

Nr. 401

Jeldrik Axmann

Maximum consensus localization using LiDAR

Nr. 401

Maximum consensus localization using LiDAR

Von der Fakultät für Bauingenieurwesen und Geodäsie
der Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des Grades

DOKTOR-INGENIEUR (Dr.-Ing.)

genehmigte Dissertation
von

Jeldrik Axmann, M.Sc.

HANNOVER 2024

Prüfungskommission

Vorsitzender:	Prof. Dr.-Ing. habil. Christian Heipke
Referent:	Prof. Dr.-Ing. Claus Brenner
Korreferenten:	Prof. Dr.-Ing. Steffen Schön
	Prof. Dr. Andreas Nüchter

Tag der mündlichen Prüfung: 10.07.2024

Abstract

The basic requirement for autonomous driving is an accurate localization of the vehicle, where not only accuracy is important, but also reliability, to provide correct information even in difficult situations, and integrity, to ensure safe operation. Reliability describes the occurrence of failures and is related to the robustness of an operation, while integrity refers to a system's ability to correctly and timely alert the user when certain error limits are exceeded.

Often, the vehicle pose is estimated under the assumption of normally distributed errors, resulting in a quadratic loss function that can be solved efficiently. However, in the presence of outliers causing deviations from the distribution function that cannot be modeled, this assumption leads to an incorrect estimate of the system state. Outliers arise from false associations caused by, for example, outdated maps, dynamic traffic participants, or rain. To reduce the susceptibility to outliers, more robust loss functions have been proposed, the most robust of which is inlier counting, also known as maximum consensus optimization, where measurements with residuals greater than a predefined threshold do not affect the result, making it ideal for robust parameter estimation.

To determine the ego-pose, autonomous vehicles are equipped with a sensor suite that combines global positioning sensors, such as GNSS receivers, and relative positioning sensors, such as cameras and LiDAR sensors. This work focuses on localization using LiDAR, which is based on the registration of point clouds. It investigates the application of the robust maximum consensus criterion and other loss functions to LiDAR-based localization and explores the application of the concept of integrity. Overall, it presents a robust and reliable localization framework, where relative positioning between a sparse 'car sensor' point cloud from a LiDAR mounted on the ego-vehicle and a dense, high resolution 'map' point cloud is performed. To address the non-convexity of the estimation problem, candidate poses within a certain parameter search space are evaluated in an exhaustive manner. This search strategy guarantees global optimality and provides full knowledge of the parameter space, which is then used to evaluate the integrity of the localization.

First, in the basic approach, the candidate poses are evaluated based on the number of measurements that match the map, known as the inlier count. The correspondences between the car sensor scan and the map are established in two ways. The first is based on projective data associations using rendered synthetic range images for each candidate pose, and the second is based on the Euclidean point-to-point distance between scan and map points after transforming the measured scan according to each candidate pose. Second, a new objective function based on Helmert's point error is proposed, where instead of simply counting matches, the constraints generated by each correspondence with respect to the unknown pose are modelled. The approach considers the uncertainty of a point-to-plane adjustment, and the corresponding score is defined as the inverse of the square of Helmert's point error, which results in a lower number of localization failures and thus higher robustness, especially in longitudinal direction while driving along straight streets. Third, in addition to the maximum consensus criterion, other well-known M-estimators and a data-adapted loss function based on a beam model for range sensors are investigated for the task of LiDAR-based localization. Fourth, by estimating discrete probability distributions, single positioning is extended to a filtering strategy. To preserve the full knowledge of the explored parameter space and to avoid the shortcomings of Monte Carlo-based filtering techniques, a histogram filter is used. It further increases the reliability and reduces the number of large localization errors. Finally, a protection level definition is proposed that exploits the knowledge of the parameter space. It enables a safe operation, and it allows to derive localization integrity levels by taking into account the true position error and an alert limit.

Keywords: LiDAR, Localization, Robust estimation, Integrity

Kurzfassung

Die Grundvoraussetzung für autonomes Fahren ist eine genaue Lokalisierung des Fahrzeugs, wobei nicht nur die Genauigkeit wichtig ist, sondern auch die Zuverlässigkeit, auch in schwierigen Situationen korrekte Informationen zu liefern, und die Integrität, um eine sichere Fahrt zu gewährleisten. Die Zuverlässigkeit beschreibt das Auftreten von Fehlern und bezieht sich auf die Robustheit der Positionierung, während sich die Integrität auf die Fähigkeit eines Systems bezieht, den Benutzer korrekt und rechtzeitig zu warnen, wenn bestimmte Fehlergrenzen überschritten werden.

Häufig wird die Position des Fahrzeugs unter der Annahme normalverteilter Fehler geschätzt, was zu einer quadratischen Verlustfunktion führt, die effizient gelöst werden kann. Bei Vorhandensein von Ausreißern, die zu nicht modellierbaren Abweichungen von der Verteilungsfunktion führen, führt diese Annahme jedoch zu einer falschen Positionsschätzung. Ausreißer entstehen durch falsche Assoziationen, die z. B. durch veraltete Karten, andere Verkehrsteilnehmer oder Regen verursacht werden. Um die Anfälligkeit für Ausreißer zu verringern, wurden robustere Verlustfunktionen vorgeschlagen, von denen die robusteste die Konsensmaximierung ist, bei der Messungen mit Residuen, die größer als ein vordefinierter Grenzwert sind, das Ergebnis nicht beeinflussen, was diese ideal für eine robuste Parameterschätzung macht.

Zur Bestimmung der Ego-Position sind autonome Fahrzeuge sowohl mit globalen Positionssensoren, wie GNSS-Empfängern, als auch mit relativen Positionssensoren, wie Kameras und LiDAR-Sensoren, ausgestattet. Diese Arbeit konzentriert sich auf die Lokalisierung mittels LiDAR, die auf der Registrierung von Punktwolken basiert. Die Arbeit untersucht die Anwendung des robusten *Maximum Consensus*-Kriteriums und anderer Verlustfunktionen auf die LiDAR-basierte Lokalisierung und erforscht die Anwendung des Konzepts der Integrität. Insgesamt wird ein robustes und zuverlässiges Lokalisierungsverfahren vorgestellt, bei dem die relative Positionierung zwischen einer spärlichen Punktwolke eines am Ego-Fahrzeug montierten LiDARs und einer dichten, hochauflösenden Punktwolke einer Karte erfolgt. Um der Nicht-Konvexität des Schätzproblems zu begegnen, werden die möglichen Posen innerhalb eines bestimmten Parametersuchraums erschöpfend bewertet. Diese Suchstrategie garantiert globale Optimalität und liefert eine vollständige Kenntnis des Parameterraums, die dann zur Bewertung der Integrität der Lokalisierung verwendet wird.

Im ersten, grundlegenden Ansatz werden die potenziellen Posen anhand der Anzahl der Messungen bewertet, die mit der Karte übereinstimmen, des sogenannten 'Inlier Counts'. Die Korrespondenzen zwischen dem Fahrzeugsensor-Scan und der Karte werden auf zwei Arten hergestellt. Die erste basiert auf projektiven Datenassoziationen unter Verwendung von gerenderten synthetischen Distanzbildern für jede Kandidatenposition und die zweite basiert auf dem euklidischen Punkt-zu-Punkt-Abstand zwischen Scan- und Kartenpunkten nach der Transformation des gemessenen Scans entsprechend jeder Kandidatenposition. Für den zweiten Lokalisierungsansatz wird eine neue Zielfunktion auf der Grundlage des Helmert'schen Punktfehlers vorgeschlagen, bei der anstelle der einfachen Konsens-Zählung die durch jede Korrespondenz erzeugten Bedingungen in Bezug auf die unbekannte Pose modelliert werden. Der Ansatz berücksichtigt die Unsicherheit einer Punkt-zu-Ebenen-Ausgleichung, wobei für die Bewertung einer Pose die Inverse des Quadrats des Helmert'schen Punktfehlers verwendet wird, was zu einer geringeren Anzahl von Lokalisierungsfehlern und damit zu einer höheren Robustheit führt, insbesondere in Längsrichtung beim Fahren auf geraden Straßen. Drittens werden neben der Konsensmaximierung auch bekannte M-Schätzer und eine datenangepasste Verlustfunktion, die auf einem Strahlenmodell für Distanzsensoren basiert, für die Aufgabe der LiDAR-basierten Lokalisierung untersucht. Viertens wird die Einzelpositionierung durch die Berechnung diskreter Wahrscheinlichkeitsverteilungen zu einer Filterstrategie erweitert. Um die vollständige Kenntnis des untersuchten Parameterraums zu bewahren und die Defizite von Monte-Carlo-basierten Filtertechniken zu vermeiden, wird ein His-

togrammfilter verwendet. Dadurch wird die Zuverlässigkeit weiter erhöht und die Zahl der großen Lokalisierungsfehler verringert. Abschließend wird eine Sicherheitslevel-Definition vorgeschlagen, die das Wissen über den Parameterraum nutzt. Sie ermöglicht eine sichere Fahrt und erlaubt die Ableitung von Integritätsniveaus für die Lokalisierung unter Berücksichtigung des wahren Positionsfehlers und einer Alarmgrenze.

Schlagworte: LiDAR, Lokalisierung, Robuste Schätzung, Integrität

Contents

1. Introduction	1
1.1. Motivation	1
1.2. Problem statement and research questions	2
1.3. Contributions	4
1.4. Structure of the thesis	7
2. Fundamentals	9
2.1. Robust estimation techniques	9
2.1.1. Problem definition	9
2.1.2. Maximum consensus formulation	10
2.1.3. M-estimators	12
2.1.4. Beam model for range sensors	15
2.2. State estimation filtering	17
2.2.1. Bayes filter	17
2.2.2. Kalman filter	19
2.2.3. Histogram filter	20
2.3. LiDAR sensor and point cloud acquisition	21
2.3.1. Scan image representation and rendering	23
2.3.2. Point cloud registration	25
2.4. Map representations	26
2.4.1. Topological maps	27
2.4.2. Point cloud maps	27
2.4.3. Occupancy grid maps	29
2.4.4. Normal distributions transform (NDT) and surfel maps	32
2.4.5. Mesh representation using the truncated signed distance function (TSDF)	34
2.5. Basic concepts of integrity	37
3. Related work	39
3.1. Point cloud registration	39
3.1.1. Correspondence-based methods	39
3.1.2. Simultaneous pose and correspondence (SPC) methods	41
3.2. LiDAR-based localization	43
3.3. Discussion	47

4. Methodological contributions	49
4.1. Methodology of maximum consensus localization	50
4.1.1. Basic concept	50
4.1.2. Maximum consensus in the observation space	52
4.1.3. Maximum consensus in the Euclidean space	58
4.2. Optimization based on point-to-plane adjustment score	61
4.3. Optimization based on common loss functions	66
4.4. State estimation filtering	69
4.5. Protection level estimation and integrity evaluation	75
5. Experiments	81
5.1. Objectives and evaluation metrics	81
5.2. Used sensors and data	84
5.3. Results	88
5.3.1. Maximum consensus localization based on the inlier count and point-to-plane adjustment score	89
5.3.2. Localization using different loss functions	102
5.3.3. Maximum consensus filter	107
5.3.4. Protection level estimation and integrity evaluation	110
5.4. Summary and discussion	117
6. Conclusion and outlook	121
6.1. Conclusion	121
6.2. Outlook	123
Appendices	127
A. Experiments	129
A.1. Localization using different loss functions	129
List of Figures	131
List of Tables	135
List of Acronyms	137
Nomenclature	139
Bibliography	143
Acknowledgements	151
Curriculum Vitae	153

1. Introduction

The introduction of this thesis begins with an explanation of the motivation behind the work, followed by the problem statement and research questions. The contributions are then outlined, and the structure of the thesis is introduced.

1.1. Motivation

Even though advanced driver assistance systems have been developed and safety standards have been improved, car traffic remains a means of transportation that causes a high number of fatalities. In 2022, 2,406,465 traffic accidents were recorded in Germany, out of which 289,672 resulted in personal injury, including 2,788 cases of death. The majority of these accidents were caused by human mistakes (Statistisches Bundesamt, 2023). Although the number of fatalities has declined over recent years, the goal remains to further reduce the number of traffic accidents and, consequently, personal injuries. A way to significantly improve traffic safety might be the full autonomy of vehicles. The technical potential of the various interacting components of an autonomous vehicle, such as localization, object detection and tracking, motion planning and collision avoidance, as well as information exchange with other vehicles, is enormous, however, so are the associated challenges regarding a trustworthy and safe operation.

The fundamental requirement for autonomous driving applications is the accurate knowledge of the ego-vehicle’s pose, consisting of position and heading. To determine their pose, autonomous vehicles are typically equipped with a sensor suite that combines global positioning sensors, such as Global Navigation Satellite System (GNSS) receivers, and relative positioning sensors, such as cameras and Light Detection And Ranging (LiDAR) sensors. Due to multi-path effects, accurate global positioning up to lane-level accuracy is especially challenging in highly urbanized areas. This results in the need for relative positioning sensors, which enable precise positioning with respect to a map based on a rough initial pose, e.g. provided by GNSS. Compared to cameras, LiDAR sensors are not dependent on illumination and provide accurate range measurements, making them the ideal sensor for precise positioning.

The demands for the localization are high because the vehicle must keep its lane and avoid collisions with other traffic participants. This results in accuracy requirements of a very few decimeters for the lateral and longitudinal position (Reid et al., 2019). Besides the accuracy of the localization, its reliability to function properly also in challenging environments and integrity to ensure a safe operation are two key requirements. Reliability describes the fault occurrence and is tied to the robustness of an operation. The higher the robustness of the positioning system, the lower is the number of faults, and the higher is the reliability (Verma et al., 2015). Integrity is defined as ‘a measure of the trust that can be placed in the correctness of the information supplied by the total system’. It refers to a system’s capacity to correctly and timely alert the user when specific error bounds are surpassed. The concept is used in aviation and it is also applicable to autonomous vehicle localization (International Civil Aviation Organization, 2006; Wörner et al., 2016).

In this work, the LiDAR-based localization of autonomous vehicles will be investigated. Special focus is on robust estimation methods for highly reliable positioning. Additionally, the application of the concept of integrity on LiDAR-based localization is examined.

1.2. Problem statement and research questions

LiDAR-based localization is usually tied to point cloud registration, with the corresponding algorithms being based on two steps: First, associations between a point cloud acquired by a LiDAR sensor mounted on the ego-vehicle and a map are established. Second, the errors of these associations are minimized using a loss function, which is to be in accordance with the error distribution of the measurements. Alternating between these two phases in an iterative loop leads to the well-known Iterative Closest Point (ICP) algorithm (Besl and McKay, 1992). Often, the assumption of normally distributed errors is employed, resulting in a quadratic loss function which can be solved efficiently using the well-known least squares method. Even though the selection of parametric distributions for modelling of uncertainties enables efficient computations, in the presence of outliers, these distributions lead to a wrong estimate of the system state. Outliers arise from wrong associations and cause deviations from the distribution function, which cannot be modeled. Common reasons for incorrect associations are wrong assumptions about the pose of the ego-vehicle, e.g. caused by an inaccurate, initial GNSS positioning, changing environments and consequently outdated maps, e.g. caused by renovations and scaffolding, and the presence of other dynamic traffic participants (c.f. Fig. 1.1).



Figure 1.1.: Common reasons for incorrect associations: Wrong assumptions about the pose of the ego-vehicle caused by an inaccurate, initial GNSS positioning in highly urbanized areas (left), the presence of other dynamic traffic participants (top right), and changing environments caused by renovations and scaffolding (bottom right).

To reduce the susceptibility towards outliers, more robust loss functions have been proposed. Well-known examples are the L1 norm and the loss functions introduced by Huber, Cauchy, Geman-McClure, Welsch, and Tukey, which vary in the influence of large residuals on the estimated state. The most robust loss function is the count of inliers, also referred to as maximum consensus optimization, where measurements with residuals larger than a predefined threshold do not have any influence on the outcome. To visualize the different weighting of large residuals and the resulting impact on an estimate, various loss functions are plotted in Fig. 1.2a and the corresponding estimation results for a 1D estimation problem in the presence of outliers are shown in Fig. 1.2b. The

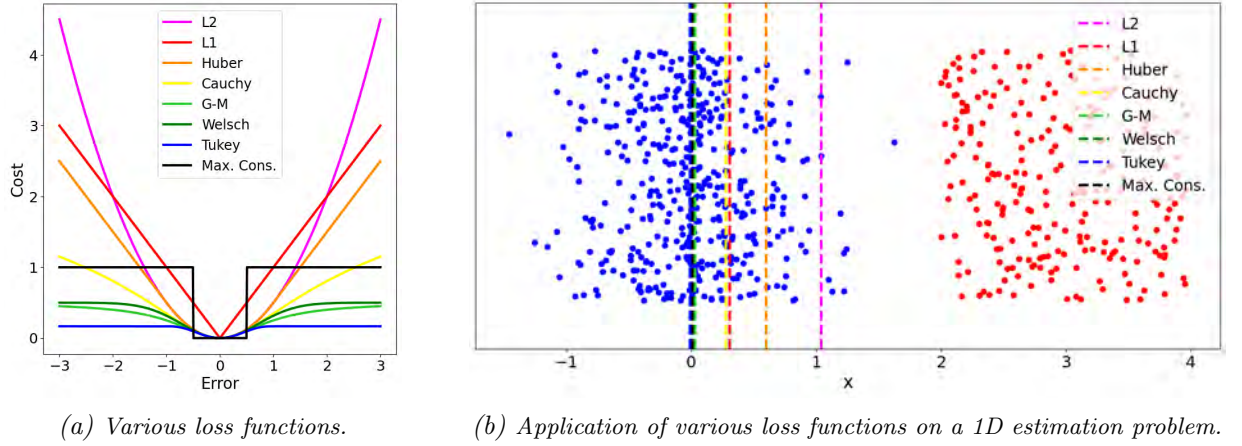


Figure 1.2.: Various loss functions (a) and the corresponding estimation results for a 1D estimation problem in the presence of outliers (b): The smaller the influence of outliers on the estimation result, the smaller is the deviation from the true solution located at 0. Outliers are shown in red, inliers are shown in blue.

smaller the influence of outliers on the estimation result, the smaller is the deviation from the true solution, which is located at 0.

Due to the exclusion of measurements beyond a specific error threshold, maximum consensus techniques are ideally-suited for robust parameter estimation in the presence of outliers. To cope with the inherently high computational effort, increasingly sophisticated methods have been proposed recently (Chin and Suter, 2017). The arguably most famous maximum consensus strategy is Random Sample Consensus (RANSAC) (Fischler and Bolles, 1981), where candidate models are instantiated based on randomly drawn samples and compared by the size of their consensus sets. RANSAC is a heuristic approach based on the assumption that within a specific number of trials at least once the candidate model is only based on inliers. For little outlier contaminated data, it usually provides decent results in a reasonable amount of time, which is why it has been applied successfully in many applications and many variants have been introduced. However, in strongly outlier contaminated data, the advantage of lower computational effort dwindles away, and due to its randomized selection there is no guarantee to obtain the globally optimal solution.

