2025 IEEE/SICE International Symposium on System Integration (SII)
January 21-24, 2025, Munich, Germany

Visual-Lidar Odometry with Orientation Correction towards Millimeter-Level Localization for Indoor Construction Robots

Hiroki Hisatsugu[1], Hisahiro Koshizuka[2], Moritoyo Hori[3] and Kenjiro Yamamoto[4]

***Abstract*— We propose a high-precision odometry method that fuses visual-lidar odometry with orientation information estimated by ceiling line features, aiming to achieve millimeter-level localization required at construction sites. Visual-Lidar Odometry and Mapping has demonstrated top-class accuracy with an error of 50 mm in benchmarks. However, when evaluated in environments with large open walls, such as indoor construction sites with few shape and image features, the method was found to be less accurate primarily due to errors in the orientation estimation. To address this issue, we propose a method that combines the use of ceiling line features to estimate the robot's orientation angle, thereby improving the accuracy of the odometry position estimates. We evaluated the accuracy using a real robot in an environment similar to a construction site and achieved an average error of 6.2 mm (straight-line route), 7.6 mm (rectangular route), surpassing the accuracy of conventional methods. These results confirm the effectiveness of utilizing orientation correction from ceiling line features.**

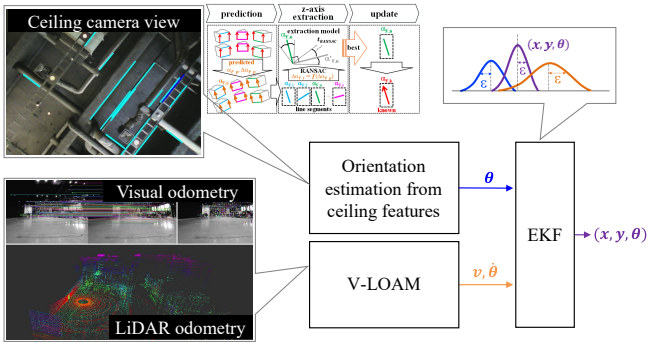


Fig. 1: Outline of the proposed method. We propose an approach that uses an extended Kalman filter to fuse V-LOAM with orientation estimation using ceiling line features.

I. INTRODUCTION

Autonomous Mobile Robots (AMR) are becoming increasingly popular and are used in many applications such as logistics, in-warehouse transportation, and in-store catering robots. AMRs utilize SLAM technology with LiDAR[1][2] and cameras[3], requiring centimeter-level high-precision localization to accomplish their tasks. In particular, indoor construction robots require localization accuracy at the millimeter-level for accurate equipment installation.

Achieving highly accurate localization in wide area environments with few shape features and partial exposure to the outdoors, such as in-building construction sites, is challenging due to the limited number of extractable and reliable features from the environment.

The in-building construction site environment lacks both interior and exterior walls, exposing structural elements such as pillars, ceiling beams, piping, and flooring materials. Additionally, the brightness of the environment constantly changes due to external light. At construction sites, it is difficult to obtain high-precision maps in advance due to the drastic changes in the environment.

Assuming such an environment, we investigated methods to achieve highly accurate localization using both a camera and LiDAR mounted on the robot. Previous research has used Visual SLAM[4][5], LiDAR SLAM[6][7], and IMU fusion[8]. However, high-precision IMUs, which are needed to reduce the effects of drift, are expensive. In the case of visual-only methods, there is concern that the accuracy will vary depending on the environment due to differences in the quality of feature points[9]. In scenes with few image feature points and a large proportion of low-quality features, as targeted in this paper, higher accuracy can be achieved by using LiDAR SLAM, which relies on highly reliable shape features such as pillars. This complements Visual SLAM, which depends on image features.

Lidar Odometry and Mapping in Real-time (LOAM)[6] is a top-performing LiDAR-based SLAM algorithm, capable of correcting point cloud distortion and estimating positions in real-time. This has been further extended by Visual-Lidar Odometry and Mapping (V-LOAM)[7], which uses monocular image information integrated with LiDAR processing and is robust to rapid motion. These methods have achieved top-class accuracy of 50 mm in the KITTI dataset benchmark[10].

The performance evaluation in an environment similar to a construction site using high-precision LiDAR-based SLAM, V-LOAM revealed that in environments with few image and shape feature points, the quantity and quality of features that can be tracked by image and LiDAR odometry is limited, which increases the localization error. Error analysis showed that the error in orientation estimation associated with the robot's turning was the dominant factor among the localization errors.

