METHODS

Essential Error Analysis of Self-Localization From Map Structure in Urban Environment

YUKI ENDO[1], (Member, IEEE), AND SHUNSUKE KAMIJO[2], (Senior Member, IEEE)
[1]Graduate School of Information Science and Technology, The University of Tokyo, Tokyo 153-8505, Japan
[2]Interfaculty Initiative in Information Studies, The University of Tokyo, Tokyo 153-8505, Japan

Corresponding author: Yuki Endo (endou@kmj.iis.u-tokyo.ac.jp)

ABSTRACT Lane-level self-localization is a critical task in the field of autonomous driving. Map-based self-localization is commonly employed to achieve lane-level accuracy in urban settings. However, it is known that in certain locations, such as narrow roads and intersections, map-based self-localization poses challenges, leading to failures of autonomous driving tasks. Potential errors in specific locations can be estimated before driving by leveraging the geometrical structures of the environment and can then be utilized to enhance the safety of autonomous driving operations. In contrast to traditional self-localization error estimation methods, which often rely on statistical analyses or regression specific to a particular map format, this research focuses on identifying fundamental errors in self-localization within various map formats. The proposed method employs a general formulation of environmental representation, specifically normal distributions, and a closed-form uncertainty approximation of optimal solutions, enabling the identification of essential self-localization errors in an environment. The results obtained through this method are not only valuable for autonomous driving tasks but also contribute to discussions on the quality of specific map formats. Experiments highlight locations within an urban environment that are susceptible to significant errors in self-localization based on the proposed metric. Furthermore, a comparison of self-localization errors associated with specific map formats using the proposed metric reveals that essential errors in algorithms can be estimated. The discussion presented in this paper reveals the patterns of geometrical features in the urban environment that are likely to result in self-localization errors.

INDEX TERMS Autonomous driving, self-localization, high-definition maps, uncertainty analysis.

I. INTRODUCTION

Self-localization is a task to estimate a pose (i.e., translation and rotation) of the self-vehicle in an environment. Lane-level self-localization is crucial in autonomous driving systems, where it finds applications in control, scene understanding, and more [1]. Reliable self-localization with lane-level accuracy is needed for practical driving in diverse environmental conditions.

Traditional self-localization methods rely on global navigation satellite systems (GNSS) and inertial navigation systems (INS) [2], [3]. However, in urban settings, GNSS-based methods are prone to have multipath interference, reflections, and non-line of sight (NLOS) problems due to

The associate editor coordinating the review of this manuscript and approving it for publication was Heng Wang.

tall buildings, leading to significant errors [2], [4]. Similarly, INS-based methods are susceptible to cumulative errors.

For achieving lane-level accuracy in urban environments, high-definition (HD) maps are utilized in today's self-localization [5], [6], [7], [8], [9], [10]. HD maps are constructed in the target environment before driving and provide geometrical information about the environment for self-localization and other autonomous driving tasks. By leveraging HD maps, self-localization strategies can be formulated to align light detection and ranging (LiDAR) scans with the map's geometrical structures. The process of aligning two point clouds is known as point cloud registration (PCR). For example, the iterative closest point (ICP) algorithm [11] is a well-known PCR method that iteratively refines the pose by minimizing the sum of Euclidean distances between LiDAR scan points and their

corresponding points in the map point cloud. In addition to point clouds, there are many map formats available for self-localization, and the HD map formats employed have led to a variety of PCR issues in autonomous driving.

It is known that self-localization errors occur even with an HD map in real environments. The error sources of self-localization are considered map representation and essential errors in the environment. For example, in the map representation problem, point cloud maps have redundancy and noise, resulting in uncertainty of self-localization. Each map format has advantages and disadvantages, and the best map format for self-localization is still under discussion.

On the other hand, it is known that certain locations in the environment pose challenges to self-localization. Such essential errors in the environment do not depend on map formats. For example, considering a narrow street surrounded by tall buildings, self-localization will confuse the actual positions because LiDAR scans look almost the same at different locations along the longitudinal direction. Such errors are inherent characteristics of the environment and cannot be improved by any algorithm or map format. Recognizing essential errors before driving is vital for ensuring safety in autonomous driving operations. Estimation of essential errors is crucial for safe operations in autonomous driving. Because the essential errors are errors that cannot be overcome with map representation, they should be considered for safe operations in autonomous driving. Also, gaps between the actual errors on a specific map format and the essential errors can be used to discuss the map format efficiency for each location. Additionally, theoretically supported essential error analysis is necessary for convincing the discussion.

Several studies have been conducted to know possible errors in self-localization before driving, such as statistical-based [12] and factor-based [13], [14]. For instance, Akai et al. estimate uncertainty empirically by conducting numerous experiments at various locations in advance [12]. However, they focus on specific algorithms, and conducting numerous experiments is expensive in practice. Conversely, Javanmardi et al. took a different approach, utilizing geometrical map structures to estimate errors before driving [14]. They enumerated possible geometrical features that could impact self-localization accuracies (called factors) and employed regression analysis on the factors to fit actual errors in a specific map format. However, their approach is not theoretically supported, and some experiments are necessary for the regression for each environment and the map format.

This paper proposes a method to estimate essential self-localization errors in the environment. Self-localization is considered to find the best-fitted pose with a map. The essential errors can be estimated with an objective function to represent how the transformed scan and the environment fit and an uncertainty estimation of an optimum pose on the function. The concept of this evaluation is shown in Figure 1. For example, in narrow streets with tall buildings, possible scans towards the longitudinal direction look similar, resulting in similar values of the objective function and the
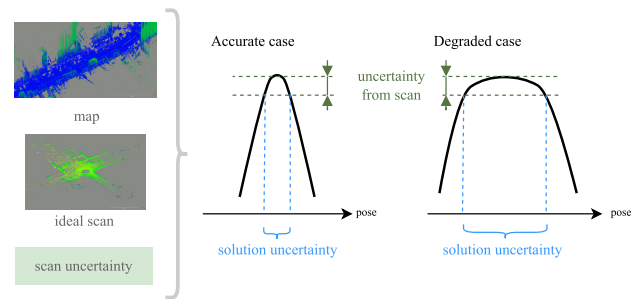
FIGURE 1. The concept to estimate optimization uncertainty from map structure.

uncertainty of the optimal pose for the longitudinal direction. This uncertainty is intrinsic to the environment and cannot be fully mitigated by any optimization algorithm. Therefore, it is clear that function shape verification is useful for evaluating self-localization.

This paper is the first study to reveal essential errors analytically and evaluation in the urban environment. As a result, the essential error patterns in an urban environment were evaluated on the proposed metric and discussed.

