensors mounted on a vehicle.

For accurate online mapping, online simultaneous localization and mapping (SLAM) can be utilized because of its role in estimating an ego motion and constructing a map. However, unlike general SLAM, which is designed for structured and object-rich environments, SLAM for unpaved roads faces specific challenges of mapping object-poor and unstructured environments. Scan matching, a critical step for light detection and ranging (LiDAR) SLAM, is typically employed to align consecutive LiDAR scans. Iterative closest point (ICP), generative ICP, and normal distributions transform are three representative methods for scan matching (Besl & McKay, 1992; Biber & Straßer, 2003; Segal et al., 2009). However, these methods have limitations in object-poor and less-structured environments, when applied to object-poor and less-structured environments. This problem is mainly due to the minimal differences in structure observed between consecutive LiDAR scans in such environments. To address this issue, we propose the use of a camera for image-based feature extraction, and descriptor-based matching to match each scene. This approach is more suitable for these types of environments than LiDAR scan matching, as image-based methods rely primarily on extracting features from two-dimensional (2D) images by calculating differences in brightness values around each pixel, rather than relying heavily on 3D shape information. This method enables the extraction of features from even poorly structured surfaces and enhances the accuracy of the generated map.

In addition to the camera, the inclusion of LiDAR in our approach was motivated by three crucial factors. First, the utilization of LiDAR aids in enhancing the accuracy and robustness of SLAM's motion estimation. This is achieved by directly estimating the 3D position of image features using the surrounding LiDAR data. Second, LiDAR point cloud data are highly effective in representing the shape of the road surface through direct 3D measurements. This advantage stands in contrast to the maps generated by feature-based visual SLAM methods, where the maps primarily consist of individual visual feature points. Third, LiDAR intensity, which is the ratio of the strength of the reflected laser to that of the emitted laser and is dependent on the material properties of the objects, can be an effective tool to visualize anomalies, such as ruts and potholes, which is difficult to detect using other methods.

Dust generated from unpaved roads is another major problem that can significantly affect the performance of LiDAR. To address this problem, we propose a novel motion estimation method called “LiDAR intensity-weighted motion estimation.” This method reduces the impact of dust on the motion estimation by incorporating LiDAR intensity values into the point-to-point matching cost on each scene.

Furthermore, we employed the real-time kinematic global navigation satellite system (RTK-GNSS) to address drift from the LiDAR-visual odometry. To smooth the gap between the GNSS positioning data and the estimated pose, pose graph optimization (PGO) is applied.

Lastly, to visualize the unevenness of the road surface on the map, the plane parameters from the ground are extracted using random sample consensus (RANSAC) plane segmentation (Fischler & Bolles, 1981). Subsequently, the map is color-coded based on the distance from the estimated ground plane. This approach provides a clear and intuitive visualization of surface variations, aiding in the analysis of road conditions.

This paper is structured as follows. In Section 2, we introduce related work on inspection systems for road surface and SLAM. In Section 3, we provide a system overview of the proposed method. In Section 4, we evaluate the proposed method through field experiments. In Section 5, we present the conclusions, including future work.

2 | RELATED WORK

2.1 | Road surface inspection system

Conventionally, road surface conditions have been evaluated by relying on visual inspection of experts (Saeed et al., 2020). Semi-automatic monitoring approaches have also been used, such as utilizing unmanned aerial vehicles (UAVs) or vehicles equipped with Global Positioning System (GPS) and additional sensors. El-Wakeel et al. (2017) proposed a monitoring system that focuses on classifying road surface anomalies, including manholes, potholes, and transverse cracks, along with determining their positions, based on the acceleration changes by a vehicle equipped with a GPS and three-axis accelerometer. Zhang (2008) proposed a system that calculates the 3D geometry and classifies the distress of an unpaved road surface by postprocessed road surface images captured from a UAV equipped with GPS and an inertial measurement unit (IMU).

However, these methods primarily focus on classifying surface anomalies without providing quantitative measurements. This limitation makes it difficult to evaluate the severity of the anomalies, even when their type and location are identified. To overcome this challenge, the construction of a road surface map using sensing data becomes imperative. This approach allows for a quantitative evaluation of the road surface condition by calculating the area or height of the anomalies represented on the map. Azevedo et al. (2013) proposed a road surface mapping system using a stereo camera mounted on a vehicle. In this method, the road surface position relative to the camera is calculated using stereo vision, and a road surface map is acquired using image reprojections and an online-trained road segmentation model. However, because the map derived from this method is constructed as a 2D grid map, the 3D information of the unevenness of the road surface could be lost in this 2D map.

2.2 | 3D mapping

The combination of RTK-GNSS, IMU, and 3D LiDAR is one of the commonly used methods for 3D mapping (Aldibaja et al., 2022; Fernández et al., 2010). However, IMU measurements are prone to sensor noise and bias, which can result in inaccuracies and drift in the orientation of the mapping agent (Ahmad et al., 2013). This, in turn, can negatively impact the quality of the 3D map (Fernández et al., 2010). To deal with these limitations, SLAM, Visual odometry, and LiDAR odometry techniques could be employed to construct 3D map.

Zhang and Singh (2014) proposed low-drift and real-time LiDAR odometry and mapping (LOAM). LOAM leverages edge and planar features extracted from LiDAR scans, based on the calculated smoothness value derived from adjacent points on the same scan line. By using features extracted in a sparse manner, LOAM simultaneously improves the computational efficiency and accuracy compared to the ICP-based method (Besl & McKay, 1992; Rusinkiewicz & Levoy, 2001). Shan et al. (2020) proposed tightly coupled LiDAR-inertial odometry via smoothing and mapping (LIO-SAM). LIO-SAM achieved high accuracy and frequency of pose estimation compared to the LOAM by effectively fusing high-frequency IMU measurements with LOAM-based mapping using IMU preintegration (Forster et al., 2016) and pose graph (Kaess et al., 2012). Another approach for high-frequency LiDAR-inertial odometry is features from accelerated segment test (FAST)-LIO2 (Xu et al., 2022; Xu & Zhang, 2021), which also fuses IMU measurements with LiDAR odometry but employs a tightly coupled iterated Kalman filter (Qin et al., 2020). However, these methods are heavily reliant on 3D structures obtained from their LiDAR scans, making them unsuitable for structure-less environments where we are the primary focus of this work. Additionally, these methods can be negatively affected by factors that impact the quality of LiDAR measurements, such as dust and snow particles in the air. To deal with this problem, several filter algorithms were proposed to directly remove airborne particles from point cloud by using point-to-point distance and LiDAR intensity (Afzalaghaeinaeini et al., 2022; Park et al., 2020). However, these algorithms involve iterative outlier searching processes using 3D K-D tree (Bentley, 1975), which are time-consuming. As a result, incorporating these methods into online mapping is not suitable due to their computational requirements.

Mur-Artal et al. (2015) proposed ORB SLAM, a monocular visual SLAM method that mainly used oriented features from accelerated segment test and rotated binary robust independent elementary features (ORB) (Rublee et al., 2011). ORB SLAM emphasizes high accuracy, high robustness, and time efficiency while also incorporating a loop closure technique for drift correction (Mur-Artal et al., 2015). ORB SLAM2 (Mur-Artal & Tardos, 2017) and ORB SLAM3 (Campos et al., 2021) expanded the capabilities of ORB SLAM to support not only a monocular camera but also stereo, RGB-D cameras, and monocular-inertial systems. However, the map obtained from ORB SLAM consists of sparse ORB features, rendering it unsuitable for mapping anomalies, which are the main focus of this work. Engelhard et al. (2011) proposed 3D SLAM using an RGB-D camera, capable of constructing dense 3D map with high accuracy in online. However, the RGB-D camera is not suitable for outdoor environments due to its susceptibility to sunlight interference and limited depth measurement range.

To deal with the limitations of RGB-D SLAM, various LiDAR-visual odometry methods were proposed because LiDAR offers superior performance compared to RGB-D cameras in terms of measurement range and robustness to external lighting conditions. One closely related concept to the proposed method in this work is real-time depth-enhanced monocular odometry (DEMO), which extracts features from the monocular camera image and estimates their depth using LiDAR (Zhang et al., 2014). The main difference between DEMO and the proposed method lies in their management of features. DEMO simultaneously handles both 2D and 3D features, whereas the proposed method simplifies the optimization process by managing only 3D features interpolated from LiDAR points. Furthermore, DEMO does not account for LiDAR degeneration caused by airborne particles, which is a core contribution of this work.

There are also impressive works on LiDAR-visual odometry fused with IMU (Shan et al., 2021; Zheng et al., 2022). These methods achieved high accuracy compared to pure visual odometry or LiDAR odometry approaches and effectively handled environments with partially containing LiDAR odometry degeneration. However, these methods were not tested extensively in environments dominated by LiDAR odometry degeneration.

3 | PROPOSED METHOD

In this section, we propose a novel SLAM system called three-dimensionalized feature-based LiDAR-visual SLAM (TFB odometry) for the online mapping of unpaved road surfaces. The system diagram is depicted in Figure 1. TFB odometry consists of four threads: the sensor preprocessing thread, the motion estimation thread, the PGO thread, and the visualization thread. Each thread is executed using multithreading to enhance the computational efficiency and ensure online mapping.

In the sensor preprocessing thread, the point cloud data from LiDAR, grayscale images from the camera, and positioning data of the GNSS are preprocessed. Following preprocessing, the image features are extracted using the ORB algorithm (Rublee et al., 2011) and three-dimensionalized using a novel interpolation concept.

In the motion estimation thread, motion estimation is performed using point-to-point matching of the features extracted from images. To mitigate the effect of airborne particles, the matching cost is weighted using LiDAR intensity. In the PGO thread, PGO is performed using the poses estimated from the motion estimation and positions obtained from the GNSS to correct drift and refine the motion.

In the visualization thread, TFB odometry generates three maps: the odometry map, which is constructed based on the estimated poses from the motion estimation thread; the global map, which is constructed based on the globally optimized poses from the PGO thread and color-coded with the LiDAR intensity; and the ground map, which is color-coded with the distances between each map point and estimated ground on the global map. Although the odometry map is based on nonglobally optimized poses obtained