In this work, the aim is to investigate the applicability of the maximum consensus criterion for the purpose of vehicle localization using LiDAR data. An exhaustive search strategy is pursued, which on the one hand guarantees global optimality and on the other hand yields full knowledge about the parameter space, which opens up possibilities for an assessment of the localization solution. In general, the computational complexity of a maximum consensus optimization problem grows exponentially in the number of variables, however, if limited to a low-dimensional problem, the optimization can be solved with manageable computational cost.

In autonomous driving, besides a high robustness, the localization of a vehicle is supposed to provide integrity ensuring a trustworthy operation. A key attribute to determine a system's integrity level is the Protection Level (PL), which describes an upper bound on the position error. The research question pursued is how to estimate an error bound around our pose that guarantees that it contains the true pose up to a certain probability. The goal is to provide not only a robust localization, but also guarantees for the localization solution.

To summarize, the following research questions are formulated:

- How to apply the maximum consensus criterion to the task of vehicle localization?

- How can maximum consensus techniques improve autonomous vehicle localization, and what are the advantages and limitations of a maximum consensus-based localization?
- Which measures can be taken to cope with existing limitations and improve the maximum consensus-based localization? What alternative objective functions can increase the localization reliability?
- How do other known and data-adapted loss functions perform in comparison to the maximum consensus criterion?
- How to extend the localization based on the maximum consensus criterion or other loss functions into a filter solution?
- How to bridge the gap between maximum consensus localization and an integrity system based on confidence or protection levels, which guarantee up to a certain probability to contain the true pose?

1.3. Contributions

In this work, a framework for robust and reliable localization using LiDAR data is introduced. Based on a rough initial pose, e.g. provided by means of GNSS or a filter prediction, relative positioning between a sparse ‘car sensor’ point cloud from a LiDAR mounted on the ego-vehicle and a dense, high resolution ‘map’ point cloud from a mobile mapping campaign is conducted. To this end, around the initial pose, a search space is defined and discretized into candidate poses, which are then evaluated in an exhaustive manner. To contain the computational complexity, which grows exponentially with the number of parameters, the pose estimation is limited to 3 Degrees of Freedom (DoF), namely the 2D position in the xy -plane and the heading angle θ . Determining the height z is generally less challenging because it can be presumed that the vehicle is traveling on the road surface. Regarding the angles, both roll and pitch are usually very small during regular driving conditions and can be measured with a high degree of reliability.

For the evaluation of the candidate poses, in the basic approach, the number of measurements matching the map is determined, which is synonymously referred to as inlier count or consensus set. This is done following two strategies, which differ in the way of establishing correspondences and to which it is related using the terms *observation space* and *Euclidean space* in the course of this thesis. In the observation space, for each candidate pose, measured ranges are predicted and if the deviation to the actual measured range is smaller than a predefined inlier threshold, the inlier count of the corresponding candidate pose is incremented by one. In the Euclidean space, the measured scan is rotated and shifted according to each candidate pose and if the xy -distance between a scan point and a map point is smaller than a predefined inlier threshold, the inlier count of the candidate pose is likewise incremented by one. For both strategies, the candidate pose yielding the highest consensus set is considered as the localization solution. The result of this exhaustive strategy is the evaluated search space, which can be represented as a stack of evaluated xy -grids, as for illustrative purposes shown in Fig. 1.3. Each layer of the stack contains the positional grid of a certain heading angle. In this example, the highest count of inliers and consequently the localization solution is at pose $(x, y, \theta) = (0\text{ m}, 0\text{ m}, 0^\circ)$, which corresponds to the reference pose.

Besides the count of inliers, a new objective function based on Helmert’s point error is proposed. Instead of simply counting matches, the constraints generated by each correspondence with respect to the unknown pose are modelled. To this end, the uncertainty of a point-to-plane adjustment is considered, where the score is defined as the inverse of the square of Helmert’s point error. Hence,

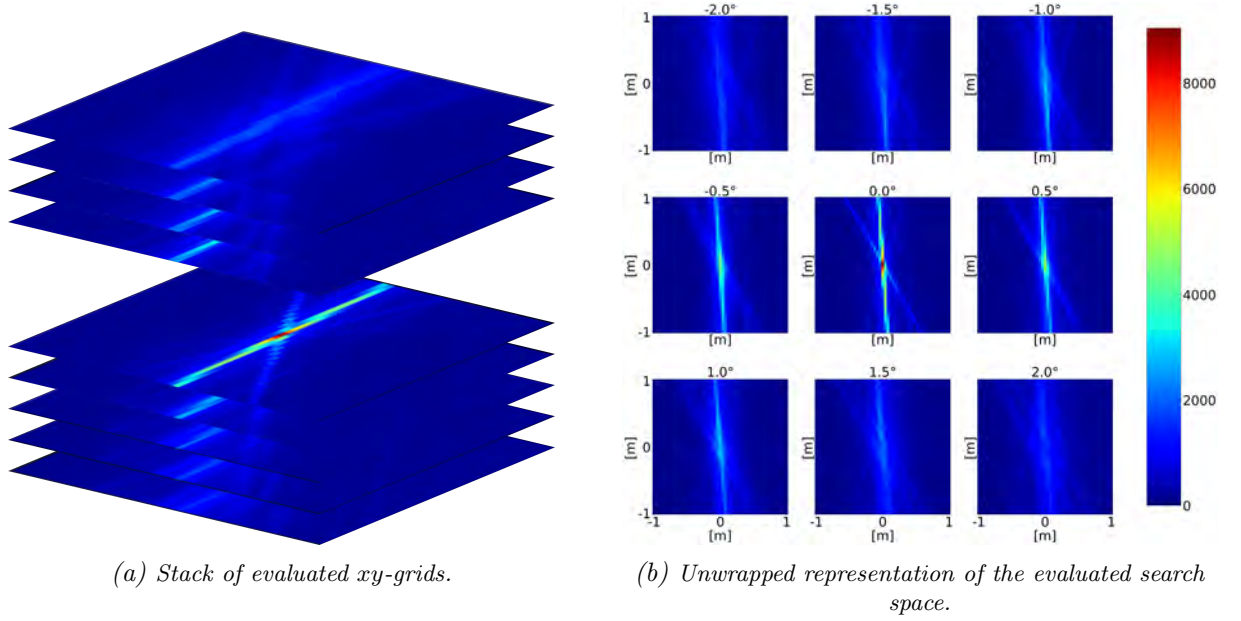


Figure 1.3.: Evaluated search space represented as stack of evaluated xy -grids (a) with the xy -grid highlighted, which contains the highest count, and represented unwrapped (b). Each layer of the stack contains the positional grid of a certain heading angle. In this example, the highest count of inliers and consequently the localization solution is at pose $(x, y, \theta) = (0\text{ m}, 0\text{ m}, 0^\circ)$, which corresponds to the reference pose.

the lower the uncertainty of the point-to-plane adjustment, the smaller is Helmert’s point error and the higher is the score. This has the effect that candidate poses that have constraints only in one direction and thus are undetermined in the other direction obtain a significantly smaller score than candidate poses that are constrained in all directions. This leads to a reduced number of localization failures and hence a higher robustness as well as reliability, especially in longitudinal direction while driving along straight streets.

To investigate not only the maximum consensus criterion as the most robust form of loss function, the localization framework is generalized to utilize any loss function. Common, well-known M-estimators are applied to the task of LiDAR-based localization and examined with respect to robustness and accuracy. In addition to those popular loss functions, the application of a data-adapted loss function based on a beam model for range sensors is tested.

With the objective to further increase the reliability and reduce the number of large localization errors, the single positioning is extended into a filter strategy. In order to preserve the full knowledge of the search space and to prevent shortcomings of Monte Carlo-based filter techniques, such as particle deprivation, a histogram filter is employed. The proposed filter approach is based on discrete probability distributions, which enables the filter to utilize any loss function.

Finally, to provide not only a robust but also a trustworthy localization, a protection level definition is proposed, which makes use of the full knowledge about the investigated search space. Based on the relations between the estimated protection level, the true position error, and an alert limit, the integrity levels of the localization are derived.

To visualize the various components of the proposed localization framework and their connections, an overview of the framework is shown in Fig. 1.4.

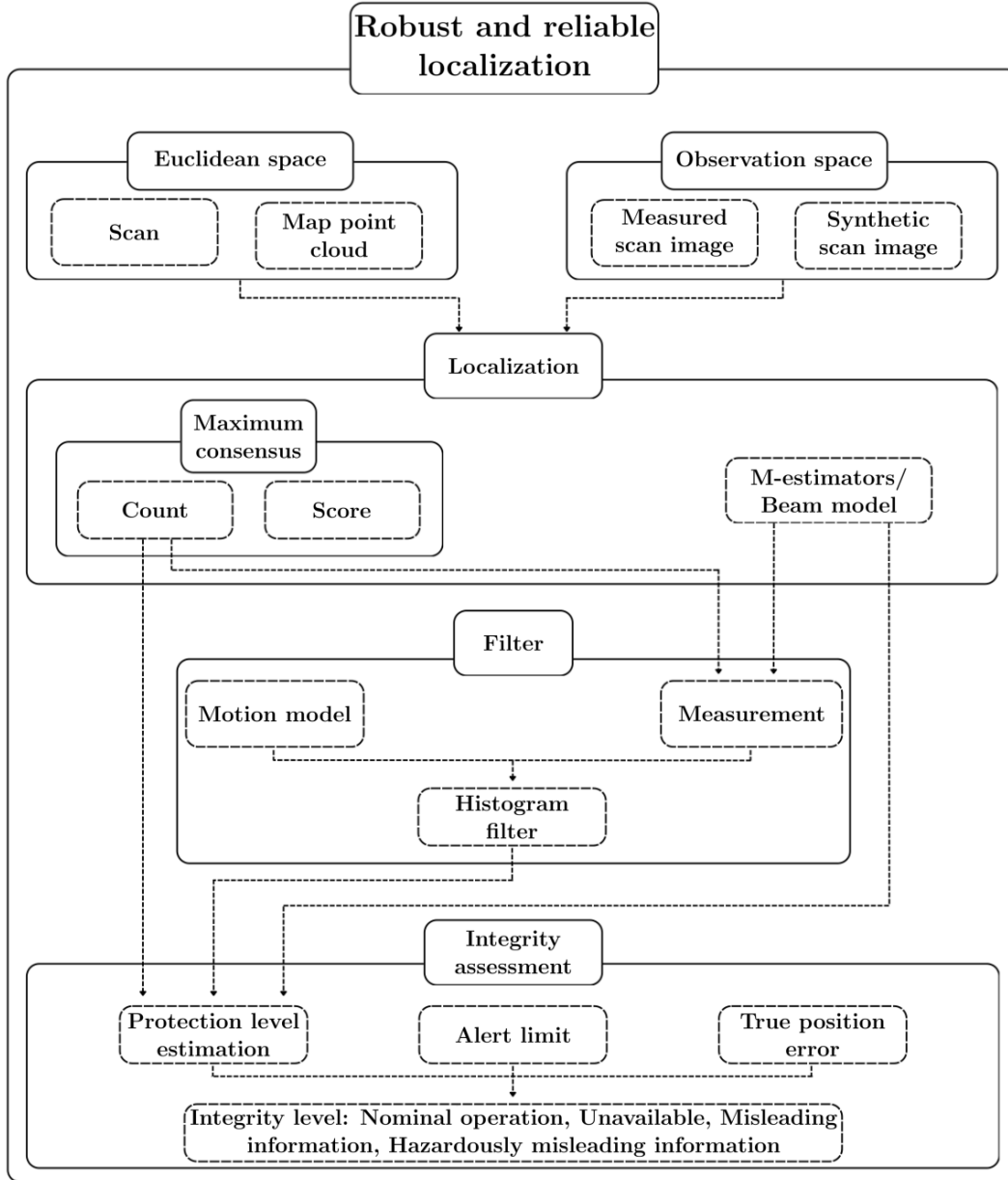


Figure 1.4.: Overview of the robust and reliable localization framework.

To summarize, the main contributions are:

- A framework for robust localization using LiDAR data, which is applicable to various objective functions, such as the maximum consensus count and point-to-plane adjustment score, as well as known M-estimators and data-adapted loss functions.
- An extension of the stand-alone, epoch-wise localization into a filter for improved robustness and reliability.
- An estimation of Protection Levels that utilizes the full knowledge of the investigated search space and allows for an integrity assessment of the localization solution.

1.4. Structure of the thesis

Chapter 2 introduces the basics necessary for understanding the work. The topics covered include robust estimation techniques, state estimation filtering, LiDAR sensors and point cloud acquisition, map representations, and basic concepts of integrity.

In Chapter 3, the related work on point cloud registration, which is divided into correspondence-based methods and simultaneous pose and correspondence (SPC) methods, as well as on LiDAR-based localization is presented. In the final discussion, the research gap is identified.

Chapter 4 presents the methodological contributions with respect to the localization framework shown in Fig. 1.4. First, the basic method of maximum consensus localization in both observation and Euclidean space is explained, followed by the optimization based on the novel point-to-plane adjustment score and the optimization based on common loss functions. Then, the filter approach and finally the protection level estimation and integrity evaluation are introduced.

Chapter 5 presents the experiments. It first explains the objectives and evaluation metrics and introduces the sensors and data used. Afterwards, the evaluation results are presented, and subsequently discussed.

Chapter 6 provides a conclusion on the findings and an outlook on both possible improvements and new research directions.

2. Fundamentals

In this chapter, the basic concepts are provided, which are essential to comprehend the following sections of this thesis. At first, in Sec. 2.1, the generally applicable notions of robust estimation are introduced. Following this, and in Sec. 2.2, the universal principles of state estimation filtering are presented. In Sec. 2.3, the operating principle of a LiDAR sensor and the point cloud acquisition are briefly explained including the representation of LiDAR scans as scan images and the concept of point cloud registration. Afterwards, in Sec. 2.4, various kinds of map representations are introduced. Finally, in Sec. 2.5, basic concepts of integrity are presented.

2.1. Robust estimation techniques

In this section, the fundamentals of robust estimation are presented. In Sec. 2.1.1, a line estimation problem and the conventional least squares solution are considered. In Sec. 2.1.2, the maximum consensus formulation is introduced, and in Sec. 2.1.3, the approach using M-estimators is explained. Finally, in Sec. 2.1.4, the beam model for range finders proposed by Thrun et al. (2005) is presented.

2.1.1. Problem definition

Given the assumption that errors are independent and identically distributed (i.i.d.) Gaussian, the optimal solution is found by minimizing the sum of squared errors, referred to as least squares estimation. However, real-world measurements are not only noisy, but also contaminated by outliers, which violate the fundamental assumption and lead to a wrong least squares result. Consequently, in robotics, for a broad variety of optimization tasks, which are based on real-world data - independent of the type of sensor - robust fitting techniques are required.

For illustration, consider a two-dimensional line fitting problem for a set of noisy, outlier-contaminated measured points $\mathbf{P} = (p_i, q_i)$. The mathematical definition for points on the line is given as $q = m \cdot p + c$, with m being the slope and c being the intercept. Note that only q is stochastic and that the residual from a point (p_i, q_i) to the line is not the perpendicular distance, but the q -axis parallel, vertical distance and determined as $r_i = m \cdot p_i + c - q_i$.

For a conventional least squares estimation, the parameter vector contains the unknowns which are to be determined, in this example $\mathbf{x} = [m, c]^T$. Given this, the general linear model can be defined as

$$\mathbf{b} = \mathbf{A}\mathbf{x}, \quad (2.1)$$

with $\mathbf{A}_i = [p_i, 1]$ and $b_i = q_i$. Following the conventional least squares approach, the vector of parameters \mathbf{x} is estimated by minimizing the sum of squared errors

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{r}\|_2^2, \quad (2.2)$$

$$\text{where } \mathbf{r} = \mathbf{A}\mathbf{x} - \mathbf{b} = \begin{bmatrix} p_1 & 1 \\ p_2 & 1 \\ \vdots & \vdots \\ p_N & 1 \end{bmatrix} \cdot \begin{bmatrix} m \\ c \end{bmatrix} - \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_N \end{bmatrix}. \quad (2.3)$$

Calculating the gradient and setting it to 0 leads to the following, well-known formulation to solve the estimation:

$$\mathbf{x}^* = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}. \quad (2.4)$$

In Fig. 2.1, an example for the described line fitting problem is visualized. The blue points are sampled with normally distributed noise from the black line, which has the parameters $\mathbf{x} = [0.5, 2]^T$ that are to be estimated, and the red points are outliers. The magenta line is estimated using least squares, and it deviates significantly from the sought, true line. This is due to that fact that the outliers violate the fundamental assumption of least squares that errors are i.i.d. Gaussian. Since all measurements are taken into consideration in standard least squares, already a single outlier can heavily bias the solution making the estimation highly susceptible towards outliers. Consequently, it is not robust, and in the presence of outliers, a more robust strategy has to be employed, which bounds the influence of outlying correspondences.

2.1.2. Maximum consensus formulation

The goal to find a robust estimate of \mathbf{x} requires an estimation approach, which tolerates outlier by significantly reducing their influence or weight on the estimate. The most rigorous form in this regard is maximum consensus, which fully disables associations with residuals larger than a constant, user-defined inlier threshold ϵ . For a specific point i , the residual is defined as

$$r_i = |\mathbf{A}_i \mathbf{x} - b_i|. \quad (2.5)$$

The core of the maximum consensus technique is its robust objective function, also referred to as robust criterion, which is defined by the consensus Ψ :

$$\Psi(\mathbf{x}) = \sum_{i=1}^N \mathbb{I}(|\mathbf{A}_i \mathbf{x} - b_i| \leq \epsilon), \quad (2.6)$$

where $\mathbb{I}(\cdot)$ is the indicator function, which returns one, if the condition is true (if the residual is smaller or equal to the distance threshold ϵ), and 0 otherwise. The maximum consensus estimate is then determined as

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmax}} \Psi(\mathbf{x}). \quad (2.7)$$

In words, using the maximum consensus criterion, for various parameter sets, the number of measurements supporting each candidate model by having a residual $r_i \leq \epsilon$ is counted, while fully ignoring all measurements with $r_i > \epsilon$. Eventually, the set of parameters yielding the highest consensus is defined as estimation result.