To improve orientation accuracy, this paper proposes the

[1]Hiroki Hisatsugu and [4]Kenjiro Yamamoto are with the Robotics Research Department, Connective Automation Innovation Center, Center for Sustainability, Hitachi, Ltd., Japan
[2]Hisahiro Koshizuka is with the Sales Department, Channel Business Promotion Center, Hitachi Channel Solutions, Corp., Japan
[3]Moritoyo Hori is with the Products Development Department, Channel Business Promotion Center, Hitachi Channel Solutions, Corp., Japan
[1] hiroki.hisatsugu.dn@hitachi.com
[2] hisahiro_koshizuka2@hitachi-ch.com
[3] moritoyo_hori@hitachi-ch.com
[4] kenjiro.yamamoto.bq@hitachi.com

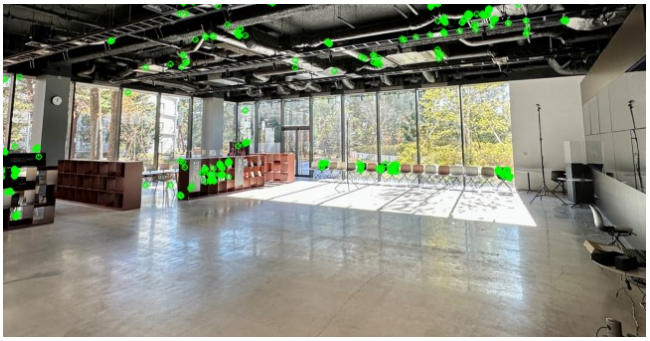
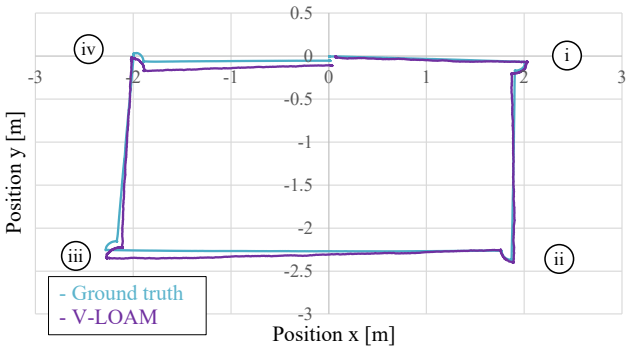
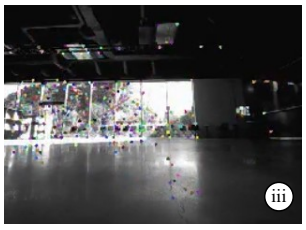


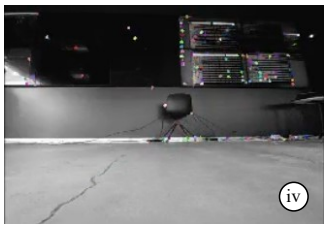
Fig. 2: An example of environment similar to the target. The extraction of image feature points indicated by the color green from scene images reveals a paucity of stable feature points.



(a) Localization result



(b) Image feature points are only outside the window



(c) Wall surfaces have few feature points

Fig. 3: V-LOAM localization results and front camera images of rectangular route experiments.

method shown in Fig. 1. We address the orientation error problem by using a camera to capture images of the ceiling, based on the assumption that structural straight-line elements are arranged in a fixed pattern. This allows the estimation of orientation using the features of ceiling pipes and beams, thereby supplementing V-LOAM. While there is prior work on SLAM using only the view of the ceiling[11][12], our approach increases accuracy by combining orientation using the view of the ceiling and translational movement using the horizontal view, compensating for each other's shortcomings.

The contributions of this paper are as follows:

- By focusing on the assumption that the structures have a pattern in their arrangement, we proposed a method to fuse V-LOAM with orientation estimation using line recognition of ceiling piping and beam features.
- The proposed method was compared with conventional methods using an actual robot, demonstrating superior accuracy.

II. Existing Techniques and Issues

A. V-LOAM under low stable feature environment

V-LOAM is based on high-precision LiDAR odometry operating at a 1 Hz cycle. It is complemented by high-speed (60 Hz) image odometry, which provides robustness against sudden movements despite being less accurate. However, in environments with few stable features such as construction sites that require millimeter-level accuracy, there is a challenge of rotation and translation errors.

As an example of the environment targeted in this paper, an environment similar to a construction site scene is shown in Fig. 2. The extraction of Oriented FAST and Rotated BRIEF (ORB) image feature points from scene images indicated by the green points in Fig. 2 reveals a paucity of stable feature points. Most of the feature points identified are unstable, changing their positions due to bright ambient light or changes in viewpoint, rather than clear positions of nearby floors or edges of pillars.