II. RELATED WORK
This section describes related work about map-based self-localization and evaluation techniques.

A. MAP FORMATS FOR SELF-LOCALIZATION
As the most basic map utilized in self-localization for autonomous driving, a point cloud map, comprising a set of three-dimensional points, is widely adopted. There are some algorithms for self-localization using a point cloud map. The most famous algorithm is the iterative closest point (ICP) [11]. ICP optimizes the self-vehicle pose by minimizing the sum of the Euclidean distances between map points and LiDAR scan points. However, there are a lot of redundant points for self-localization in a point cloud map. For example, a flat wall is represented as a dense point cloud, but the geometrical information is only essentially a plane. This redundancy will cause a storage size problem in automated vehicles' limited storage systems. Also, noises in a point cloud map will affect self-localization accuracy. Some algorithms based on the ICP algorithm have been proposed, such as point-to-plane ICP [15] and generalized-ICP [16]. They extend the point cloud with normal information belonging to each point.

As a more sophisticated map format, the normal distribution (ND) map is typically employed in self-localization for autonomous driving. The normal distribution transformation (NDT) [17] algorithm achieves accurate self-localization with smaller map sizes than the point cloud map. An ND map consists of a set of normal distributions where each normal distribution is defined in a voxel. With an ND map, the NDT algorithm optimizes the self-vehicle's pose by maximizing

the sum of the likelihood of scan points within neighboring normal distributions.

Conversely, certain map formats are tailored to specific environments. For example, a vector-ND map [18], [19] incorporates geographic information system (GIS) data for self-localization in urban environments. Although this format employs the same self-localization algorithm as NDT, the results will differ as it focuses on voxels related to vector information. Various maps based on vector information have been developed, catering to specific aspects of the urban environment. Examples include traffic light maps [20], pole [21], [22], and lines [23].

A common characteristic among these formats is that there are locations where self-localization cannot identify the pose well, regardless of the chosen map format. This challenge arises because errors are inherent in the environment itself.

B. SELF-LOCALIZATION EVALUATION BEFORE DRIVING

There are several research to evaluate HD maps for self-localization before actual driving.

Akai et al. proposed an experimental-based error evaluation [12]. Their approach involves conducting evaluations of self-localization algorithms multiple times in the target environment before actual driving.

One approach involves a factor-based method that enumerates geometrical factors influencing self-localization accuracy [14]. For example, the feature ratio measures the proportion of each geometrical primitive (i.e., 1D, 2D, and 3D) in the environment. The authors assume that potential self-localization errors can be evaluated with a regression on these factor values. Experimental results demonstrated a correlation between the NDT algorithm and the errors estimated by the factor-based method.

In our previous study addressing the problem without conducting experiments in the target environment, we focused on objective function shapes [24]. Initially, the objective function shape is estimated at each location with synthetically generated LiDAR scans. Subsequently, evaluating the objective function shape involves assessing the condition number for the covariance matrix fitted to the shape. The criterion can capture the difficulty of iterative optimization of self-localization on the objective function at the location. However, the computational cost is large because the matching score for estimating the objective function shape will be repeated at each candidate of the poses. Also, the evaluation is not theoretically supported.

Typically, evaluations in this domain rely on heuristic or experimentally obtained rules. In contrast, this paper introduces a theoretically derived uncertainty for optimizing an objective function with map structures. The results from this proposed evaluation can serve as a baseline for essential self-localization errors in the environment without the need for specific rules, as seen in the previous work.

C. UNCERTAINTY ESTIMATION OF OPTIMAL SOLUTION

In GNSS, the possible positioning errors are modeled as a dilution of precision (DoP) [25]. A DoP assumes noisy observations for satellites with known positions and estimates a covariance matrix for the analytical optimum. However, it is hard to estimate such a covariance matrix analytically in map-matching cases because the exact correspondences are unknown.

On the other hand, several papers focus on estimating a covariance matrix of the estimation for specific algorithms, not for the map evaluation. The covariance matrix estimation for the ICP algorithm is widely studied. A Hessian matrix of the objective function is typically used to estimate a covariance matrix, known as Laplace approximation [26]. Aoki et al. considered the Laplace approximation with the effects of initial guess distribution for covariance matrix estimation [27]. Their uncertainty estimation is specialized for an algorithm (i.e., ICP) and used for the Kalman filter or some filtering algorithm. These analyses are algorithm-specific settings and do not model the uncertainty of self-localization inherent in the map.

On the other hand, the proposed model reasonably analyzes the errors inherent in the environment by utilizing an approximation theorem of uncertainty independent of the algorithm's properties. Also, a prior distribution of scans is assumed using a set of normal distributions, resulting in an adequate representation of the environment for self-localization, which can be a basis for all the map representations.

III. PRELIMINARY

Firstly, we introduce a theorem as the foundational theory for approximating the covariance matrix of the optimal solution with respect to an objective function. Secondly, the objective function is defined as the sum of the likelihood of a LiDAR scan to a set of normal distributions.

A. POSE PARAMETERIZATION

A self-localization result indicates a transformation from a local coordinate system of the LiDAR mount on the self-vehicle to the map coordinate system. The result is written in a six-degree vector comprising both a translation part and a rotation part.

1) TRANSFORMATION MATRIX

The six-degree vector is called a pose parameter, and it is written with $\boldsymbol{T} = [\boldsymbol{t}^\top \ \boldsymbol{\theta}^\top]^\top \in \mathbb{R}^6$. Here, $\boldsymbol{t} = [x \ y \ z]^\top \in \mathbb{R}^3$ is the translation part and $\boldsymbol{\theta} = [w_1 \ w_2 \ w_3]^\top \in \mathbb{R}^3$ is the rotation part. With the parameters, the transformation matrix $\boldsymbol{T} \in \mathrm{SE}(3)$ from a local coordinate system to the map coordinate system can be drawn as below,

$$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{R}(\theta) & \boldsymbol{t} \\ \boldsymbol{0}^\top & 1 \end{bmatrix} \tag{1}$$

where $\boldsymbol{R} : \mathbb{R}^3 \rightarrow \mathrm{SO}(3)$ is a function to convert a rotation vector $\boldsymbol{\theta}$ to a rotation matrix.

2) ROTATION MATRIX

The rotation vector coefficients, w_x, w_y, w_z here, means the coefficients of the vector space $\mathfrak{so}(3)$, which is the Lie algebra of SO(3). The element in the Lie algebra $\mathfrak{so}(3)$ for the $\boldsymbol{\theta}$ can be written as below,

$$\boldsymbol{\theta}^{\wedge} = [\boldsymbol{\theta}]_{\times} \tag{2}$$

$$= \begin{bmatrix} 0 & -w_z & w_y \\ w_z & 0 & -w_x \\ -w_y & w_x & 0 \end{bmatrix} \tag{3}$$

where $[\cdot]_{\times}$ is a projection to a skew-symmetric matrix. Then, the corresponding rotation matrix $\boldsymbol{R}(\boldsymbol{\theta}) \in$ SO(3) can be written via the exponential function as below.

$$\boldsymbol{R}(\boldsymbol{\theta}) = \exp(\boldsymbol{\theta}^{\wedge}) \tag{4}$$

3) POINT TRANSFORMATION AND JACOBIAN MATRICES

The three-dimensional point transformation uses $\cdot$ like the following equation.