Figure 2.1 illustrates the advantage of the maximum consensus estimation for the line fitting problem. While the magenta line estimated using least squares is significantly biased by the

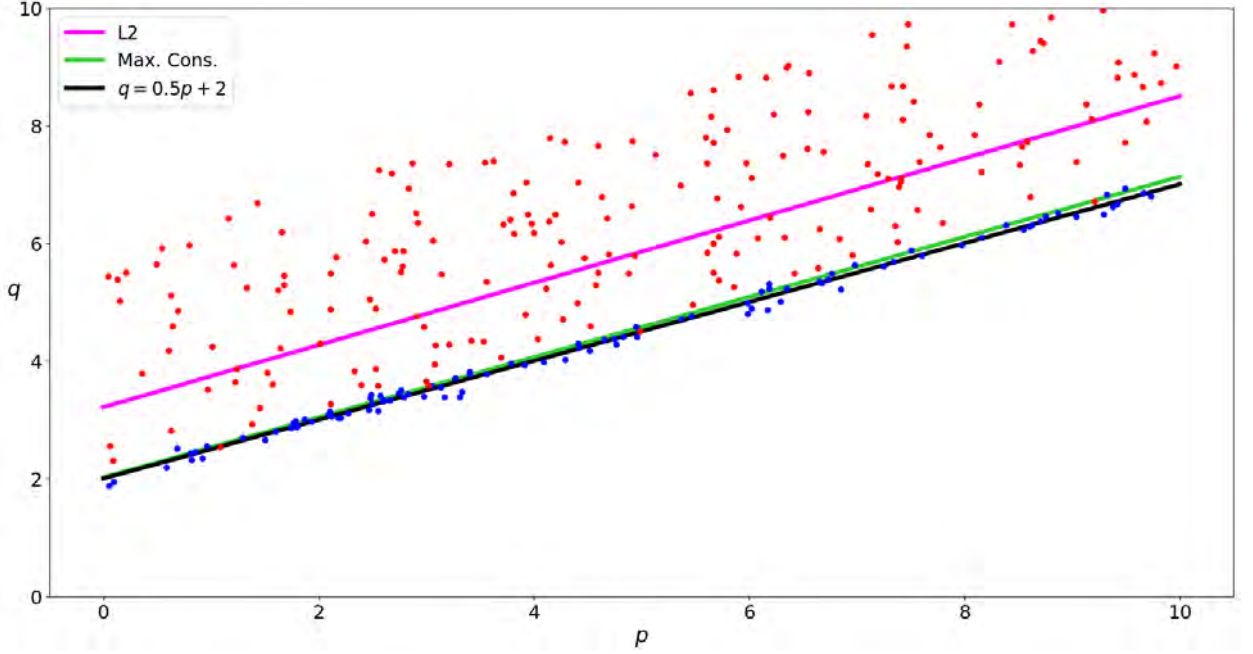


Figure 2.1.: Illustration of a line fitting problem in the presence of outliers: The 100 blue points are inliers sampled with normally distributed noise from the black line, which has the parameters $\mathbf{x} = [0.5, 2]^T$ that are to be estimated, and the 200 red points are outliers. The magenta line deviates significantly from the sought, true line because it is estimated using least squares, which is highly susceptible towards outliers. The green line, instead, is close to the true line because it is determined robustly using maximum consensus, which finds the set of parameters showing the highest agreement with the data while excluding measurements with residuals larger than the inlier threshold.

outliers, the green line determined using maximum consensus is close to the true line as points with residuals larger than the inlier threshold ϵ do not have any influence on the maximum consensus estimate. This represents the crucial difference to standard least squares, where the weights of correspondences are unbounded and even growing with larger residuals, which is why the existence of only a single outlier can cause a heavily biased estimate. As the maximum consensus approach finds the optimal line by simply counting the number of points with a residual value smaller than or equal to ϵ for various line parameters, the method only works if there is no second, false hypothesis that is supported by more inliers.

Instead of maximizing the consensus, the count of outliers can be minimized. Furthermore, maximum consensus fitting can be used as outlier identification by considering the measurements which are inconsistent with the final estimate ($r_i > \epsilon$) as outliers. The logical subsequent step is usually a final least squares adjustment on the remaining inlier set to refine the estimation result.

Since maximum consensus optimization is applied to a broad variety of problems, a generalized formulation of the maximum consensus problem is introduced as

$$\boldsymbol{\omega}^* = \underset{\boldsymbol{\omega} \in \Omega}{\operatorname{argmax}} \Psi(\boldsymbol{\omega}), \quad (2.8)$$

with $\boldsymbol{\omega}$ being a set of parameters from the domain Ω that defines the functional model. The consensus function is defined accordingly as

$$\Psi(\boldsymbol{\omega}) = \sum_{i=1}^N \mathbb{I}(r_i(\boldsymbol{\omega}) \leq \epsilon), \quad (2.9)$$

with $r_i : \Omega \mapsto \mathbb{R}^+$ being a nonnegative function that yields the residual value of each measurement.

Maximum consensus methods can be categorized in exact and approximate approaches. The notion of ‘solving’ a maximum consensus problem refers to obtaining the *exact* or *global* solution, and the notion of ‘approximating’ refers to obtaining an *approximate* solution. For a maximum consensus problem, an *exact* or *global* solution is defined as parameter vector ω^* with $\Psi(\omega^*) \geq \Psi(\omega)$ for all other $\omega \in \mathbb{R}^d$, and an *approximate* solution is defined as parameter vector $\hat{\omega}$ with $\Psi(\hat{\omega}) > \alpha \cdot \Psi(\omega^*)$, where α is referred to as ‘performance guarantee’ or ‘approximation factor’. An α -approximation algorithm yields a solution in polynomial time, where the consensus of the solution is within the factor α of the consensus of the optimal solution. For maximization problems, such as maximum consensus, $\alpha < 1$ and for minimization problems, $\alpha > 1$. In general, ω^* is not unique, as, on the one hand, $\Psi(\omega)$ may have multiple separate peaks, and on the other hand, a ‘peak’ is often a plateau in the consensus function, which means that an interval of parameters ω yields the same maximum consensus (Chin and Suter, 2017; Williamson and Shmoys, 2011).

A *local* solution to the optimization problem that $f(\omega)$ with $\omega \in \mathbb{R}^d$ is to be maximized is defined as a point $\hat{\omega}$ in a local neighborhood N with $f(\hat{\omega}) \geq f(\omega)$ for all $\omega \in N$. According to this definition, when using maximum consensus as optimization scheme, any ω represents a local solution as $\Psi(\omega)$ is piecewise constant. Even though many optimization algorithms determine locally optimal solutions, in maximum consensus it is consequently not meaningful. Moreover, as it lacks gradient information, the piecewise constant nature of $\Psi(\omega)$ impedes utilizing gradient-based optimization techniques for consensus maximization (Chin and Suter, 2017). Furthermore, maximum consensus problems are intractable, or more precisely NP-hard optimization problems, which means that it is not possible to solve the optimization in polynomial time. A proof can be found in (Chin et al., 2018).

2.1.3. M-estimators

M-estimators represent a generalized form of the maximum likelihood estimator that allow to use various kinds of loss functions. Whereas in conventional least squares, the sum of squared errors $\sum_i r_i^2$ is minimized and residuals have an unbounded influence on the estimate, the goal of M-estimators is to limit this influence by substituting the squared loss with another, less increasing, and consequently more robust loss function. This leads to the following, general formulation of the estimation problem:

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \sum_{i=1}^N \rho(|A_i \mathbf{x} - b_i|), \quad (2.10)$$

where $\rho(r)$ is a loss function, which is to be symmetric and positive-definite, with a single minimum at $r = 0$. Its shape is of fundamental importance as it defines the influence of each residual on the estimate. For $\rho(r) = r^2/2$, the M-estimator is the conventional least squares estimator, where only a single outlier may result in a strongly biased estimate. Consequently, to obtain a more robust behavior, the chosen loss functions have to increase less strongly than quadratic. A selection of well-known loss functions is presented below, with the corresponding plots and formulas shown in Fig. 2.2.

If $\rho(r) = |r|$, the loss function corresponds to the absolute deviation, which in fact is increasing less strongly, yielding a reduced influence of large errors, however, it is still linearly growing, which makes this loss function not entirely robust. Huber’s loss function is a combination of L2 and L1 as it shows a parabolic behavior for a range below a certain threshold around zero, with $|r| \leq c$, and a linear increase beyond that threshold $|r| > c$. Consequently, as for L1, it does not entirely remove the impact of significant outliers. The remaining loss functions Cauchy, Geman-

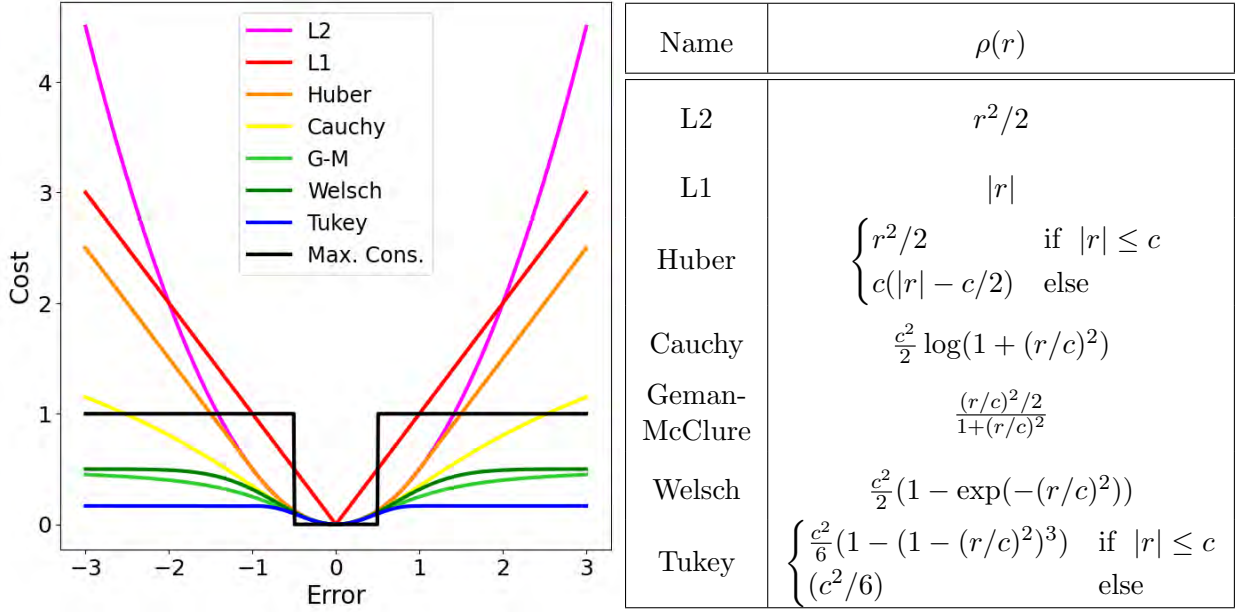


Figure 2.2.: Plots and formulas of various well-known loss functions. For the plots, the inlier threshold ϵ of the maximum consensus loss function is selected as $\epsilon = 0.5$, and the parameter c is selected as $c = 1$ for all other loss functions.

McClure and Welsch more and more reduce the impact of large residuals, while Tukey's biweight function sets the loss to a constant for $|r| > c$ and thereby even eliminates the influence on the estimate entirely. Due to their non-convex shape, they are capable of significantly reducing, or even entirely suppressing, the impact of large errors caused by outliers, however, at the same time the estimation cannot be solved using classical gradient descent approaches and a unique solution cannot be guaranteed. The parameter c shows an analogy to the inlier threshold ϵ in the maximum consensus function, as it is used to tune the shape and define the transitions of the loss functions. It is typically selected according to the standard deviation of the measurements σ .

Next, the application of the different loss functions is considered using a simple estimation example to show their influence on the estimation and convexity or concavity. In Fig. 2.3, for a 1D estimation problem in the presence of outliers, both the estimation results and the corresponding loss curves are visualized for the loss functions introduced above. The blue points are inliers sampled with normally distributed noise, and the red points are outliers, where the points are spread along the y -axis only for illustration purposes. The smaller the influence of outliers on the estimation result, the smaller is the deviation from the true solution, which is located at 0. Consequently, the L2 loss is the furthest away from the desired solution, followed by Huber, L1, and Cauchy, all of which also show a significant deviation. On the other hand, Geman-McClure, Welsch, Tukey, and maximum consensus lead to the correct estimate. Regarding the loss curves, L2, L1, and Huber have only one minimum, while the loss curves of Geman-McClure, Welsch, Tukey, and maximum consensus exhibit local minima in the area of the outliers. Therefore, estimation problems based on L2, L1, and Huber can be solved by gradient descent, whereas those based on Geman-McClure, Welsch, Tukey, and maximum consensus cannot because it may get stuck in a local minimum. In this example, Cauchy also has only one minimum, but in principle it is also susceptible to local minima. The quadratic L2 loss is strictly convex and can be solved very efficiently in closed form. The L1 and Huber losses are convex; however, they cannot be solved in closed-form but can still be solved efficiently using convex optimization (Boyd and Vandenberghe, 2004). On the other hand,

the Cauchy, Geman-McClure, Welsch, Tukey, and maximum consensus losses are not convex and therefore NP-hard, which means that they cannot be solved efficiently.

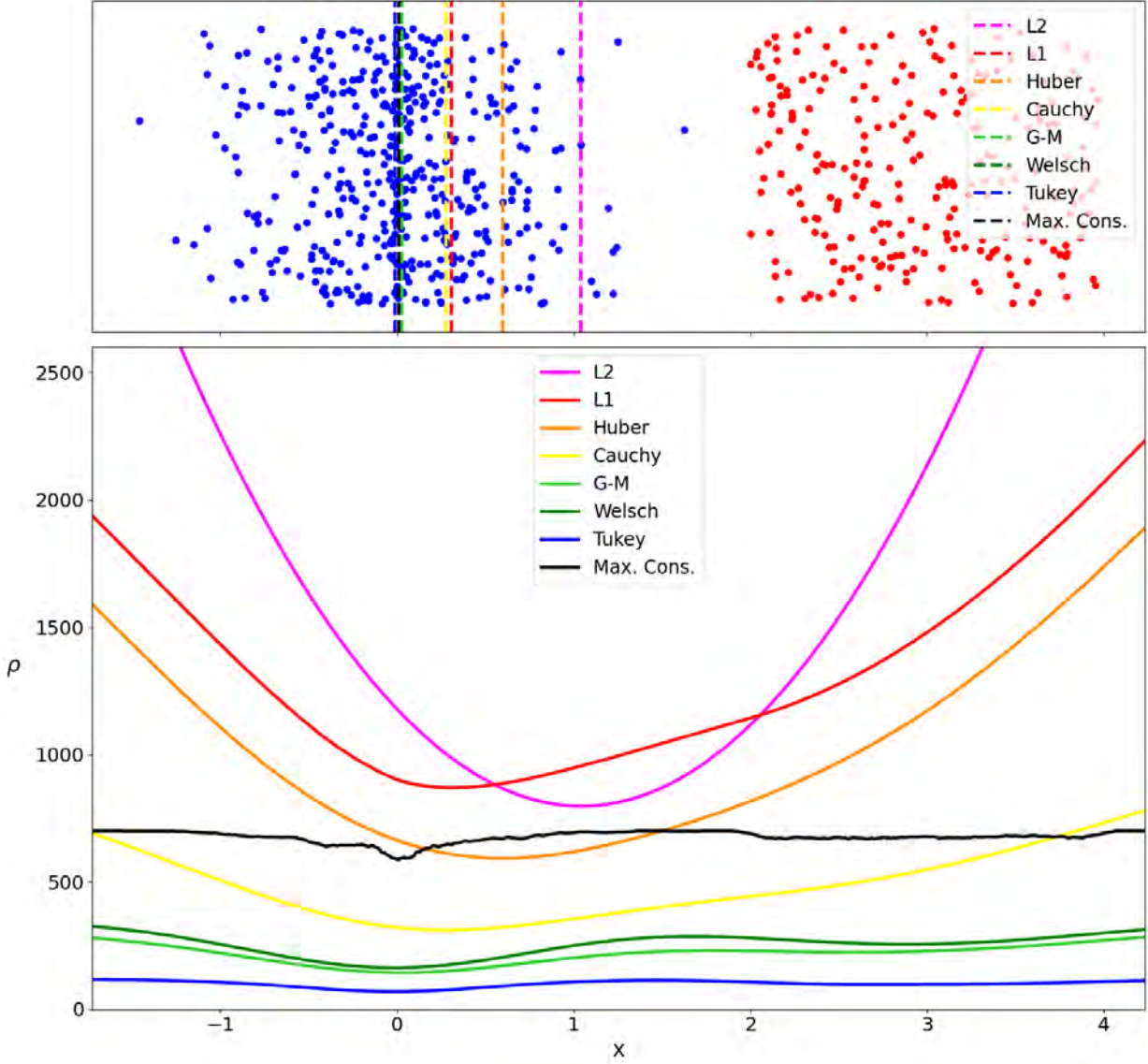


Figure 2.3.: The figures illustrate a 1D estimation problem in the presence of outliers and the corresponding estimation results (top) and loss curves (bottom) of the loss functions introduced above. The blue points are inliers sampled with normally distributed noise, and the red points are outliers, where the points are spread along the y-axis only for illustration purposes. The smaller the influence of outliers on the estimation result, the smaller is the deviation from the true solution, which is located at 0.

The standard approach to solve the estimation problem (cf. Eq. 2.10) is the iteratively reweighted least squares (IRLS) (Björck, 1996), which alternates between estimating weights $w_i^{(t)}$ based on the current parameters $\mathbf{x}^{(t)}$ and solving a weighted least squares estimation to update the parameters:

$$\mathbf{x}^{(t+1)} := \underset{\mathbf{x}}{\operatorname{argmin}} \sum_{i=1}^N w_i^{(t)} (\mathbf{A}_i \mathbf{x} - b_i)^2, \quad w_i^{(t)} := \frac{\rho'(|\mathbf{A}_i \mathbf{x}^{(t)} - b_i|)}{|\mathbf{A}_i \mathbf{x}^{(t)} - b_i|} \quad (2.11)$$

with ρ' being the first derivative of ρ , also referred to as influence function as it describes the influence of a measurement on the estimated parameter. If initial parameters are known, the

weights can be computed first, otherwise the weights are typically set to one and a set of parameters is first estimated using least squares. After reaching convergence, a local minimum is obtained, but it is not guaranteed to be the global minimum.

The maximum consensus problem (cf. black line in Fig. 2.2) can also be formulated as an M-estimator:

$$\rho(r) = \begin{cases} 0 & \text{if } |r| \leq c \\ 1 & \text{else} \end{cases}. \quad (2.12)$$

This is also referred to as L0 loss. However, the maximum consensus function does not fulfill all required properties of an M-estimator as, for example, it does not provide a unique minimum. Furthermore, the influence function ρ' is not always defined, and where it is defined, it equals 0. Overall, due to its piece-wise constant characteristic, gradient-based optimization is complicated and IRLS is not capable of solving the maximum consensus problem (Zhang, 1997; Chin and Suter, 2017).

Finally, it is important to note that an optimal solution of a robust M-estimator may not coincide with the optimal solution of a maximum consensus problem, and consequently, a conventional M-estimator cannot be utilized as a substitute for solving maximum consensus (Chin and Suter, 2017).