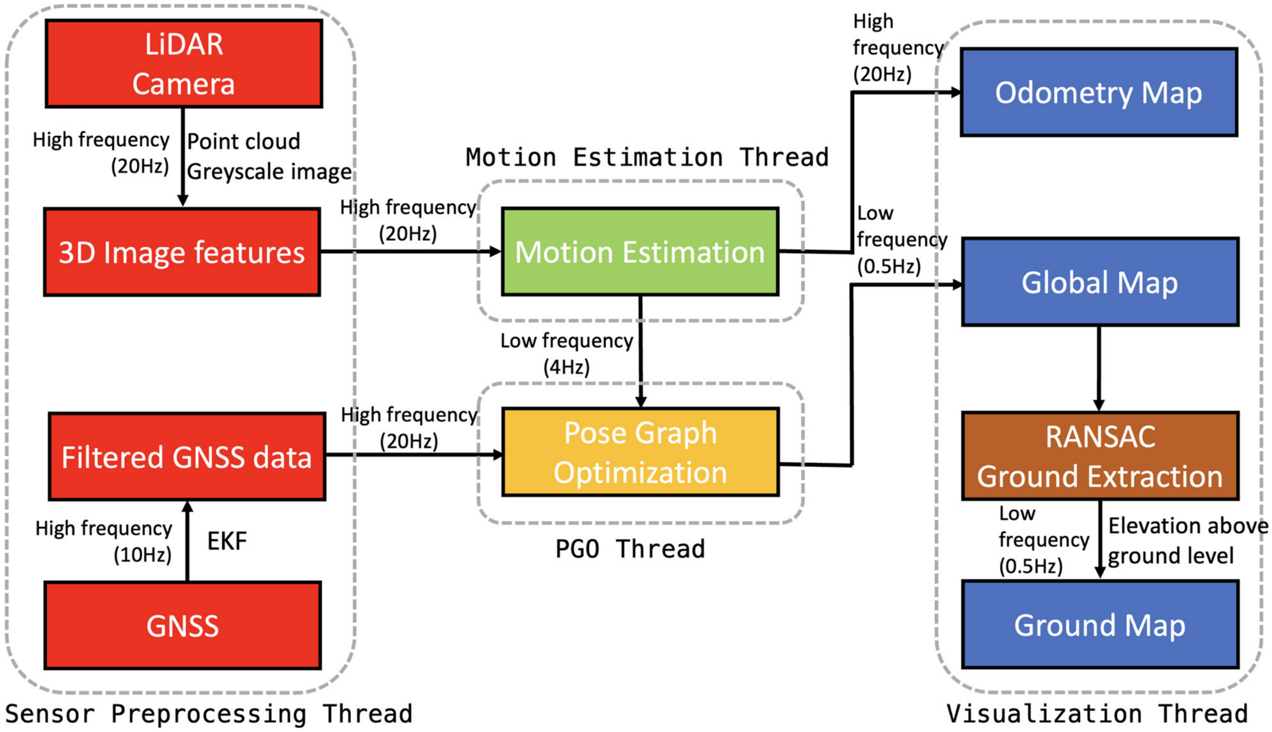

FIGURE 1 System diagram of three-dimensionalized feature-based LiDAR-visual simultaneous localization and mapping (TFB odometry). 3D, three-dimensional; EKF, extended Kalman filter; GNSS, global navigation satellite system; LiDAR, light detection and ranging; ORB, oriented features from accelerated segment test and rotated binary robust independent elementary features; PGO, pose graph optimization; RANSAC, random sample consensus. [Color figure can be viewed at wileyonlinelibrary.com]

from frame-to-frame motion estimation only, it is constructed with high frequency (20 Hz in this work). Meanwhile, the global and ground maps are constructed with low frequency (0.5 Hz in this work) due to the large computational time of PGO, which allows the globally optimized and refined poses.

3.1 | Feature extraction and matching

To associate the LiDAR point cloud and the corresponding camera image, we project the LiDAR point cloud onto a 2D image plane using the pinhole camera model (Brown, 1992). Subsequently, the 2D LiDAR point cloud is cropped according to the image region for region association.

In this paper, we adopted the ORB algorithm as a feature extractor and descriptor. This algorithm enables real-time performance and robustness to changes in viewpoint and illumination. Note that alternatives such as SuperPoint (DeTone et al., 2018) and scale invariant feature transform (Lowe, 2004) face limitations in real-time performance, while origin FAST (Rosten & Drummond, 2006) and Harris corner (Harris & Stephens, 1988) have constraints in handling significant changes in viewpoint and illumination.

Before the feature extraction, we applied a contrast-limited adaptive histogram equalization (CLAHE) filter (Reza, 2004) to the image to mitigate the impact of varying lighting conditions and enhance the consistency of feature extraction. We also utilized the feature extractor of ORB SLAM, which partitions the image into a predetermined number of grids and extracts at least five features from each grid by adjusting the detector threshold to ensure a uniform distribution (Mur-Artal et al., 2015).

To obtain 3D coordinates of the ORB features, we propose a novel interpolation concept. This concept infers the 3D coordinates of an ORB feature $[x_s, y_s, z_s]$ using the three closest LiDAR points $[x_1, y_1, z_1]$, $[x_2, y_2, z_2]$, and $[x_3, y_3, z_3]$ to the ORB feature, as shown in Figure 2. In the proposed method, the point cloud obtained from LiDAR is first projected onto the image plane using the pinhole camera model, and the ORB features are associated with the projected point cloud in the same frame. Subsequently, the projected point cloud is stored in a 2D K-D tree (Bentley, 1975), and the three closest projected points $[u_1, v_1]$, $[u_2, v_2]$, and $[u_3, v_3]$ to the image feature $[u_s, v_s]$ are determined. Before interpolation, the image feature conditioned on the distance between the feature and the third nearest projected point falling below a predetermined threshold was only employed. After that, we assumed a triangle formed by $[u_1, v_1]$, $[u_2, v_2]$, and $[u_3, v_3]$. We also placed weights $[m_1, m_2, m_3]$ at each vertice, where the sum of the weights is 1. Assuming that the centroid of the mass of the triangle is located at the feature point, the weights $[m_1, m_2, m_3]$ are formulated as follows:

$$\begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} = \begin{bmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ 1 & 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} u_s \\ v_s \\ 1 \end{bmatrix}. \quad (1)$$

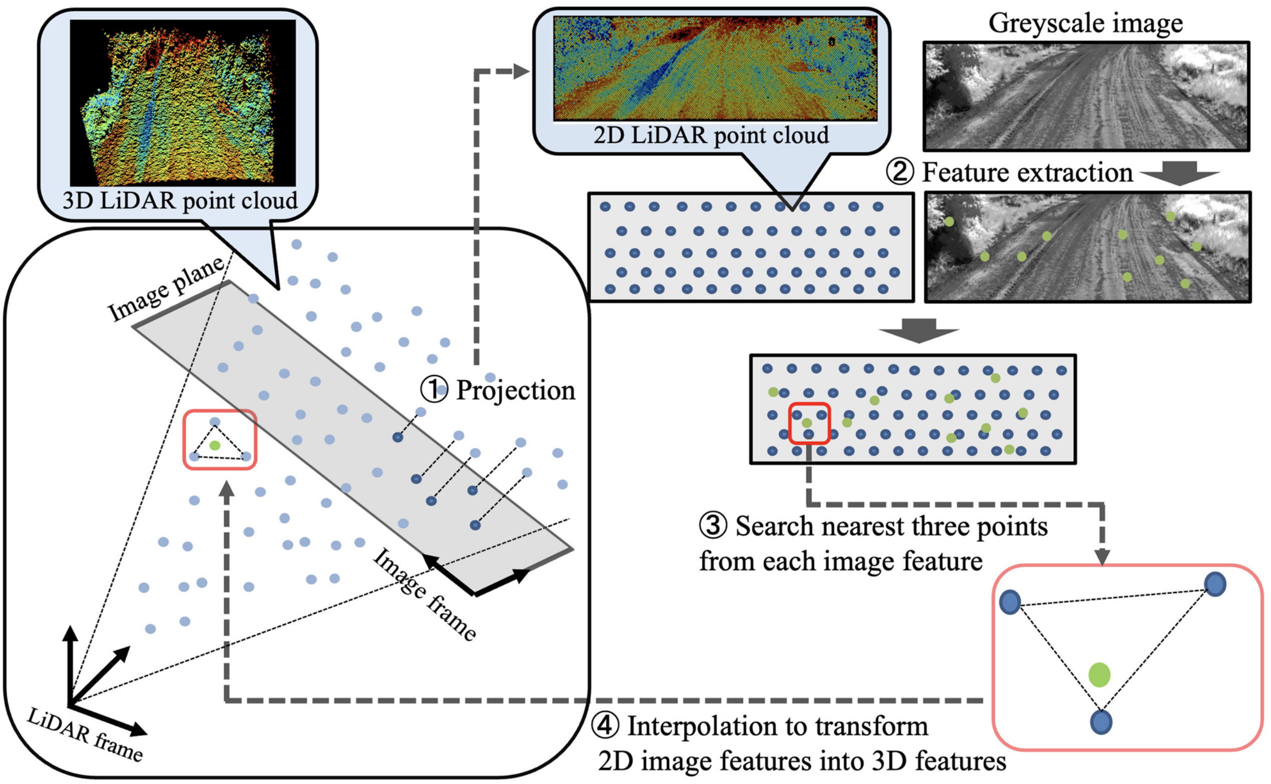

FIGURE 2 Illustration of the proposed interpolation to estimate the 3D coordinates of image features. The blue dots denote LiDAR points, whereas the green dots denote the ORB features extracted from the image. 2D, two-dimensional; 3D, three-dimensional; LiDAR, light detection and ranging; ORB, oriented features from accelerated segment test and rotated binary robust independent elementary features. [Color figure can be viewed at wileyonlinelibrary.com]

Assuming that the centroid of mass of $[x_1, y_1, z_1]$, $[x_2, y_2, z_2]$, and $[x_3, y_3, z_3]$ is also located at $[x_s, y_s, z_s]$, $[x_s, y_s, z_s]$ is interpolated as follows:

$$\begin{bmatrix} x_s \\ y_s \\ z_s \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix}. \qquad (2)$$

When the matrix in (1) is not invertible or when the determinant value of the matrix is close to 0, we formulated an interpolation as follows:

$$\begin{bmatrix} x_s \\ y_s \\ z_s \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{bmatrix} \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}. \qquad (3)$$

Figure 3 shows an example of the original ORB features from an image of an unpaved road and the corresponding 3D points interpolated using the proposed concept. Note that our three-dimensionalized *features* ensure consistency and robustness through the tight coupling of LiDAR and image information, both of which are indispensable.

3.2 | LiDAR intensity-weighted motion estimation

The proposed method performs image feature matching between consecutive frames using the interpolated features and their descriptors. By leveraging this feature-matching information, the motion of the mapping agent is estimated. Let $\mathbf{x}_{(n,k)}$ be the 3D coordinates of the kth matching point in the frame n, and $\mathbf{t}$ and $\mathbf{R}$ are the 3 × 1 translation vector and 3 × 3 rotation matrix between frame n and $n + 1$, respectively. The motion estimation based on the point-to-point matching is formulated as follows:

$$(\mathbf{R}, \mathbf{t})^* = \underset{(\mathbf{R},\mathbf{t})}{\operatorname{argmin}} \left(\sum_k \left\| \mathbf{x}_{(n,k)} - (\mathbf{R}\mathbf{x}_{(n+1,k)} + \mathbf{t}) \right\| \right). \qquad (4)$$

However, the LiDAR performance can be significantly affected by the dust from unpaved roads, leading to a degradation of LiDAR accuracy and LiDAR-based motion estimation (Goodin et al., 2013; Ren et al., 2021). Likewise, also in our experiments, dust affected the LiDAR performance, as shown in Figure 4. In the proposed method, leveraging the insight that the LiDAR intensity of the dust is lower than that of the other points (Afzalaghaeinaeini et al., 2022), the LiDAR intensity-weighted motion estimation is proposed for an accurate estimation of motion even in the presence of dust.