In this paper, the accuracy of V-LOAM was evaluated in an environment shown in Fig. 2 and the localization result is shown in Fig. 3. The average error between the ground truth and the estimated position was 39 mm (maximum error 112 mm), and the average orientation error was 1.84 degrees. Analysis of the error trends showed that while the error in translation was small, the error in orientation estimation after turning corners was significant. As the error in orientation estimation increases, the error in the position along the route also increases due to error accumulation in odometry. Therefore, orientation estimation accuracy is crucial.

Possible error factors during turning include the following: When the robot turns, the feature points projected on the camera plane move largely in a horizontal direction, increasing the apparent speed of feature point movement. Since the movement per frame is large, blur is likely to occur, leading to feature point mismatching, especially in scenes with few stable feature points. To address this issue, a promising approach is to track feature points on nearby ceilings that exhibit no large motion from the perspective of the sensor. The hypotheses regarding error factors are discussed later in the discussion section.

B. Robot orientation estimation from ceiling line features based on Stata Center Frame

Consider the use of ceiling information as a feature that can be effectively utilized for location estimation indoors at construction sites.

Indoor artificial structures such as walls, ceilings, plumbing, and lighting are typically arranged in a regular pattern or orthogonally. Based on this world assumption, we apply the Stata Center Frame (SCF) Localization[12], which uses the linear components of structures and lighting landmarks to estimate the position of a robot moving indoors on

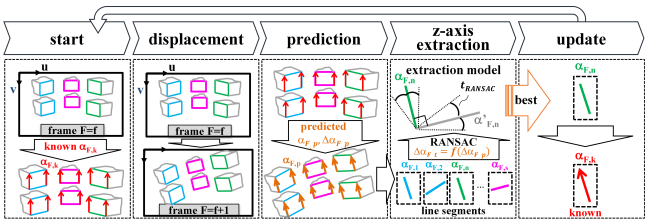


Fig. 4: Flow of orientation estimation based on Stata Center Frame (SCF)[12].

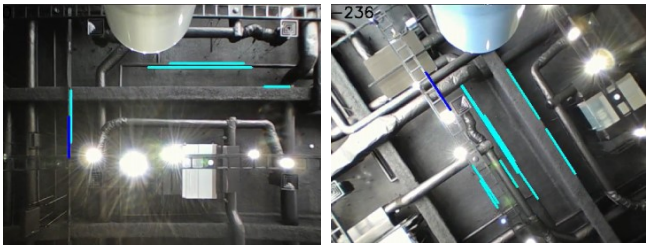


Fig. 5: Example of orientation estimation based on SCF. Light blue lines are the extracted straight lines. Dark blue line is the estimated orientation line.

a plane. SCF-Localization is more accurate in orientation estimation than ceiling feature point tracking by V-LOAM because it performs line based orientation angle estimation. The orientation estimation logic is described in Fig. 4 and follows.

1) **Start:** Begin from the initial state where the robot's absolute orientation is known.
2) **Displacement:** The robot moves and captures a new image of the ceiling.
3) **Prediction:** Extract the edges of the captured image where there are significant changes in brightness, and from these edges, extract multiple straight-lines (light blue lines in Fig. 5).
4) **Z-axis extraction:** Randomly sample a certain number of straight-lines, calculate the directional variation using the sampled lines (including orthogonal line segments), select the sample with the least angular variation, and extract the direction using RANSAC processing[13] (dark blue line in Fig. 5). In this process, outliers with straight edges other than the ceiling are also removed.
5) **Update:** Calculate the robot's orientation after removing noise from the extracted orientation using a time series filter. These steps are used to continuously estimate the orientation of a robot moving in space.

The average orientation angle error for this method is 0.55 deg, which is high accuracy compared to the accuracy of Visual SLAM and LiDAR SLAM when turning. The challenge with this method is that although the orientation estimation is highly accurate, the translation error is large, and the realized error of localization is about 120 mm[12].

TABLE I: Comparison of rotation and translation errors of V-LOAM and SCF-Localization when running a rectangular route

Method	Rotation error	Translation error
V-LOAM	△ (1.84 deg)	✓ (39 mm)
SCF-Localization	✓ (0.55 deg)	△ (120 mm)

III. Proposed Method

The relationship between rotation and translation error based on prior method comparisons is shown in Tab. I. V-LOAM uses LiDAR odometry, which provides high translational accuracy but struggles with rotational accuracy. Conversely, SCF-Localization offers high rotational accuracy for turning but low translational accuracy. Therefore, this paper proposes a method to improve position estimation accuracy by correcting V-LOAM results using orientation information obtained from ceiling line features.

