

# Enhanced Integrity of Lidar Localization: A Study on Feature Extraction Techniques

Kana Nagai, Boris Pervan, *Illinois Institute of Technology*

## BIOGRAPHY

**Kana Nagai** is currently a Ph.D. candidate and Research Assistant in Mechanical and Aerospace Engineering at Illinois Institute of Technology (IIT). She received her B.S. and M.S. in Mechanical Engineering from IIT and B.E. in Architecture from Hokkaido University, Japan.

**Dr. Boris Pervan** is a Professor and Frank Gunsaulus Faculty Fellow in Mechanical and Aerospace Engineering at the Illinois Institute of Technology (IIT), where he conducts research on high integrity navigation systems. Prior to joining the faculty at IIT, he was a spacecraft mission analyst at Hughes Aircraft Company (now Boeing) and a postdoctoral research associate at Stanford University. He received his B.S. from the University of Notre Dame, M.S. from the California Institute of Technology, and Ph.D. from Stanford University.

## ABSTRACT

This study develops a methodology to minimize extraction faults in landmark-based Lidar localization using a map. The analysis begins by addressing the integrity risks associated with Lidar navigation systems in self-driving cars and identifying the causes of incorrect extractions. Example Lidar-fault integrity requirements are introduced to establish a basis for data point selection using the Neyman-Pearson criterion. Weighted least squares estimation, combined with a chi-square residual test, is then applied for landmark extraction. Experimental results confirm that the methodology reduces landmark extraction errors to centimeter-level.

## I. INTRODUCTION

The concept of driverless cars offers exciting prospects for the future of urban transportation but also raises serious concerns when human lives could be at stake. Consequently, safety is the top priority in these developments (American Automobile Association, 2023; The National Science and Technology Council and the United States Department of Transportation, 2020). In urban environments, the proximity of people, vehicles, and structures leads to higher accident and fatality rates than in rural areas (The National Highway Traffic Safety Administration, 2023). The congestion in these areas necessitates stricter safety requirements (EUSPA, 2021; Reid et al., 2019) but can undermine the accuracy of global navigation satellite systems (GNSS). Tall buildings and narrow streets interfere with satellite signals, complicating navigation (Chapman et al., 2002; Nagai et al., 2020). Accordingly, existing GNSS integrity risk evaluation methods cannot be directly applied where driverless cars operate.

The principle of integrity underlies aviation navigation safety requirements (Davis and Kelly, 1993; RTCA (Firm). SC-159, 2004; Grimes, 2007), and this concept has been applied to ground autonomous vehicles recently (Zhu et al., 2018; Cosmen-Schortmann et al., 2008). If the integrity risk, which measures the unreliability of information from a navigation system, is high, the vehicle using the system might be in danger. Fault-free integrity evaluates whether the navigation system's position error exceeds allowable bounds in the absence of faults. Faults are errors in sensor measurements that may elude immediate detection, eventually leading to a position error. Both fault-free and fault integrity can be quantified probabilistically, and safety evaluations function by comparing these risks to integrity requirements (RTCA (Firm). SC-159, 2004).

We examined the fault-free integrity of navigation systems for autonomous vehicles in urban environments, where GNSS signals are disrupted frequently. Despite integrating tactical grade inertial sensors (INS) (Titterton and Weston, 2004), odometers (Gao et al., 2006; Hazlett et al., 2011), zero velocity updates (Grejner-Brzezinska et al., 2001), and vehicle kinematic constraints (Godha and Cannon, 2007) with GNSS using a tightly coupled extended Kalman filter (EKF) (Falco et al., 2017), our findings emphasized the need for alternative position reference updates (Nagai et al., 2024). We suggested enhancing the system with local ranging capabilities provided by light detection and ranging (Lidar) as one of the options.

Lidar, an active remote sensing technology, utilizes laser beams to detect objects within its field of view and measures distances and directions to those objects and the intensities of the laser returns from them. It computes the ranges by measuring the time it takes the emitted laser beam to return after bouncing off the target. The emission angles of the beams determine the directions. The number of signal photons, representing the amount of light that returns to the sensor, measures intensities.

Various autonomous driving applications utilize Lidar, including lane detection, object detection, tracking, segmentation,

mapping, and localization (Benedek et al., 2021). Simultaneous Localization and Mapping (SLAM) operates by creating a map in an unknown environment without prior knowledge while simultaneously estimating its position. SLAM is valuable to create a high-definition map in GNSS unavailable areas but experiences drift errors which necessitate post-processing (Jo et al., 2018; Keitaanniemi et al., 2023).

Since our challenge concerns integrity evaluation, which requires accurate positioning free from drift errors, we assume the navigation system already has a *prior* digital map. Map-based Lidar localization can achieve centimeter-level accuracy under fault-free conditions (Levinson et al., 2007; Wan et al., 2018). One popular approach is scan matching, which divides the local space into small cells to align the Lidar scan with the map (Levinson and Thrun, 2010). However, because of the large number of data points in a typical scan and unavoidable scene resolution limitations on both the scanner and the map, quantitative evaluations of integrity risk will be extremely difficult, if not impossible. For this reason, we restrict our study to landmark-based localization.

The landmark-based localization method begins by detecting *landmarks*, which are specific physical features whose positions are recorded in a database known as a *landmark map*. Subsequently, the sensor measures their distances and angles, and the measured landmarks are matched with the information in the map (Levinson et al., 2007). The system can estimate a vehicle’s position by combining these measurements with the landmark locations defined in the map. Although all objects in urban environments are candidates for landmarks, we specifically focus on extracting pole-like landmarks (e.g., street lamps) because of their distinct shapes, semi-permanent locations, weather resistance, defined centroids, and relative ubiquity, as demonstrated in Gim et al. (2021); Sefati et al. (2017); Tee-Ann and Chi-Min (2015).

Landmark-based localization relies on two intermediary procedures for positioning from point clouds, feature extraction and data association, both having the potential to introduce faults (Bar-Shalom and Fortmann, 1988). Feature extraction process involves identifying pre-defined landmarks from point clouds. Incorrect extraction (IE) faults happen when the system mistakenly identifies points of undefined ‘obstacles’ as landmarks. However, even if the extraction works correctly, uncertainty regarding the source of each measurement can lead to an incorrect association (IA), resulting in a mismatch between the measurement observations and their corresponding landmark locations in the map.

Association faults known to compromise navigation integrity (Bar-Shalom et al., 2009; Pirovano et al., 2020), and methods to quantify association risk evaluation have been developed. For example, the method proposed in (Joerger et al., 2016) determines the upper bounds on the risk. By integrating an IMU for accurate state prediction and setting a minimum separation distance between mapped landmarks, we can ensure tight bounds on association risk (Hassani and Joerger, 2023). However, feature extraction risk was not addressed in this prior work.