$$\boldsymbol{T} \cdot \boldsymbol{p} := \boldsymbol{R}(\boldsymbol{\theta})\boldsymbol{p} + \boldsymbol{t} \tag{5}$$

where $\boldsymbol{T} \in$ SE(3) and $\boldsymbol{p} \in \mathbb{R}^3$ are a transformation matrix and a point on a local coordinate system, respectively. Then, a Jacobian matrix of the transformation, $\frac{\partial}{\partial \boldsymbol{T}}(\boldsymbol{T} \cdot \boldsymbol{p}) \in \mathbb{R}^{3 \times 6}$ and $\frac{\partial}{\partial \boldsymbol{p}}(\boldsymbol{T} \cdot \boldsymbol{p}) \in \mathbb{R}^{3 \times 3}$, can be derived as follows [28].

$$\frac{\partial}{\partial \boldsymbol{T}}(\boldsymbol{T} \cdot \boldsymbol{p}) = [\boldsymbol{I}_{3 \times 3} \quad -\boldsymbol{R}(\boldsymbol{\theta})[\boldsymbol{p}]_{\times}] \tag{6}$$

$$\frac{\partial}{\partial \boldsymbol{p}}(\boldsymbol{T} \cdot \boldsymbol{p}) = \boldsymbol{R}(\boldsymbol{\theta}) \tag{7}$$

B. COVARIANCE MATRIX APPROXIMATION

Our previous work tried to assess uncertainty from the function shape directly [24]. On the other hand, this paper focuses on an analytical approximation of the covariance matrix of the optimal solution for an objective function. Censi [29] derived an approximation of a covariance matrix of the optimal point of an objective function via the implicit function theorem.

Theorem 1: Let us assume an optimization problem of a parameter $\boldsymbol{x}$ with respect to a target function $L(\boldsymbol{T}, \boldsymbol{z})$ with an observation $\boldsymbol{z}$. If the algorithm $A(\boldsymbol{z})$ which outputs the optimal solution $\hat{\boldsymbol{x}}$ is differentiable, the covariance matrix of the optimal solution can be approximated as

$$\text{cov}(\hat{\boldsymbol{x}}) \approx \left(\frac{\partial^2 L}{\partial \boldsymbol{x}^2}\right)^{-1} \frac{\partial^2 L}{\partial \boldsymbol{z} \partial \boldsymbol{x}} \text{cov}(\boldsymbol{z}) \left(\frac{\partial^2 L}{\partial \boldsymbol{z} \partial \boldsymbol{x}}\right)^{\top} \left(\frac{\partial^2 L}{\partial \boldsymbol{x}^2}\right)^{-1}$$

We assume $\boldsymbol{z} \in \mathbb{R}^{3N}$ as a LiDAR scan and a parameter $\boldsymbol{x} \in \mathbb{R}^6$ as a pose parameter. Specifically, as described in the section III-A, the (translation + rotation vector) parametrization is used for the pose parameter in this article.

IV. PROPOSED EVALUATION METRIC

This section provides an overview of the proposed map evaluation metric. The proposed method does not require an iterative process in the evaluation process once synthetic LiDAR scans are constructed.

The subsequent sections delve into the specifics of the proposed method, covering synthetic LiDAR generation and the formulation for the resultant evaluation.

A. ERRORS OF MAP-BASED SELF-LOCALIZATION

Most map-based self-localization algorithms employ optimization to address the question of "what is the best pose to align a LiDAR scan with the surrounding geometrical structures of a map?". The latent error source for self-localization in the environment can be considered as the uncertainty of optimum solutions on an objective function representing the fitness score between the scan and the environment.

Figure 1 illustrates this error source originating from the environment. With the inherent noise in LiDAR scans, these uncertainties can be considered as a range of possible optimum solutions. In other words, this range in which function values are indistinguishable is determined by the noise range of a LiDAR scan.

Now, let us explore the considerations of a map format and the shape of an objective function for map evaluation and derive uncertainty from the objective function.

B. NORMAL DISTRIBUTION ENCODING OF MAPS

We employ a prior distribution of map structures, assuming a combination of normal distributions. This assumption is grounded in the concept that a LiDAR scan is derived from a set of normal distributions.

Initially, the entire three-dimensional space is subdivided into voxels. Subsequently, a normal distribution is estimated using the points contained within each voxel. The objective function to be minimized for a LiDAR pose $\boldsymbol{x}$ and a LiDAR scan $\boldsymbol{P} = \{\boldsymbol{p}_n\}_{n=1,\cdots,N}$ can be formulated as follows:

$$L(\boldsymbol{T}, \boldsymbol{P}) = \sum_{n=1}^{N} \sum_{v \in \mathcal{N}(\boldsymbol{p}_n)} l_n^v \tag{8}$$

$$l_n^v = \exp\left(-\frac{1}{2}(\boldsymbol{r}_n^v)^{\top} \boldsymbol{\Sigma}_v^{-1}(\boldsymbol{r}_n^v)\right), \tag{9}$$

$$\boldsymbol{r}_n^v = \boldsymbol{T} \cdot \boldsymbol{p}_n - \boldsymbol{\mu}_v \tag{10}$$

where $\{(\boldsymbol{\mu}_v, \boldsymbol{\Sigma}_v)\}_{v=1,\cdots,V}$ is a set of the normal distribution parameters, and $\mathcal{N}(\boldsymbol{p}_n)$ is a set of neighbor map component indices around the point.

Figure 2 compares the shapes of objective functions between point-to-point (P2P)-based and normal distribution-based objective function definitions at a narrow street with tall buildings. The function shape of P2P distances exhibits local minimums, leading to an overestimation of uncertainty. In contrast, the normal distribution-based function shape is designed to smooth out environmental effects. This design proves to be more reasonable for capturing self-localization uncertainty in a narrow street with tall buildings.

In this study, the voxelization is carried out with a small voxel size and a wide search radius for each point. A small voxel makes a detailed representation of geometrical features.

(a) P2P objective function   (b) ND objective function
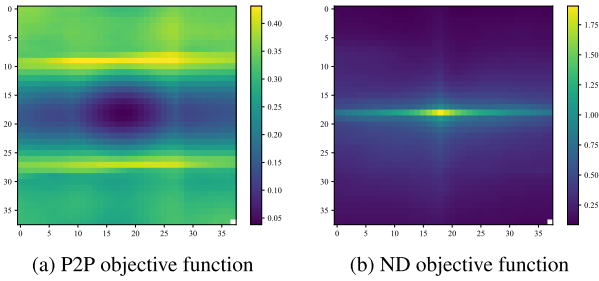
FIGURE 2. Objective function shape comparison at a narrow street. (location (i) in Figure 6) Note that the goodness of values is invested in each other.

Also, a wide search radius can imitate the reassignment in the registration process. This makes the representation of objective function more reasonable for general self-localization algorithms.