Before conducting the motion estimation, a straightforward mismatch removal is applied to extract reliable matching points. First, a similarity score between the two descriptions of the matching features is computed using the hamming distance. Then, only keypoints with a higher similarity score than the threshold, $S_{\min}$, are

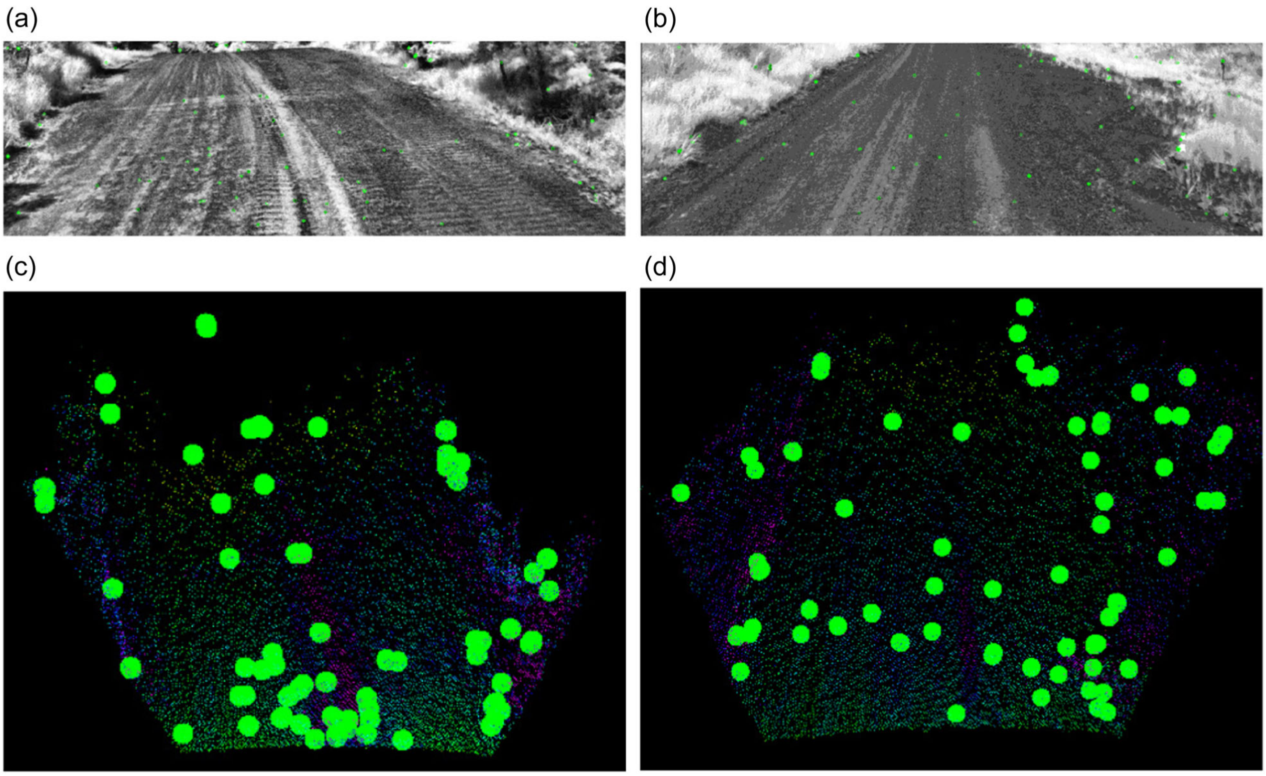

FIGURE 3 Two examples of interpolated three-dimensional feature points for each image. (a) and (b) are grayscale images of unpaved roads. (c) and (d) are LiDAR data of each scene. Green points of (a) and (b) are ORB features. Green points of (c) and (d) are interpolated points by the proposed method. (a) A grayscale image of the unpaved road in scene 1, (b) a grayscale image of the unpaved road in scene 2, (c) interpolated feature points in scene 1, and (d) interpolated feature points in scene 2. LiDAR, light detection and ranging; ORB, oriented features from accelerated segment test and rotated binary robust independent elementary features. [Color figure can be viewed at wileyonlinelibrary.com]

extracted. Second, only keypoints for which the Euclidean distance of the matching 3D features are less than a threshold, $D_{\text{threshold}}$, are extracted. $D_{\text{threshold}}$ is determined as three times the displacement of the previous interframe motion estimation, assuming that there is no rapid acceleration or deceleration of the mapping agent.

The LiDAR intensity of the 3D image feature points is also interpolated using the LiDAR intensity of the three nearest LiDAR points and the weights $[m_1, m_2, m_3]$. Leveraging the matching features filtered with the mismatching removal method, and the corresponding LiDAR intensities $l_{(n,k)}$ and $l_{(n+1,k)}$ of the kth matching, the cost function $\mathbf{f}$ for the LiDAR intensity-weighted motion estimation and corresponding optimization problem can be formulated as follows:

$$\mathbf{f}(\mathbf{R}, \mathbf{t}) = \sum_{k} \left\| \min(l_{(n,k)}, l_{(n+1,k)})(\mathbf{x}_{(n,k)} - (\mathbf{R}\mathbf{x}_{(n+1,k)} + \mathbf{t})) \right\|, \tag{5}$$

$$(\mathbf{R}, \mathbf{t})^{*} = \underset{(\mathbf{R},\mathbf{t})}{\operatorname{argmin}}(\mathbf{f}(\mathbf{R}, \mathbf{t})). \tag{6}$$

By multiplying the original cost function by the LiDAR intensity, the impact of the dust points with low LiDAR intensity is reduced, while the general points with higher intensity have a larger influence on the optimization process for motion estimation.

The nonlinear optimization problem (6) is solved with the Levenberg–Marquardt algorithm (Marquardt, 1963) as follows:

$$\mathbf{T} \leftarrow \mathbf{T} - (\mathbf{J}^{\top}\mathbf{J} + \lambda \operatorname{diag}(\mathbf{J}^{\top}\mathbf{J}))^{-1}\mathbf{J}^{\top}\mathbf{f}(\mathbf{T}), \tag{7}$$

where $\mathbf{T} = [\mathbf{R}|\mathbf{t}]$ is an $SE(3)$ transformation matrix that consists of $\mathbf{R}$ and $\mathbf{t}$, $\mathbf{J} = \frac{\partial \mathbf{f}}{\partial \mathbf{T}}$ is the Jacobian matrix of $\mathbf{f}$, and λ is a damping factor.

3.3 | PGO of TFB odometry

To correct the drift errors in motion estimation and smooth the gap between the GNSS positioning data and the estimated pose, we employed a maximum a posteriori (MAP) estimation approach. MAP estimation for TFB odometry could be formulated as follows:

$$(\mathbf{x}_0, \ldots, \mathbf{x}_n)^{*} = \underset{(\mathbf{x}_0, \ldots, \mathbf{x}_n)}{\operatorname{argmax}} \left(P(\mathbf{x}_0) \prod_{j=0}^{n} (P(\bar{\mathbf{p}}_{(j-1,j)}|\mathbf{x}_{j-1}, \mathbf{x}_j)) \prod_{i=0}^{n} (P(\bar{\mathbf{g}}_j|\mathbf{x}_j)) \right), \tag{8}$$

where $\bar{\mathbf{p}}_{(\mathbf{j-1,j})}$ denotes the transformation measurements from the motion estimation between frame $j-1$ and j, $\bar{\mathbf{g}}_{\mathbf{j}}$ is the GNSS positioning measurements in frame j and $\mathbf{x}_{\mathbf{j}}$ is the estimated pose in frame $j-1$. Assuming that a likelihood $P(\bar{\mathbf{p}}_{(\mathbf{j-1,j})}|\mathbf{x}_{\mathbf{j-1}}, \mathbf{x}_{\mathbf{j}})$ and $P(\bar{\mathbf{g}}_{\mathbf{j}}|\mathbf{x}_{\mathbf{j}})$ follows the multivariate Gaussian distribution, the likelihoods and residual are formulated as follows:

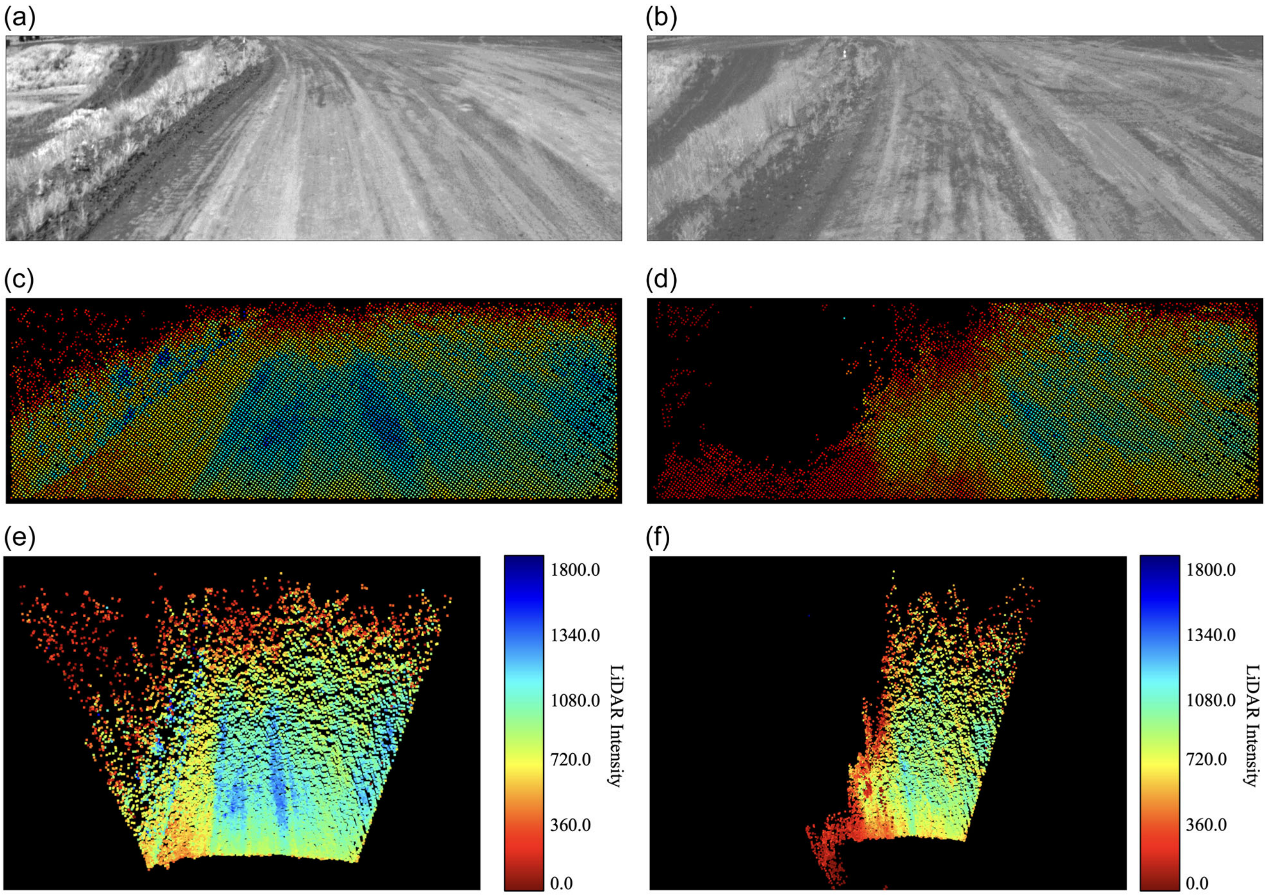

FIGURE 4 Comparison of normal sensor data on an unpaved road and that disturbed by dust in the air. (a) Grayscale image (normal), (b) grayscale image (dust), (c) projected LiDAR data (normal), (d) projected LiDAR data (dust), (e) 3D LiDAR point cloud (normal), and (f) 3D LiDAR point cloud (dust). 2D, two-dimensional; 3D, three-dimensional; LiDAR, light detection and ranging. [Color figure can be viewed at wileyonlinelibrary.com]

$$\begin{aligned} P(\bar{\mathbf{p}}_{(j-1,j)}|\mathbf{x}_{j-1},\mathbf{x}_j) &= \lambda_1 \exp\left(-\frac{1}{2}\mathbf{e}_{\mathbf{p}}(\mathbf{x}_{j-1},\mathbf{x}_j)^\top \mathbf{P}_{(j-1,j)}\mathbf{e}_{\mathbf{p}}(\mathbf{x}_{j-1},\mathbf{x}_j)\right), \\ \mathbf{e}_{\mathbf{p}}(\mathbf{x}_{j-1},\mathbf{x}_j) &= \mathbf{f}_{\mathbf{v}}\left(\bar{\mathbf{p}}^{-1}_{(j-1,j)}\left(\mathbf{X}^{-1}_{j-1}\mathbf{X}_j\right)\right), \end{aligned} \tag{9}$$

$$\begin{aligned} P(\bar{\mathbf{g}}_j|\mathbf{x}_j) &= \lambda_2 \exp\left(-\frac{1}{2}\mathbf{e}_{\mathbf{g}}(\bar{\mathbf{g}}_j,\mathbf{x}_j)^\top \mathbf{G}_j\mathbf{e}_{\mathbf{g}}(\bar{\mathbf{g}}_j,\mathbf{x}_j)\right), \\ \mathbf{e}_g(\bar{\mathbf{g}}_j,\mathbf{x}_j) &= \bar{\mathbf{g}}_j - \mathbf{x}_j, \end{aligned} \tag{10}$$

where $\mathbf{P}_{(j-1,j)}$ and the $\mathbf{G}_j$ denote 6 × 6 and 3 × 3 fisher information matrix of each measurement, respectively, $\mathbf{X}_j$ denotes the $SE(3)$ transformation matrix of $\mathbf{x}_j$, $\mathbf{f}_{\mathbf{v}}$ is a function that transforms the $SE(3)$ transformation matrix to the 6 × 1 vector including a translation and Euler angles, and λ_1 and λ_2 are constants of each multivariate Gaussian distribution.