This paper describes a methodology to minimize extraction faults in landmark-based Lidar localization. The analysis begins by addressing the integrity risks of Lidar positioning for self-driving cars and identifying the causes of incorrect extractions. Example Lidar-fault integrity requirements are introduced to establish target risk levels for the data point selection method. Subsequently, our proposed point filtering process consists of two stages: coarse and fine, using the Neyman-Pearson Lemma, and then weighted least squares estimation combined with a chi-square test.

Following this introduction, Section II discusses the causes of Lidar faults along with Lidar measurement models. Section III defines these faults and formulates the integrity risk equation. Section IV reviews prior work, including fault-free position error and data association risk. Section V introduces the extraction methodology aimed at minimizing extraction faults. Section VI presents the experimental results, demonstrating the effectiveness of the proposed method. Finally, Section VII summarizes the work our provides our conclusions.

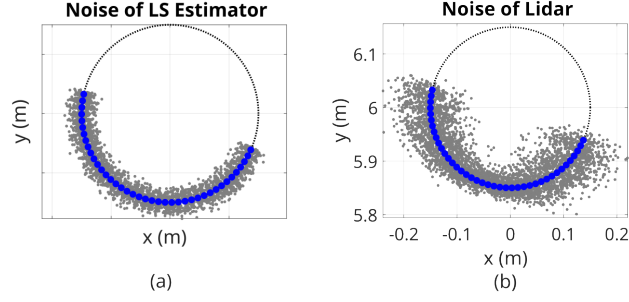
## II. LIDAR LOCALIZATION

### 1. Feature Extraction Procedure

Lidar localization starts by filtering points from a point cloud that outline a cylindrical landmark; this process is defined as *point selection*. The centroid, representing the landmark’s position in the horizontal plane, is then estimated by fitting these points onto the perimeter of a circle with known diameter; this process is defined as *landmark extraction*. Feature extraction refers to the combination of point selection and landmark extraction. The most common landmark extraction method involves minimizing the sum of squared distances from each point to the centroid, using least squares estimation (Umbach and Jones, 2003). However, while this method assumes a uniform error distribution along the radial direction, Lidar measurement errors are not uniformly distributed, as shown in Figure 1.

We address this issue by developing a centroid estimator using weighted least squares. The Lidar measurements of cylindrical objects are described as follows:

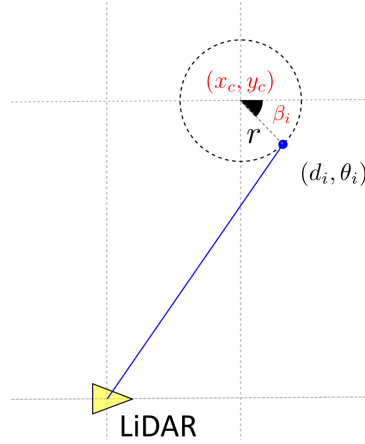
$$d_i = \sqrt{(x_c + r \cos \beta_i)^2 + (y_c - r \sin \beta_i)^2} + v_d \quad (1)$$



**Figure 1:** The least squares estimator in (Umbach and Jones, 2003) assumes that measurement errors follow normal distributions along the radial direction at any location (a). However, the Lidar sensor noise simulation for a Velodyne VLP-16 reveals that the noise is unevenly distributed along the radial direction (b).

$$\theta_i = \arctan \left( \frac{y_c - r \sin \beta_i}{x_c + r \cos \beta_i} \right) + v_\theta \quad (2)$$

where  $d_i$  and  $\theta_i$  are the measurements of range and angle, respectively, to points  $i = 1, 2, \dots, n$ .  $x_c$  and  $y_c$  are states representing the unknown position of the cylinder centroid in Cartesian coordinates centered at the Lidar,  $r$  is the known radius of the landmark from the map,  $\beta_i$  is a state representing the unknown angle between the x-axis in the body frame and the measurement point on the cylinder, and  $v_d$  and  $v_\theta$  represent the measurement noise for range and angle. Figure 2 illustrates the notation and the positional relationships among the Lidar measurements, the landmark centroid, and the Lidar location.



**Figure 2:** This diagram illustrates the positional relationships among the Lidar measurements, the unknown landmark centroid, and the Lidar location.

After linearization, Equations (1) and (2) can be expressed in the following matrix form:

$$\underbrace{\begin{bmatrix} d_1 - d_1^* \\ \vdots \\ d_n - d_n^* \\ \theta_1 - \theta_1^* \\ \vdots \\ \theta_n - \theta_n^* \end{bmatrix}}_z = \mathbf{H} \underbrace{\begin{bmatrix} \delta x_c \\ \delta y_c \\ \delta \beta_1 \\ \vdots \\ \delta \beta_n \end{bmatrix}}_x + \underbrace{\begin{bmatrix} v_d \\ \vdots \\ v_d \\ v_\theta \\ \vdots \\ v_\theta \end{bmatrix}}_v \quad (3)$$

where

$$\mathbf{H} = \begin{bmatrix} \left. \frac{\partial d_1}{\partial x_c} \right|_* & \left. \frac{\partial d_1}{\partial y_c} \right|_* & \left. \frac{\partial d_1}{\partial \beta_1} \right|_* & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \left. \frac{\partial d_n}{\partial x_c} \right|_* & \left. \frac{\partial d_n}{\partial y_c} \right|_* & 0 & \cdots & \left. \frac{\partial d_n}{\partial \beta_n} \right|_* \\ \left. \frac{\partial \theta_1}{\partial x_c} \right|_* & \left. \frac{\partial \theta_1}{\partial y_c} \right|_* & \left. \frac{\partial \theta_1}{\partial \beta_1} \right|_* & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \left. \frac{\partial \theta_n}{\partial x_c} \right|_* & \left. \frac{\partial \theta_n}{\partial y_c} \right|_* & 0 & \cdots & \left. \frac{\partial \theta_n}{\partial \beta_n} \right|_* \end{bmatrix}.$$

The matrix calculation for the weighted least squares estimator is

$$\hat{\mathbf{x}} = (\mathbf{H}^\top \mathbf{V} \mathbf{H})^{-1} \mathbf{H}^\top \mathbf{V} \mathbf{z} \quad (4)$$

$$\hat{\mathbf{P}} = (\mathbf{H}^\top \mathbf{V} \mathbf{H})^{-1} \quad (5)$$

The estimated centroid (i.e.,  $\hat{x}_c$  and  $\hat{y}_c$ ) is defined as the measurement of the landmark position relative to the Lidar. These states can also be transformed into polar coordinates,

$$\hat{d} = \sqrt{\hat{x}_c^2 + \hat{y}_c^2} \quad (6)$$

$$\hat{\theta} = \arctan \left( \frac{\hat{y}_c}{\hat{x}_c} \right) \quad (7)$$

which can then be interpreted as virtual range and angle measurements from the Lidar to the cylinder center, whose absolute location is known from the map. When Lidar measurements, denoted by  $d_i$  and  $\theta_i$  in Equations (1) and (2), are selected from the cylindrical part of the landmark, the landmark centroid, denoted by  $\hat{d}$  and  $\hat{\theta}$ , can be accurately estimated. However, if the points originate from obstacles, the estimates will not correspond to the centroid, potentially leading to feature extraction faults.