Specifically, the image information from a monocular camera mounted upward on the robot detects ceiling patterns, piping, and beam features as lines, and estimates the change in orientation between frames. By fusing this orientation information with the translation information from V-LOAM, the limitations of both methods are compensated, achieving highly accurate localization.

To correct the position with highly accurate orientation information, we use an extended Kalman filter (EKF)[14], a type of sequential Bayesian filter. The EKF operates in three main steps: prediction, observation, and update.

The state vector is defined as:

$$\boldsymbol{x} = \begin{bmatrix} x & y & \theta & v & \dot{\theta} \end{bmatrix}^\top \tag{1}$$

where x and y represent the robot's position, θ is the orientation angle, v is the linear velocity, and $\dot{\theta}$ is the angular velocity.

The state transition model predicts the robot's state based on its current state. The actual state transition with process noise is expressed as:

$$\boldsymbol{x}_k = \boldsymbol{f}(\boldsymbol{x}_{k-1}) + \boldsymbol{w}_{k-1} \tag{2}$$

where

$$\boldsymbol{f}(\boldsymbol{x}) = \begin{bmatrix} x + v\cos(\theta)\Delta t \\ y + v\sin(\theta)\Delta t \\ \theta + \dot{\theta}\Delta t \\ v \\ \dot{\theta} \end{bmatrix} \tag{3}$$

and $\boldsymbol{w}_{k-1} \sim \mathcal{N}(\boldsymbol{0}, \mathbf{Q})$ represents process noise with covariance $\mathbf{Q}$.

The Jacobian of $\boldsymbol{f}(\boldsymbol{x})$ with respect to $\boldsymbol{x}$ is:

$$\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}} = \begin{bmatrix} 1 & 0 & -v\sin(\theta)\Delta t & \cos(\theta)\Delta t & 0 \\ 0 & 1 & v\cos(\theta)\Delta t & \sin(\theta)\Delta t & 0 \\ 0 & 0 & 1 & 0 & \Delta t \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{4}$$

In the observation step, we perform two types of sensor observations, which we denote using subscripts 1 and 2. The observation models are described below.

a) Orientation Angle (θ): The orientation angle is estimated using SCF. The observation function is:

$$\boldsymbol{h}_1(\boldsymbol{x}_k) = \begin{bmatrix}\theta_k\end{bmatrix} \tag{5}$$

The actual observation is:

$$\boldsymbol{z}_{1,k} = \boldsymbol{h}_1(\boldsymbol{x}_k) + \boldsymbol{v}_{1,k} \tag{6}$$

where $\boldsymbol{v}_{1,k} \sim \mathcal{N}(\boldsymbol{0}, \mathbf{R}_1)$. The Jacobian of $\boldsymbol{h}_1(\boldsymbol{x}_k)$ is:

$$\frac{\partial \boldsymbol{h}_1(\boldsymbol{x}_k)}{\partial \boldsymbol{x}_k} = \begin{bmatrix}0 & 0 & 1 & 0 & 0\end{bmatrix} \tag{7}$$

b) Linear and Angular Velocities ($v, \dot{\theta}$): The linear and angular velocities are obtained from V-LOAM. The observation function is:

$$\boldsymbol{h}_2(\boldsymbol{x}_k) = \begin{bmatrix}v_k \\ \dot{\theta}_k\end{bmatrix} \tag{8}$$

The actual observation is:

$$\boldsymbol{z}_{2,k} = \boldsymbol{h}_2(\boldsymbol{x}_k) + \boldsymbol{v}_{2,k} \tag{9}$$

where $\boldsymbol{v}_{2,k} \sim \mathcal{N}(\boldsymbol{0}, \mathbf{R}_2)$. The Jacobian of $\boldsymbol{h}_2(\boldsymbol{x}_k)$ is:

$$\frac{\partial \boldsymbol{h}_2(\boldsymbol{x}_k)}{\partial \boldsymbol{x}_k} = \begin{bmatrix}0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1\end{bmatrix} \tag{10}$$

The update step follows the standard EKF framework as described in [14], where the Kalman gain is computed, and the state vector and covariance are updated probabilistically based on the observations.

IV. Experiments

To evaluate the proposed technology, an experimental data set was created in a wide area indoor environment with few stable feature points, such as a construction site. The dataset included the robot's ground truth (position and orientation) acquired by a motion capture system, and simultaneously recorded information from 3D LiDAR, a front-facing wide-angle camera, and a ceiling-facing wide-angle camera. The motion capture system is capable of obtaining ground truth values with a repeatability of 0.2 mm, enabling the evaluation of positional errors at the millimeter-level.