To efficiently solve the MAP problem, a PGO is performed using the pose graph, as shown in Figure 5. To process the GNSS data, we first transformed the geographic coordinates (latitude, longitude, and altitude) to a local Cartesian coordinate frame. Subsequently, the GNSS data were noise-filtered by the EKF-based method (Moore & Stouch, 2016). Owing to the high computational cost of the PGO, only keyframes selected from the entire motion estimation frames are inserted in the pose graph. The motion estimation keyframe is selected under the following conditions:

- After the previous keyframe was selected, the duration exceeds 0.25 s (4 Hz).
- The displacement or rotation between the current frame and the previous keyframe exceeds a specific displacement or rotation from the previous keyframe.

By selectively inserting keyframes based on the specific conditions, rather than using all the frames, the calculation efficiency of the optimization is increased, while, during scenes with rapid changes, the accuracy of calculation is increased by densely inserting keyframes.

After inserting the keyframes into the pose graph, the MAP problem of the pose graph is solved at intervals of 0.5 Hz by iSAM2 (Kaess et al., 2012), which achieves good computational efficiency in large-scale environments using a Bayes tree. After that, using the optimized keyframe poses and the corresponding point cloud data, the global map is constructed and visualized. Since the resulting global map is constructed only using optimized keyframes, the map

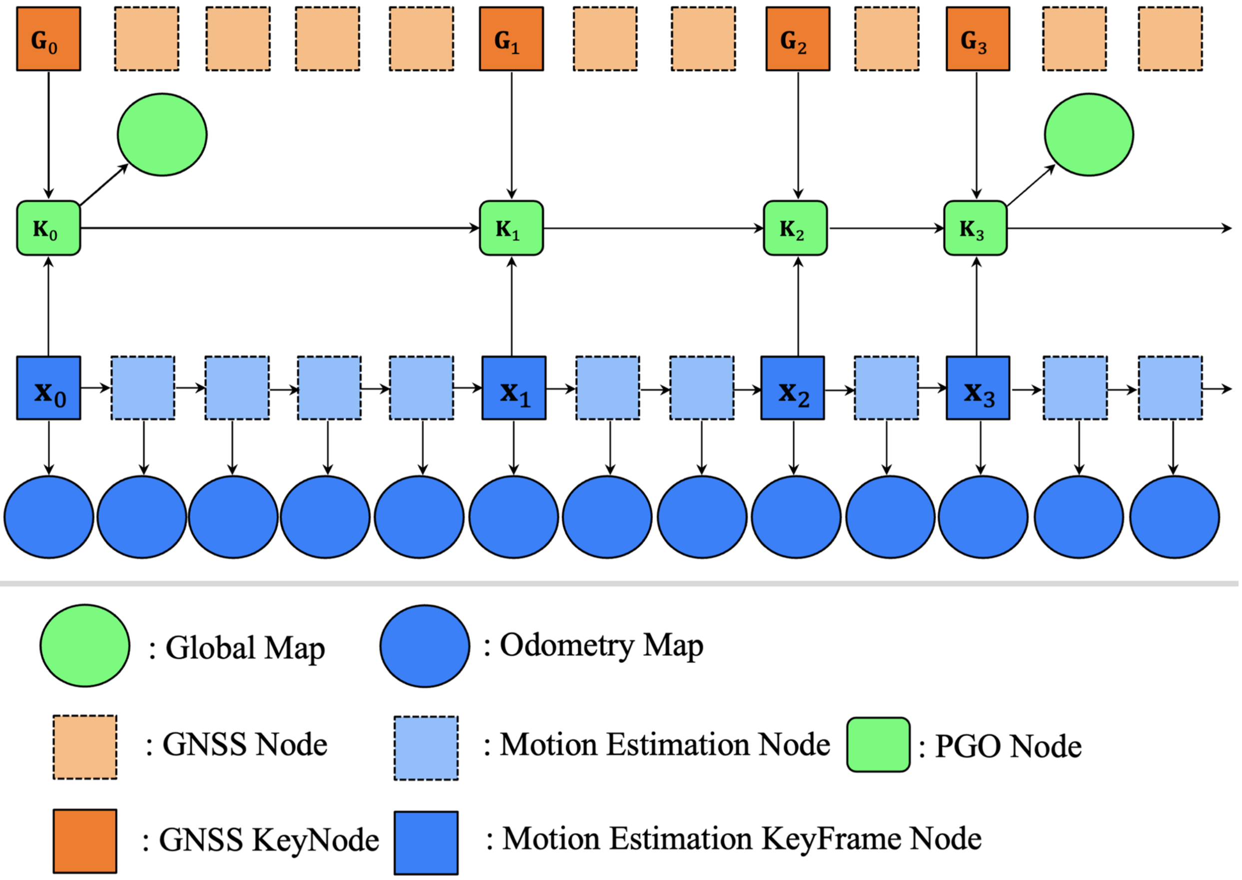

FIGURE 5 Example of a pose graph constructed using the proposed method. **x** denotes a pose estimated using the frame-to-frame motion estimation. **K** denotes a pose optimized by the PGO. **G** denotes GNSS positioning data. GNSS, global navigation satellite system; PGO, pose graph optimization. [Color figure can be viewed at wileyonlinelibrary.com]

volume is decreased, and the accuracy is improved, compared to the odometry map that uses all the frames.

3.4 | Ground plane estimation

In this study, we constructed a ground map color-coded with the distance between an estimated ground plane and each point of the global map.

To estimate the ground plane parameters, the point cloud is filtered using a predefined region of interest to avoid approximating a nonground plane as a ground plane by removing points other than around the ground. Subsequently, the plane parameters are extracted by applying RANSAC (Fischler & Bolles, 1981) plane segmentation algorithm, which consists of the following three steps repeated for a predetermined number of iterations, $N_{\text{threshold}}$:

(1) Three points $\mathbf{x}_1, \mathbf{x}_2$, and $\mathbf{x}_3$ are randomly extracted from the keyframe point cloud P_{net}, and the plane parameter $(\mathbf{n}_t, d_t)$ is temporarily obtained using $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ as follows:

$$\mathbf{n}_t = \frac{(\mathbf{x}_3 - \mathbf{x}_1) \times (\mathbf{x}_2 - \mathbf{x}_1)}{\| (\mathbf{x}_3 - \mathbf{x}_1) \times (\mathbf{x}_2 - \mathbf{x}_1) \|}, \quad d_t = \mathbf{n}_t \cdot \mathbf{x}_1. \tag{11}$$

(2) The distance r between the estimated plane and each point p_i is calculated as follows:

$$r = \| \mathbf{n}_t \cdot p_i + d_t \|, \tag{12}$$

if the distance r is smaller than a predetermined inlier threshold, r_{inlier}, the point is inserted in the point cloud, $\mathbf{P}_l$.

(3) The arithmetic mean of distance between the estimated plane and each point in $\mathbf{P}_l$, E_l is calculated as follows:

$$E_l = \frac{\sum_{\mathbf{p} \in \mathbf{P}_l} \left| \mathbf{n}_t \cdot \mathbf{p} + d_t \right|}{\sum_{\mathbf{p} \in \mathbf{P}_l} (1)}, \tag{13}$$

where $E_{\text{threshold}}$ denotes the minimum of all E_l values before. If E_l is smaller than $E_{\text{threshold}}$, a new $E_{\text{threshold}}$ is determined as E_l, and the estimated plane parameter $(\mathbf{n}_{\text{plane}}, d_{\text{plane}})$ is replaced with $(\mathbf{n}_t, d_t)$.

Based on the estimated plane parameter of each keyframe, the points are color-coded according to the distance between the estimated plane and each point. The resulting ground map can visualize the height variations on the ground across the mapped area, with different colors indicating different levels of elevation. This map can be useful for various applications, such as terrain analysis and road inspection.

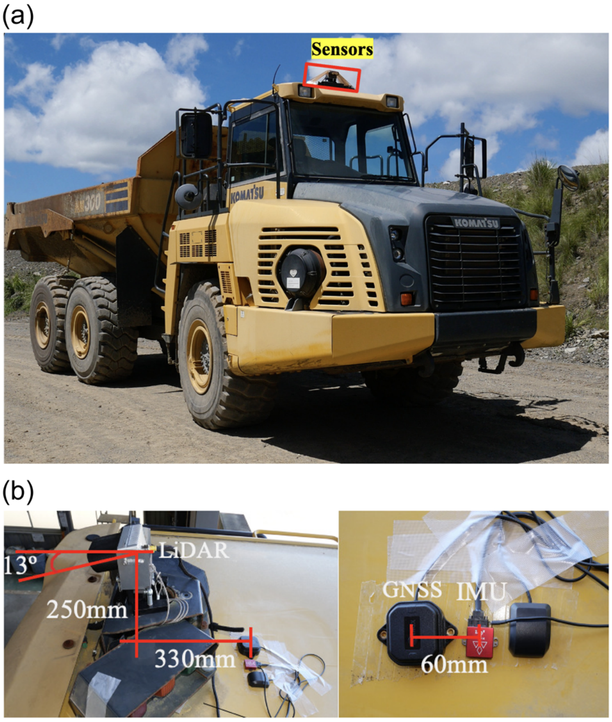

FIGURE 6 Suite of sensors, including a camera, LiDAR, IMU, and GNSS, attached to the cabin of a Komatsu HM300 truck for field experiments. (a) Komatsu, HM300 articulated dump truck and (b) sensors attachment. GNSS, global navigation satellite system; IMU, inertial measurement unit; LiDAR, light detection and ranging. [Color figure can be viewed at wileyonlinelibrary.com]

4 | FIELD EXPERIMENT

The proposed method was evaluated on a Komatsu HM300 truck outfitted with a XenoLidar Xact solid-state LiDAR, ublox ZED-F9P RTK-GNSS receiver, ublox ANN-MB-00 GNSS antennae, and VectorNav VN-200 IMU. LiDAR provides the data at a rate of 20 Hz with an angular resolution of $0.3^\circ \times 0.3^\circ$, a field of view of $60^\circ \times 20^\circ$, and a range of up to 50 m. The LiDAR also includes a built-in camera with a data output rate of 20 Hz and a resolution of 1536 × 512. The RTK-GNSS has a data output rate of 10 Hz and can achieve centimeter-level accuracy. Throughout the field experiments, the RTK-GNSS status consistently remained at FIX, and the position covariance remained remarkably low, measuring below $0.5\,\mathrm{m}^2$. The IMU was set to a data output rate of 125 Hz to collect the acceleration, angular velocity, and heading information of the truck. The sensors were mounted on the cabin of the truck, as shown in Figure 6. The LiDAR was installed at an angle of 13° below the horizontal to primarily capture the road surface. The proposed method was executed on a robot operation system (Quigley et al., 2009) using an Ubuntu Linux computer equipped with an Intel i9-12900k CPU and 32 GB RAM.

The field experiment was conducted on a sunny day at noon on unpaved roads with dust occasionally generated during driving. To evaluate the performance of the proposed method in mapping road surface anomalies, we prepared artificial ruts and a pothole as anomalies on the driving route, as shown in Figure 7.