C. COVARIANCE MATRIX APPROXIMATION FOR NORMAL DISTRIBUTION MAPS

The hessian matrices are derived in this section for the approximation (8). First, the gradient vector for the l_n^v in (8) can be derived as below (n, v is omitted for simplicity).

$$\frac{\partial l}{\partial \boldsymbol{T}} = \frac{\partial}{\partial \boldsymbol{T}} \exp\left(-\frac{1}{2}\boldsymbol{r}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{r}\right) \tag{11}$$

$$= \frac{\partial}{\partial \boldsymbol{T}}\left(-\frac{1}{2}\boldsymbol{r}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{r}\right)l \tag{12}$$

$$= -\left(\boldsymbol{r}^\top \boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{r}}{\partial \boldsymbol{T}}\right)l \tag{13}$$

Then, the hessian matrices of (8) can be drawn as below.

$$\frac{\partial^2 l}{\partial \boldsymbol{T}^2} = -\left(\boldsymbol{J}_T^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{J}_T\right)l - \left(\boldsymbol{r}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{J}_T\right)^\top \frac{\partial l}{\partial \boldsymbol{T}} \tag{14}$$

$$= -l\left(\boldsymbol{J}_T^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{J}_T - (\boldsymbol{J}_T^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{r}\boldsymbol{r}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{J}_T)\right) \tag{15}$$

where $\boldsymbol{J}_T = \partial \boldsymbol{r}/\partial \boldsymbol{T}$.

$$\frac{\partial^2 l}{\partial \boldsymbol{p}\partial \boldsymbol{T}} = -(\boldsymbol{J}_T^\top \boldsymbol{\Sigma}_v^{-1}\boldsymbol{J}_p)l - (\boldsymbol{r}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{J}_T)^\top \frac{\partial l}{\partial \boldsymbol{p}} \tag{16}$$

$$= -l\left(\boldsymbol{J}_T^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{J}_p - (\boldsymbol{J}_T^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{r}\boldsymbol{r}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{J}_p)\right) \tag{17}$$

where $\boldsymbol{J}_p = \partial \boldsymbol{r}/\partial \boldsymbol{p}$.

The Jacobian matrices $\boldsymbol{J}_T, \boldsymbol{J}_p$ can be drawn as below.

$$\boldsymbol{J}_T = \frac{\partial \boldsymbol{r}}{\partial \boldsymbol{T}} = [\boldsymbol{I}_{3\times3} \quad -\boldsymbol{R}(\boldsymbol{\theta})[\boldsymbol{p}]_\times] \tag{18}$$

$$\boldsymbol{J}_p = \frac{\partial \boldsymbol{r}}{\partial \boldsymbol{p}} = R(\boldsymbol{\theta}) \tag{19}$$

D. SYNTHETIC LIDAR SCAN GENERATION

In the covariance approximation, a model LiDAl scan is needed to imitate an average (i.e., ideal) LiDAR scan. For the model scan, a synthetic LiDAR scan is generated to mimic the ideal LiDAR scan at a location. Importantly, our evaluation metric does not necessitate actual LiDAR scans, as synthetic scans are generated from an existing point cloud map.

The process to generate a LiDAR scan with a point cloud map follows these steps:

- assumes a LiDAR model (e.g., VLP-16, Velodyne), specifying parameters such as resolution for azimuth, altitude, and range.
- generates a ray from a position towards the map.
- samples three-dimensional points along the ray at a defined invertal.
- identifies the nearest point in the point cloud map for each sampled point, starting from a near point on the ray.
- accepts a point as a LiDAR point if the distance to the nearest point is smaller than a specified threshold.
- repeats the sampling process for all rays.

Subsequently, the proposed metric models actual noisy effects, including motion blur, occlusion, and sensing noises in cov($\boldsymbol{z}$).

V. EXPERIMENTS

This section details the experiments conducted to reveal how the proposed metric evaluates real urban environments. It gives us an insight into the essential error sources of self-localization in real environments. Through these experiments, we can analyze where self-localization is poor.

Furthermore, the experiments include a comparison with factor values by Javanmardi et al., [13]. These factors highlight potential features of geometrical structures in a map that contribute to self-localization errors. This comparative analysis convinces the conclusion from the geometrical properties of the surroundings.
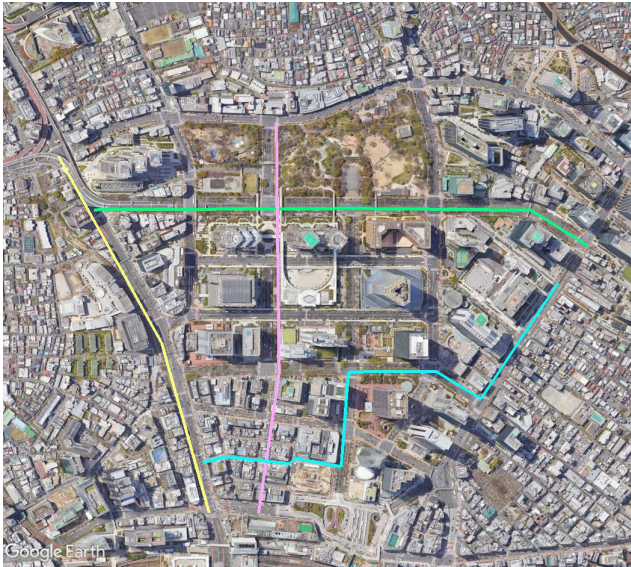
A. STUDY AREA

The experiments were conducted with data from an urban environment, Shinjuku, Tokyo, Japan. An appearance of the environment and paths we used are shown in Figure 3.

The environments have a lot of specific features in urban environments, such as tall buildings, trees, and intersections.

B. VEHICLE SETTINGS AND IMPLEMENTATION

A point cloud map for the environment was collected using previous work [30].

For the model of a LiDAR for self-localization, VLP-16 from Velodyne is assumed in the experiments. We use a voxel size of 0.5 meters for the normal distribution map construction. The search radius of the nearest neighbor voxels for a query point is set to 4.0 meters to capture features around the point. The uncertainty of the scan noise, cov($\boldsymbol{z}$), is assumed to be an isotropic distribution for each direction. For the noise level of LiDAR scans, we assume $\sigma = 0.3$ for each direction. The sigma value has linear increasing effects on the estimated covariance matrix. It means that the value will not matter if we want to know the error levels in the environment. The effects of the voxel size are discussed in Section. V-F.

FIGURE 3. Study area and four experimental paths in Shinjuku, Tokyo, Japan.

C. ESTIMATED ERROR RESULT

Figure 4 shows the error levels for the lateral and longitudinal directions with the point cloud map. As you can see, there are some specific regions where estimation will fail for a direction of longitudinal, lateral, or both. Through all paths, lateral errors are relatively lower than longitudinal errors.

D. COMPARISON WITH ACTUAL ERRORS OF SELF-LOCALIZATION ALGORITHMS

To make our method more convincing for self-localization evaluation, the comparison between the actual errors of some algorithms and the estimated errors. Actually, self-localization results from ICP and NDT and the model standard deviations from the proposed method for each direction were compared.

1) IMPLEMENTATION

Iterative optimization in ICP and NDT tends to be stacked in local minimums but can be overcome with better settings or optimization algorithms. To ignore such effects for the error analysis, ICP and NDT were conducted in a grid search manner. It means that the criterion values (target function values of ICP and NDT) for each 5 centimeters were evaluated, and the candidates with top or bottom 10 % from the best criterion value are regarded as estimation.