## 2. Lidar Position Estimation

The estimated centroid in Equations (6) and (7) is used for vehicle position estimation. The Lidar models estimating vehicle position are

$$\hat{d}^j = \sqrt{(x^j - x_N)^2 + (y^j - y_N)^2} + v_{\hat{d}} \quad (8)$$

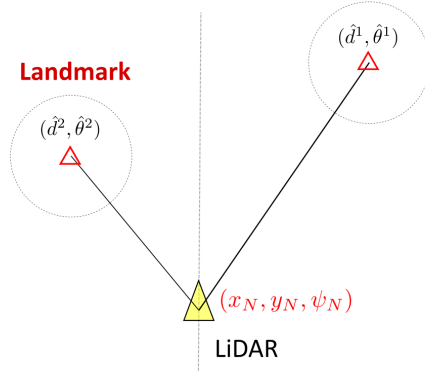
$$\hat{\theta}^j = \arctan \left( \frac{y^j - y_N}{x^j - x_N} \right) - \psi_N + v_{\hat{\theta}} \quad (9)$$

where  $\hat{d}^j$  and  $\hat{\theta}^j$  represent the  $j$ -th extracted landmark, with index  $j = 1, 2, \dots$  denoting the specific landmark.  $x^j$  and  $y^j$  are the known coordinates of the landmark from the map,  $x_N$ ,  $y_N$ , and  $\psi_N$  are states representing, respectively, the vehicle's along-track and cross-track position coordinates and yaw angle, all with respect to Navigation (i.e., map) frame, and  $v_{\hat{d}}$  and  $v_{\hat{\theta}}$  represent the white Gaussian noise (Figure 3). These models can be integrated with other sensors using an EKF for vehicle position estimation (Nagai et al., 2024). However, unlike GNSS, Lidar signals do not carry pseudo-random codes to identify the source of emission (or reflection, in the case of Lidar), which means the sources of the measurements are not immediately identifiable. Instead, the extracted landmark centroid measurements (i.e.,  $\hat{d}^j$  and  $\hat{\theta}^j$ ) in Equations (8) and (9) are typically associated with pre-defined landmark positions in the map (i.e.,  $x^j$  and  $y^j$ ) by means of nearest neighbor selection. This process, however, can sometimes result in data association faults.

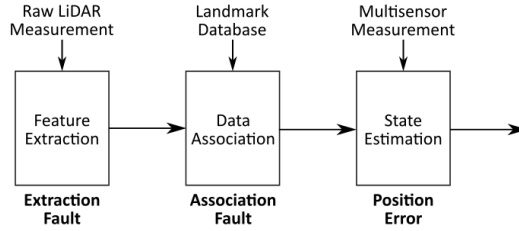
## III. INTEGRITY RISK

### 1. Fault Definitions

We now develop an integrity risk equation that quantifies navigation safety probabilistically. Integrity risk in Lidar positioning is influenced by fault-free position errors, data association faults, and feature extraction faults. Figure 4 illustrates the Lidar positioning process, showing that data association takes place after the feature extraction process.



**Figure 3:** Diagram illustrating the Lidar localization with the states,  $x_N, y_N, \psi_N$ , and landmark extractions,  $\hat{d}^j$  and  $\hat{\theta}^j$ .



**Figure 4:** The positioning process for a Lidar-based navigation system. Faults can occur during the feature extraction and data association processes, threatening state estimation integrity.

We use detection theory to identify faults by applying a binary hypothesis to each procedure; we call this approach the successive hypothesis test. The binary hypothesis testing problem for feature extraction is

$$\begin{cases} H_0 : \text{correct extraction (CE)} \\ H_1 : \text{incorrect extraction (IE)}. \end{cases} \quad (10)$$

$P(CE)$  defines the probability that the extracted feature *corresponds* to a landmark position on the map, and  $P(IE)$  defines the probability that the extracted feature *does not correspond* to a landmark position on the map.

Since data association follows feature extraction, the hypotheses in the association process involve conditional probabilities. The successive binary hypothesis problem for data association given correct extraction is

$$\begin{cases} H_0 : \text{correct association (CA) given CE} \\ H_1 : \text{incorrect association (IA) given CE.} \end{cases} \quad (11)$$

$P(CA|CE)$  denotes the probability that the correctly extracted feature is associated with the *correct* landmark position, and  $P(IA|CE)$  indicates the probability that the correctly extracted feature is associated with the *incorrect* landmark position.

Given incorrect extraction, we have

$$\begin{cases} H_0 : \text{CA given IE} \\ H_1 : \text{IA given IE.} \end{cases} \quad (12)$$

$P(CA|IE)$  denotes the probability that the incorrectly extracted feature is associated with the *correct* landmark position.  $P(IA|IE)$  represents the probability that an incorrectly extracted feature is associated with the *incorrect* landmark position. In our analysis,  $P(CA|IE)$  is set to zero because an incorrectly extracted feature can never be associated correctly. In addition, we conservatively assume that an incorrectly extracted feature will always be associated to *some* mapped landmark, which will clearly always be incorrect. The overview of the fault definitions using probability form is in Table 1.

**Table 1:** Fault Definitions of Lidar Positioning System

Feature Extraction	Data Association
$P(H_0) = P(CE)$	$P(H_0) = P(CA CE)$ $P(H_1) = P(IA CE)$
$P(H_1) = P(IE)$	$P(H_0) = P(CA IE) = 0$ $P(H_1) = P(IA IE) = 1$

## 2. Integrity Risk Equation

Integrity risk is represented by probability of hazardously misleading information (HMI). With the fault definitions in Table 1, the upper bound is described as follows:

$$P(HMI) \leq 1 - (1 - P(HMI|CA, CE))P(CA|CE)P(CE). \quad (13)$$

The derivation of the equation is provided in the Appendix. Equation (13) is the probability at a single epoch, and the integrity risk at any epoch  $n$  is

$$P(HMI_n) \leq 1 - (1 - P(HMI_n|CA_N, CE_N))P(CA_n|CE_N)P(CE_n) \quad (14)$$

where  $N$  denotes all time increments from time epoch 1 to  $n$ .  $P(CA_n|CE_N)$  and  $P(CE_n)$  can be calculated by the following equations (Joerger et al., 2016).

$$P(CA_n|CE_N) = \prod_{l=1}^n P(CA_l|CA_{L-1}, CE_N), \quad L = 1, \dots, l \quad (15)$$

$$P(CE_n) = \prod_{l=1}^n P(CE_l|CE_{L-1}), \quad L = 1, \dots, l \quad (16)$$

We aim to reduce the integrity risk defined in Equation (14), which accounts for position estimation error, association faults, and extraction faults, to meet the integrity requirements for driverless vehicles in urban environments.

## 3. Integrity Requirements

It is necessary to reduce integrity risk according to the requirements for self-driving vehicles, although specific quantitative criteria are not yet defined. We focus on creating a methodology to assess Lidar integrity performance, which allows us to proceed without detailed knowledge of these criteria. However, to generate useful quantitative outcomes, we will need to make some reasonable assumptions.