2.1.4. Beam model for range sensors

For a conventional least squares estimation, where the solution is found by minimizing the sum of squared residuals, the quadratic loss function originates from the assumption that the residuals are normally distributed. In turn, applying the L1 loss function corresponds to the assumption that the errors are Laplace distributed. In fact, the distribution of the residuals of the measurements determines the loss function to be applied and every likelihood function $\mathcal{L}(r)$ can be expressed as a loss function $\rho(r)$ and vice versa based on the relations

$$\mathcal{L}(r) = e^{-\rho(r)} \quad \text{and} \quad \rho(r) = -\log \mathcal{L}(r). \quad (2.13)$$

Instead of employing a symmetric, positive-definite loss function with a single minimum, as it is done using M-estimators (cf. Sec. 2.1.3), a loss function can also be derived from a likelihood distribution, which is based on residuals from a real-world measurement. A way to define such a likelihood distribution is the beam model for range finders proposed by Thrun et al. (2005), which is a mixture density that considers the following four possible outcomes for an expected range value z^{k*} obtained by ray casting in a given map and the actual measured range given by the sensor z^k :

- The measurement of the correct range subject to local measurement noise, which is modeled as a Gaussian (cf. Fig. 2.4a)

$$p_{hit}(z^k | z^{k*}) = \begin{cases} \eta \mathcal{N}(z^k | z^{k*}, \sigma_{hit}^2) & \text{if } 0 \leq z^k \leq z_{max} \\ 0 & \text{else,} \end{cases} \quad (2.14)$$

$$\text{with } \mathcal{N}(z^k | z^{k*}, \sigma_{hit}^2) = \frac{1}{\sqrt{2\pi\sigma_{hit}^2}} e^{-\frac{(z^k - z^{k*})^2}{2\sigma_{hit}^2}},$$

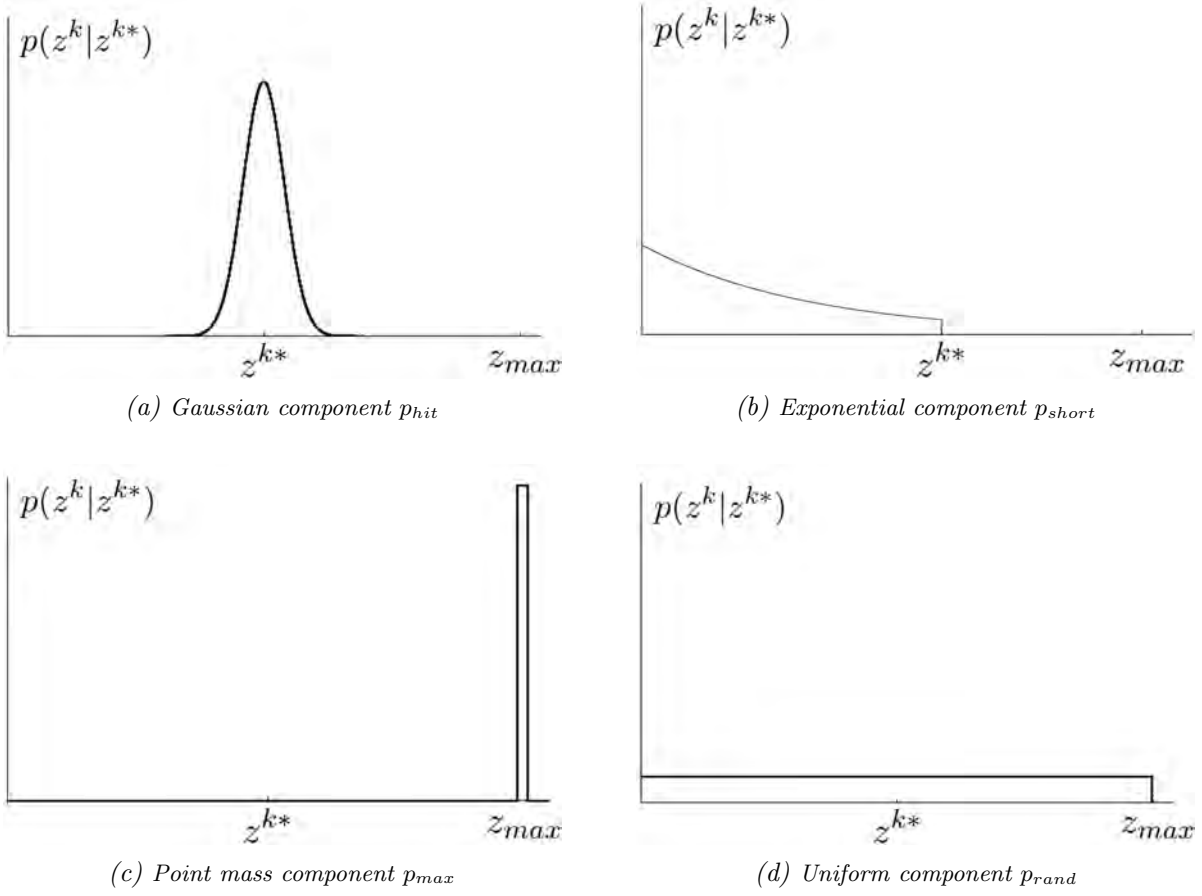


Figure 2.4.: Components of the beam model for range sensors. The horizontal axes represent the actual measurement z^k and the vertical axes the likelihood $p(z^k | z^{k*})$.

- a measurement of an unexpected obstacle that is in the foreground and not contained in the map (e.g. other dynamic traffic participants or parked cars), which is modeled as a decaying exponential distribution (cf. Fig. 2.4b)

$$p_{short}(z^k | z^{k*}) = \begin{cases} \eta \lambda_{short} e^{-\lambda_{short} z^k} & \text{if } 0 \leq z^k \leq z^{k*} \\ 0 & \text{else,} \end{cases} \quad (2.15)$$

- a sensor failure or exceeding the maximum measuring range, where the sensor returns the maximum range value, modeled as a point-mass distribution (cf. Fig. 2.4c)

$$p_{max}(z^k | z^{k*}) = \begin{cases} 1 & \text{if } z = z_{max} \\ 0 & \text{else,} \end{cases} \quad (2.16)$$

- and a not explainable, random measurement, which is modeled as a uniform distribution over the entire range (cf. Fig. 2.4d)

$$p_{rand}(z^k | z^{k*}) = \begin{cases} \frac{1}{z_{max}} & \text{if } 0 \leq z^k < z_{max} \\ 0 & \text{else.} \end{cases} \quad (2.17)$$

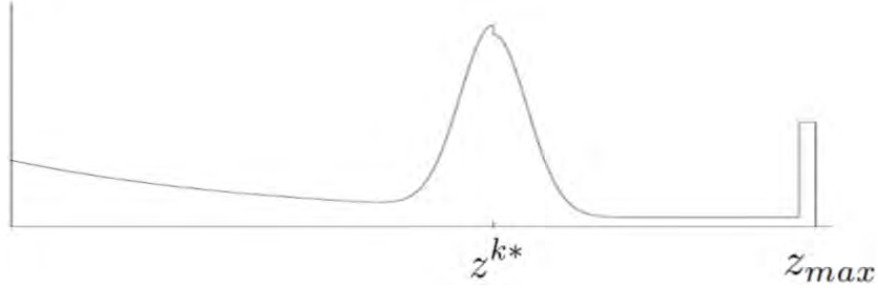


Figure 2.5.: Physical beam model of range sensors represented by the mixture distribution $p(z^k | z^{k*})$.

In Fig. 2.4, for each of the four components of the beam model the likelihood $p(z^k | z^{k*})$ is visualized. The four weights w_{hit} , w_{short} , w_{max} , and w_{rand} determine the proportion of each basic case to the mixture distribution, which is defined as:

$$p(z^k | z^{k*}) = \begin{pmatrix} w_{hit} \\ w_{short} \\ w_{max} \\ w_{rand} \end{pmatrix}^T \cdot \begin{pmatrix} p_{hit}(z^k | z^{k*}) \\ p_{short}(z^k | z^{k*}) \\ p_{max}(z^k | z^{k*}) \\ p_{rand}(z^k | z^{k*}) \end{pmatrix}, \quad (2.18)$$

with $w_{hit} + w_{short} + w_{max} + w_{rand} = 1$.

In Fig. 2.5, the mixture distribution $p(z^k | z^{k*})$ based on the four basic models is visualized. Overall, the unknown intrinsic parameters of the beam model Θ , which define the shape of the mixture density $p(z^k | z^{k*})$, comprise the standard deviation σ_{hit} of the Gaussian part, the parameter λ_{short} of the exponential part, and the four weights. Given a set of measurements, the intrinsic parameters can be determined using the expectation maximization strategy proposed in Thrun et al. (2005). In an iterative manner, the intrinsic parameters Θ are estimated by maximizing the likelihood $p(\mathbf{Z} | \mathbf{X}, \mathcal{M}, \Theta)$ of the measured ranges $\mathbf{Z} = \{z_i\}$, where the reference poses $\mathbf{X} = \{\mathbf{x}_i\}$, in which the range measurements were conducted, and the map \mathcal{M} together yield the expected range values.

2.2. State estimation filtering

If measurements from various sensors are available for the localization of the vehicle, suitable approaches are required to fuse the information and to estimate an accurate and reliable pose. To this end, commonly Bayesian filters are used, with the Kalman filter being the most popular variant, which originates from the Bayes filter in case of Gaussian distributed states and measurements. For nonlinear systems, it is adapted to the so-called extended Kalman filter (EKF). Besides the Kalman filter and its variants, nonparametric filters, such as histogram and particle filters, are widely used, which cover the state space with uniformly and randomly distributed candidate states, respectively.

2.2.1. Bayes filter

The distribution of a state \mathbf{x} conditioned on a set of observations \mathbf{Z} , with $\mathbf{Z} = \{z_1, z_2, \dots, z_k\}$ and k being the number of observations, is expressed by the conditional probability $p(\mathbf{x} | \mathbf{Z})$. The fundamental relation between the conditional variable $p(\mathbf{x} | \mathbf{Z})$ and its ‘inverse’ $p(\mathbf{Z} | \mathbf{x})$, which is also called *likelihood*, is defined by the Bayes rule or synonymously by the Bayes theorem (Bishop, 2006):

$$p(\mathbf{x} | \mathbf{Z}) = \frac{p(\mathbf{Z} | \mathbf{x}) p(\mathbf{x})}{p(\mathbf{Z})}, \quad (2.19)$$

where $p(\mathbf{x})$ is the a priori probability distribution of the state, which is also referred to as *prior*, and $p(\mathbf{Z})$ is the normalization term, which is determined by

$$p(\mathbf{Z}) = \int p(\mathbf{Z}|\mathbf{x}) p(\mathbf{x}) d\mathbf{x}. \quad (2.20)$$

If the state \mathbf{x} is conditioned on multiple variables, e.g. \mathbf{Z} and \mathbf{u} , it can be extended as follows:

$$p(\mathbf{x}|\mathbf{Z},\mathbf{u}) = \frac{p(\mathbf{Z}|\mathbf{x},\mathbf{u}) p(\mathbf{x}|\mathbf{u})}{p(\mathbf{Z}|\mathbf{u})}. \quad (2.21)$$

Since the pose of a robot is usually not measured directly, a robot needs to estimate it based on the measured data. In probabilistic robotics, the inferred state of a robot and its environment is called *belief*, which describes the robot's internal knowledge and is represented by a conditional probability distribution.

A belief refers to the posterior probability for a state conditioned on all available information. Hence, for a state \mathbf{x}_t at time t , the belief $bel(\mathbf{x}_t)$ is defined as

$$bel(\mathbf{x}_t) = p(\mathbf{x}_t|\mathbf{Z}_{1:t},\mathbf{u}_{1:t}), \quad (2.22)$$

where the conditional probability $p(\mathbf{x}_t|\mathbf{Z}_{1:t},\mathbf{u}_{1:t})$ is the posterior, with $\mathbf{Z}_{1:t}$ being all past measurements and $\mathbf{u}_{1:t}$ being all past controls. Calculating the belief at time t before obtaining the new set of measurements \mathbf{Z}_t , directly after executing the control \mathbf{u}_t , leads to the *prediction* denoted as:

$$\overline{bel}(\mathbf{x}_t) = p(\mathbf{x}_t|\mathbf{Z}_{1:t-1},\mathbf{u}_{1:t}). \quad (2.23)$$

It is the first of two steps in a probabilistic filter, which is only based on the previous state posterior $bel(\mathbf{x}_{t-1})$ and the control \mathbf{u}_t . In the second step, the belief $bel(\mathbf{x}_t)$ is calculated from the prediction $\overline{bel}(\mathbf{x}_t)$ taking into account the new set of measurements \mathbf{Z}_t , which is called *measurement update* or synonymously *correction*.

In a Bayes filter, the calculation of the belief happens in a recursive manner. Hence, besides the latest control \mathbf{u}_t and latest set of measurements \mathbf{Z}_t , the belief $bel(\mathbf{x}_{t-1})$ from the previous epoch $t-1$ serves as input to determine the new belief $bel(\mathbf{x}_t)$ at time t . One filter iteration includes the prediction (cf. Eq. 2.24a) and the correction (cf. Eq. 2.24b), between which is continuously alternated and which are calculated as follows:

$$\overline{bel}(\mathbf{x}_t) = \int p(\mathbf{x}_t|\mathbf{u}_t, \mathbf{x}_{t-1}) bel(\mathbf{x}_{t-1}) d\mathbf{x}_{t-1} \quad (2.24a)$$

$$bel(\mathbf{x}_t) = \eta p(\mathbf{Z}_t|\mathbf{x}_t) \overline{bel}(\mathbf{x}_t). \quad (2.24b)$$

For the task of robot localization, the state \mathbf{x}_t corresponds to the robot's pose, and the control information \mathbf{u}_t could be, for example, given as motor ticks. The state transition probability $p(\mathbf{x}_t|\mathbf{u}_t, \mathbf{x}_{t-1})$ specifies how the state evolves over time while solely depending on robot controls \mathbf{u}_t , and the measurement probability $p(\mathbf{Z}_t|\mathbf{x}_t)$ defines the probabilistic law that specifies how probable the set of measurements \mathbf{Z}_t is, given the state \mathbf{x}_t . Together, both probabilities specify the robot's dynamical stochastic system. An illustration of the relations between consecutive states, control inputs, and measurements, which are defined by those probabilities, is given in Fig. 2.6, which shows a dynamic Bayes network.

The prediction (cf. Eq. 2.24a) corresponds to the integral over the product of the prior belief $bel(\mathbf{x}_{t-1})$ and the state transition probability $p(\mathbf{x}_t|\mathbf{u}_t, \mathbf{x}_{t-1})$. In the subsequent correction, the

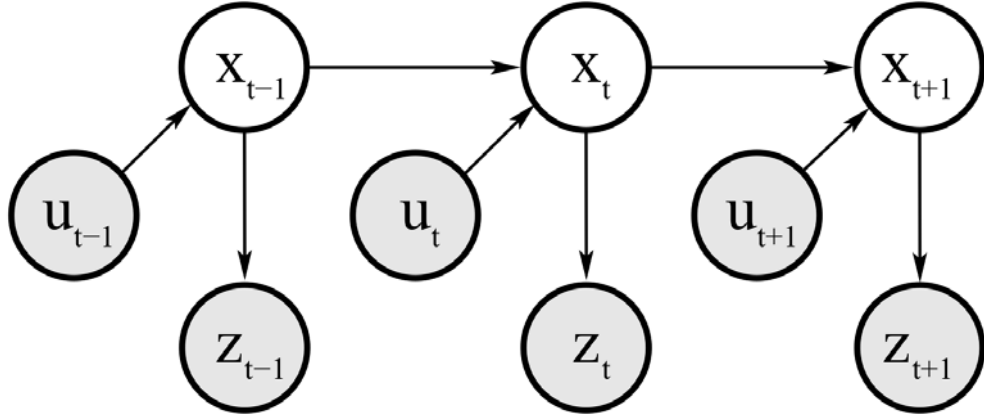


Figure 2.6.: The dynamic Bayes network illustrating the relations between consecutive states, control inputs, and measurements, which are defined by the state transition probability and measurement probability. The illustration is adapted from Thrun et al. (2005).

prediction $\overline{bel}(\mathbf{x}_t)$ is multiplied by the measurement probability $p(\mathbf{Z}_t|\mathbf{x}_t)$, and the product is normalized using the normalization factor η . As the Bayes filter is a recursive algorithm, it requires an initialization $bel(\mathbf{x}_0)$ for the epoch $t = 0$, which can be a point mass distribution if \mathbf{x}_0 is known, a uniform distribution if no initial information is present, or any other distribution in between if partial knowledge about \mathbf{x}_0 exists (Thrun et al., 2005).

2.2.2. Kalman filter

The Kalman filter proposed by Kalman (1960) is an implementation of the Bayes filter, if it is applied to a linear, Gaussian system. The belief at time t is represented by the parameter of a multi-dimensional normal distribution, which are the mean $\boldsymbol{\mu}_t$ and the covariance matrix $\boldsymbol{\Sigma}_t$. The motion equation to predict the state \mathbf{x}_t at time t is the linear function

$$\mathbf{x}_t = \mathbf{A}_t \mathbf{x}_{t-1} + \mathbf{B}_t \mathbf{u}_t + \boldsymbol{\epsilon}_t, \quad (2.25)$$

with \mathbf{x}_t being the state vector of dimension n and \mathbf{u}_t the control vector of dimension m . The matrix \mathbf{A}_t of dimension $n \times n$ models the influence of the previous state on the current state, and the matrix \mathbf{B}_t of dimension $n \times m$ models the influence of the control on the current state. Additionally, a zero mean, Gaussian noise is introduced by the vector $\boldsymbol{\epsilon}_t$ of dimension n with the associated covariance matrix \mathbf{R}_t of dimension $n \times n$ (Kalman, 1960; Thrun et al., 2005).

The predicted state vector $\bar{\boldsymbol{\mu}}_t$ and its corresponding covariance matrix $\bar{\boldsymbol{\Sigma}}_t$ are determined as follows:

$$\bar{\boldsymbol{\mu}}_t = \mathbf{A}_t \boldsymbol{\mu}_{t-1} + \mathbf{B}_t \mathbf{u}_t, \quad (2.26a)$$

$$\bar{\boldsymbol{\Sigma}}_t = \mathbf{A}_t \boldsymbol{\Sigma}_{t-1} \mathbf{A}_t^T + \mathbf{R}_t. \quad (2.26b)$$

The matrix \mathbf{C}_t of dimension $k \times n$ models the measurements, which can be expressed as

$$\mathbf{z}_t = \mathbf{C}_t \mathbf{x}_t + \boldsymbol{\delta}_t, \quad (2.27)$$

with \mathbf{z}_t being the measurement vector of dimension k , and $\boldsymbol{\delta}_t$ being zero-mean multivariate Gaussian noise with the covariance matrix \mathbf{Q}_t .