For the other setting, NDT was conducted with a voxel size of 2.0 meters.

2) COMPARISON RESULT

Figure 5 shows the comparison results between modeled errors and actual errors in Seq.1. As seen in the figure, the estimated errors can track the actual self-localization errors. Notably, it is common among the self-localization

algorithms. It means that our method can track essential errors of self-localization.

Since the estimated errors indicate essential errors in the environment, gaps between actual and estimated errors may indicate the effects of choices in map formats for each algorithm. Such gaps can be used for the discussion of compression difficulties of the point cloud map in each map format.

On the other hand, there are locations with gaps between the higher estimated errors and the lower actual errors, such as around 400 meters in the cumulative distance. The location is at an intersection in Seq.1, resulting in a smaller number of normal distributions for covariance estimation. The sparsity of distributions might make the estimation confusing for self-localization error estimation.

E. QUALITATIVE EVALUATION OF ESTIMATED COVARIANCE MATRICES AND DISCUSSION

Figure 6 shows estimated covariance matrices in the specific locations. Results are shown on narrow streets with tall buildings, wider streets with trees, and intersections.
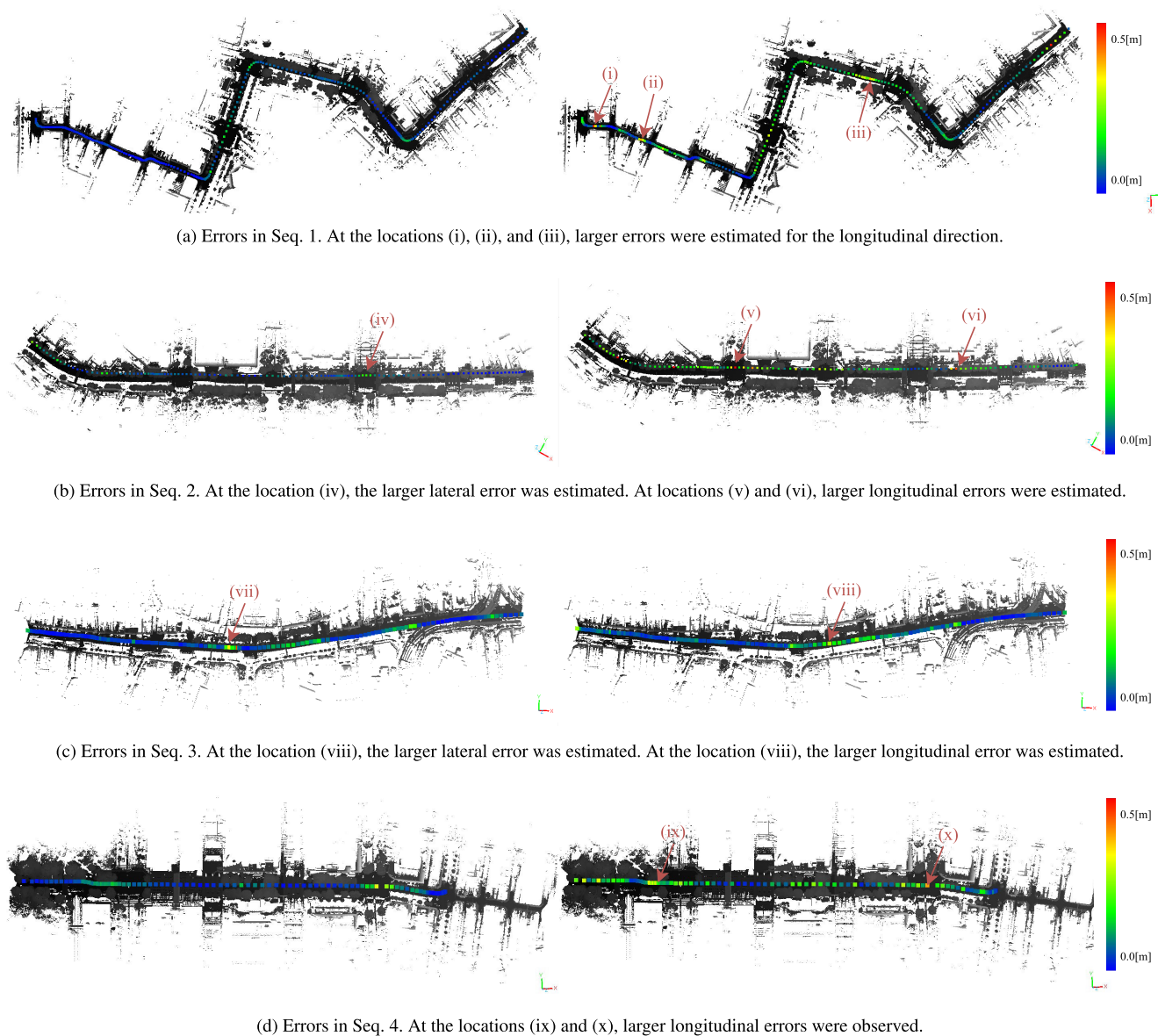
Lateral errors tend to be high on streets with far buildings (e.g., (vii) and (ix)). Sparse scan points for far buildings will result in uncertainty toward the lateral direction because the building walls are essential for lateral estimation. Longitudinal errors tend to be high on narrow streets with tall buildings (e.g., (i) and (ii)). This is because there are few features to identify longitudinal positions in the environment. On the other hand, the lateral errors are low because of sufficient planar features for the lateral direction. Also, on streets with a lot of trees, longitudinal errors tend to be high (e.g., (iii), (vi), (viii), and (ix)). This is because mismatching inter near pole-like features (i.e., tree trunks) will create uncertainty in the longitudinal direction. At intersections ((iv), (vii), and (x)), lateral, longitudinal, or both errors are high. It is well known that there are few features at intersections, resulting in such large self-localization errors.

F. PARAMETER EFFECTS

We examined the parameter effects for the estimation. Especially, the voxel size effects were examined in a narrow area with tall buildings as Figure 7. As we see, the larger the voxel size, the estimation tends to ignore the environment. This is because large voxels will ignore the distinctive features in the environment for self-localization. A smaller voxel size can capture local geometry in detail, resulting in the accurate modeling of the environment with normal distributions. From the aforementioned discussion, we used 0.5 meters of the voxel size to track the effects of distinctive features in the surroundings.

G. FACTOR ANALYSIS

Finally, factor values will be compared in locations where estimated errors are either high or low. These factors have been proposed by Javanmardi et al. for map assessment for self-localization [13]. A factor measures a geometrical

(a) Errors in Seq. 1. At the locations (i), (ii), and (iii), larger errors were estimated for the longitudinal direction.

(b) Errors in Seq. 2. At the location (iv), the larger lateral error was estimated. At locations (v) and (vi), larger longitudinal errors were estimated.

(c) Errors in Seq. 3. At the location (viii), the larger lateral error was estimated. At the location (viii), the larger longitudinal error was estimated.