To ensure integrity, a vehicle may compute a protection level that bounds its horizontal positioning error within a specified probability of exceedance. *Fault-free* integrity is evaluated by determining whether the protection level exceeds a required alert limit, which defines a position domain containment volume (RTCA (Firm). SC-159, 2004). Alternatively, the probability of exceeding the alert limit can be computed directly. This probability, denoted by  $P(HMI_n|CA_N, CE_N)$  in Equation (14), can be calculated using the variance of the position state:

$$P(HMI_n | CA_N, CE_N) = 2\Phi \left[ -\frac{AL}{\bar{\sigma}_n} \right] \quad (17)$$

where  $\Phi[\cdot]$  is the standard normal cumulative distribution function (CDF),  $AL$  is the alert limit, and  $\bar{\sigma}_n$  is a standard deviation of the position estimate error under consideration (e.g.,  $\bar{\sigma}_x$  or  $\bar{\sigma}_y$ ).

*Faults*, which are errors in sensor measurements that evade immediate detection, can also contribute to position inaccuracies. The specific requirements for allowable faults remain undefined; however, for our current purposes, we assume that the probability allocation for association and extraction risk is  $2 \times 10^{-7}$  per exposure time. This probability corresponds to Equations (15) and (16):

$$P(CA_n|CE_N) = \prod_{l=1}^n \underbrace{(P(CA_l|CA_{l-1}, CE_N))}_{\triangleq (1-P(IA_l))} > 1 - 2 \times 10^{-7} \quad (18)$$

$$P(CE_n) = \prod_{l=1}^n \underbrace{(P(CE_l|CE_{l-1}))}_{\triangleq (1-P(IE_l))} > 1 - 2 \times 10^{-7}. \quad (19)$$

The equations indicate that the risk is the cumulative effect of individual moments. Assuming a self-driving event operates for 1 hour with a Lidar measurement frequency of 1 Hz, the allowable association and extraction integrity risks are each allocated  $10^{-11}$  per moment:

$$P(IA_l) < 10^{-11} \quad (20)$$

$$P(IE_l) < 10^{-11}. \quad (21)$$

Table 2 provides an overview of the self-driving car integrity requirement assumptions in our work.

**Table 2:** Example of integrity requirements for driverless vehicles in urban environments

	upper (EUSPA, 2021)	lower (Reid et al., 2019)
availability	> 99.9%	(> 99.9%)
probability of exceedance	(< $10^{-7}$ per moment)	< $10^{-8}$ per moment
protection level	$5\sigma_{\text{position}}$	$6\sigma_{\text{position}}$
lateral and longitudinal alert limit ( $AL$ )	(< 0.5 m)	< 0.3 m
the lateral and longitudinal maximum allowable $\sigma_n$	< 0.1 m	< 0.05 m
allowable association and extraction integrity risk	(< $10^{-7}$ per exposure time, < $10^{-11}$ per moment)	

( ) represents a value used in our analysis but not specified in the cited paper.

## IV. PRIOR WORK

### 1. Fault-free Integrity

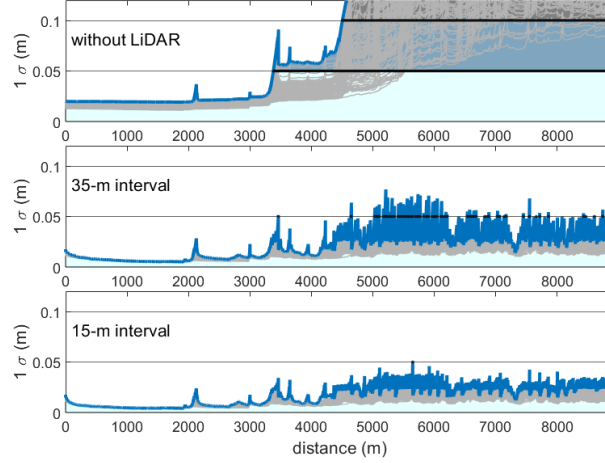
We utilize a multi-sensor integrated navigation system, including INS, Lidar, GNSS, zero velocity update (ZUPT), wheel speed sensors, and vehicle kinematic constraints, combined in an EKF for precise positioning, to assess the fault-free capability of self-driving cars in urban environments (Nagai et al., 2024).

Figure 5 shows the position error standard deviation along the track ( $\bar{\sigma}_x$ ) simulated in downtown Chicago, where GNSS signals are frequently compromised. Without Lidar, the system does not achieve the integrity requirements due to insufficient accuracy. However, when Lidar positioning is utilized with landmark intervals between 15 and 35 m, the position error falls below the maximum allowable position error standard deviation (Nagai et al., 2024). The landmark density determines the Lidar position reference update rate, and the results demonstrate that narrower landmark intervals are beneficial for accurate positioning.

### 2. Association Risk

The Lidar positioning system requires pairing each measurement with its corresponding landmark location on the map, as shown in Equations (8) and (9). Unlike pseudo-random noise codes in GNSS, Lidar signals lack a tagging system to identify the source of each measurement. Bar-Shalom and Fortmann (1988) employed the nearest neighbor algorithm, which selects the optimal pair based on the smallest Mahalanobis distance. While this approach can successfully associate Lidar measurements with map information, it may result in incorrect associations due to measurement noise and state estimate errors, especially when landmarks are close to each other.

The probability of association risk in Equation (15) can be calculated using a Chi-squared distribution, and the lower bound risk equation is derived in (Joerger et al., 2016; Hassani and Joerger, 2023). We demonstrated that a landmark spacing of a few meters was sufficient to maintain the probability below  $10^{-11}$  (Nagai et al., 2023), thereby meeting the integrity requirements. While a dense landmark distribution can minimize positional errors, as discussed in Section IV-1, a more spaced-out arrangement is necessary to reduce the risk of incorrect associations. Thus, landmark density plays a crucial role in ensuring the safety of driverless vehicles with Lidar navigation, emphasizing the need for careful selection of landmarks.

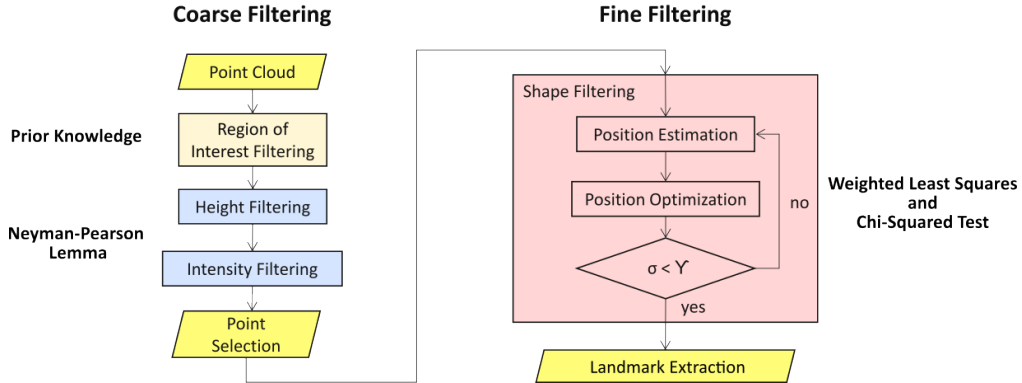


**Figure 5:** Position error along the track simulated in downtown Chicago without Lidar, with Lidar at the 35-meter intervals, and at the 15-meter intervals.