For the measurement update, the so-called *Kalman-Gain* matrix \mathbf{K} is introduced, which is calculated based on the covariance matrix of the predicted state and the covariance matrix of the measurement, as stated in Eq. 2.28a. In the update step, the Kalman gain matrix weights according to the uncertainties whether the new state should follow rather the prediction or rather the measurement. The belief $bel(x_t)$ is represented by $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$, which are calculated following Eq. 2.28b and Eq. 2.28c, respectively. The difference between the actual measurement \mathbf{z}_t and the expected measurement $\mathbf{C}_t \bar{\boldsymbol{\mu}}_t$ is called *innovation* (Thrun et al., 2005).

$$\mathbf{K}_t = \bar{\boldsymbol{\Sigma}}_t \mathbf{C}_t^T (\mathbf{C}_t \bar{\boldsymbol{\Sigma}}_t \mathbf{C}_t^T + \mathbf{Q}_t)^{-1} \quad (2.28a)$$

$$\boldsymbol{\mu}_t = \bar{\boldsymbol{\mu}}_t + \mathbf{K}_t (\mathbf{z}_t - \mathbf{C}_t \bar{\boldsymbol{\mu}}_t) \quad (2.28b)$$

$$\boldsymbol{\Sigma}_t = (\mathbf{I} - \mathbf{K}_t \mathbf{C}_t) \bar{\boldsymbol{\Sigma}}_t \quad (2.28c)$$

Detailed information on the mathematical derivations of the Kalman filter can be found in Kalman (1960) and Thrun et al. (2005).

2.2.3. Histogram filter

The histogram filter belongs to the category of so-called *nonparametric filters*, which do not depend on a specific functional form for the posterior distribution. They are well-suited for the representation of multimodal beliefs, which are required in case of data association problems that lead to separate, distinct hypotheses. Consequently, compared to the Kalman filter, the histogram filter is not restricted to Gaussian distributions. Yet, the ability to represent arbitrary distributions is accompanied by a higher computational cost. In general, however, using nonparametric methods, the number of parameters can be chosen in accordance to the complexity of the distribution, and in case of low complexity, only a small number of parameters is required.

Applying a histogram filter, the state space is divided into a finite number of bins or regions, where the posterior of each region corresponds to a single probability. Thus, a continuous density is approximated by a histogram of piece-wise constant probabilities, and with an increasing number of bins, the discrete approximation tends to converge to the correct posterior. Besides the histogram filter, the second most famous, nonparametric filter technique is the particle filter, which represents the posterior by a finite set of state samples randomly drawn from the posterior distribution (Thrun et al., 2005).

In a histogram filter, the discrete representation of a continuous random variable \mathbf{X}_t is typically obtained by decomposition using a multi-dimensional grid, resulting in finitely many grid cells. Each of these multi-dimensional grid cells, denoted as $\mathbf{x}_{k,t}$, only carries a probability $p_{k,t}$ and is often represented by the mean state. This also applies to the localization approaches proposed in this work where the candidate poses are the center points of the multi-dimensional grid cells. The granularity of the grid serves as a parameter to adjust between accuracy and computational efficiency, where small grid cells result in little approximation errors, however, at the expense of large computational cost and vice versa (Thrun et al., 2005).

The histogram filter is obtained by representing the general Bayes filter formulation stated in Eq. 2.24 in a discrete fashion and substituting the integration over a continuous distribution with a sum over discrete states. This leads to the following relations for the prediction (cf. Eq. 2.29a)

and measurement correction (cf. Eq. 2.29b), which are conducted for each individual candidate state:

$$\bar{p}_{k,t} = \sum_i p(\mathbf{X}_t = \mathbf{x}_k | \mathbf{u}_t, \mathbf{X}_{t-1} = \mathbf{x}_i) p_{i,t-1} \quad (2.29a)$$

$$p_{k,t} = \eta p(\mathbf{Z}_t | \mathbf{X}_t = \mathbf{x}_k) \bar{p}_{k,t}, \quad (2.29b)$$

with \mathbf{x}_i and \mathbf{x}_k being individual states (Thrun et al., 2005). The inputs of the filter are the belief of the previous epoch $t - 1$, which is the discrete probability distribution over all grid cells $\{p_{i,t-1}\}$, the latest control \mathbf{u}_t , and the latest set of measurements \mathbf{Z}_t . In the prediction, the probability $\bar{p}_{k,t}$ of being in state \mathbf{x}_k is estimated while only considering the control, and the probabilities $\{\bar{p}_{k,t}\}$ of all states form the predicted belief at time t . In the update step, the probability $p_{k,t}$ of being in state \mathbf{x}_k is corrected based on the set of measurements, and the probabilities of all states $\{p_{k,t}\}$ form the new belief.

2.3. LiDAR sensor and point cloud acquisition

Light Detection And Ranging (LiDAR) sensors, also referred to as laser scanners, provide accurate range measurements and thereby enable to capture the three-dimensional environment. By measuring the elapsed time between the emission of a laser light beam and the detection of its reflection, the distance from the sensor to the reflecting object can be determined:

$$d = c \frac{\Delta t}{2}, \quad (2.30)$$

with c being the speed of light and Δt being the elapsed time. This principle is called Time-of-Flight (ToF), on which conventional pulsed LiDARs rely (Li et al., 2022).

To measure the range using LiDAR, depending on the application, light with a wavelength of 0.4-1.6 μm is used (Beraldin et al., 2010). In the automotive and robotics field, typically near-infrared light with a wavelength around 905 nm is employed, which also applies to the LiDAR sensors used in this work (Velodyne LiDAR, Inc., 2016; Hesai Technology Co., Ltd., 2021).

The application of LiDAR measurements can be divided in three categories:

- **Aerial or airborne** laser scanning refers to the process of capturing an area from the air with the LiDAR being mounted on an airplane, a helicopter, or, increasingly, a UAV.
- **Static terrestrial** laser scanning refers to a measurement conducted from one or multiple static points of view. To capture the environment in three dimensions, the sensor has to rotate around two axes while measuring.
- **Kinematic or dynamic terrestrial** laser scanning refers to capturing the environment using a LiDAR mounted on a moving platform, which can be installed on bikes or cars, or even be carried by a person. However, the principal use case are cars in regular traffic, where two main applications are mapping of the environment, e.g. using expensive, high end LiDAR mobile mapping systems, and localization based on maps using LiDAR measurements acquired by a low-cost automotive sensor, which is the application investigated in this work.

LiDAR sensors are commonly classified both based on their sensing principle, which specifies how the range is determined from the transmitted and received signal, and their mechanical structure, which can be either mechanically rotating or solid-state. In (Li et al., 2022), LiDAR sensors are classified into the following three categories based on their ranging technology:

- **Pulsed ToF** is the prevalent LiDAR principle, where a light pulse is first emitted, then reflected by the target, and finally detected by the sensor’s receiver. The elapsed time between the emission and reception leads to the range measurement (cf. Eq. 2.30). Due to the high power of the pulse, the pulsed ToF principle is suitable for sensing longer distances. It is employed in all LiDARs used in this work.
- **Amplitude-modulated continuous wave (AMCW) ToF** LiDARs utilize an amplitude-modulated optical signal instead of an optical pulse. In the AMCW method, the range is determined based on the phase shift between the transmitted modulated light and the received light. Since the AMCW principle does not rely on a pulse, but rather on a modulated signal, it is only applicable to sense moderate distances.
- **Frequency-modulated continuous wave (FMCW)** LiDARs emit an optical signal with modulated frequency and then analyze the beat signal between emitted and reflected signal. The frequency modulation enables the detection of the Doppler effect. As a result, the relative velocity of the measured target can be determined in addition to its distance.

With respect to the mechanical structure, traditional LiDAR sensors rely on a mechanically rotating system, which is usually a mirror. To capture the three dimensional environment, it is necessary to emit the laser beams in different directions along two spatial axes. In terrestrial laser scanning, this is achieved by rotating a mirror around one axis and simultaneously rotating the housing around a second axis. In the case of mobile platforms, besides a rotating mirror, the third dimension is realized by the movement of the platform and the installed LiDAR sensor (Beraldin et al., 2010). For the deflection of the laser beam, various rotating mirror constructions exist. Detailed information on these can be found in (Beraldin et al., 2010). Using automotive LiDARs, a 3D coverage of the environment is achieved by not only deploying one laser-detector-pair, but an array of laser-detector-pairs (cf. Fig. 2.7). These laser-detector-pairs are arranged at fixed angular intervals along the rotating axis enabling the acquisition of 3D data (Velodyne LiDAR, Inc., 2016; Hesai Technology Co., Ltd., 2021).

While mechanically rotating LiDARs enable laser scanning with a wide Field of View (FoV), their mechanical structure leads to a bulky size and a susceptibility to vibrations. In contrast, solid-state LiDARs operate without any mechanical components. Currently, various implementations of solid-state LiDARs exist, where the three most prevalent are (Li et al., 2022): *Flash LiDARs* utilizing a photodetector (PD) array, which enables to capture the entire target scene in a single shot, *Microelectromechanical systems-based (MEMs) LiDARs* using MEMS mirrors, which offer advantages in terms of compactness and weight, and *Optical Phased Array-based (OPA) LiDARs*, which rely on integrated photonics technology and provide a similarly compact sensor.

Independent of the LiDAR model, the mechanical structure and the sensing principle, the sensor provides the point measurements in spherical coordinates comprising range r , azimuth α (horizontal angle), and elevation β (vertical angle) with respect to its internal coordinate system. In Fig. 2.7, the structure of a mechanically rotating LiDAR is illustrated. The vertical FoV covered by 32 scan layers, the sensor coordinate system, and the definition of the azimuth angle are visualized.

The conversion into Cartesian coordinates is as follows (Velodyne LiDAR, Inc., 2016):

$$\mathbf{p} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} r \cos \beta \sin \alpha \\ r \cos \beta \cos \alpha \\ r \sin \beta \end{pmatrix}. \quad (2.31)$$

Note that the definition of the angles may differ for various LiDAR manufacturers and therefore also the stated conversion rule. Besides the spherical coordinates of a point, LiDAR sensors measure

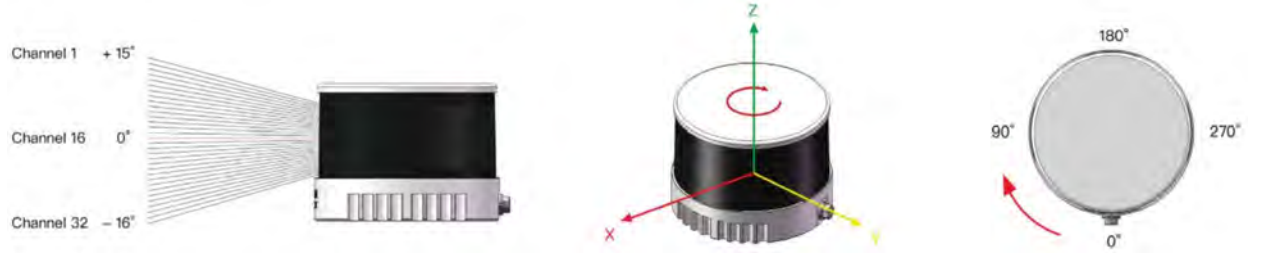


Figure 2.7.: Structure of a mechanically rotating LiDAR: On the left, the vertical FoV covered by 32 scan layers is visualized, in the middle, the sensor coordinate system is depicted, and on the right, the definition of the azimuth angle is shown. The illustration is adapted from Hesai Technology Co., Ltd. (2021).

the amplitude of the reflected laser beam, which is typically stored as *intensity* value and which can provide additional utilizable information about the measured object.

The measurements are typically conducted in a high frequency, with modern LiDARs capturing several million points per second. After converting them into Cartesian coordinates, they can be visualized as so-called *3D-point clouds*. An example is shown in Fig. 2.8, where the points are colored according to the measured intensity value.

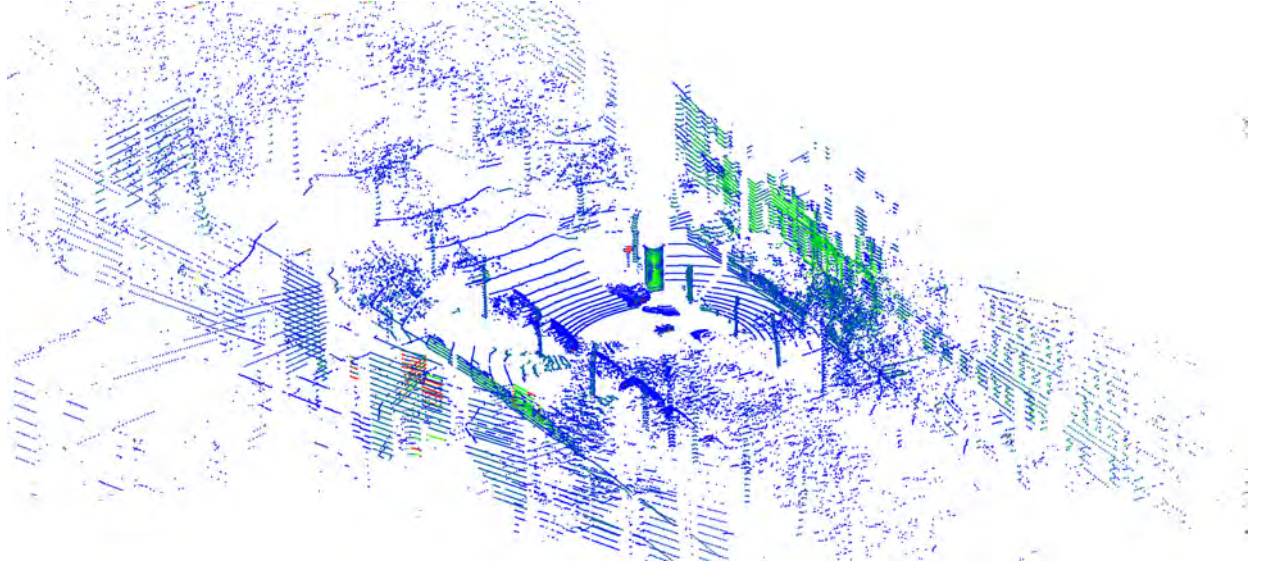


Figure 2.8.: 3D LiDAR point cloud with over 60000 points acquired using a mechanically rotating, pulsed ToF, automotive LiDAR sensor with 32 scan layers. The points are colored according to the measured intensity value. Various objects such as trees, facades, street signs and parked cars are easily identifiable.

In this work, two mechanically rotating, pulsed ToF automotive LiDARs are used for the task of vehicle localization, providing 16 and 32 scan layers, respectively. Detailed information on the sensors is given in Sec. 4.1.2. In the following, the term LiDAR refers to a mechanically rotating LiDAR sensor as used in the context of this work.

2.3.1. Scan image representation and rendering

Besides the representation as 3D point cloud (cf. Fig 2.8), LiDAR scans can be illustrated and exploited as range or scan image, which are used synonymously in this manuscript. The range image can be considered as an immediate representation of a scan as it can be almost directly derived from the measured raw data. For a rotating LiDAR, it contains the measured distances

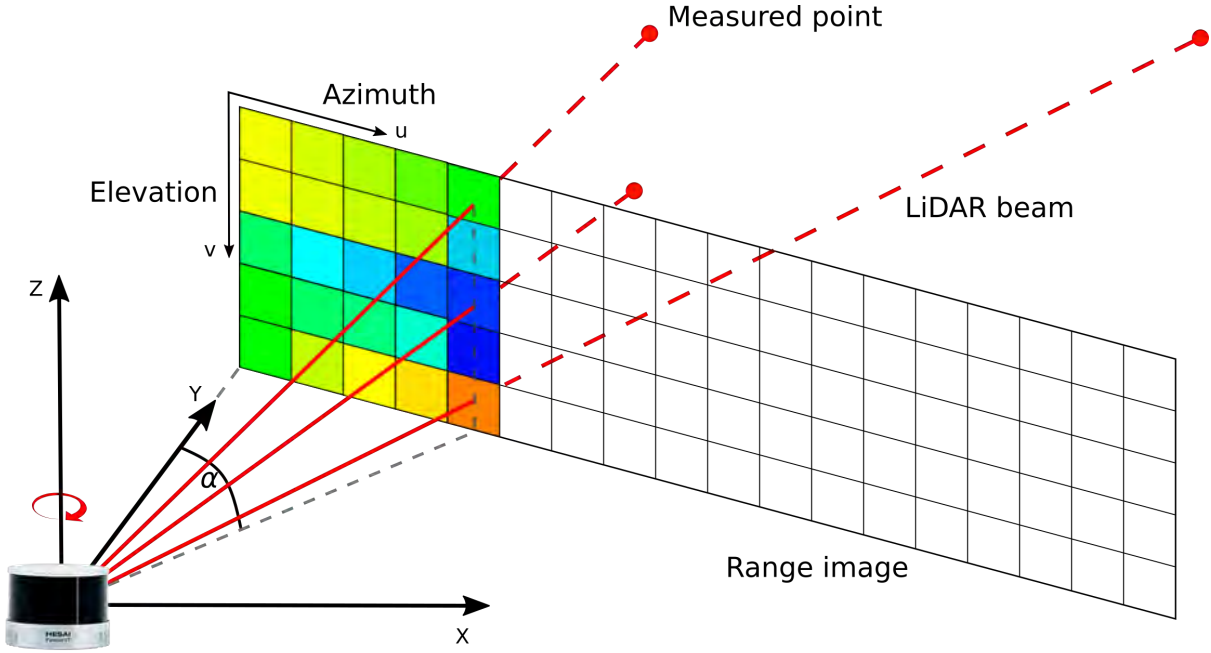


Figure 2.9.: Basic principle how a range image is created: For one 360° measurement rotation, the determined ranges from the laser beams are assigned to the corresponding pixels and the pixel color encodes the range value. The illustration is adapted from Fan et al. (2021).

of one full 360° scan as pixel values, with the scan layer defining the row and the azimuth angle defining the column of the corresponding pixel. Figuratively speaking, during one rotation, the LiDAR fills the 2D image in azimuthal direction by assigning to each pixel the measured range of the corresponding laser beam. A visualization of the principle is given in Fig. 2.9.

Given the number of scan layer n_l and the horizontal resolution $\Delta\alpha$ of the scanner, the resulting range image has the dimension $n_l \times \frac{360^\circ}{\Delta\alpha}$, with n_l defining the height and $\frac{360^\circ}{\Delta\alpha}$ the width. An illustration of a measured range image is given in Fig. 2.10, where for the sake of better visibility only a section of the full 360° image is shown. On the left side, the 32 scan layers of the LiDAR are visualized, whose measured range values are practically rolled off when displayed as a scan image. The layer color encodes the layer number from 0 to 31, and the pixel color encodes the measured range. In the scan image, objects such as parked cars, trees, and buildings are discernible based on the depth jumps.

The LiDAR sensor provides an azimuth angle α and an integer number l for the layer, where the numbering of the layers from top to bottom or vice versa is varying between different manufacturers. Hence, the relation to map the azimuth α and the layer l to pixel coordinates is defined as:

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \frac{\alpha}{\Delta\alpha} \\ f_{LiDAR}(l) \end{pmatrix}, \quad (2.32)$$

with f_{LiDAR} being a sensor specific function to map l to the corresponding row of the scan image while ensuring that the layer with the highest inclination obtains $v = 0$ and the layer with the lowest inclination obtains v_{max} .