Fig. 6 shows the prototype robot used for the experiments and dataset acquisition. The robot base is a differential two-wheeled rover (V-stone Megarover 3.0), equipped with a LiDAR (Velodyne HDL-32E) necessary for V-LOAM, a forward-facing monocular camera (Buffalo BSW200MBK), and an upward-facing monocular camera (Buffalo BSW200MBK) used for orientation estimation using the ceiling. The camera has a field of view (FOV) of 120 degrees and a resolution of 640×480 pixels.

Sensor data were recorded by the front camera at 30 Hz, the 3D LiDAR at 10 Hz, and the ceiling camera at 10 Hz. The internal parameters of the cameras and the external parameters between the cameras and LiDAR were calibrated and known. Spherical retroreflective markers were mounted

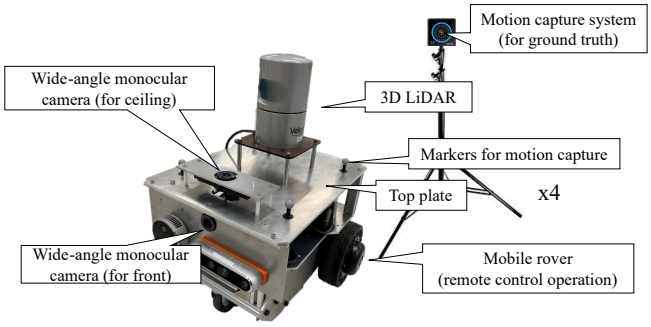


Fig. 6: Experimental robot with sensors.

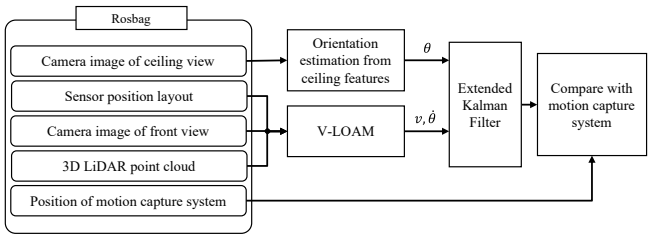


Fig. 7: System diagrams for evaluation.

at four locations on the top plate to measure the ground truth of the robot's position and orientation using the Optitrack PrimeX13 motion capture system. The robot's position and orientation information, measured by the motion capture system, is received by the ROS program on the robot through the Wi-Fi network. This data is recorded in a rosbag along with information from the onboard sensors.

For the accuracy evaluation of the algorithm, the recorded sensor data is input into the localization process using V-LOAM[15] and orientation estimation by ceiling, and the estimated state of the robot is compared with the ground truth. Fig. 7 shows the diagram of the evaluation system. The error evaluation between the ground truth and the estimated position is calculated using the running trajectory. Since the ground truth and the estimated position have different sampling rates, the Euclidean distance of the estimated position from the nearest ground truth value is evaluated as the error. Orientation errors are compared over time series. Since the ground truth value is affected by communication delay, in the region where the traverse speed is fast, the calculated orientation error value is estimated to be larger than the actual error. Therefore, the error in this region is ignored as a reference value. The robot moves at a maximum translational speed of 0.1 m/s and a turning speed of 0.5 rad/s, and is manually operated by remote control.

For millimeter-level accurate robot localization at construction sites, an approach in which the robot communicates with a laser measurement instrument fixed to the ground to measure its absolute position is usually used. However, the robot is required to estimate the millimeter-level position even in situations where the laser measurement instrument's optical beam is temporarily occluded by a pillar or other

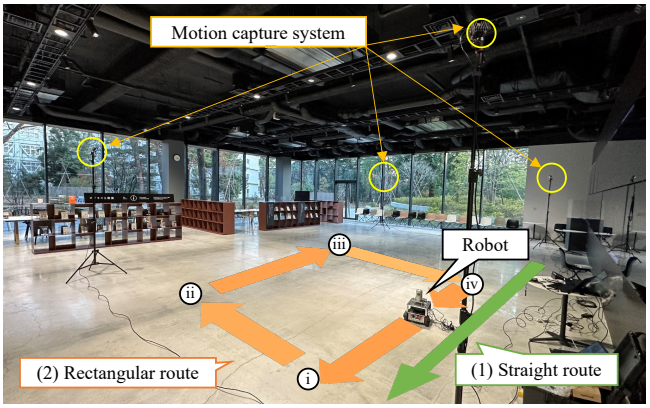


Fig. 8: Experimental route and setup for capturing ground truth data in an experimental environment.