## V. FEATURE EXTRACTION

### 1. Methodology for Point Selection

Estimating landmark positions from large volumes of Lidar point cloud while minimizing extraction faults is challenging. We implement a coarse-to-fine approach to address this issue. The method involves identifying landmarks from scan segments and then refining the focus to point scales (Tee-Ann and Chi-Min, 2015; Golovinskiy et al., 2009). We apply coarse filtering for point selection, identifying points from the point cloud that belong to landmarks. Afterward, fine filtering removes any outliers that pass through the coarse filtering. Figure 6 provides an overview of our feature extraction algorithm based on this approach.

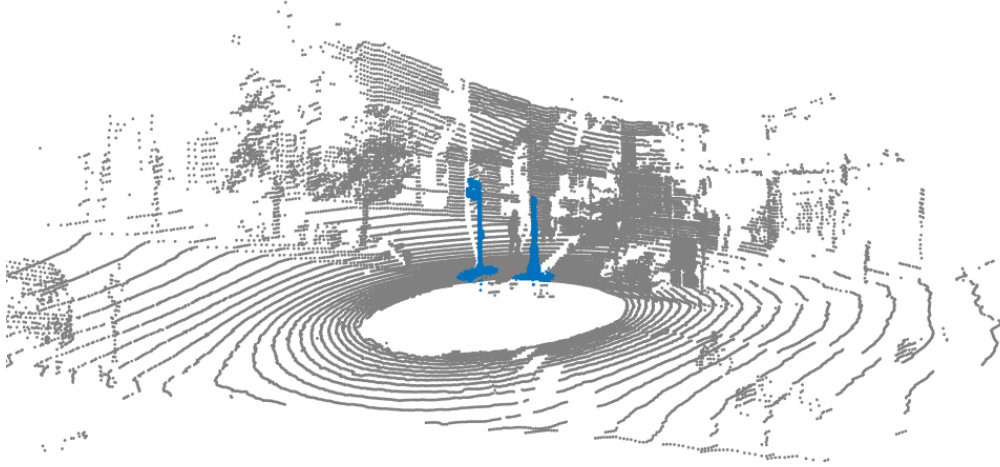


**Figure 6:** The coarse-to-fine filtering approach is used for feature extraction. Hypothesis tests are applied at each stage of the filtering process.

The coarse filtering process begins with selecting points near landmarks using a map, which serves as prior knowledge. For instance, the extraction region could be defined as the area within a 1-meter radius of the landmark’s center, as illustrated in Figure 7.

After filtering (i.e., keeping only) points near mapped landmarks, height and intensity filters are applied. We determine whether these points belong to landmarks using a binary hypothesis test utilizing the Neyman-Pearson criterion. Given the prevalence of more obstacles than landmarks in the environment, the two hypotheses are defined as follows:

$$\begin{cases} H_0 : \text{obstacle} \\ H_1 : \text{landmark.} \end{cases} \quad (22)$$



**Figure 7:** Example of region-of-interest filtering: the extraction region is defined as a 1-meter radius around the landmark's center.

The likelihood functions, represented by the probability density of the normal distribution, are expressed as

$$\begin{cases} p(z|H_0) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{1}{2\sigma_0^2}(z - \mu_0)^2\right) \\ p(z|H_1) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{1}{2\sigma_1^2}(z - \mu_1)^2\right) \end{cases} \quad (23)$$

where  $\sigma$  represents the standard deviation of intensity (or height), and  $\mu$  is the mean, both of which can be obtained from publicly available data or experiments.

The likelihood ratio test is given by

$$\Lambda(z) = \frac{p(z|H_1)}{p(z|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \quad (24)$$

where  $\gamma$  is a threshold determined by integrity risk requirements discussed in Section III. The test has four possible outcomes:

1. pick  $H_0$  given  $H_0 \rightarrow$  correct
2. pick  $H_0$  given  $H_1 \rightarrow$  no-extraction (NE)
3. pick  $H_1$  given  $H_1 \rightarrow$  correct
4. pick  $H_1$  given  $H_0 \rightarrow$  incorrect extraction (IE).

Only decision (4) poses an integrity threat.

Using the Neyman-Pearson Lemma, the incorrect extraction risk per moment can be calculated as

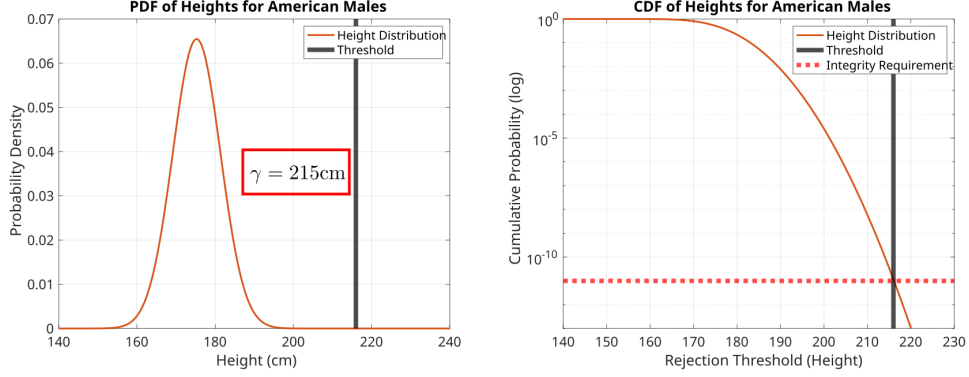
$$P(IE_l) = \int_{\{z:\Lambda(z)>\gamma\}} p(z|H_0) dx. \quad (25)$$

and the no-extraction event per moment can be calculated as

$$P(NE_l) = \int_{\{z:\Lambda(z)<\gamma\}} p(z|H_1) dx. \quad (26)$$

## 2. Application of Neyman-Pearson Hypothesis Test

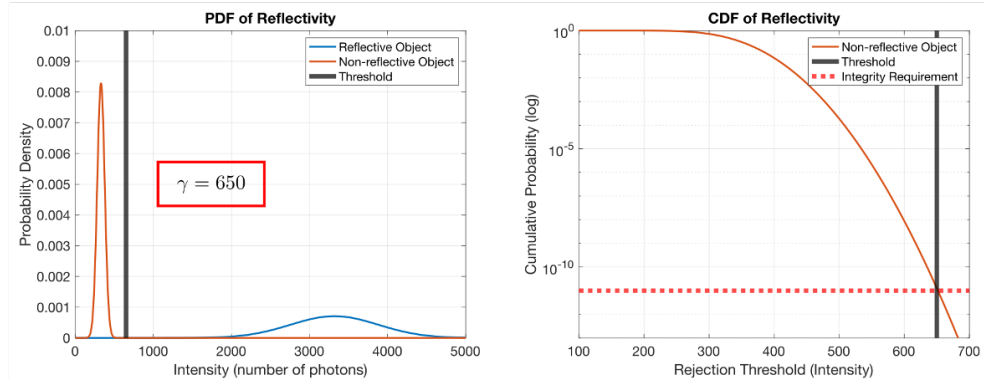
This Neyman-Pearson hypothesis test can be applied to both height and intensity filtering. According to U.S. Department of Health and Human Services (2021), the average height of American males is 175 cm with a standard deviation of 6 cm.