Due to the unstructured nature of a 3D point cloud, particularly in large-scale contexts such as autonomous driving applications, the processing and direct utilization of 3D data is often challenging for methods using neural networks. However, the use of 2D scan images enables more efficient processing and a more compact representation without sacrificing raw data information,

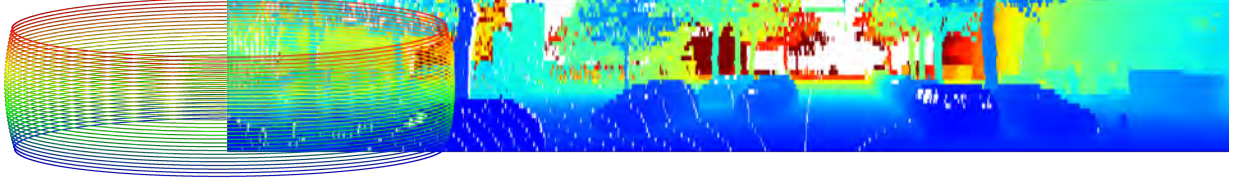


Figure 2.10.: Illustration of a measured range image: For the sake of visibility only a section of the 360° range image is shown. On the left side, the 32 scan layers of the LiDAR are visualized, whose measured range values are practically rolled off when displayed as a scan image. The layer color encodes the layer number from 0 to 31, and the pixel color encodes the measured range. In the scan image, objects such as parked cars, trees, and buildings are discernible based on the depth jumps.

which allows to exploit well-established 2D convolutional neural networks that have originally been employed for tasks based on images. In addition to the range, other information obtained from the LiDAR, such as intensity, can be represented as 2D images and fed into a neural network as additional channels.

In this work, the representation as scan images is used to efficiently evaluate candidate poses based on the maximum consensus criterion and other loss functions for the task of vehicle localization. To this end, synthetic scan images are rendered on the GPU for a set of candidate positions, using triangular meshes to represent the environment. By comparing the synthetic scan images with the measured scan image, using the maximum consensus criterion or other objective functions, the localization solution is found.

2.3.2. Point cloud registration

Vehicle localization is usually tied to point cloud registration as matching the LiDAR scan acquired by the sensor on the ego-vehicle to a map leads to the pose of the vehicle. In the following, the fundamentals for point cloud registration are presented. An overview about various state of the art approaches is provided in Sec. 3.

Given N putative correspondences $D = \{(\mathbf{p}_i, \mathbf{q}_i)\}_{i=1}^N$ from two sets of 3D points, the point cloud registration problem can be formulated as

$$\mathbf{R}^*, \mathbf{t}^* = \underset{\mathbf{R} \in \text{SO}(3), \mathbf{t} \in \mathbb{R}^3}{\operatorname{argmin}} \sum_{i=1}^N \|\mathbf{R}\mathbf{p}_i + \mathbf{t} - \mathbf{q}_i\|_2^2, \quad (2.33)$$

with \mathbf{R} being a 3×3 rotation matrix of the special orthogonal group $\text{SO}(3)$ and \mathbf{t} a 3×1 translation vector. The residual function corresponds to the Euclidean distance between the points \mathbf{p}_i , after applying the rigid transformation, and the associated points \mathbf{q}_i : $r_i(\mathbf{R}, \mathbf{t}) = \|\mathbf{R}\mathbf{p}_i + \mathbf{t} - \mathbf{q}_i\|_2$. Hence, point cloud registration aims to find the optimal rigid transformation \mathbf{T} consisting of the rotation matrix \mathbf{R} and translation \mathbf{t} to align the point clouds (Chin and Suter, 2017).

In general, 3D point cloud registration methods can be divided into two categories: *Correspondence-based* approaches and *correspondence-free* approaches, which are also called *Simultaneous Pose and Correspondence* approaches (Yang et al., 2021).

In **correspondence-based methods**, corresponding points are either known or established using feature descriptors (Rusu et al., 2009; Drost et al., 2010; Tombari et al., 2013). If the correspondences are known and do not contain any outliers, the optimal transformation can be estimated in closed form as proposed by Horn (1987). However, if the correspondences are determined based on feature descriptors, they typically exhibit a large portion of outliers as 3D feature matching is

error-prone. In real-world data, outlier rates often exceed 95 percent, requiring robust methods that can still solve the registration problem (Parra Bustos and Chin, 2018).

To robustly estimate the rigid transformation \mathbf{T} , the point cloud registration problem can be formulated as a maximum consensus problem (cf. Eq. 2.8 and 2.9), where $\omega = (\mathbf{R}, \mathbf{t})$ and Ω is the space of all rigid transformations, which is known as the special Euclidean group $SE(3)$ (Chin and Suter, 2017):

$$\begin{aligned} \mathbf{R}^*, \mathbf{t}^* &= \operatorname{argmax}_{\mathbf{R} \in SO(3), \mathbf{t} \in \mathbb{R}^3} \Psi(\mathbf{R}, \mathbf{t}) \\ \Psi(\mathbf{R}, \mathbf{t}) &= \sum_{i=1}^N \mathbb{I}(\|\mathbf{R}\mathbf{p}_i + \mathbf{t} - \mathbf{q}_i\|_2 \leq \epsilon). \end{aligned} \quad (2.34)$$

Arguably the most famous approach for robust registration is Random Sample Consensus. While RANSAC proves highly effective in applications with little noise and few outliers, its performance diminishes in the presence of high outlier rates, leading to slow convergence and reduced accuracy, as it becomes more difficult to find a candidate model which does not contain a single outlier. Furthermore, as RANSAC relies on randomly instantiating candidate models, it cannot give any guarantees to find the optimal solution. To this end, global methods based on maximum consensus have been developed, which guarantee to yield the globally optimal solution. Specific approaches can be found in Sec. 3.1.1.

Other approaches to solve an outlier contaminated estimation rely on the above mentioned M-estimators, which replace the quadratic loss function with robust loss functions that are less sensitive to outliers.

Simultaneous pose and correspondence (SPC) methods, also referred to as correspondence-free approaches, alternate between the two steps of establishing correspondences and minimizing the errors by estimating the best transformation. To improve the efficiency of determining correspondences, *kd-tree* spatial data structures are commonly utilized (Bentley, 1975). The most famous SPC technique is the Iterative Closest Point algorithm introduced by Besl and McKay (1992), which iteratively finds point-to-point correspondences based on the minimum Euclidean distance and minimizes the errors using least squares. This strategy, however, has significant drawbacks: On the one hand, it can converge to a local minimum, which makes a good initialization inevitable, and on the other hand, it can be heavily biased by outliers, which requires measures to robustify the estimation. To this end, many further developments of the ICP have been proposed, some of which are discussed in Sec. 3.1.2. In addition, to more realistically model the matching of surfaces, improved approaches have been proposed that do not establish point-to-point, but point-to-plane or plane-to-plane correspondences (Chen and Medioni, 1992; Zhang, 1994; Segal et al., 2009).

2.4. Map representations

The localization of autonomous vehicles requires a representation of the environment, commonly referred to as a *map*. These maps can either be known a priori or be created simultaneously, which then corresponds to a Simultaneous Localization and Mapping (SLAM) problem. In the context of this work, the focus is on localization based on a priori known maps, where, however, there exists a broad variety of different types. In the following, different map representations are presented.

2.4.1. Topological maps

In a topological map, besides the position of map objects, their spatial relationships and connections are represented. In fact, the focus is on the relative relationships between objects rather than their exact, absolute locations. Topology refers to a collection of rules that allows to model geometric relationships, which are essential to ensure data integrity and enable data management operations (Esri, 2023). Spatial relationships are defined between geographic elements, which are points, lines, and polygons, using basic topological features, which are nodes, areas and edges. By defining these basic elements along with mutual relations to the adjacent features, the geometries can be represented using a graph of topological elements. Hence, a layer of polygons can be either described as a collection of geographic elements or graph of topological features enabling operations both based on coordinates and based on topology. Spatial relationships can be analyzed and spatial operations conducted, as, for example, merging neighboring polygons with the same attribute value by dissolving existing boundaries (Esri, 2023).

Topological maps are particularly useful for tasks where the exact geometry of the features is less important than their relationships or connections, such as route planning, network analysis, and spatial reasoning. The advantage of topological maps is that they can be stored efficiently since only positions and features of entire objects need to be stored, resulting in little required storage per object. Hence, with respect to the task of navigation, the run time of algorithms based on topological maps is directly depending on the number of objects in the map. The disadvantage of topological maps is that objects cannot be immediately derived from sensor data, which requires an additional processing and feature extraction (Langerwisch, 2014).

2.4.2. Point cloud maps

3D georeferenced data is nowadays used for a broad range of applications, such as documentation and change detection, creation of 3D city models, control in construction, and vehicle navigation, in industry, administration as well as science (Kutterer, 2010). Georeferenced object information can be obtained from 3D point clouds acquired during airborne or terrestrial laser scanning, where the terrestrial laser scanning can be either conducted in static mode or in kinematic mode. In either case, the 3D point cloud ‘maps’ describe the environment using for each point a 3D Cartesian coordinate, possibly an intensity value, and, if any annotation has been carried out, even a class label. Even though static terrestrial scanning yields point clouds of good geometric quality due to the stable scanner, the data acquisition is too time consuming to be applied in large scale. In kinematic mode, the effort for point cloud acquisition is significantly reduced and road scenes can be observed much faster compared to static mode. Whereas in static laser scanning, all points of one scan refer to the same static coordinate system (before georeferencing), in dynamic laser scanning, the LiDAR moves continuously and each individual scan point refers to its unique 3D Cartesian coordinate system. To transform all the points into a joint, global reference system, such as an Earth-centered, Earth-fixed (ECEF) coordinate system like WGS 84, both the positions and the orientations of the scanner must be accurately observed using additional GNSS and Inertial Measurement Unit (IMU) equipment. Furthermore, a precise time synchronization between the imaging sensors, such as LiDAR and camera, and the positioning unit needs to be ensured. The implementations of such kinematic systems are called Mobile Mapping System (MMS). They exist in various designs and sensor combinations, and are typically mounted on the roof of a passenger car (Kutterer, 2010). In Fig. 2.11, a RIEGL VMX-250 MMS is shown. The system comprises two RIEGL VQ-250 lasers scanners, four cameras, an Applanix POS LV GNSS/IMU system, an odometer attached to a wheel, and a control unit for operation, processing, and storage. Each RIEGL VQ-250 LiDAR acquires up to 300,000 points per second with a range accuracy of 10 mm (RIEGL Laser Measurement Systems GmbH, 2012).



Figure 2.11.: RIEGL VMX-250 mobile mapping system mounted on the roof of a Volkswagen T5 minibus. The system comprises two RIEGL VQ-250 lasers scanners, four cameras, an Applanix POS LV GNSS/IMU system, an odometer attached to a wheel, and a control unit for operation, processing, and storage (RIEGL Laser Measurement Systems GmbH, 2012).

While driving, both laser scanners are continuously measuring in a 360° 2D profile and due to their tilted orientation the resulting point cloud shows a crossing scan pattern. Besides the angular resolution and turning rate of the LiDARs, the velocity of the vehicle directly affects the resolution of the acquired point cloud, where the higher the velocity, the lower is the resolution, and vice versa. The 3D point cloud is obtained by transforming the points from the scanner coordinate system into a global coordinate system based on the simultaneously observed vehicle trajectory. Consequently, a crucial requirement for mobile mapping is to observe and provide the positions and orientations of the scanners with high temporal resolution and high accuracy in a global reference system. This is especially difficult in urban environments where due to high-rise buildings GNSS signals are blocked or reflected, leading to Non-Line-Of-Sight (NLOS) and multipath effects. To obtain more accurate trajectory information, in a post processing, Real-Time Kinematic positioning (RTK) correction data, e.g. in Germany provided by the satellite positioning service SAPOS® (GeoBasis-DE, SAPOS, 2023), is taken into account (Kutterer, 2010). However, it should be noted that RTK cannot prevent inaccuracies caused by multipath effects.

To further improve the accuracy of the map data, point clouds of the same location gathered during multiple measurement runs can be aligned using adjustment techniques, e.g. as proposed by Brenner (2016). Such a final post-processing yields a map with a standard deviation below two centimeters. In addition, the shifts obtained from the point cloud adjustment can be applied as corrections to refine the initial trajectory, which results in a highly precise fit between the final trajectory and the adjusted map point cloud.

The publicly available *LUCOOP: Leibniz University Cooperative Perception and Urban Navigation Dataset* provides, among other data, a map point cloud and the corresponding reference trajectory acquired by the RIEGL VMX-250 MMS shown in Fig. 2.11 (Axmann et al., 2023a).

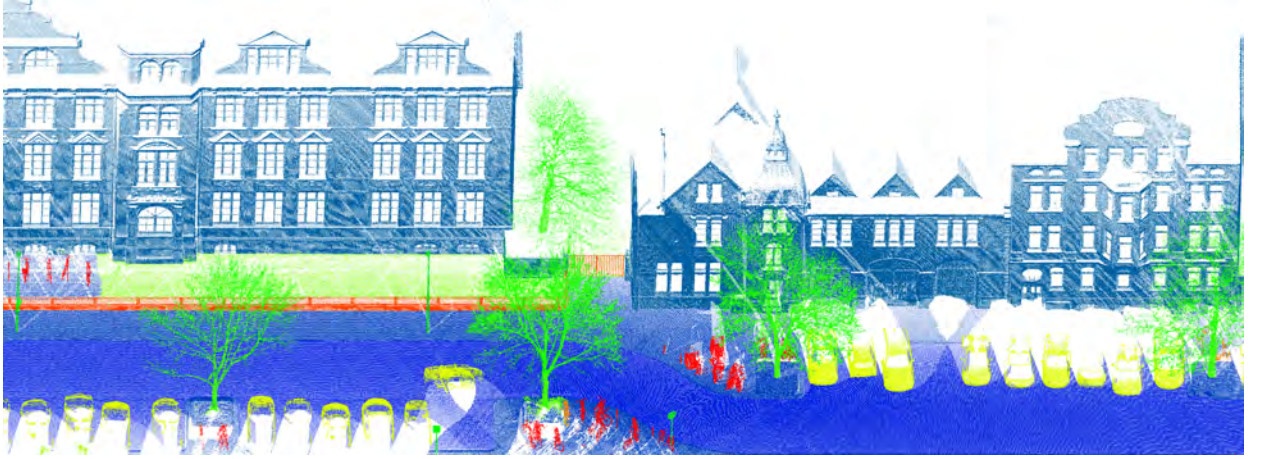


Figure 2.12.: Map point cloud acquired by the RIEGL VMX-250 MMS and colored according to manually annotated labels. The object classes comprise road, sidewalk, building, and wall in blueish colors, street light, street sign, tree crown, tree trunk, bush, and other vegetation in greenish colors, fence, pole, and bike in reddish colors, and cars in yellow color.

Additionally, these map point clouds can be enriched with point-wise class labels, either by manual annotation or by training and employing convolutional neural networks, such as *KPConv* proposed by Thomas et al. (2019). Given the class labels, objects, which may cause wrong associations in the localization of an autonomous vehicle, as it applies to parked cars or vegetation, can be easily removed from the map. In Fig. 2.12, a map point cloud is visualized, which was acquired by the RIEGL VMX-250 MMS and colored according to manually annotated labels. The object classes comprise road, sidewalk, building, and wall in blueish colors, street light, street sign, tree crown, tree trunk, bush, and other vegetation in greenish colors, fence, pole, and bike in reddish colors, and cars in yellow color.

From technical perspective, there are many similarities between mobile mapping and autonomous navigation based on SLAM, however, in mobile mapping, the map is the main product, whereas in SLAM, the generated point cloud map is only a by-product, which serves the main navigation task (Kutterer, 2010). In this work, the point cloud maps are previously generated in a separate mapping step according to the described procedure and are a priori known in the application of vehicle localization.

2.4.3. Occupancy grid maps

The goal of occupancy grid mapping is to generate reliable maps from noisy measurement data, and the corresponding graphical model is shown in Fig. 2.13. In an occupancy grid, the map is represented as a regular grid, and each grid cell corresponds to a binary random variable, which indicates the occupancy of the grid cell's location. The posteriors for these random variables are estimated approximately using respective mapping algorithms. Often, occupancy grid mapping is utilized in a post-processing, e.g., after a SLAM problem has been solved using an alternative method and the estimated poses are considered to be correct. A common use case for these maps are 2D floor plans, which represent only a slice of the 3D environment, as the extension of occupancy grid maps to three dimensions comes along with a significant increase in computational cost (Thrun et al., 2005).

The goal is to compute the posterior $p(m | \mathbf{z}_{1:t}, \mathbf{x}_{1:t})$, with m being the map, $\mathbf{z}_{1:t}$ being all measurements until time t , and $\mathbf{x}_{1:t}$ being all poses of the robot. Occupancy grid maps partition a continuous space into a fine-grained grid, with the grid cell of index i being denoted as \mathbf{m}_i , and

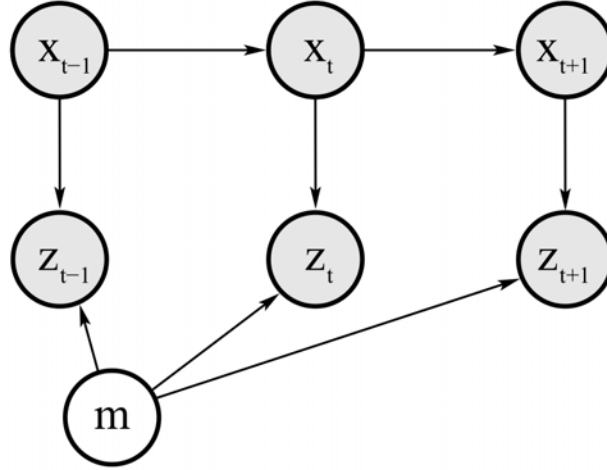


Figure 2.13.: Graphical representation of a mapping problem with the goal to generate the map m , given the premise that the robot poses \mathbf{x} and the measurements \mathbf{z} are known. The figure is taken from Thrun et al. (2005).

the finite set of grid cells representing the full map being denoted as $m = \{\mathbf{m}_i\}$. Furthermore, the probability of an occupied grid cell is denoted by $p(\mathbf{m}_i)$, and the state of occupancy is indicated by a binary value, with ‘0’ referring to free and ‘1’ referring to occupied.

The calculation of the posterior probability $p(m | \mathbf{z}_{1:t}, \mathbf{x}_{1:t})$ over all possible maps is intractable due to its dimensionality, which is why the map estimation problem is decomposed into estimating $p(\mathbf{m}_i | \mathbf{z}_{1:t}, \mathbf{x}_{1:t})$ independently for each grid cell \mathbf{m}_i , and approximating the posterior using factorization (Thrun et al., 2005):

$$p(m | \mathbf{z}_{1:t}, \mathbf{x}_{1:t}) = \prod_i p(\mathbf{m}_i | \mathbf{z}_{1:t}, \mathbf{x}_{1:t}). \quad (2.35)$$

The decomposition of the problem is convenient, but it neglects dependencies between neighboring grid cells. Nevertheless, for the sake of tractability, the factorization is used and the occupancy probability of each grid cell is estimated individually. This results in binary estimation problems over the entire occupancy grid, which are solved using binary Bayes filters with static state, which means that the environment is assumed to be static and to be estimated from a sequence of observations.