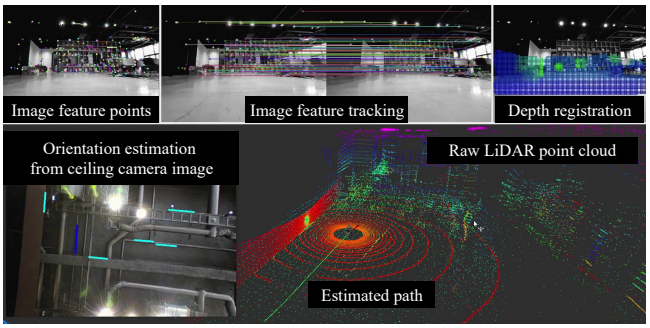


Fig. 9: Processed sensor data during the experiment.

obstruction. Therefore, in this experiment, we confirm that the proposed method of odometry can estimate the position with millimeter-level accuracy even when the robot moves about 10 m from a known initial position. The following two types of route patterns were used to compare cases involving straight route and turning: (1) Straight 10 m route and (2) Rectangular (2×4 m) circumferential route.

We evaluated the position errors of V-LOAM and our proposed method in the test routes (1) and (2) in the environment shown in Fig. 8. The experimental environment was a 10×10 m indoor area with daylight coming in from the outside, with few shape features and partly open to the outdoors, with an average brightness of 600 lumens. The ceiling has exposed pipes and beams.

V. Results

A. Straight (10 m)

Fig. 9 shows the processed sensor data during the experiment. The display screen presents, from left to right in the upper row, the processing results of extracted image feature points, frame-to-frame motion estimation by tracking feature points, and the results of depth map registration. The bottom row on the left shows the straight-line extraction for orientation estimation by the ceiling. The bottom row on the right shows the raw LiDAR point clouds and estimated robot's position.

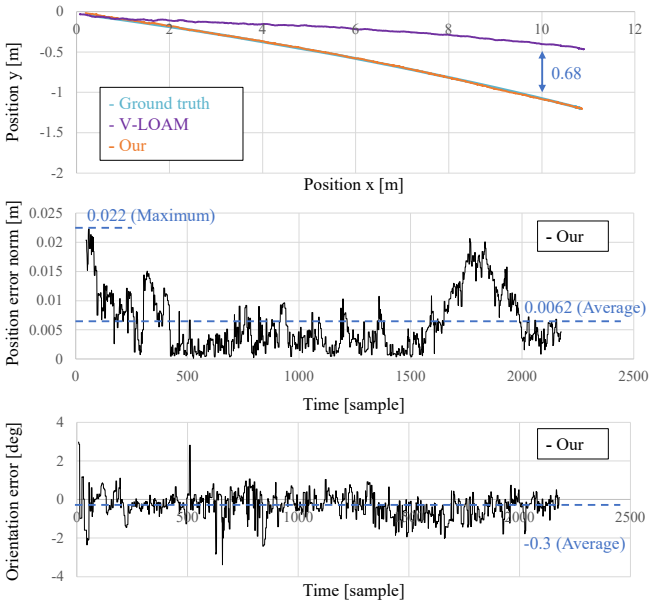


Fig. 10: Result of the 10 m straight route.

Fig. 10 shows the experimental results. The graph on the top illustrates the robot's trajectory. The ground truth (light blue) and the results of our method (orange) are almost identical, whereas the V-LOAM results (purple) show a deviation of 0.68 m in the y-axis direction at the 10 m mark. The time series graph in Fig. 10 displays the error (difference from the ground truth) of our method.

The middle graph shows the time series plot of the position error, and the bottom graph shows the time series plot of the orientation error. From the graphs, the errors of the estimated position after fusion compared to the ground truth are: mean error of 6.2 mm (maximum error of 22 mm) and mean orientation error of -0.3 degrees.

B. Rectangle (2×4 m)

In order to evaluate positional accuracy in moves involving turnings, experiments were conducted on rectangular-shaped routes. The results are shown in Fig. 11. The graph on the top illustrates the course, which shows the robot's trajectory when it started from near the origin of the coordinates and followed the right-hand course around the rectangle manually. The light blue line represents the ground truth obtained by motion capture, the purple line indicates the result from V-LOAM, and the orange line shows the result of our method. The time series graph shows the error of our fusion method; the middle graph shows the time series plot of the position error, and the bottom graph shows the time series plot of the orientation error. As described above, the part surrounded by yellow color in Fig. 11 is a turning region where the estimated error is larger than the actual error, so it is excluded from the calculation of the average.