(d) Errors in Seq. 4. At the locations (ix) and (x), larger longitudinal errors were observed.

FIGURE 4. Estimated errors with map point clouds. Lateral (left) and longitudinal (right) errors are plotted.

feature in the surroundings as a value. For example, feature ratios indicate how primitive shapes are the majority in surroundings (i.e., line, plane, and others). These factors were previously identified to have an impact on self-localization accuracy.

By comparing the factor values with the estimated errors, the geometrical specifics can be explained with the factor values. This comparative analysis will offer insights into the reasons behind the estimated errors in the environment.

1) FACTOR COMPUTATION
We compare three factors, such as 1d, 2d, and 3d feature ratio.

The factors are defined using geometrical features of normal distribution voxels and those layouts. Specifically, the

voxel size of 2.0 meters is adopted in their proposal, and we adopt the same setting with them.

The following sections describe the definition of factors roughly. Please refer to the original paper [13] for the details.

2) FEATURE RATIO
The 1d, 2d, and 3d feature ratio measures the ratio of each feature in the surroundings. Each voxel in the environments for factor computation is classified into three classes, such as 1d, 2d, and 3d, based on their eigenvalues of the covariance matrix. Let $\sigma_1 \geq \sigma_2 \geq \sigma_3 \in \mathbb{R}$ be the squared rooted eigenvalues for the covariance matrix of a voxel. A metric
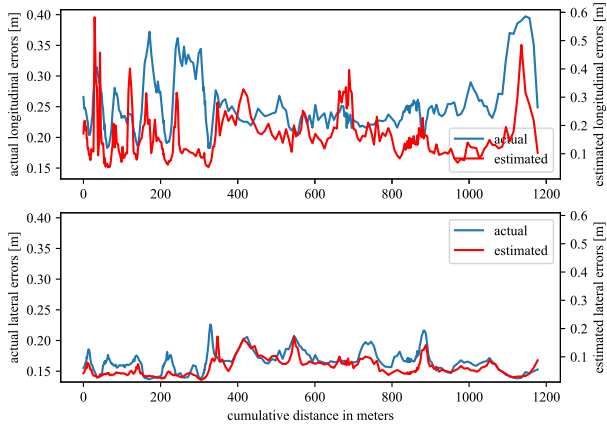
(a) Comparison with ICP
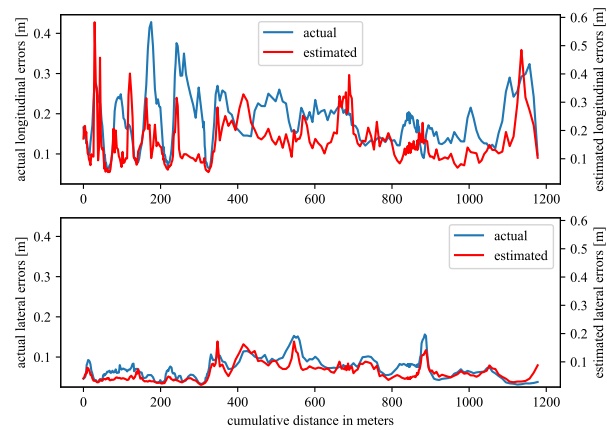
(b) Comparison with NDT

FIGURE 5. Comparison between model errors and actual self-localization errors.

a_1, a_2, a_3 is defined as follows.

$$a_1 = \frac{\sigma_1 - \sigma_2}{\sigma_1}, a_2 = \frac{\sigma_2 - \sigma_3}{\sigma_1}, a_3 = \frac{\sigma_3}{\sigma_1} \qquad (20)$$
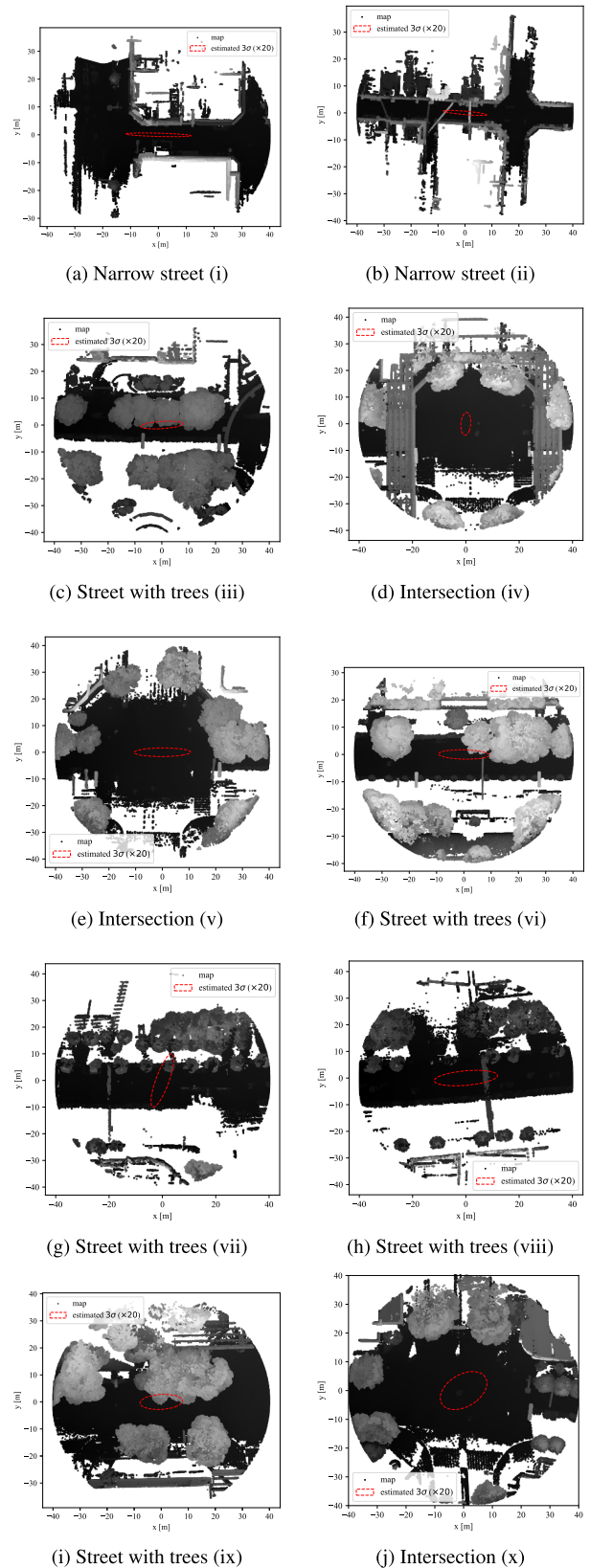
Then, the voxel is classified as 1d when $a_1 \geq a_2, a_3$, 2d when $a_2 \geq a_1, a_3$, and otherwise 3d.

The ratio of the number of each feature is used as a factor value.