In the implementation of the occupancy grid mapping, instead of representing the occupancy by $p(\mathbf{m}_i | \mathbf{z}_{1:t}, \mathbf{x}_{1:t})$, the log odds ratio is used:

$$l_{t,i} = \log \frac{p(\mathbf{m}_i | \mathbf{z}_{1:t}, \mathbf{x}_{1:t})}{1 - p(\mathbf{m}_i | \mathbf{z}_{1:t}, \mathbf{x}_{1:t})}, \quad (2.36)$$

which prevents numerical problems for probabilities close to zero or one, and which can be easily converted back into probabilities:

$$p(\mathbf{m}_i | \mathbf{z}_{1:t}, \mathbf{x}_{1:t}) = 1 - \frac{1}{1 + \exp\{l_{t,i}\}}. \quad (2.37)$$

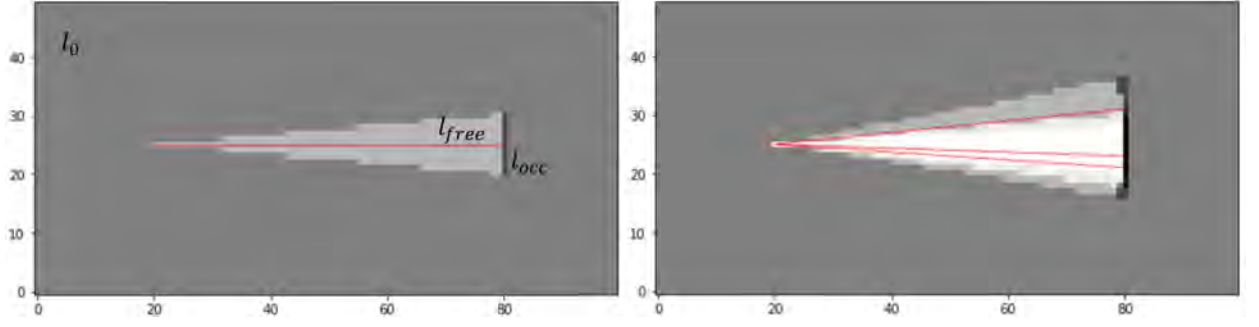


Figure 2.14.: Illustration of a simple inverse sensor model $l(\mathbf{m}_i | \mathbf{z}_t, \mathbf{x}_t)$, which determines the probabilities of the grid cells of the map given a measurement. The darker a grid cell, the more likely it is occupied. On the left, visualized for a single ray, the occupancy value $l_{free} < l_0$ is assigned to the grid cells traversed by the sensor cone, the occupancy value $l_{occ} > l_0$ is assigned to the grid cells of the measured range \mathbf{z}_t where the ray hits an obstacle, and the prior l_0 is assigned to all remaining grid cells. On the right, the effect of adding two more rays is visualized. The figures are taken from Brenner (2023).

In the occupancy grid mapping algorithm, in one iteration, for all grid cells, which are within the sensor cone of observation \mathbf{z}_t , the occupancy values are updated according to

$$l_{t,i} = l_{t-1,i} + l(\mathbf{m}_i | \mathbf{z}_t, \mathbf{x}_t) - l_0, \quad (2.38)$$

where $l(\mathbf{m}_i | \mathbf{z}_t, \mathbf{x}_t)$ implements the inverse sensor model $p(\mathbf{m}_i | \mathbf{z}_t, \mathbf{x}_t)$ as log odds ratio

$$l(\mathbf{m}_i | \mathbf{z}_t, \mathbf{x}_t) = \log \frac{p(\mathbf{m}_i | \mathbf{z}_t, \mathbf{x}_t)}{1 - p(\mathbf{m}_i | \mathbf{z}_t, \mathbf{x}_t)}, \quad (2.39)$$

and l_0 is the constant prior, which is defined as (Thrun et al., 2005):

$$l_0 = \log \frac{p(\mathbf{m}_i = 1)}{p(\mathbf{m}_i = 0)} = \log \frac{p(\mathbf{m}_i)}{1 - p(\mathbf{m}_i)}. \quad (2.40)$$

Both the inverse sensor model $l(\mathbf{m}_i | \mathbf{z}_t, \mathbf{x}_t)$ and the prior of occupancy l_0 are represented in log odds form. In summary, for cell i at time t , an update is determined by adding to the previous log odds ratio $l_{t-1,i}$ the log odds ratio of the current measurement, which is determined by the inverse sensor model, and subtracting the occupancy prior l_0 . Hence, the calculations for the log odds map only rely on addition and subtraction.

For all other grid cells, outside of the perceptual field of measurement \mathbf{z}_t , the occupancy values remain unchanged:

$$l_{t,i} = l_{t-1,i}. \quad (2.41)$$

With respect to the inverse sensor model $l(\mathbf{m}_i | \mathbf{z}_t, \mathbf{x}_t)$, given a measurement, it determines the occupancy values and probabilities of the grid cells of the map, respectively. It assigns the occupancy value $l_{free} < l_0$ to the grid cells traversed by the sensor cone, the occupancy value $l_{occ} > l_0$ to the grid cells of the measured range \mathbf{z}_t where the ray hits an obstacle, and the prior of occupancy l_0 to all the remaining grid cells. An illustration is given in Fig. 2.14, where the darkness of the grid cells indicates their probability of being occupied. Detailed information on the implementation of the inverse sensor model can be found in (Thrun et al., 2005).

2.4.4. Normal distributions transform (NDT) and surfel maps

In the Normal Distributions Transform (NDT) map representation proposed by Biber and Strasser (2003), the object space is partitioned into grid cells, similar to occupancy grid maps. However, while occupancy grid maps store the probability of being occupied for each cell, NDT map representations store a normal distribution for each cell, which is defined by the mean and covariance matrix of all contained points. The additional information inside each grid cell allows to use larger and consequently fewer grid cells, compared to occupancy grid maps. While many scan matching techniques require establishing correspondences, e.g. between points, the cell-wise normal distributions of an NDT map locally model the probability of measuring a point, resulting in a piecewise continuous and differentiable probability density that enables to use Newton's algorithm for scan matching and obviates the need for finding explicit correspondences (Biber and Strasser, 2003).

To generate the NDT map representation, for each cell containing at least three points, the empirical mean $\boldsymbol{\mu}$ and the empirical covariance matrix $\boldsymbol{\Sigma}$ are calculated as follows:

$$\boldsymbol{\mu} = \frac{1}{n} \sum_i \mathbf{p}_i, \quad \boldsymbol{\Sigma} = \frac{1}{n} \sum_i (\mathbf{p}_i - \boldsymbol{\mu})(\mathbf{p}_i - \boldsymbol{\mu})^T. \quad (2.42)$$

The probability of measuring a point \mathbf{p}_j in the cell with index k is then defined by the cell's normal distribution $N(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$:

$$p(\mathbf{p}_j) \sim \exp\left(-\frac{1}{2}(\mathbf{p}_j - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{p}_j - \boldsymbol{\mu}_k)\right). \quad (2.43)$$

Due to the separation into grid cells, artifacts appear at the borders. To reduce these discretization effects, the partitioning into grid cells is conducted for four overlapping grids: the first one is initially placed, the second one is shifted by half a grid cell horizontally, the third one is shifted by half a grid cell vertically, and the fourth one is shifted by half a grid cell horizontally and vertically. If conducted in 3D, this strategy results in eight shifts. Consequently, the computational cost are four and eight times as high, respectively. Afterwards, the probability densities for all shifts are summed up (Biber and Strasser, 2003). In Fig. 2.15, on the left, the partitioning into grid cells and the cell-wise normal distributions are visualized for the initially placed grid, and in the middle, they are visualized exemplarily for the horizontally shifted grid. Furthermore, the sum of the densities from all four grid positions is shown on the right.

Given the application of point cloud registration, the target point cloud or map is represented as NDT according to Eq. 2.42 and the discretization described. After applying an initial transformation on the source point cloud, for each point \mathbf{p}_j , the corresponding grid cell is determined, which is denoted with index k . Given this index, a score is calculated for each point, which is defined as the non-normalized value of the normal distribution (cf. Eq. 2.43):

$$\text{score}_j := \exp\left(-\frac{1}{2}(\mathbf{p}_j - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{p}_j - \boldsymbol{\mu}_k)\right). \quad (2.44)$$

Finally, the overall score for the tested transformation parameters is computed as the sum of all point scores:

$$\text{score} := \sum_j \text{score}_j. \quad (2.45)$$

Following an exhaustive search strategy, the overall score can be computed for a finite set of candidate transformations around the presumably correct transformation, and the solution corresponds

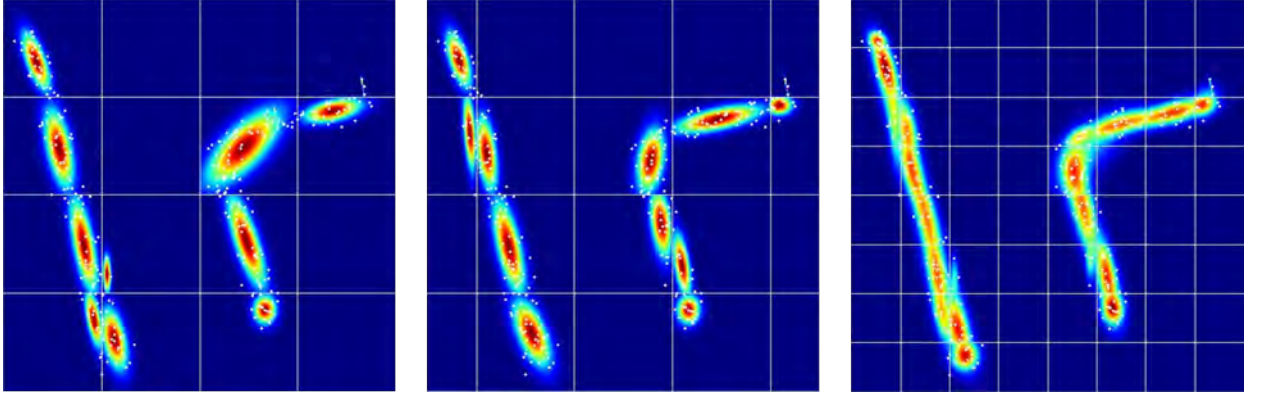


Figure 2.15.: Illustration of the NDT map generation process: On the left, the partitioning into grid cells and the cell-wise normal distributions are visualized for the initially placed grid, and in the middle, they are visualized exemplarily for the horizontally shifted grid. Furthermore, on the right, the sum of the densities from all four grid positions is shown. The figures are taken from Brenner (2023).

to the one yielding the highest score. However, this approach comes along with high computational effort, and due to the analytic score function, it is not necessary to evaluate all possible transformations with respect to their score. Instead, in an iterative procedure, new transformation parameters can be estimated while optimizing the score function using Newton’s algorithm. More precisely, the negative score function is minimized using the gradient and the Hessian matrix, which are the first and second order derivatives of the score function with respect to the transformation parameters, respectively. Following this approach, the solution is typically found within a few iterations, however, due to the non-convexity of the problem, the true solution cannot be found if the initial transformation is too inaccurate, which makes a good initialization inevitable. Detailed information on Newton’s algorithm and the used derivatives can be found in the original publication from Biber and Strasser (2003).

Surfel maps

Using NDT, the representation based on first and second moments inside each grid cell introduces a bias, as points from a planar surface, which are not identical with μ_i , have a non-zero Mahalanobis distance d_M , even if they are located exactly on the plane. In Fig. 2.16, the shortcoming regarding the representation of planar structures is visualized. The formulation of the NDT score based on the Mahalanobis distance is as follows:

$$\begin{aligned} \text{score}_j &:= \exp\left(-\frac{1}{2}d_M^2\right), \\ \text{with } d_M^2 &= (\mathbf{p}_j - \mu_i)^T \Sigma_i^{-1} (\mathbf{p}_j - \mu_i). \end{aligned} \quad (2.46)$$

Surfel maps tackle this drawback by computing a Principal Component Analysis (PCA) inside each cell instead of the normal distribution, and by using the distance to the principal axis (2D) or the principal plane (3D) instead of the Mahalanobis distance (Huang and Anguelov, 2010; Stückler and Behnke, 2014). The result is an improved representation of planar structures, as shown in Fig. 2.17.

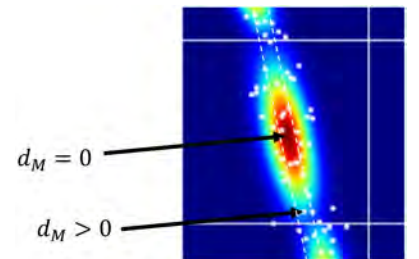


Figure 2.16.: Visualization of the shortcoming regarding the representation of planar structures induced by the Mahalanobis distance of the NDT. The figure is taken from Brenner (2023).

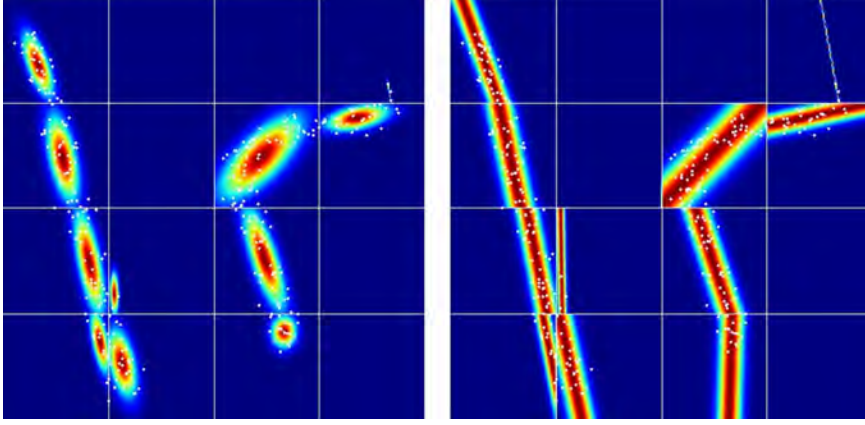


Figure 2.17.: NDT map (left) and surfel map (right): Due to the distance definition with respect to the major axis of the PCA, the surfel map yields an improved representation of planar structures. The figures are taken from Brenner (2023).

2.4.5. Mesh representation using the truncated signed distance function (TSDF)

The Truncated Signed Distance Function (TSDF) was proposed by Curless and Levoy (1996) for the reconstruction of surfaces from aligned range images. Range images contain depth values in a regularly sampled, grid-like FoV and can be generated using depth cameras, but also using LiDARs as introduced in Sec. 2.3.1. In their work, Curless and Levoy defined a list of desirable properties that an algorithm for surface reconstruction should fulfill:

- *Representation of range uncertainty*, which accounts for the fact that range image data typically shows an error distribution that is asymmetric.
- *Utilization of all range data*, which enables to integrate redundant observations to reduce measurement noise.
- *Incremental and order independent updating*, which means that a surface reconstruction can be obtained after integrating only a single, new scan, and a different order of scans does not produce a different result, allowing for parallelization.
- *Time and space efficiency*, which requires an efficient model representation and fast range image integration, to obtain a practical algorithm.
- *Robustness* towards outliers to prevent holes and self-intersecting surfaces.
- *No restrictions on topological type*, which prohibits assumptions about the object's genus in the algorithm that may lead to good results for a certain type of problem, but at the same time cause a poor generalization to other problems.
- *Ability to fill holes in the reconstruction*, which requires a method that automatically and plausibly fills the surface in case of an incomplete reconstruction, for example caused by occlusions.

Their method, which successively integrates range images, meets all of these requirements. The core of the approach is a cumulative weighted signed distance function, which allows to combine new range images with already integrated data in a simple additive scheme (Curless and Levoy, 1996).

In detail, an implicit function $D(\mathbf{x})$ is defined, which represents the weighted signed distance from a point \mathbf{x} to the closest surface along the measurement ray from the sensor. The space containing

the surface or model, which is to be reconstructed, is partitioned into a voxel grid and for each discrete grid location \mathbf{x} , a cumulative signed distance $D(\mathbf{x})$ and a cumulative weight $W(\mathbf{x})$ are calculated as:

$$D(\mathbf{x}) = \frac{\sum w_i(\mathbf{x}) d_i(\mathbf{x})}{\sum w_i(\mathbf{x})}, \quad (2.47a)$$

$$W(\mathbf{x}) = \sum w_i(\mathbf{x}) \quad (2.47b)$$

with $D(\mathbf{x})$ being the weighted mean of signed distance functions $d_i(\mathbf{x})$ and weight functions $w_i(\mathbf{x})$ that are derived from the range images with index i . The weights reflect the varying uncertainty across the triangulated range surface, which mainly depends on the incidence angle between the measuring ray and the surface. It applies the smaller the angle, the higher is the uncertainty.

The cumulative signed distance function $D_i(\mathbf{x})$ and the cumulative weight function $W_i(\mathbf{x})$ can be updated incrementally by integrating the next range image with index $i + 1$. To do this, the following relations are to be applied:

$$D_{i+1}(\mathbf{x}) = \frac{W_i(\mathbf{x}) \cdot D_i(\mathbf{x}) + w_{i+1}(\mathbf{x}) \cdot d_{i+1}(\mathbf{x})}{W_i(\mathbf{x}) + w_{i+1}(\mathbf{x})}, \quad (2.48a)$$

$$W_{i+1}(\mathbf{x}) = W_i(\mathbf{x}) + w_{i+1}(\mathbf{x}) \quad (2.48b)$$

The algorithm applied to three-dimensional data can be broken down into the following steps:

- 1 All voxel weights are initialized as zero.
- 2 For each acquired range image, triangles are created on the sampled lattice by connecting the nearest neighbors. To prevent triangulation across discontinuities, triangles with an edge length over a specific threshold are discarded.
- 3 For each vertex, a weight is calculated according to the dot product between the normal vector of the vertex and the measurement direction of the sensor. For grazing angles between the measurement direction and the surface, the utilized weighting scheme yields small weights, reflecting the higher uncertainty.
- 4 Given a triangular mesh with weighted vertices, the voxel grid is updated, which means that the two functions $D(\mathbf{x})$ and $W(\mathbf{x})$ are recalculated at, and only at discrete grid locations. Moreover, the update happens only for voxels which are in the vicinity of the isosurface $D(\mathbf{x}) = 0$. This restriction corresponds to a truncation of correspondences with large residuals. It ensures robustness towards outliers, reduces the computational effort as only a part of the voxel grid needs to be updated, and leads to the name *truncated* signed distance function (TSDF). Furthermore, to be able to model thin structures and to avoid that opposing surfaces interfere with each other, the truncation range is smaller behind surfaces. Considering the range image with index $i + 1$, for each grid position within the truncation range, the update happens according to the following steps:
 - 4a To compute the signed distance value d_{i+1} , a ray is casted from the sensor through the selected grid location and voxel, respectively, and intersected with the triangular mesh of the current range image generated in step 2. The signed distance d_{i+1} is then calculated from the grid location to the point where the ray intersects the triangular mesh.
 - 4b To compute the weight w_{i+1} for the update, the weights at the vertices of the intersected triangle, which have been computed in step 3, are linearly interpolated.