In this section, we present the evaluation of the proposed method, which is divided into four parts. First, we evaluated the localization accuracy of the proposed method by comparing it to that of other SLAM methods. Second, we assessed the accuracy of the road surface anomaly mapping by comparing it with the GNSS/IMU mapping conventionally used for road surface mapping. In this paper, GNSS/IMU mapping refers to the mapping process that utilizes the position and orientation derived from the EKF-based fusion of GNSS and IMU data as described by Moore and Stouch (2016). Third, we evaluate the mapping ability in dusty environments by comparing the proposed method with conventional point-to-point matching and other filter algorithms.

4.1 | Evaluation of the localization accuracy

To evaluate the localization accuracy, we conducted experiments using five sequences, as shown in Figure 8. In each sequence, we evaluated the performance of our proposed method by comparing it with that of LOAM (Zhang & Singh, 2014), LIO-SAM (Shan et al., 2020), FAST-LIO2 (Xu et al., 2022), ORB SLAM2 (the monocular-version) (Mur-Artal & Tardos, 2017), ORB SLAM3 (the monocular-version and monocular-inertial version) (Campos et al., 2021), and LVI-SAM (Shan et al., 2021).

For ORB SLAM2 and ORB SLAM3, we corrected the trajectory scale using the GNSS trajectory because the monocular camera SLAM often suffers from the scale problem. For a fair comparison, the proposed method did not use the RTK-GNSS data fusion to correct the drift. Instead, we used the RTK-GNSS positioning data as the ground truth for the evaluation.

We evaluated the methods using the absolute trajectory error (ATE) and relative trajectory error (RTE) metrics, which were calculated with respect to the ground truth. The root-mean-square error of ATE and RTE from frame 0 to n are defined as follows:

$$E_{\mathrm{ate}} = \frac{1}{n}\sum_{j=0}^{n}(\|\hat{\mathbf{x}}_j - \hat{\mathbf{g}}_j\|), \tag{14}$$

$$E_{\mathrm{rte}} = \frac{1}{n}\sum_{j=1}^{n}(\|(\hat{\mathbf{x}}_j - \hat{\mathbf{x}}_{j-1}) - (\hat{\mathbf{g}}_j - \hat{\mathbf{g}}_{j-1})\|), \tag{15}$$

where $\hat{\mathbf{x}}_j$ is the estimated position in frame j, and $\hat{\mathbf{g}}_j$ is the ground truth position in frame j.

The evaluation results are presented in Table 1. LOAM, LIO-SAM, and FAST-LIO2 were degraded, and their localization failed in all sequences because of the inconsistency of the extracted features and scarcity of the objects on the road surface. Moreover, ORB SLAM2 faced tracking failures of the ORB features in sequences 2 and 3. Similarly, the monocular version of ORB SLAM3 experienced failures in sequences 1, 3, and 4, while the monocular-inertial version of ORB SLAM3 failed in all sequences. LVI-SAM encountered localization failures in sequences 2, 4, and 5 due to the failure of its visual SLAM subsystem.

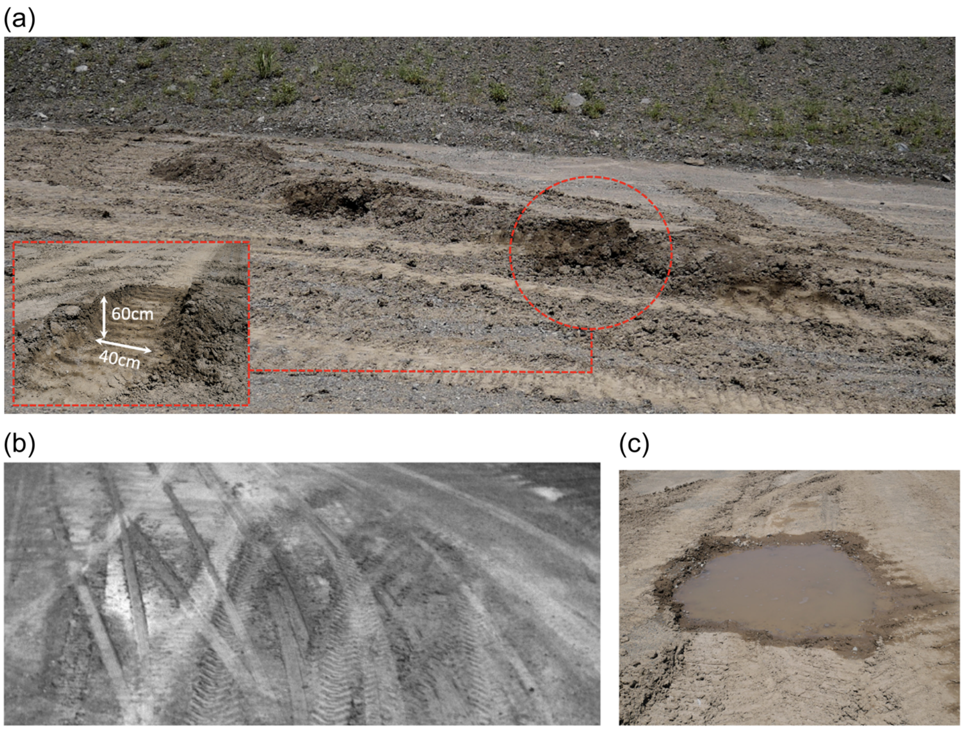

FIGURE 7 Three types of prepared anomalies on the driving route: big rut, small rut, pothole. (a) Big rut with a height of 0.6 m and wheel marks with a width if 0.4 m, (b) small rut, and (c) pothole. [Color figure can be viewed at wileyonlinelibrary.com]

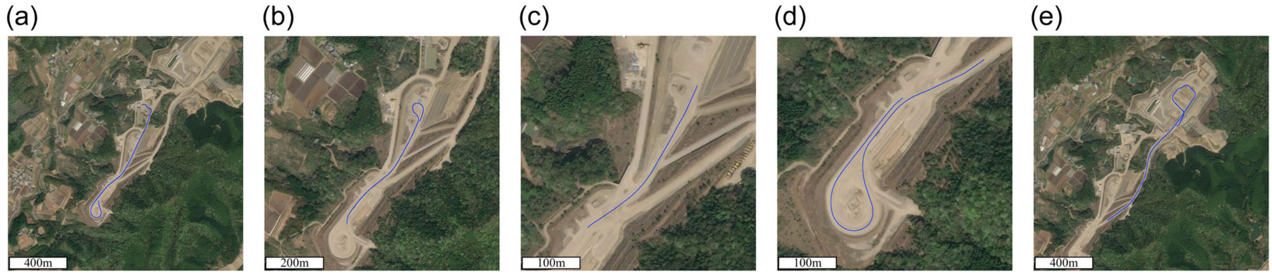

FIGURE 8 GNSS trajectories (blue line) of the tested driving sequences overlaid on the satellite image. (a) Sequence 1, (b) Sequence 2, (c) Sequence 3, (d) Sequence 4, and (e) Sequence 5. GNSS, global navigation satellite system. [Color figure can be viewed at wileyonlinelibrary.com]

These failures are attributed to the limitations of visual SLAM in handling rapidly changing lighting conditions in outdoor environments, leading to inconsistency in the extracted features. Additionally, the use of triangulation, a common method for estimating the 3D coordinates of image features, is ineffective in environments with purely linear motion. In contrast, the proposed method employed a CLAHE filter to effectively handle changing lighting conditions and utilized direct matching of 3D points obtained from LiDAR data for motion estimation. This approach eliminated the reliance on perspective-n-point pose computation or triangulation, commonly used in ORB SLAM and other visual SLAM methods. As a result, the proposed method achieved successful localization in all sequences, distinguishing it from the other compared methods. Additionally, the proposed method demonstrated lower RTE values compared to compared methods across all sequences.

Furthermore, to evaluate the sufficiency of the features involved in localization, Figure 9 presents the 3D feature points utilized in localization. The results indicate that a minimum of 40 features were utilized in every sequence. These features exhibited a uniform distribution due to the grid-based feature extraction and were refined using the proposed mismatch rejection, which considers both the similarity score between feature descriptions and the Euclidean distance between matched 3D features. For these reasons, the participated features in these experiments can be considered sufficient for 3D localization and mapping.

4.2 | Evaluation on the road surface mapping

To evaluate the performance of the road surface mapping, we compared the proposed method with the GNSS/IMU mapping used

TABLE 1 Absolute trajectory error (ATE) and relative trajectory error (RTE) between the ground truth and the trajectory of different methods (ORB SLAM2, monocular version of ORB SLAM3, LVI-SAM, and proposed method).

				ATE (m)			RTE (m)		
Sequence	Length (m)	Time (s)	Method	Maximum	RMSE	STD	Maximum	RMSE	STD
1	1123.0	219.7	ORB SLAM2	76.26	**24.40**	11.47	40.69	11.76	7.82
			ORB SLAM3	N/A	N/A	N/A	N/A	N/A	N/A
			LVI-SAM	65.83	40.09	19.57	18.34	10.46	5.39
			Proposed	**37.19**	26.13	7.20	**7.66**	**4.67**	2.41
2	524.61	88.42	ORB SLAM2	N/A	N/A	N/A	N/A	N/A	N/A
			ORB SLAM3	32.57	23.40	9.78	12.44	2.97	2.49
			LVI-SAM	N/A	N/A	N/A	N/A	N/A	N/A
			Proposed	**16.78**	**6.99**	2.77	**5.83**	**2.77**	1.56
3	258.18	39.80	ORB SLAM2	N/A	N/A	N/A	N/A	N/A	N/A
			ORB SLAM3	N/A	N/A	N/A	N/A	N/A	N/A
			LVI-SAM	25.02	19.26	7.14	13.47	7.98	3.39
			Proposed	**4.11**	**1.38**	0.79	**6.98**	**6.09**	0.39
4	665.3	100.0	ORB SLAM2	95.07	24.14	17.61	37.46	11.84	7.61
			ORB SLAM3	N/A	N/A	N/A	N/A	N/A	N/A
			LVI-SAM	N/A	N/A	N/A	N/A	N/A	N/A
			Proposed	**22.22**	**11.05**	4.18	**6.86**	**3.90**	1.90
5	1537.2	374.5	ORB SLAM2	**56.94**	27.51	12.12	53.38	9.86	6.57
			ORB SLAM3	115.37	54.34	31.73	92.79	28.48	19.28
			LVI-SAM	N/A	N/A	N/A	N/A	N/A	N/A
			Proposed	74.17	**25.29**	12.28	**6.06**	**3.00**	1.14

Note: LOAM, LIO-SAM, FAST-LIO2, and ORB SLAM3 (monocular-inertial) failed in all sequences. The bold values indicate the best performance in the comparison. N/A indicates the failure of SLAM.

Abbreviations: FAST-LIO, fast direct LiDAR-inertial odometry; LiDAR, light detection and ranging; LIO-SAM, LiDAR-inertial odometry via smoothing and mapping; LOAM, LiDAR odometry and mapping; LVI-SAM, lidar-visual-inertial odometry via smoothing and mapping; ORB, oriented features from accelerated segment test and rotated binary robust independent elementary features; RMSE, root-mean-square error; SLAM, simultaneous localization and mapping; STD, standard deviation.

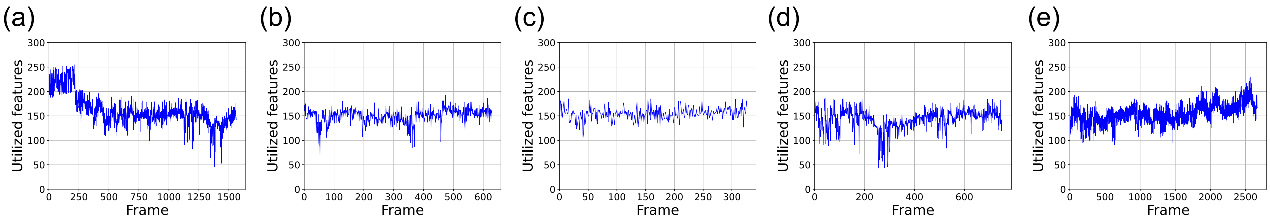

FIGURE 9 Three-dimensionalized feature points involved in the motion estimation and mapping after mismatching removal. (a) Sequence 1, (b) Sequence 2, (c) Sequence 3, (d) Sequence 4, and (e) Sequence 5. [Color figure can be viewed at wileyonlinelibrary.com]

in conventional road surface mapping. The evaluation was carried out in two approaches. First, we compared the performance of each method by examining the map of the region containing the road surface anomalies as shown in Figure 7. Second, we determined the performance of each method by examining the connectivity of the mapped ground, because a well-connected ground is crucial for accurately identifying road anomalies and properly mapping surface unevenness.