**Figure 8:** Height filtering: we can define the detection threshold as 215 cm, based on  $10^{-11}$  fault integrity requirement.

By combining the height distribution function with the integrity requirement for extraction faults, we can define the detection threshold as 215 cm (Figure 8). This result indicates that the probability of a man exceeding a height of 215 cm is only  $10^{-11}$ .

The height data for American males is publicly available, but the intensity distribution for reflective and non-reflective objects was previously unknown. We conducted experiments to obtain these distributions in Nagai et al. (2024), and Figure 9 presents the intensity decision threshold based on the experimental results. When the intensity exceeds the detection threshold of 650, measured in signal photons, the system can detect reflective objects.



**Figure 9:** Intensity filtering: we can define the detection threshold as 650, based on  $10^{-11}$  fault integrity requirement.

### 3. Methodology for Landmark Extraction

Once the relevant points are selected, the system estimates the landmark's centroid, as described in Equation (4). Estimation determines the best value based on available data but does not serve to exclude outliers passing through height and intensity filtering. Fine filtering identifies outliers by analyzing deviations from the circular shape of pole-like landmarks using Chi-square tests. We call the integration of estimation and detection landmark extraction, and this approach improves the estimation of the landmark's centroid.

We calculate the Mahalanobis distance after estimating the centroid using Equation (4). By applying the estimation results to Equations (1) and (2), the Lidar measurement model is

$$\underbrace{\begin{bmatrix} d_i \\ \theta_i \end{bmatrix}}_{z_i} = \underbrace{\begin{bmatrix} \sqrt{(\hat{x}_c + r \cos \hat{\beta}_i)^2 + (\hat{y}_c - r \sin \hat{\beta}_i)^2} \\ \tan^{-1} \frac{\hat{y}_c - r \sin \hat{\beta}_i}{\hat{x}_c + r \cos \hat{\beta}_i} \end{bmatrix}}_{h(\hat{x}_i)} + \underbrace{\begin{bmatrix} v_d \\ v_\theta \end{bmatrix}}_{\Lambda} \quad (27)$$

where  $d_i$  and  $\theta_i$  are Lidar measurements of point  $i$ ,  $\hat{x}_c, \hat{y}_c, \hat{\beta}_i$  are the state estimates,  $r$  is the known radius from a map, and  $v_d, v_\theta$  are white Gaussian measurement noises acquired from Lidar specifications. The residuals between the measurements

and estimations are

$$\gamma_i = z_i - h(\hat{x}_i). \quad (28)$$

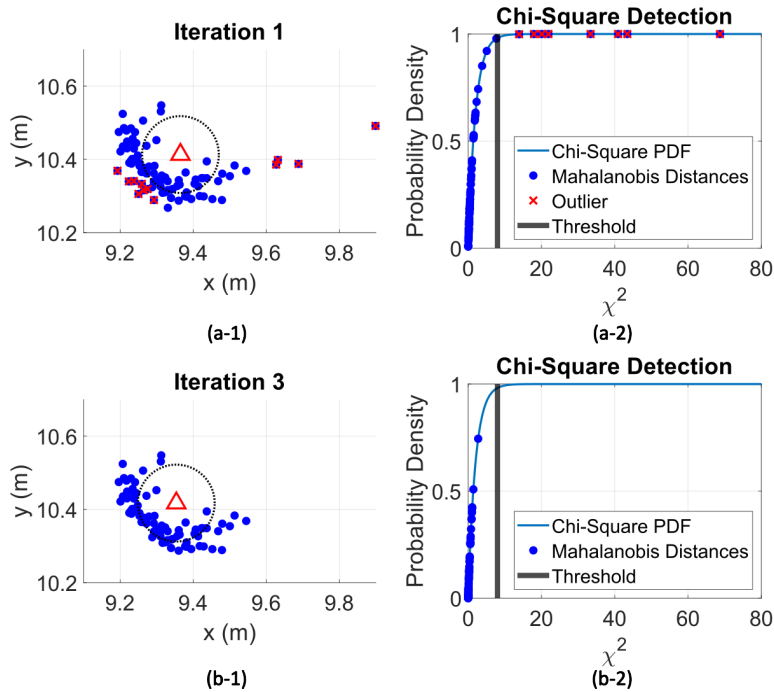
The residuals represent the comparison of measurements from the real point to the ideal point on the circle centered at the estimated position. The normalization of Equation (28) gives the square of the Mahalanobis distance:

$$y_i^2 = \gamma_i^T \Lambda^{-1} \gamma_i. \quad (29)$$

A chi-square test screens outliers by comparing the Mahalanobis distance of the point to a threshold. The threshold is set at 8, given 2 degrees of freedom (i.e.,  $d_i$  and  $\theta_i$ ), which corresponds to 3 standard deviations (98%). After excluding outliers, the landmark's centroid is re-estimated until all measurement points fall within the threshold.

#### 4. Demonstration of Chi-square Hypothesis Test

Figure 10 (a-1) presents an experimental data points, estimated centroid, and outliers. All points in this figure are Lidar measurements of a landmark, selected through coarse filtering, and the red triangle represents the estimated centroid. Outliers, detected by the chi-square test, are marked with red crosses. Figure 10 (a-2) presents the chi-square test results, with points representing the Mahalanobis distance for each point in (a-1). Several points exceed the threshold, corresponding to the red-crossed points shown in (a-1). In each iteration, the algorithm estimates the centroid of the circle, detects outliers and removes them from the measurements. Once all points fall within the threshold, as observed in Figure 10 (b-1, b-2), the iteration stops, and the landmark extraction is completed.



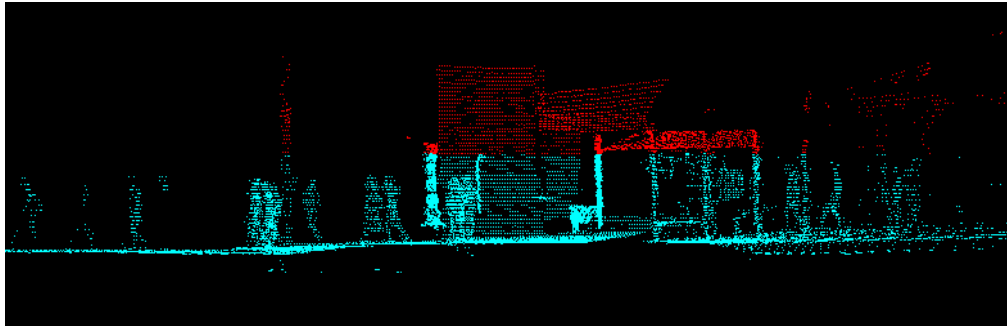
**Figure 10:** This figure illustrates how our algorithm operates using experimental LiDAR point measurements. It includes outlier detection and centroid re-estimation at both the initial and final iterations. The process ensures that all points used for estimation are within three standard deviations of the LiDAR measurement error.