From the graphs, the errors of the estimated position of our fusion method compared to the ground truth were: average error of 7.6 mm (maximum error of 25 mm) and average orientation error of 0.97 degrees. The results for V-LOAM

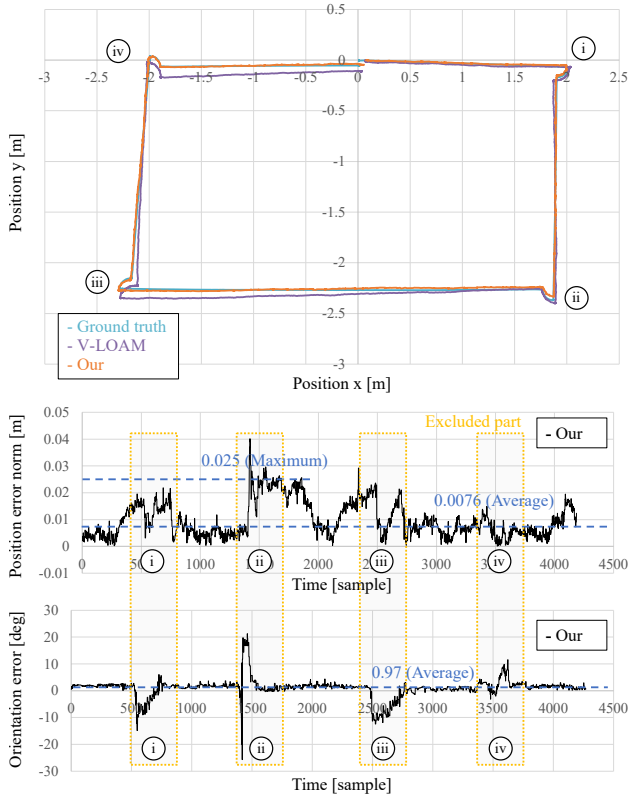


Fig. 11: Result of the rectangular route.

TABLE II: Average error results of two types of tests.

Test type	Av. error [mm]	Av. error [deg]
(1) Straight OUR	6.2	-0.3
(1) Straight V-LOAM	346	1.02
(2) Rectangle OUR	7.6	0.97
(2) Rectangle V-LOAM	39	1.84

showed an average error of 39 mm (maximum error of 112 mm) and an average orientation error of 1.84 degrees.

VI. Discussions

The results are summarized in Tab. II. In both types of experiments, the average and maximum position errors of our method were less than 50 mm. The accuracy was improved by applying our fusion method compared to the estimation results using V-LOAM (average error: 39 mm to 7.6 mm, maximum error: 112 mm to 25 mm). This confirms the usefulness of the method utilizing ceiling line information. This method's advantage is attributed to the ceiling line information, which does not accumulate errors, effectively mitigating the accumulation of orientation errors, thereby significantly enhancing accuracy.

The results of the straight route in scene (1) demonstrated that the estimated position trajectory was closer to the ground truth than the V-LOAM estimation results. By integrating the method proposed in this paper, which estimates the absolute orientation during motion from ceiling images, the orientation error can be reduced, and the estimated position can be made closer to the ground truth.

Fig. 3 shows the image input from the front camera at the turning point on the route in scene (2). The extracted features are marked in the image. In this experiment, the error increased significantly at locations (iii) and (iv). The reasons are explained below. Near the window at position (iii), there are few stable feature points except those outside the window. Additionally, at position (iv), a uniform wall surface is captured, indicating a scarcity of feature points on the wall surface. Since V-LOAM fuses image odometry and LiDAR odometry, it is susceptible to large positional estimation errors in image odometry.

The following hypotheses can be made regarding error factors of V-LOAM in the environments with few stable features. In environments that are monotonous in color and shape and have few stable features, where the retrievable image and shape features are located far away, the positioning accuracy of the feature points will be poor and unstable. When the robot turns, the feature points projected on the camera plane move in a horizontal direction, increasing the apparent speed of feature point movement. As a result, a large movement per frame may cause distortion and blurring if the shutter speed is insufficient, leading to feature point mismatching. Tracking becomes more difficult at higher rotation speeds, and if the frame rate is insufficient, the possibility of losing corresponding feature points from the camera's FOV increases. V-LOAM's evaluation[7] noted that tracking is more likely to be lost when turning with camera configurations that have a limited FOV, such as wide-angle lenses rather than fisheye lenses. In addition, exposure changes are also a barrier to feature tracking. In a rectangular scene, the brightness appears to change each time a corner is turned. Accurate orientation estimation is crucial because a larger estimation error of the orientation will result in a larger position error in the subsequent route in odometry due to error accumulation.

The results show that the trajectory of translational displacement is close to the ground truth, but the orientation differs. Thus, fusion with the method proposed in this paper, which estimates the absolute orientation during motion from ceiling images, could reduce orientation errors and bring the estimated position closer to the ground truth.