The big rut, small rut, and pothole were mapped in online by each method. For the big rut, the shape of the rut and width of the truck

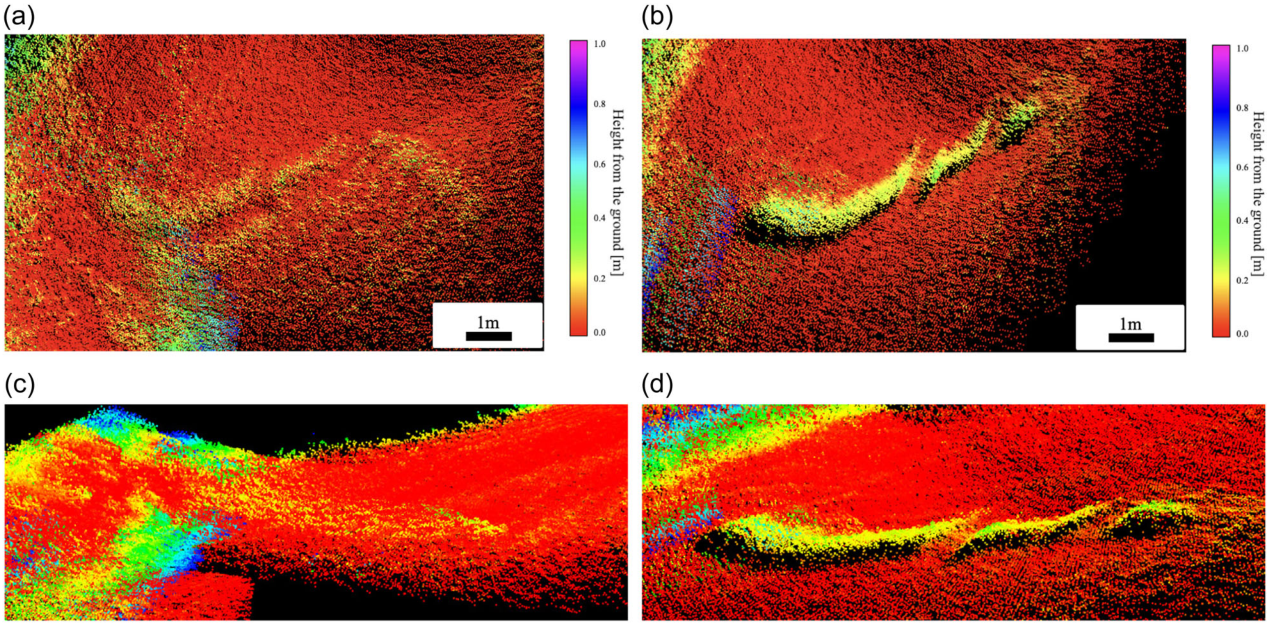

FIGURE 10 Comparison of the mapping performance of the prepared big rut between the proposed method and GNSS/IMU mapping. The maps shown are color-coded according to the distance between the map point and the estimated ground plane. (a) GNSS/IMU mapping, top view, (b) proposed, top view, (c) GNSS/IMU mapping, oblique view, and (d) proposed, oblique view. GNSS, global navigation satellite system; IMU, inertial measurement unit. [Color figure can be viewed at wileyonlinelibrary.com]

traces were described with higher accuracy using the proposed method than that using the GNSS/IMU mapping, as shown in Figure 10. For the small rut, which is presented as a yellow-colored region in the ground map, the proposed method described the unevenness of the rut with higher accuracy than that of the GNSS/IMU mapping, as shown in Figure 11. The global map colored according to the LiDAR intensity shows the well-connected presentation of the small rut using the proposed method, whereas the rut was cut off on the map constructed by the GNSS/IMU mapping. For the pothole, the region of water in the pothole is originally shown as a no-data region because the laser beam of the LiDAR rarely reflects water. As a result, the pothole region is depicted with the background color of the map. The proposed method described the pothole with higher accuracy than that of the GNSS/IMU mapping because fewer points were included as potholes in the proposed method, as shown in Figure 12. Therefore, all prepared anomalies were described with higher accuracy using the proposed method than the GNSS/IMU mapping.

To evaluate the connectivity of the ground on the map, we proposed a novel evaluation method as follows:

1. The checkpoints are set at the same intervals on the estimated trajectory from the obtained map.
2. At each checkpoint location, a 6 m × 6 m × 6 m cube is set. Then, the ground plane is extracted from the point cloud within this cube using a plane segmentation algorithm.
3. The average distance between the estimated plane and each point in the cube is calculated to examine the connectivity of the mapped ground.

Lower average distances indicate that the points on the map are closer together on the plane, suggesting their better connectivity. whereas, if sequential frames of the point cloud are not connected properly, the points on the map will be further away from the estimated plane, as depicted in Figure 13.

The evaluation was performed on sequence 5, as shown in Figure 14. The proposed method obtained lower average distances than that of the GNSS/IMU mapping, regardless of whether the GNSS data were used or not in the proposed method. The only exception was that the values increased rapidly at around the 200th and 430th checkpoint in the case of the proposed method not using GNSS. These checkpoints correspond to the areas with a high dust content, which negatively affects the motion estimation and increases the mismatch in the generated map.

As a result, excluding the dust region, the average distances between the estimated plane and each point were less than 0.4 m in the proposed method, indicating that the proposed method can sufficiently conduct the mapping anomalies with heights greater than 0.4 m.

Our method provides a ground map visualized with distance from the ground and a global map visualized with LiDAR intensity. As shown in Figure 10, the ground map can effectively highlight anomalies with substantial height differences, such as large ruts, pits, and slopes. On the other hand, the global map is useful for visualizing anomalies with large differences in material properties, such as potholes with water and dirt ruts, as shown in Figure 11.

One of the main goals of our mapping approach is to achieve online execution with several advantages, such as high-speed map updates, which are crucial for supporting motion planning and online

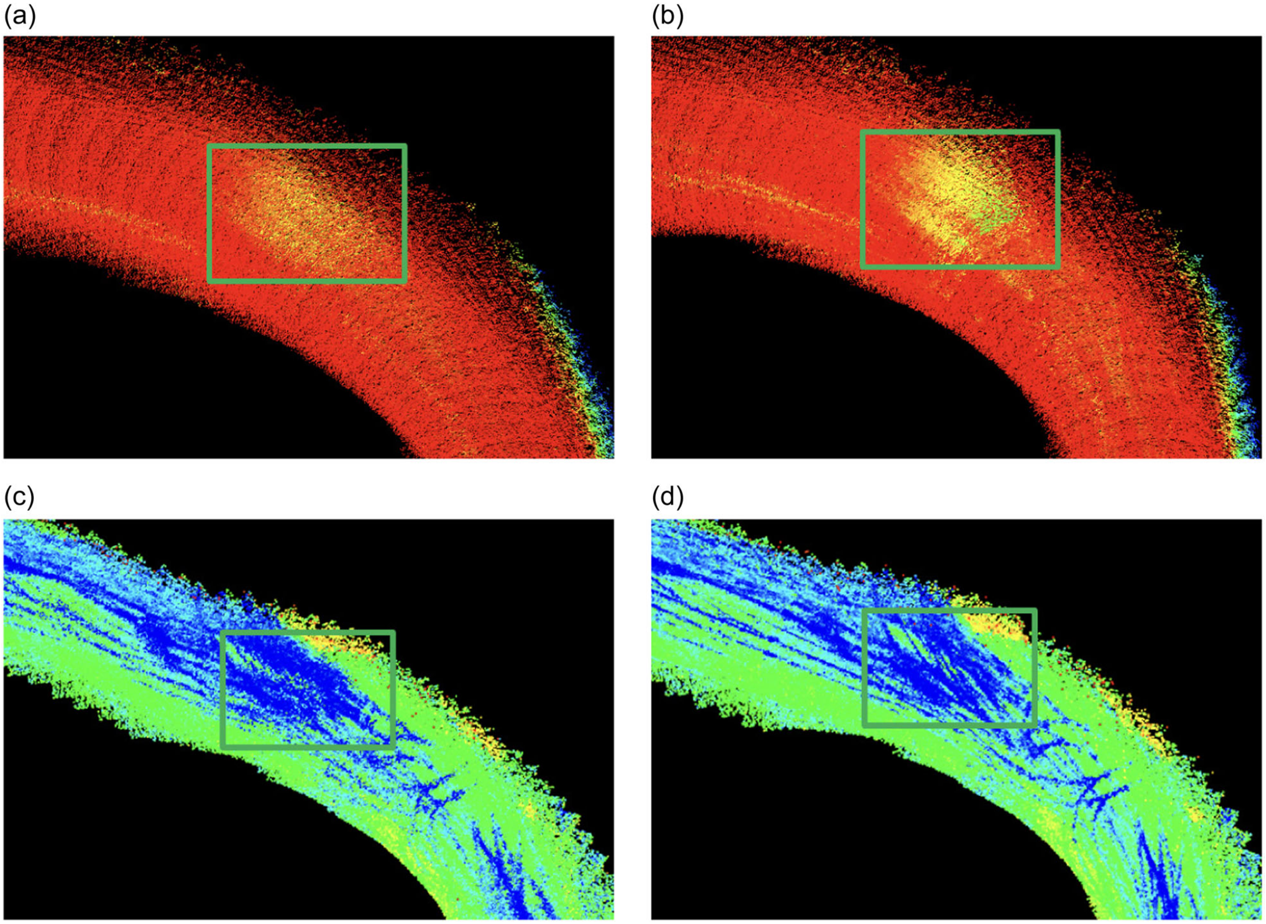

FIGURE 11 Comparison of the mapping performance of the prepared small rut using the proposed method and GNSS/IMU mapping. The maps in (a) and (b) are color-coded according to the distance between each map point and the estimated ground plane. The maps in (c) and (d) are color-coded based on the LiDAR intensity of each corresponding map point. The green rectangle denotes a region of the small rut on the map. (a) GNSS/IMU mapping, top view, (b) proposed, top view, (c) GNSS/IMU mapping, top view (intensity), and (d) proposed, top view (intensity). GNSS, global navigation satellite system; IMU, inertial measurement unit; LiDAR, light detection and ranging. [Color figure can be viewed at wileyonlinelibrary.com]