3) RESULTS

Figure 8 shows the statistical results of Welch's t-test for comparing factor values at locations with higher estimated lateral errors ($\geq$ 0.1 meters and lower ones ($<$ 0.1). The criterion is determined by the lane-level requirement in self-localization. As the results indicate, there are significant differences in factor values between locations.

The feature ratio indicates that the surrounding geometrical structures. From the result, the 3d feature ratio is lower when the self-localization errors are lower. The 1D feature ratio

(a) Narrow street (i)

(b) Narrow street (ii)

(c) Street with trees (iii)

(d) Intersection (iv)

(e) Intersection (v)

(f) Street with trees (vi)

(g) Street with trees (vii)

(h) Street with trees (viii)

(i) Street with trees (ix)

(j) Intersection (x)

FIGURE 6. Visualization of estimated covariance matrices. Note that it shows 20 times trust regions. The points indicate the neighbor point cloud. The colors of the point cloud become brighter the higher height.
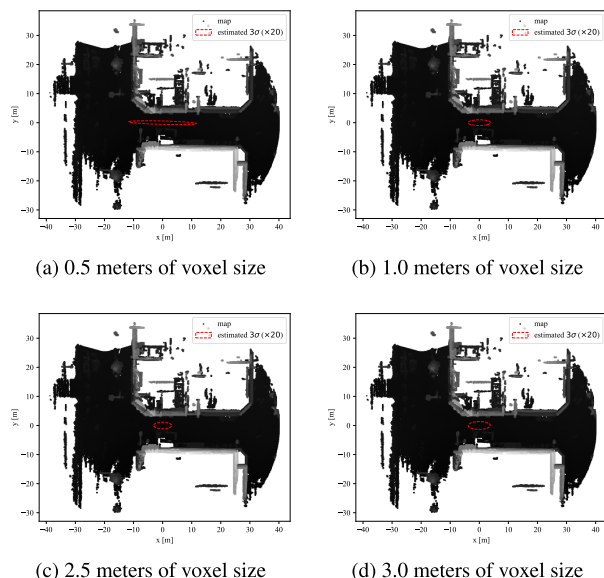
(a) 0.5 meters of voxel size

(b) 1.0 meters of voxel size

(c) 2.5 meters of voxel size

(d) 3.0 meters of voxel size

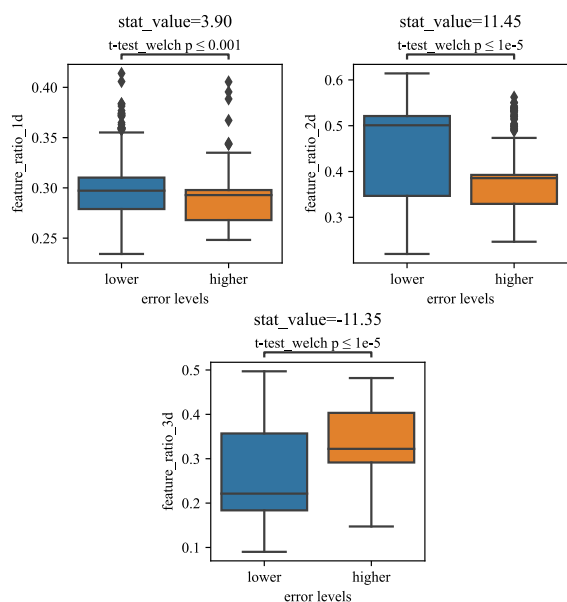**FIGURE 7.** Parameter effects for covariance estimation.

FIGURE 8. Statistical test of factor values between places with higher and lower estimated errors.

and the 2D feature ratio are higher when the self-localization errors are lower. In conclusion, self-localization tends to be accurate when the 1d feature ratio and the 2d feature ratio are relatively higher. These results are reasonable for the common knowledge of self-localization. As seen in location (vii) and (ix), 2D features are required for the lateral accuracy. Also, 1D features (pole-like objects) are essential to localize both lateral and longitudinal directions. On the other hand, the 3D feature spread out for each direction in three-dimensional space, resulting in less information about self-localization in the surroundings.

VI. CONCLUSION

We introduce a novel evaluation technique to assess the high-definition map's potential capabilities for self-localization in real-world environments. Unlike previous methods that rely on empirical or rule-based error estimation using geometrical features, our approach directly estimates potential self-localization errors based on theoretical covariance derived from the assumption that the map comprises normal distributions. This theoretical modeling provides a rough representation of the environment's geometric features, enabling the estimation of latent errors without the need for experiments in the target environment before driving.

Our method involves deriving a covariance matrix approximation analytically, incorporating likelihood calculations for a hypothetical scan against normal distribution maps. To accommodate differences in sensor models, we employ ideal LiDAR scan generation using pre-existing point cloud maps. Moreover, small voxels and wide correspondence searching are used to enhance the ability of our method to capture correspondence switching beyond voxels.

We conducted experiments in an urban environment in Shinjuku, Tokyo, Japan. In the environment, we estimated self-localization errors that vary in different environmental settings. Notably, errors tend to be higher in the longitudinal direction on narrow streets surrounded by tall buildings, while lateral errors are pronounced in areas with wide roads. High error levels in both lateral and longitudinal directions were observed at intersections, aligning with intuitive expectations and prior analyses. Also, factors [13], [14] are introduced to describe the estimated self-localization errors from the geometrical feature aspects. Observed factors reason

The proposed method effectively captures essential errors in the environment, providing valuable insights into the environment for self-localization purposes before driving. Autonomous driving systems can leverage our method for safe operations, particularly in areas where higher errors are estimated. Furthermore, we believe that our estimation of essential errors can contribute to discussions on HD map formats from a compression perspective.

REFERENCES

[1] G. Bresson, Z. Alsayed, L. Yu, and S. Glaser, "Simultaneous localization and mapping: A survey of current trends in autonomous driving," *IEEE Trans. Intell. Vehicles*, vol. 2, no. 3, pp. 194–220, Sep. 2017.

[2] N. Zhu, J. Marais, D. Bétaille, and M. Berbineau, "GNSS position integrity in urban environments: A review of literature," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 9, pp. 2762–2778, Sep. 2018.

[3] B. Shi, M. Wang, Y. Wang, Y. Bai, K. Lin, and F. Yang, "Effect analysis of GNSS/INS processing strategy for sufficient utilization of urban environment observations," *Sensors*, vol. 21, no. 2, p. 620, Jan. 2021.

[4] Y. Gu, L.-T. Hsu, and S. Kamijo, "GNSS/Onboard inertial sensor integration with the aid of 3-D building map for lane-level vehicle self-localization in urban canyon," *IEEE Trans. Veh. Technol.*, vol. 65, no. 6, pp. 4274–4287, Jun. 2016.

[5] B. E. Soorchaei, M. Razzaghpour, R. Valiente, A. Raftari, and Y. P. Fallah, "High-definition map representation techniques for automated vehicles," *Electronics*, vol. 11, no. 20, p. 3374, Oct. 2022.

[6] S. Kuutti, S. Fallah, K. Katsaros, M. Dianati, F. Mccullough, and A. Mouzakitis, "A survey of the state-of-the-art localization techniques and their potentials for autonomous vehicle applications," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 829–846, Apr. 2018.