## VI. EXPERIMENT

### 1. Height Filtering

The first experiment investigates the effects of height filtering to exclude measurement points from people on a street. We hypothesize that a person taller than 215 cm exists with a probability of only  $10^{-11}$ , meaning such an individual is unlikely to exist. To test this, point clouds were collected on State Street, Chicago, IL, U.S., over 8 km using an Ouster OS-1 64. The experiment was conducted in August 2024, on a Sunday. The data showed more than 500 people on the street, and none were

taller than 215 cm. The tallest individual was a child on his father’s shoulders, considered an abnormal case. These findings support the hypothesis that 215 cm works as an effective threshold to exclude points corresponding to people. Further research is needed to explore abnormal cases, such as children on shoulders. If such cases occur frequently, we may need to consider a higher threshold.

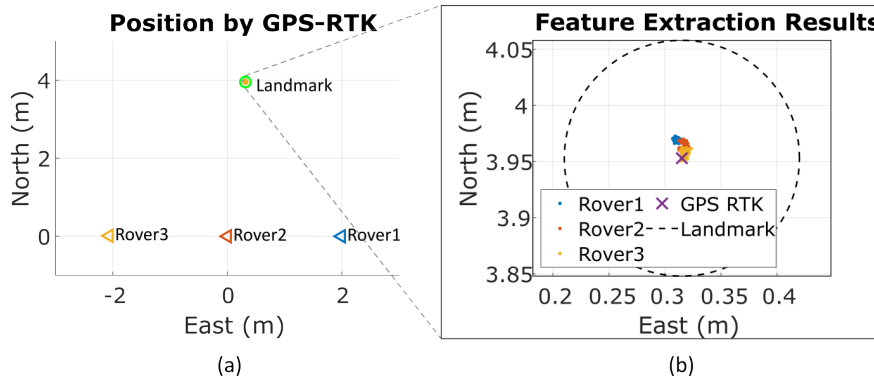


**Figure 11:** Example of point clouds from the height experiment. Red dots represent points taller than 215 cm, while blue dots represent points shorter than that. There are many people, but none exceed the 215 cm threshold.

## 2. Landmark Extraction

We conducted a landmark extraction experiment in August 2024 on the Illinois Institute of Technology campus in an open-sky environment to validate our methodology. GPS RTK was used to obtain truth reference positions of a cylindrical landmark relative to three stationary rover positions shown in Figure 12 (a). Point clouds of a landmark were captured for 100 frames at each of the three rover positions using a Velodyne VLP-16. We utilized this data to estimate the centroid employing our landmark extraction methodology.

Figure 12 (b) presents the centroid estimation results, displaying the landmark’s perimeter with the center evaluated using RTK GPS. The results are color-coded based on the measurements acquired from each rover position. All landmark extraction results fall within 1 cm, confirming that the methodology effectively minimizes the risk of extraction faults.



**Figure 12:** Landmark extraction points, with colors corresponding to each rover position. All landmark extraction results are within 1 cm, validating the methodology.

## VII. CONCLUSIONS

This study proposes a methodology to reduce extraction faults in map-based landmark Lidar localization. It begins by examining the integrity risks in Lidar navigation systems for self-driving cars and identifying the factors behind incorrect extractions. Example Lidar-fault integrity requirements are introduced, guiding a data selection process based on the Neyman-Pearson criterion. Landmark extraction uses weighted least squares with a chi-square test to reject excessive Lidar noise events and outliers. Experimental results demonstrate that the methodology effectively reduces extraction faults, achieving centimeter-level accuracy in landmark positioning. Future studies will focus on extending experiments to dynamic vehicle scenarios, using an intensity filter and incorporating real urban environments.

## ACKNOWLEDGEMENTS

This article is based on work supported by the Center for Assured and Resilient Navigation in Advanced TransportatION Systems (CARNATIONS) under the US Department of Transportation (USDOT)'s University Transportation Center (UTC) program (Grant No. 69A3552348324). Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the sponsors.

## REFERENCES

- American Automobile Association (2023). Fear of self-driving cars on the rise. <https://newsroom.aaa.com/2023/03/aaa-fear-of-self-driving-cars-on-the-rise/>.
- Bar-Shalom, Y., Daum, F., and Huang, J. (2009). The probabilistic data association filter. *IEEE Control Systems Magazine*, 29(6):82–100.
- Bar-Shalom, Y. and Fortmann, T. (1988). *Tracking and Data Association*. Mathematics in science and engineering. Academic Press.
- Benedek, C., Majdik, A., Nagy, B., Rozsa, Z., and Sziranyi, T. (2021). Positioning and perception in LIDAR point clouds. *Digital Signal Processing*, 119:103193.
- Chapman, L., Thornes, J. E., and Bradley, A. V. (2002). Sky-view factor approximation using GPS receivers. *International Journal of Climatology*, 22(5):615–621.
- Cosmen-Schortmann, J., Azaola-Sáenz, M., Martínez-Olague, M., and Toledo-López, M. (2008). Integrity in urban and road environments and its use in liability critical applications. In *2008 IEEE/ION Position, Location and Navigation Symposium*, pages 972–983.
- Davis, J. M. and Kelly, R. J. (1993). RNP tunnel concept for precision approach with GNSS application. In *Proceedings of the 1993 Annual Meeting of The Institute of Navigation*, pages 135–154.
- EUSPA (2021). Report on road user needs and requirements. <https://www.gsc-europa.eu>.
- Falco, G., Pini, M., and Marucco, G. (2017). Loose and tight GNSS/INS integrations: Comparison of performance assessed in real urban scenarios. *Sensors*, 17(2):255.
- Gao, J., Petovello, M., and Cannon, M. (2006). Development of precise GPS/INS/wheel speed sensor/yaw rate sensor integrated vehicular positioning system. In *Proceedings of the 2006 National Technical Meeting of The Institute of Navigation*, pages 780–792.
- Gim, J., Ahn, C., and Peng, H. (2021). Landmark attribute analysis for a high-precision landmark-based local positioning system. *IEEE Access*, 9:18061–18071.
- Godha, S. and Cannon, M. (2007). GPS/MEMS INS integrated system for navigation in urban areas. *GPS Solutions*, 11(3):193–203.
- Golovinskiy, A., Kim, V. G., and Funkhouser, T. (2009). Shape-based recognition of 3D point clouds in urban environments. In *2009 IEEE 12th International Conference on Computer Vision*, pages 2154–2161. IEEE. <https://doi.org/10.1109/ICCV.2009.5459471>.
- Grejner-Brzezinska, D. A., Yi, Y., and Toth, C. K. (2001). Bridging GPS gaps in urban canyons: The benefits of ZUPTs. *NAVIGATION: Journal of the Institute of Navigation*, 48(4):216–226.
- Grimes, J. (2007). Global positioning system precise positioning service performance standard. *Department of Defense, United States of America*.
- Hassani, A. and Joerger, M. (2023). Analytical and empirical navigation safety evaluation of a tightly integrated lidar/imu using return-light intensity. *NAVIGATION: Journal of the Institute of Navigation*, 70(4).
- Hazlett, A., Crassidis, J., Fuglewicz, D., and Miller, P. (2011). Differential wheel speed sensor integration with GPS/INS for land vehicle navigation. *AIAA Guidance, Navigation, and Control Conference*, page 6577.
- Jo, K., Kim, C., and Sunwoo, M. (2018). Simultaneous localization and map change update for the high definition map-based autonomous driving car. *Sensors*, 18(9).
- Joerger, M., Jamoom, M., Spenko, M., and Pervan, B. (2016). Integrity of laser-based feature extraction and data association. In *2016 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, pages 557–571. IEEE.