This paper verified the reproducibility of accuracy improvement through sensor fusion in two running routes in scenes with few stable features and ambient light, where V-LOAM errors are significant. The method of estimating absolute orientation during motion from ceiling images was found to contribute significantly to accuracy improvement. However, this method relies on linear recognition of ceiling beams and pipes. Unlike ceiling beams, which are generally arranged orthogonally, pipes are often inaccurately installed and may be placed at an angle.

In addition to evaluating the impact of errors due to vehicle speed and other construction site-like environments, a mechanism to exclude objects that seem to be inaccurate when extracting straight-lines from the ceiling is considered nec-

essary in the future. These improvements offer the prospect of enhanced accuracy for millimeter-level localization.

VII. Conclusion and Future Work

To improve the localization error in wide-open environments with few stable feature points, such as construction sites where millimeter-level accuracy is required, this paper proposed a high-precision odometry method that integrates orientation estimation using ceiling line information in addition to conventional V-LOAM point cloud and image feature point tracking. Evaluation experiments were conducted using an actual robot in a construction site-like environment characterized by wide areas, few shape features, and external light interference. Two running routes were tested: straight 10 m route and a rectangular 2×4 m route.

We confirmed the effectiveness of using ceiling information by achieving an average error of 6.2 mm for the straight 10 m route and an average error of 7.6 mm for the rectangular 2×4 m route. In both running routes, the accuracy of the proposed method exceeded that of the conventional V-LOAM method. In the future, we will further test the method in various scenes and investigate ways to improve accuracy by auto-tuning the parameters of the orientation estimation.

References

[1] M. U. Khan, S. A. A. Zaidi, and A. Ishtiaq, "A Comparative Survey of Lidar-SLAM and Lidar based Sensor Technologies," *2021 Mohammad Ali Jinnah University International Conference on Computing (MAJICC)*, 2021, pp. 1-8.

[2] L. Huang, "Review on Lidar-based SLAM Techniques," *2021 International Conference on Signal Processing and Machine Learning (CONF-SPML)*, 2021, pp. 163-168.

[3] S. Zhang, L. Zheng, and W. Tao, "Survey and Evaluation of RGB-D SLAM," *IEEE Access*, vol. 9, 2021, pp. 21367-21387.

[4] S. Sumikura, M. Shibuya, and K. Sakurada, "OpenVSLAM: A versatile visual SLAM framework," *27th ACM International Conference on Multimedia*, 2019, pp. 2292-2295.

[5] C. Campos et al., "ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial, and Multimap SLAM," *IEEE Trans. Robotics*, vol. 37, no. 6, Dec. 2021.

[6] J. Zhang and S. Singh, "LOAM: Lidar Odometry and Mapping in Realtime," *Robotics: Science and Systems Conference (RSS)*, July 2014, pp. 109-111.

[7] J. Zhang and S. Singh, "Visual Lidar Odometry and Mapping: Low Drift, Robust, and Fast," *IEEE International Conference on Robotics and Automation (ICRA)*, 2015, pp. 2174-2181.

[8] M. B. Alatise and G. P. Hancke, "A Review on Challenges of Autonomous Mobile Robot and Sensor Fusion Methods," *IEEE Access*, vol. 8, 2020, pp. 39830-39846.

[9] A. Merzlyakov and S. Macenski, "A Comparison of Modern General-Purpose Visual SLAM Approaches," *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 9190-9197.

[10] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *Int. J. Rob. Res.*, vol. 32, no. 11, 2013, pp. 1231-1237.

[11] Y. Wang, Y. Zhang, and J. Wang, "Ceiling-View Semi-Direct Monocular Visual Odometry with Planar Constraint," *Remote Sens.*, vol. 14, 2022, no. 5447.

[12] A. M. Kaneko and R. Ichinose, "Stata Center Frame: A Novel World Assumption for Self Localization," *IEEE International Conference on Robotics and Automation Engineering (ICRAE)*, 2019, pp. 29-38.

[13] C. C. Chou and C. C. Wang, "1-point Affine RANSAC for Scene Image Matching in Appearance-based Localization," *IEEE International Conference on Automation Science and Engineering (CASE)*, Taipei, Taiwan, 2014, pp. 1194-1199.

[14] T. Moore and D. Stouch, "A Generalized Extended Kalman Filter Implementation for the Robot Operating System," *Advances in Intelligent Systems and Computing (AISC)*, vol. 302, 2015.

[15] Y. Xia, "Implementation of V-LOAM," 2018. [Online] Available:*https://github.com/YukunXia/VLOAM-CMU-16833/blob/master/16833_report_V2.pdf*.