[7] E. Yurtsever, J. Lambert, A. Carballo, and K. Takeda, "A survey of autonomous driving: Common practices and emerging technologies," *IEEE Access*, vol. 8, pp. 58443–58469, 2020.

[8] H. Cai, Z. Hu, G. Huang, D. Zhu, and X. Su, "Integration of GPS, monocular vision, and high definition (HD) map for accurate vehicle localization," *Sensors*, vol. 18, no. 10, p. 3270, Sep. 2018.

[9] S. Bauer, Y. Alkhorshid, and G. Wanielik, "Using high-definition maps for precise urban vehicle localization," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2016, pp. 492–497.

[10] J. M. Kang, T. S. Yoon, E. Kim, and J. B. Park, "Lane-level map-matching method for vehicle localization using GPS and camera on a high-definition map," *Sensors*, vol. 20, no. 8, p. 2166, Apr. 2020.

[11] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-squares fitting of two 3-D point sets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-9, no. 5, pp. 698–700, Sep. 1987.

[12] N. Akai, L. Y. Morales, E. Takeuchi, Y. Yoshihara, and Y. Ninomiya, "Robust localization using 3D NDT scan matching with experimentally determined uncertainty and road marker matching," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2017, pp. 1356–1363.

[13] E. Javanmardi, M. Javanmardi, Y. Gu, and S. Kamijo, "Factors to evaluate capability of map for vehicle localization," *IEEE Access*, vol. 6, pp. 49850–49867, 2018.

[14] E. Javanmardi, M. Javanmardi, Y. Gu, and S. Kamijo, "Pre-estimating self-localization error of NDT-based map-matching from map only," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 12, pp. 7652–7666, Dec. 2021.

[15] Y. Chen and G. Medioni, "Object modelling by registration of multiple range images," *Image Vis. Comput.*, vol. 10, no. 3, pp. 145–155, Apr. 1992.

[16] A. V. Segal, D. Hähnel, and S. Thrun, "Generalized-ICP," in *Robotics: Science and Systems*, vol. 2. Seattle, WA, USA, 2009, p. 435.

[17] M. Magnusson, "The three-dimensional normal-distributions transform: An efficient representation for registration, surface analysis, and loop detection," Ph.D. thesis, Örebro Universitet, 2009.

[18] T. Izawa, Y. Endo, and S. Kamijo, "Integrated high-definition maps for self-localization with vector and polygon data," in *Proc. IEEE 25th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2022, pp. 1043–1048.

[19] Y. Endo, T. Izawa, and S. Kamijo, "Hierarchical map representation using vector maps and geometrical maps for self-localization," *IATSS Res.*, vol. 46, no. 4, pp. 450–456, Dec. 2022.

[20] C. Wang, H. Huang, Y. Ji, B. Wang, and M. Yang, "Vehicle localization at an intersection using a traffic light map," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 4, pp. 1432–1441, Apr. 2019.

[21] J. Kümmerle, M. Sons, F. Poggenhans, T. Kühner, M. Lauer, and C. Stiller, "Accurate and efficient self-localization on roads using basic geometric primitives," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2019, pp. 5965–5971.

[22] A. Schaefer, D. Büscher, J. Vertens, L. Luft, and W. Burgard, "Long-term urban vehicle localization using pole landmarks extracted from 3-D LiDAR scans," in *Proc. Eur. Conf. Mobile Robots (ECMR)*, Sep. 2019, pp. 1–7.

[23] J.-H. Im, S.-H. Im, and G.-I. Jee, "Extended line map-based precise vehicle localization using 3D LiDAR," *Sensors*, vol. 18, no. 10, p. 3179, 2018.

[24] Y. Endo and S. Kamijo, "Self-localization error evaluation from map structure based on autocorrelation values," in *Proc. IEEE 26th Int. Conf. Intell. Transp. Syst. (ITSC)*. IEEE, to be published.

[25] R. Santerre, A. Geiger, and S. Banville, "Geometry of GPS dilution of precision: Revisited," *GPS Solutions*, vol. 21, no. 4, pp. 1747–1763, Oct. 2017.

[26] S. Srinara, S. Tsai, C.-X. Lin, M.-L. Tsai, and K.-W. Chiang, "Reliable evaluation of navigation states estimation for automated driving systems," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2022, pp. 1765–1773.

[27] K. Aoki, T. Sato, E. Takeuchi, Y. Ninomiya, and J. Meguro, "Error covariance estimation of 3D point cloud registration considering surrounding environment," *J. Robot. Mechatronics*, vol. 35, no. 2, pp. 435–444, Apr. 2023.

[28] J. Sola, J. Deray, and D. Atchuthan, "A micro lie theory for state estimation in robotics," 2018, *arXiv:1812.01537*.

[29] A. Censi, "An accurate closed-form estimate of ICP's covariance," in *Proc. IEEE Int. Conf. Robot. Autom.*, Apr. 2007, pp. 3167–3172.

[30] M. Javanmardi, E. Javanmardi, Y. Gu, and S. Kamijo, "Towards high-definition 3D urban mapping: Road feature-based registration of mobile mapping systems and aerial imagery," *Remote Sens.*, vol. 9, no. 10, p. 975, Sep. 2017.

YUKI ENDO (Member, IEEE) received the M.E. degree in information engineering from The University of Tokyo, Japan, in 2021, where he is currently pursuing the Ph.D. degree in information and communication engineering. His research interests include self-localization, deep learning, and intelligent transportation systems.

SHUNSUKE KAMIJO (Senior Member, IEEE) received the B.S. and M.S. degrees in physics and the Ph.D. degree in information engineering from The University of Tokyo, Tokyo, Japan, in 1990, 1992, and 2001, respectively. He began working with Fujitsu Ltd., Tokyo, in 1992, as a Processor Design Engineer. He was an Assistant Professor, from 2001 to 2002, and he has been an Associate Professor, since 2002. His research interests include computer vision, wireless communication, their applications to intelligent transportation systems (ITS), autonomous vehicles, traffic video surveillance, traffic signal control, vehicle-to-everything (V2X) communications, and pedestrian and car navigations. He is a member of the IEEE ITS Society, the Institute of Electronics, Information and Communication Engineers (IEICE), and International Association of Traffic and Safety Sciences (IATSS). He has been a member of the Board of Governors of the IEEE ITS Society, since 2015, and the Vice President of Publication, since 2022. He has served as the Vice-Chairperson of the Program Committee for the ITS World Congress Tokyo 2013, the General Co-Chair for IEEE International Conference on Vehicular Electronics and Safety (ICVES) 2015 and IEEE International Conference on Intelligent Transportation Systems (ITSC) 2017, and the International Program Chair for IEEE Intelligent Vehicles Symposium (IV) 2017. He is the Editor-in-Chief of *IATSS Research* (Elsevier) and an Editorial Board Member of the *International Journal on Intelligent Transportation Systems Research* (Springer) and *Multimedia Tools and Applications* (Springer).

• • •