- Keitaanniemi, A., Rönholm, P., Kukko, A., and Vaaja, M. T. (2023). Drift analysis and sectional post-processing of indoor simultaneous localization and mapping (SLAM)-based laser scanning data. *Automation in Construction*, 147:104700.
- Levinson, J., Montemerlo, M., and Thrun, S. (2007). Map-based precision vehicle localization in urban environments. In *Robotics: science and systems*, volume 4, page 1.
- Levinson, J. and Thrun, S. (2010). Robust vehicle localization in urban environments using probabilistic maps. In *2010 IEEE International Conference on Robotics and Automation*, pages 4372–4378.
- Nagai, K., Chen, Y., Spenko, M., Henderson, R., and Pervan, B. (2023). Integrity with extraction faults in LiDAR-based urban navigation for driverless vehicles. In *2023 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, pages 1099–1106. IEEE. <https://doi.org/10.1109/PLANS53410.2023.10140132>.
- Nagai, K., Fasoro, T., Spenko, M., Henderson, R., and Pervan, B. (2020). Evaluating GNSS navigation availability in 3-D mapped urban environments. In *2020 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, pages 639–646.
- Nagai, K., Spenko, M., Henderson, R., and Pervan, B. (2024). Fault-free integrity of urban driverless vehicle navigation with multi-sensor integration: A case study in downtown Chicago. *NAVIGATION: Journal of the Institute of Navigation*, 71(1).
- Pirovano, L., Principe, G., and Armellin, R. (2020). Data association and uncertainty pruning for tracks determined on short arcs. *Celestial Mechanics and Dynamical Astronomy*, 132(1):6.
- Reid, T. G., Houts, S. E., Cammarata, R., Mills, G., Agarwal, S., Vora, A., and Pandey, G. (2019). Localization requirements for autonomous vehicles. *SAE International Journal of Connected and Automated Vehicles*, 2(3):1–16.
- RTCA (Firm). SC-159 (2004). *Minimum Aviation System Performance Standards for the Local Area Augmentation System (LAAS)*. RTCA.
- Sefati, M., Daum, M., Sondermann, B., Kreisköther, K. D., and Kampker, A. (2017). Improving vehicle localization using semantic and pole-like landmarks. In *2017 IEEE Intelligent Vehicles Symposium (IV)*, pages 13–19. IEEE.
- Tee-Ann, T. and Chi-Min, C. (2015). Pole-like road object detection from mobile lidar system using a coarse-to-fine approach. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(10):4805–4818.
- The National Highway Traffic Safety Administration (2023). Rural/urban comparison of motor vehicle traffic fatalities. <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813336>.
- The National Science and Technology Council and the United States Department of Transportation (2020). Ensuring american leadership in automated vehicle technologies. <https://www.transportation.gov/sites/dot.gov/files/2020-02/EnsuringAmericanLeadershipAVTech4.pdf>.
- Titterton, D. and Weston, J. (2004). *Strapdown Inertial Navigation Technology*. Radar, Sonar and Navigation. Institution of Engineering and Technology.
- Umbach, D. and Jones, K. (2003). A few methods for fitting circles to data. *IEEE Transactions on Instrumentation and Measurement*, 52(6):1881–1885.
- U.S. Department of Health and Human Services (2021). Anthropometric reference data for children and adults: United states, 2015–2018. [https://www.cdc.gov/nchs/data/series/sr\\_03/sr03-046-508.pdf](https://www.cdc.gov/nchs/data/series/sr_03/sr03-046-508.pdf).
- Wan, G., Yang, X., Cai, R., Li, H., Zhou, Y., Wang, H., and Song, S. (2018). Robust and precise vehicle localization based on multi-sensor fusion in diverse city scenes. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4670–4677. IEEE.
- Zhu, N., Marais, J., Bétaille, D., and Berbineau, M. (2018). GNSS position integrity in urban environments: A review of literature. *IEEE Transactions on Intelligent Transportation Systems*, 19(9):2762–2778.

## APPENDIX: DERIVATION OF INTEGRITY RISK EQUATION

From the law of total probability, the integrity risk presented by hazardous misleading information (HMI) is

$$P(HMI) = P(HMI|H_0)P(H_0) + P(HMI|H_1)P(H_1) \quad (30)$$

where  $P(H_0)$  and  $P(H_1)$  are the prior probabilities of each of the two mutually exclusive hypotheses.  $P(HMI|H_0)$  and  $P(HMI|H_1)$  are the corresponding risks of HMI. Considering the feature extraction procedure defined in Section III, (30) can

be replaced by the following:

$$P(HMI) = P(HMI|CE)P(CE) + P(HMI|IE)P(IE). \quad (31)$$

Since the data association process is given correct extraction in (11),  $P(HMI|CE)$  becomes

$$P(HMI|CE) = P(HMI|CA, CE)P(CA|CE) + P(HMI|IA, CE)P(IA|CE). \quad (32)$$

The position error resulting from an incorrect extraction and incorrect association is unknown, and we take the upper bound of the conditional probabilities of HMI in these cases as 1:

$$P(HMI|IE) \stackrel{\text{def}}{=} 1 \quad (33)$$

$$P(HMI|IA, CE) \stackrel{\text{def}}{=} 1 \quad (34)$$

Due to the mutually exclusive conditions presented in Equations (10) and (11), we can express  $P(IE)$  and  $P(IA|CE)$  using the following equations:

$$P(IE) = 1 - P(CE) \quad (35)$$

$$P(IA|CE) = 1 - P(CA|CE) \quad (36)$$

Combining (31) - (36), the integrity risk's upper bound is

$$P(HMI) \leq 1 - (1 - P(HMI|CA, CE))P(CA|CE)P(CE). \quad (37)$$