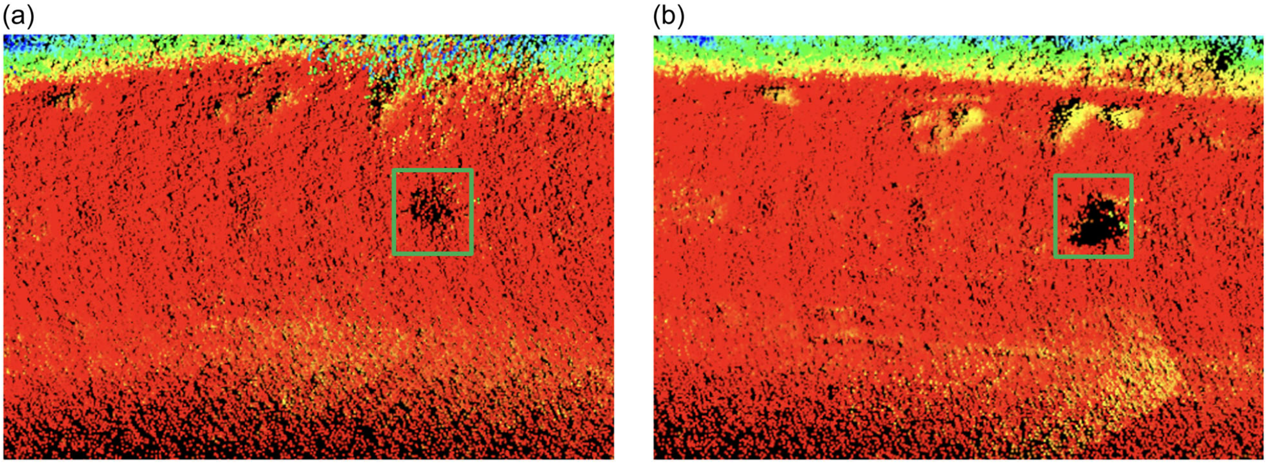

FIGURE 12 Comparison of the mapping performance of the prepared pothole using the proposed method and GNSS/IMU mapping. The maps are color-coded according to the distance between each map point and the estimated ground plane. Green rectangle denotes a pothole region on the map. (a) GNSS/IMU mapping, top view and (b) Proposed, top view. IMU, inertial measurement unit; GNSS, global navigation satellite system.

inspection of road surfaces. To evaluate the performance of the online execution of our mapping system, we measured the time consumption of each thread during the execution of sequence 5, as shown in Figure 15. The thread with the largest time consumption was the sensor preprocessing thread because the image feature extraction requires a longer calculation time. Considering the maximum operational frequency of 20 Hz (50 ms) for the camera and LiDAR used in our system, none of the threads took longer than 50 ms to execute in one cycle, as shown in Figure 15. Hence, the proposed method is applicable to the online operation.

4.3 | Evaluation of the mapping ability in dusty environments

To evaluate the mapping ability in dusty environments, the truck was driven through a dusty area, and the sensor data were recorded. The mapping performance of the proposed with and without the LiDAR intensity-weighted motion estimation was compared. As shown in Figure 16, the map generated without the LiDAR intensity-weighted method suffered from mismatching and rapid drift in the dusty region. In contrast, the proposed method, which used the LiDAR intensity-weighted method, demonstrated proper mapping even in the presence of dust. These results indicate that the proposed method has superior mapping capabilities in dusty environments, demonstrating it as a promising candidate for outdoor applications, including unpaved roads.

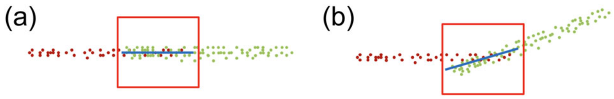

FIGURE 13 Illustration of the sequential frames with good and bad matching. The red and green dots represent the LiDAR points of former and latter frame, respectively. The red rectangle represents a 6 m × 6 m × 6 m cube on the checkpoint. The blue line represents a plane estimated using the points in the cube. (a) Good matching situation of sequential frames (side view) and (b) bad matching situation of sequential frames (side view). LiDAR, light detection and ranging. [Color figure can be viewed at wileyonlinelibrary.com]

Furthermore, we compared the proposed method with two state-of-the-art filter algorithms: statistical outlier removal (Rusu & Cousins, 2011) and low-intensity outlier removal (Park et al., 2020), which are used to directly remove airborne particles. Each filter algorithm was applied before executing the baseline method described in (4). The parameters used for the filters in our experiment were set to be the same as those described in the work by Park et al. (2020). The evaluation was conducted on sequence 5, which represents a scenario of LiDAR degeneration due to dust particles. As shown in Table 2, the proposed method outperformed both the baseline and the combination of filtering algorithms and the baseline in terms of processing time and RTE. This result highlights the effectiveness of the proposed method and its advantage over existing approaches in mitigating the impact of dust particles and improving overall performance.

5 | CONCLUSION

In this paper, we proposed TFB odometry for the online mapping of unpaved road surfaces. TFB odometry is built upon a novel interpolation method that directly estimates the 3D coordinates of image features by utilizing corresponding LiDAR points. These three-dimensionalized features enable accurate motion estimation even in

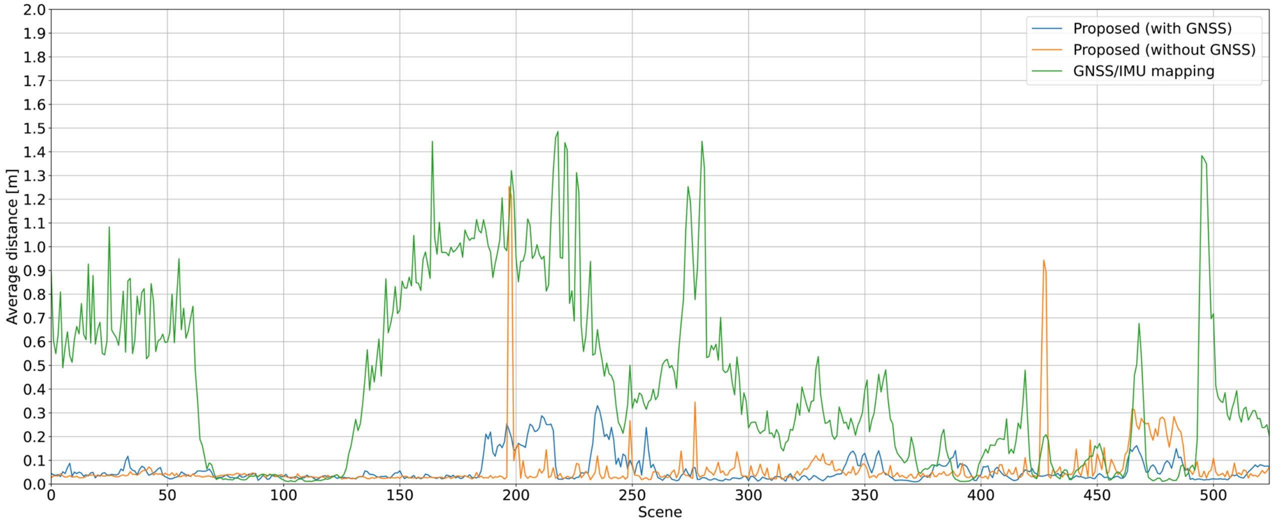

FIGURE 14 Comparison of the average distance **D** in sequence 5 between the proposed method using GNSS (blue line), the proposed method not using GNSS (orange line), and GNSS/IMU mapping (green line). The average distance **D** represents the distance between the estimated plane and each point within a cube of dimensions 6 m × 6 m × 6 m that is positioned along the estimated trajectory. GNSS, Global Navigation Satellite System; IMU, Inertial Measurement Unit. [Color figure can be viewed at wileyonlinelibrary.com]

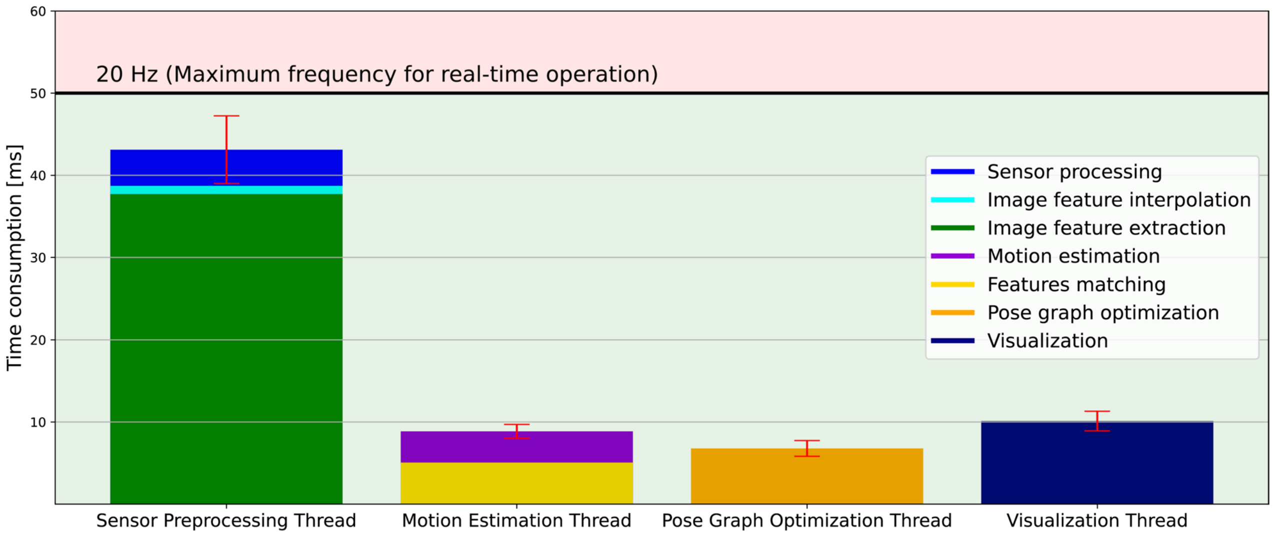

FIGURE 15 Average time consumption of threads in the proposed method. The red bar indicates the standard deviation. The distinction between online (green region) and offline (red region) execution is based on a sensor frequency of 20 Hz. GNSS, Global Navigation Satellite System; IMU, Inertial Measurement Unit. [Color figure can be viewed at wileyonlinelibrary.com]

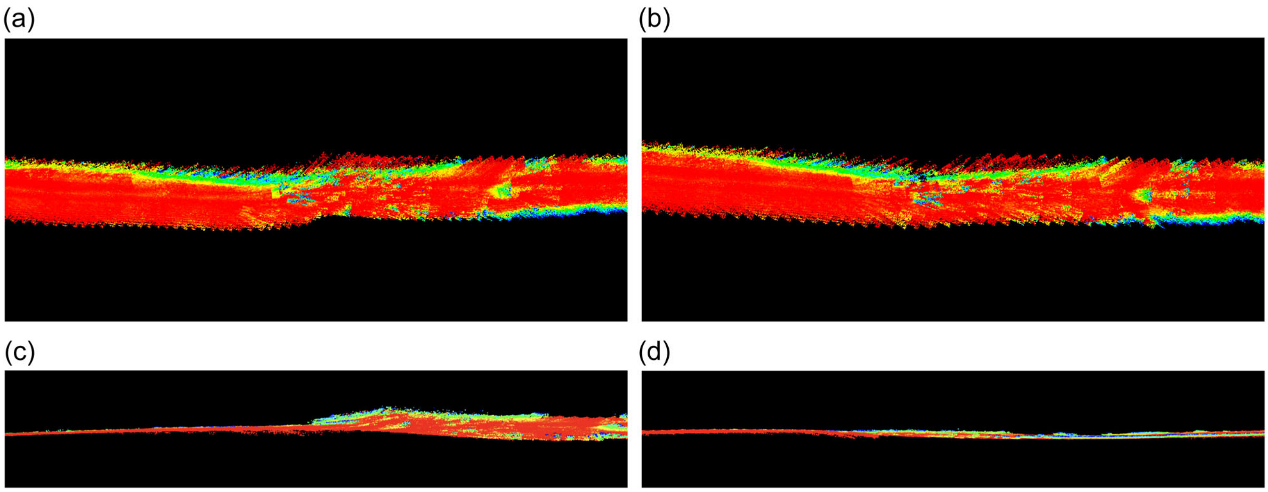

FIGURE 16 Comparison of the mapping performance in a dusty environment between the proposed method with and without the LiDAR intensity-weighted motion estimation. (a) Baseline, top view, (b) proposed, top view, (c) baseline, side view, and (d) proposed, side view. LiDAR, light detection and ranging. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Average processing time and relative trajectory error between the ground truth and the trajectory of different methods in sequence 5.

Method	Processing time (ms)		RTE (m)	
	Filtering	Opti.	Maximum	RMSE
Baseline	–	8.12	9.10	5.84
Baseline + SOR	50.4	5.34	8.91	4.44
Baseline + LIOR	28.6	6.61	8.70	6.56
Proposed	–	8.21	**6.06**	**3.00**

Note: The bold values indicate the best performance in the comparison.

Abbreviations: Filtering, filtering process; LIOR, low-intensity outlier removal; Opti., optimization process; RMSE, the root-mean-square error; RTE, relative trajectory error; SOR, statistical outlier removal.

LiDAR degenerate environments. Furthermore, TFB odometry incorporates a LiDAR intensity-weighted method, effectively countering the impact of airborne dust-induced LiDAR degeneracy. Additionally, to efficiently fuse the GNSS positioning data with estimated pose, PGO is employed.

Field experiments on an unpaved road surface were conducted to evaluate the performance of TFB odometry. The experimental results demonstrated several key advantages of the proposed method over the compared methods. First, TFB odometry demonstrated successful full mapping for all tested sequences, whereas the compared method failed in specific or all sequences due to LiDAR degradation or visual initialization failure. In addition, the proposed method exhibited higher accuracy with lower RTE values than that of the compared method. Second, TFB odometry outperformed the GNSS/IMU mapping method in accurately mapping the prepared anomalies and achieving better connectivity of the mapped road surface. Third, TFB odometry achieved precise mapping even in dusty environments compared to filtering-based methods, by employing the LiDAR intensity-weighted motion estimation. Lastly, TFB odometry ensured online mapping with all of its threads having a cycle time of less than 50 ms.

Nevertheless, certain challenges within TFB odometry still require attention. One notable limitation pertains to fast rotations, as the current method solely depends on LiDAR and camera inputs for motion estimation, without utilizing IMU data. Another limitation is the accumulated drift, particularly during prolonged localization processes.

To address these limitations, we plan to integrate IMU data into our pose graph based on IMU preintegration (Forster et al., 2016) and introduce loop closure based on bag-of-words (Gálvez-López & Tardos, 2012; Rublee et al., 2011) as our future work.

Lastly, we believe that our proposed mapping system can function as a dynamic monitoring system for unpaved roads or harsh environments. Through the dynamic monitoring of road anomalies, we envision the potential for pinpoint maintenance or automation of road maintenance.

ACKNOWLEDGMENTS

This work was carried out using the internal funding from the Komatsu MIRAI Construction Equipment Cooperative Research Center and did not involve any external funding. We also received special support from Komatsu Ltd. for conducting the field experiments.

ORCID

Junwoon Lee http://orcid.org/0000-0002-8260-7310

REFERENCES

Afzalaghaeinaeini, A., Seo, J., Lee, D. & Lee, H. (2022) Design of dust-filtering algorithms for lidar sensors using intensity and range information in off-road vehicles. *Sensors*, 22(11), 4051.

Ahmad, N., Ghazilla, R.A.R., Khairi, N.M. & Kasi, V. (2013) Reviews on various inertial measurement unit (IMU) sensor applications. *International Journal of Signal Processing Systems*, 1(2), 256–262.

Aldibaja, M., Suganuma, N., Yoneda, K. & Yanase, R. (2022) Challenging environments for precise mapping using GNSS/INS-RTK systems: reasons and analysis. *Remote Sensing*, 14(16), 4058.

Azevedo, V.B., De Souza, A.F., Veronese, L.P., Badue, C. & Berger, M. (2013) Real-time road surface mapping using stereo matching, v-disparity and machine learning. In: *The 2013 International Joint Conference on Neural Networks (IJCNN), Dallas, TX, USA*, IEEE. pp. 1–8.

Bentley, J.L. (1975) Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9), 509–517.

Besl, P. & McKay, N.D. (1992) A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2), 239–256.

Biber, P. & Straßer, W. (2003) The normal distributions transform: a new approach to laser scan matching. In: *2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA*, vol. 3. IEEE. pp. 2743–2748.

Brown, L.G. (1992) A survey of image registration techniques. *ACM Computing Surveys (CSUR)*, 24(4), 325–376.

Burningham, S. & Stankevich, N. (2005) Why road maintenance is important and how to get it done. *Transport Notes*, TRN-4, 1–10.

Campos, C., Elvira, R., Rodríguez, J.J.G., Montiel, J.M. & Tardós, J.D. (2021) ORB-SLAM3: an accurate open-source library for visual, visual-inertial, and multimap SLAM. *IEEE Transactions on Robotics*, 37(6), 1874–1890.

DeTone, D., Malisiewicz, T. & Rabinovich, A. (2018) Superpoint: self-supervised interest point detection and description. In: *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA*, IEEE. pp. 224–236.

El-Wakeel, A.S., Li, J., Rahman, M.T., Noureldin, A. & Hassanein, H.S. (2017) Monitoring road surface anomalies towards dynamic road mapping for future smart cities. In: *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Montreal, QC, Canada*, IEEE. pp. 828–832.

Engelhard, N., Endres, F., Hess, J., Sturm, J. & Burgard, W. (2011) Real-time 3D visual SLAM with a hand-held RGB-D camera. In: *RGB-D Workshop on 3D Perception in Robotics at the European Robotics Forum, Vasteras, Sweden*, vol. 180. pp. 1–15.

Fernández, A., Diez, J., de Castro, D., Silva, P.F., Colomina, I., Dovis, F. et al. (2010) ATENEA: advanced techniques for deeply integrated GNSS/INS/LiDAR navigation. In: *2010 5th ESA Workshop on Satellite Navigation Technologies and European Workshop on GNSS Signals and Signal Processing (NAVITEC), Noordwijk, Netherlands*, IEEE. pp. 1–8.

Fischler, M.A. & Bolles, R.C. (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), 381–395.

Forster, C., Carlone, L., Dellaert, F. & Scaramuzza, D. (2016) On-manifold preintegration for real-time visual-inertial odometry. *IEEE Transactions on Robotics*, 33(1), 1–21.

Gálvez-López, D. & Tardos, J.D. (2012) Bags of binary words for fast place recognition in image sequences. *IEEE Transactions on Robotics*, 28(5), 1188–1197.

Goodin, C., Durst, P.J., Prevost, Z.T. & Compton, P.J. (2013) A probabilistic model for simulating the effect of airborne dust on ground-based lidar. In: *Active and Passive Signatures IV*, vol. 8734. Baltimore, MD, USA: SPIE. pp. 83–90.

Harris, C. & Stephens, M. (1988) A combined corner and edge detector. In: *Alvey Vision Conference, Manchester, UK*, vol. 15. Alvety Vision Club. pp. 10–5244.

Kaess, M., Johannsson, H., Roberts, R., Ila, V., Leonard, J.J. & Dellaert, F. (2012) iSAM2: incremental smoothing and mapping using the Bayes tree. *The International Journal of Robotics Research*, 31(2), 216–235.

Lowe, D.G. (2004) Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60, 91–110.

Marquardt, D.W. (1963) An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2), 431–441.

Moore, T. & Stouch, D. (2016) A generalized extended Kalman filter implementation for the robot operating system. In: *Intelligent Autonomous Systems*, vol. 13. Shanghai, Chian: Springer. pp. 335–348.

Mur-Artal, R., Montiel, J.M.M. & Tardos, J.D. (2015) ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE Transactions on Robotics*, 31(5), 1147–1163.

Mur-Artal, R. & Tardos, J.D. (2017) ORB-SLAM2: an open-source SLAM system for monocular, stereo, and RGB-D cameras. *IEEE Transactions on Robotics*, 33(5), 1255–1262.

Park, J.-I., Park, J. & Kim, K.-S. (2020) Fast and accurate desnowing algorithm for lidar point clouds. *IEEE Access*, 8, 160202–160212.

Qin, C., Ye, H., Pranata, C.E., Han, J., Zhang, S. & Liu, M. (2020) LINS: a lidar-inertial state estimator for robust and efficient navigation. In: *2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France*, IEEE. pp. 8899–8906.

Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J. et al. (2009) ROS: an open-source robot operating system. In: *2009 IEEE International Conference on Robotics and Automation Workshop on Open Source Software*, *Kobe*, *Japan*, vol. 3. IEEE, p. 5.

Ren, R., Fu, H., Xue, H., Li, X., Hu, X. & Wu, M. (2021) Lidar-based robust localization for field autonomous vehicles in off-road environments. *Journal of Field Robotics*, 38(8), 1059–1077.

Reza, A.M. (2004) Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement. *Journal of VLSI Signal Processing Systems for Signal, Image and Video Technology*, 38(1), 35–44.

Rosten, E. & Drummond, T. (2006) Machine learning for high-speed corner detection. In: *Computer Vision—ECCV 2006: 9th European Conference on Computer Vision*, *Graz*, *Austria*, Springer. pp. 430–443.

Rublee, E., Rabaud, V., Konolige, K. & Bradski, G. (2011) ORB: an efficient alternative to sift or surf. In: *2011 International Conference on Computer Vision (ICCV)*, *Barcelona*, *Spain*, IEEE. pp. 2564–2571.

Rusinkiewicz, S. & Levoy, M. (2001) Efficient variants of the ICP algorithm. In: *Proceedings of the Third International Conference on 3-D Digital Imaging and Modeling*, *Quebec City*, *QC*, *Canada*, IEEE. pp. 145–152.

Rusu, R.B. & Cousins, S. (2011) 3D is here: point cloud library (PCL). In: *2011 IEEE International Conference on Robotics and Automation (ICRA)*, *Shanghai*, *China*, IEEE. pp. 1–4.

Saeed, N., Dougherty, M., Nyberg, R.G., Rebreyend, P. & Jomaa, D. (2020) A review of intelligent methods for unpaved roads condition assessment. In: *2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, *Kristiansand*, *Norway*, IEEE. pp. 79–84.

Segal, A., Haehnel, D. & Thrun, S. (2009) Generalized-ICP. In: *Robotics: science and systems*. Seattle, WA, USA: MIT Press, vol. 2. p. 435.

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)*. *Las Vegas*, *NV*, *USA*, IEEE. pp. 5135–5142.

Shan, T., Englot, B., Ratti, C. & Rus, D. (2021) LVI-SAM: tightly-coupled lidar-visual-inertial odometry via smoothing and mapping. In: *2021 IEEE International Conference on Robotics and Automation (ICRA)*, *Xi'an*, *China*, IEEE. pp. 5692–5698.

Xu, W., Cai, Y., He, D., Lin, J. & Zhang, F. (2022) FAST-LIO2: fast direct LiDAR-inertial odometry. *IEEE Transactions on Robotics*, 38(4), 2053–2073.

Xu, W. & Zhang, F. (2021) FAST-LIO: a fast, robust LiDAR-inertial odometry package by tightly-coupled iterated Kalman filter. *IEEE Robotics and Automation Letters*, 6(2), 3317–3324.

Zhang, C. (2008) An UAV-based photogrammetric mapping system for road condition assessment. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37, 627–632.

Zhang, J., Kaess, M. & Singh, S. (2014) Real-time depth enhanced monocular odometry. In: *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, *Chicago*, *IL*, *USA*, IEEE. pp. 4973–4980.

Zhang, J. & Singh, S. (2014) LOAM: lidar odometry and mapping in real-time. In: *Robotics: science and systems*, Berkeley, CA, USA. vol. 2. pp. 1–9.

Zheng, C., Zhu, Q., Xu, W., Liu, X., Guo, Q. & Zhang, F. (2022) FAST-LIVO: fast and tightly-coupled sparse-direct LiDAR-inertial-visual odometry. In: *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, *Kyoto*, *Japan*, IEEE. pp. 4003–4009.

How to cite this article: Lee, J., Kurisu, M. & Kuriyama, K. (2024) Three-dimensionalized feature-based LiDAR-visual odometry for online mapping of unpaved road surfaces. *Journal of Field Robotics*, 1–17. https://doi.org/10.1002/rob.22334