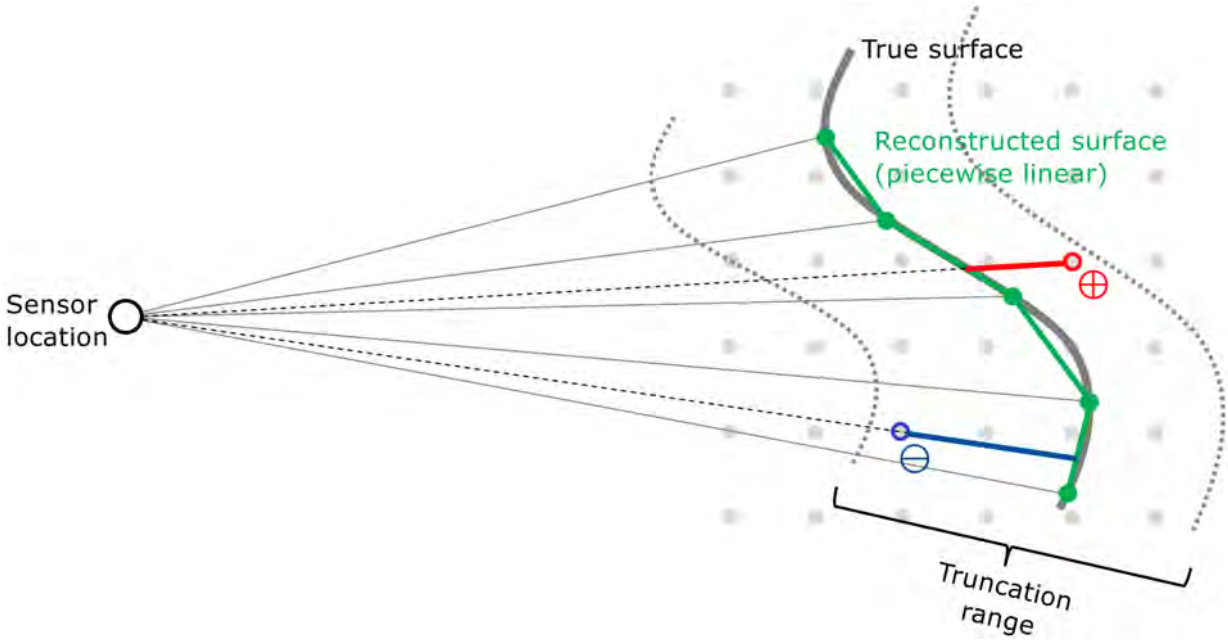


Figure 2.18.: Principle of the TSDF: The gray dots represent the grid locations, of which those within the truncation distance are considered for the update step, and the bold curve indicates the true surface that is to be reconstructed. The solid lines show the rays of the measured range image, the green dots the measured points, and the green lines between those points the piecewise linear, triangulated surface. The dotted lines represent rays casted from the sensor through grid locations and intersected with the triangulated surface of the measured points, with the red and blue lines illustrating positive and negative distance values between the grid location and the point of intersection. The figure is taken from Brenner (2023).

- 4c Given the signed distance d_{i+1} from step 4a and the weight w_{i+1} from step 4b, for a voxel within the truncation range, the cumulative signed distance function $D_{i+1}(\mathbf{x})$ and the cumulative weight function $W_{i+1}(\mathbf{x})$ are updated as stated in Eq. 2.48.

The isosurface $D(\mathbf{x}) = 0$ can be extracted from the voxel grid at any point during the integration of range images. Curless and Levoy (1996) employ the well-known Marching Cubes algorithm proposed by Lorensen and Cline (1987) to obtain the surfaces, while using a lookup table in which ambiguous cases are resolved (Montani et al., 1994).

In Fig. 2.18, the principle of the TSDF is illustrated. The gray dots represent the grid locations, of which those within the truncation distance are considered for the update step, and the bold curve indicates the true surface that is to be reconstructed. The solid lines show the rays of the measured range image, the green dots the measured points, and the green lines between those points the piecewise linear, triangulated surface (cf. step 2). The dotted lines represent rays casted from the sensor through grid locations and intersected with the triangulated surface of the measured points, with the red and blue lines illustrating positive and negative distance values between the grid location and the point of intersection (cf. step 4a).

The step of updating the voxel grid given a new range image can be conducted in two ways: Either by tracing rays into the voxel space, as described in step 4a, or by projecting the voxels within the FoV of the sensor into the image plane. Detailed information on the second approach can be found in the original publication by Curless and Levoy (1996), where it is referred to as *fast volume traversal*. Moreover, in the original article, the utilized run-length encoding, which is employed for the purpose of space efficiency, and the hole filling technique to close gaps in the model are explained.

2.5. Basic concepts of integrity

Integrity aims to increase transportation safety. It is one of the four performance criteria in GNSS navigation, besides accuracy, continuity, and availability (Zhu et al., 2018). The concept was originally introduced in the domain of aviation, where integrity was defined as follows (International Civil Aviation Organization, 2006): *A measure of the trust that can be placed in the correctness of the information supplied by the total system. Integrity includes the ability of a system to provide timely and valid warnings to the user (alerts).* This means that the system is able to determine by itself when it must not be used for the intended operation since certain error bounds with respect to the positioning are exceeded. Timely warnings are specified by the so-called *time-to-alert*, which denotes the time between surpassing the error bound and raising an alert.

Due to the growing automation of vehicles, integrity has also gained interest in the field of autonomous driving, where it is relevant to the localization task. Similar to aviation, the goal is that the vehicle warns the passengers if the estimated position uncertainty does not satisfy predefined safety requirements and exceeds a limit. However, whereas in aviation, the positioning errors in horizontal and vertical directions are relevant, for autonomous vehicles, the positioning errors along the longitudinal and lateral axes of the vehicle have to be considered (Wörner et al., 2016).

The probably most popular representation of a system's operational states concerning its integrity is the Stanford-ESA Integrity Diagram. It defines integrity levels by comparing the estimated error, the true error, and the tolerable error. An illustration is provided in Fig. 2.19 (Tossaint et al., 2007). On the horizontal axis, the True Position Error (TPE) is defined, which is commonly unknown during an operation, and on the vertical axis, the so called Protection Level (PL) is represented, which is an estimated upper bound on the position error calculated by the positioning system. A perfect localization system, which is able to exactly estimate the TPE, would only provide TPE-PL-pairs lying on the diagonal dashed line that separates the diagram into the upper left triangle, where the PL is greater than the TPE, and the lower right triangle, where the PL is less than the TPE. If the system operates within the upper left triangle, represented by the green and yellow areas, it has high integrity and a trustworthy operation can be ensured because the localization unit overestimates the position error. However, if the system operates within the lower right triangle, represented by the orange and red areas, it lacks integrity and cannot be trusted because the localization unit underestimates the position error.

The vertical and horizontal dashed lines indicate the Alert Limit (AL), which corresponds to the maximum tolerable position error. If the PL exceeds the AL, an alert will be raised indicating that the system cannot operate itself safely and the user must take control over the system and operation, respectively. If the PL is larger than the TPE, the alert limit will be exceeded in the upper left triangle, and a transition from *Nominal operation* (green area) to *System unavailable* (yellow area) will take place. In this case, as the PL is still larger than the TPE, integrity is maintained, and only continuity and availability are lost.

Misleading information refers to operational states in which the PL is less than the TPE. This is not very critical if the TPE is less than the AL (large orange area) because the maximum tolerable error is not yet exceeded and no dangerous situation exists, or if the TPE and PL exceed the AL at the same time (small orange area) because an alert is raised and the user can take over control. However, if the TPE exceeds the AL, but the PL does not, a transition from *Misleading information* (orange areas) to *Hazardously misleading information* (red area) will take place. This creates a dangerous situation as the maximum tolerable error is surpassed, while the user, due to the overly optimistic PL and the consequently missing alert, is unaware of the situation and cannot take accident preventing counter-measures.

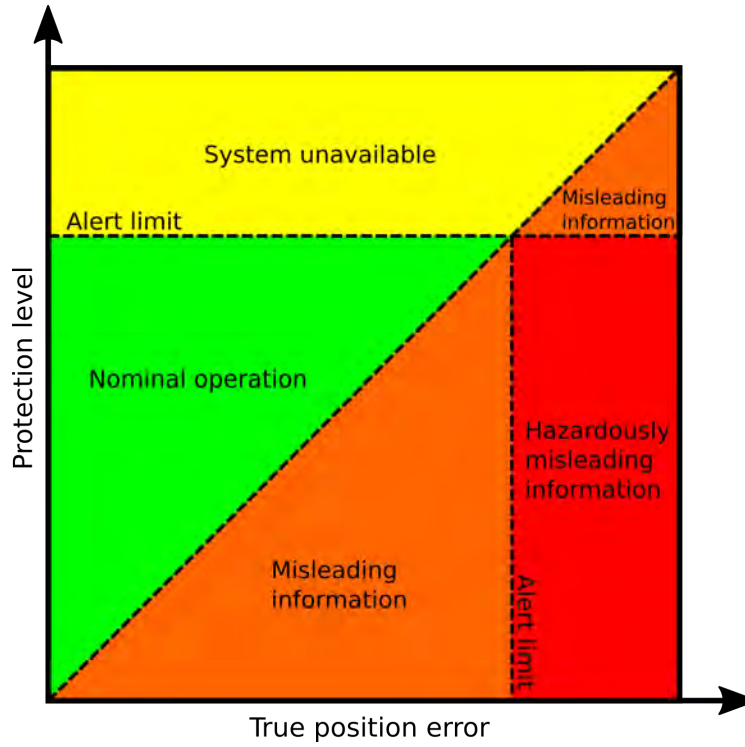


Figure 2.19.: Stanford-ESA Integrity Diagram (Tossaint et al., 2007): The integrity levels of a system are defined by the relations between the protection level (PL), which corresponds to an estimated upper bound on the position error, the true position error (TPE), and the alert limit (AL), which corresponds to the maximum tolerable error.

A central element of the concept of integrity is the *integrity risk*, which refers to the probability of the operational state being *Hazardously misleading information*, i.e., the TPE exceeding the AL without triggering an alert. In aviation, it is typically selected between 10^{-7} and 10^{-9} (Zhu et al., 2018), and for localization in road traffic, Reid et al. (2019) determine an integrity risk of 10^{-8} failures per hour of operation. Thus, in order to guarantee the safety of a system, the proportion of TPE-PL-pairs within the area of *Hazardously misleading information* must never exceed the specified integrity risk (Wörner et al., 2016).

3. Related work

In this chapter, an overview of the state-of-the-art methods is given, which are related to the different parts of the robust and reliable localization pipeline proposed in this thesis (cf. Fig. 1.4). More specifically, in Sec. 3.1, various point cloud registration techniques are presented that may not have been originally developed for localization, however, as LiDAR-based localization and point cloud registration are usually tightly coupled, they are taken into consideration. Afterwards, in Sec. 3.2, methods especially designed for LiDAR-based localization are introduced, where local map matching and global place recognition as well as relevant SLAM approaches are presented. Finally, in Sec. 3.3, the introduced state-of-the-art methods are discussed and the research gap is identified.

3.1. Point cloud registration

There exist two principles for 3D point cloud registration, based on which the corresponding methods can be categorized: Correspondence-based methods and Simultaneous Pose and Correspondence (SPC) methods, i.e. correspondence-free methods (Yang et al., 2021).

3.1.1. Correspondence-based methods

Using correspondence-based techniques, the first step is to identify distinctive feature points in both the source and the target point cloud. Afterwards, these feature points from both point clouds are assigned to each other based on their feature descriptors (Rusu et al., 2009; Drost et al., 2010; Tombari et al., 2013; Choy et al., 2019). Given the putative correspondences, a transformation between both point clouds is estimated. However, during the feature matching, typically many wrong assignments occur that easily cause outlier ratios of 95% and make robust techniques inevitable (Parra Bustos and Chin, 2018).

Assuming a registration problem with known correspondences, points subject to zero-mean Gaussian noise, and no outliers, the optimal transformation consisting of scale, rotation, and translation can be estimated in closed form according to Horn (1987) or Arun et al. (1987). Whereas in the estimation from Horn (1987) rotations are represented using quaternions, in the approach from Arun et al. (1987) they are represented using rotation matrices. If the correspondences are unknown, the registration problem becomes nonconvex, and approaches based on local optimization, such as ICP, are prone to getting trapped in local minima. To tackle this problem, Olsson et al. (2009) and Briales and Gonzalez-Jimenez (2017) propose registration approaches that utilize point-to-point, point-to-line, and point-to-plane correspondences and guarantee to find the globally optimal solution. While Olsson et al. (2009) use convex underestimators together with branch-and-bound, the method proposed in (Briales and Gonzalez-Jimenez, 2017) is based on Lagrangian dual relaxation.

In the presence of outliers, robust registration methods are required, of which RANSAC is arguably the most famous one (Fischler and Bolles, 1981). Even though it has facilitated many applications in computer vision (Hartley and Zisserman, 2004; Meer et al., 1991), due to its randomized sampling and instantiation of candidate models RANSAC lacks both efficiency and accuracy in the presence of high outlier rates (Parra Bustos and Chin, 2018). Consequently, it also does not allow any guarantees in terms of optimality.

Besides the implementation of the maximum consensus criterion, other methods rely on M-estimation to obtain higher robustness towards outliers (MacTavish and Barfoot, 2015; Black and Rangarajan, 1996; Barron, 2017; Zhou et al., 2016). In (MacTavish and Barfoot, 2015), different robust loss functions are implemented in an IRLS scheme and compared with respect to the task of camera-based visual navigation. The aggressive re-descending Geman-McClure-based loss functions show good results in terms of accuracy and convergence, enabling successful localization even for rates of 85% correspondence outliers.

In (Zhou et al., 2016), a fast global registration method based on a scaled Geman-McClure estimator is proposed. To solve the nonconvex optimization, a line process as introduced by Black and Rangarajan (1996) is utilized. In addition, graduated non-convexity is leveraged during the optimization for a more precise registration, where the shape parameter μ of the Geman-McClure function is gradually reduced. The method does not require an initialization and is highly efficient because it does not rely on correspondence updates or closest-point queries, while it is still able to reach the accuracy of SPC methods such as ICP. However, the fast global registration does not provide optimality guarantees and it is prone to fail in the presence of high outlier rates, as shown in (Yang et al., 2021).

Barron (2017) proposes a generalized loss function whose shape can be adjusted by means of a shape parameter and a scale parameter. It enables to represent several well-known single-parameter loss functions and can be utilized for various learning-based and classical vision tasks. Applied to the task of registration and compared to the method from Zhou et al. (2016), the generalized loss replaces the Geman-McClure loss and its convexity is manipulated using the shape parameter, resulting in a moderate improvement in the presence of high noise.

Enqvist et al. (2009) propose an algorithm for 3D point cloud registration that approaches the NP-hard vertex cover problem by combining the guaranteed optimal branch-and-bound and factor-2 approximation. By exploiting pairwise constraints together with graph theory, optimal correspondences are determined. The method finds a solution in a reasonable amount of time while being sufficiently robust to overcome high outlier rates and multiple hypotheses. The combinatorial problem of matching with pairwise constraints is also tackled in (Parra Bustos et al., 2019), where an efficient maximum clique method is proposed to find the pairwise consistent subset of correspondences.

Parra Bustos and Chin (2018) introduce a method called *Guaranteed Outlier Removal (GORE)*, which aims to reduce the usually high number of outliers in a set of 3D keypoint correspondences and significantly facilitate and speed up subsequent optimization steps. By utilizing geometric operations, only guaranteed outliers are removed, leaving the globally optimal solution unaffected.

Yang and Carlone (2019) decouple the estimation of rotation, translation, and scale and solve the resulting subproblems in polynomial time to obtain an overall polynomial time point cloud registration method, which they name *TEASER (Truncated least squares Estimation And SEMidefinite Relaxation)*. Moreover, the approach is highly robust as it utilizes a truncated least squares loss function that allows accurate registration in the presence of outlier rates of up to 99%. An adaptive voting scheme is used to estimate the scale and the translation, however, in a component-wise manner, and a semidefinite program (SDP) relaxation is used to estimate the rotation, which, however, scales poorly and therefore limits the method to small registration problems. In (Yang et al., 2021), advancements of the method are presented, comprising an optimality certification and a more than three orders of magnitude faster and still certifiable algorithm called *TEASER++*, where solving the SDP is replaced by utilizing GNC in the rotation estimation.

Robust feature matching between images is a crucial requirement for tasks such as image stitching, homography estimation, or fundamental matrix estimation in the presence of outliers. RANSAC

is also a common approach in this area, as it usually yields good results in a reasonable amount of time. However, the method is only heuristic and not exact, which is why no globally optimal solution can be guaranteed. This lack of optimality gave rise to many exact methods based on the maximum consensus criterion (Chin and Suter, 2017). However, the ability to provide guarantees is accompanied by high computational cost, as maximum consensus is fundamentally NP-hard, which means that the problem cannot be solved in polynomial time (Chin et al., 2018; Tzoumas et al., 2019). To limit the computational effort, many sophisticated search strategies have been proposed (Chin and Suter, 2017), which comprise branch-and-bound methods, such as (Breuel, 1992; Li, 2009; Hartley and Kahl, 2009; Zheng et al., 2011; Bazin et al., 2013; Campbell et al., 2017), and tree search strategies, such as (Chin et al., 2017; Cai et al., 2019). These techniques determine the globally optimal solution, however, they only generate a significant benefit with respect to computation time if they can prune large portions of the search space, and this is only possible if most solutions are concentrated within a small region or if only a few local maxima exist in general. If not only a very small portion, but large parts or in the worst case the entire search space has to be explored, the approaches still run in exponential time.

Olsson et al. (2008); Enqvist et al. (2012); Ask et al. (2013) propose exact methods for consensus maximization based on enumeration, which essentially equals an exhaustive search. Since the fundamental optimization problem still scales exponentially with the dimension of the parameter space, the enumeration approach becomes intractable for high-dimensional problems and is therefore only suitable for low-dimensional problems.

Non-randomized, deterministic optimization methods such as gradient ascent or descent can be employed if gradients are accessible. Le et al. (2019a) propose two deterministic, heuristic approaches utilizing a non-smooth penalty method and the Alternating Direction Method of Multipliers (ADMM), respectively. By solving convex subproblems, these approaches enhance the optimization performance, especially in the presence of higher outlier rates. Nevertheless, they inherently lack the ability to guarantee globally optimal solutions and are not applicable to non-convex problems.

3.1.2. Simultaneous pose and correspondence (SPC) methods

SPC approaches alternate between the two phases of establishing correspondences and estimating the best transformation by minimizing the errors of these correspondences. They can be categorized in local, global, and deep learning-based methods (Yang et al., 2021).

The most famous local SPC method is the Iterative Closest Point (ICP) algorithm proposed by Besl and McKay (1992), which is a common approach for point cloud registration. It iteratively alternates between finding correspondences and minimizing their residuals, where the residuals can be defined as point-to-point (Besl and McKay, 1992; Vizzo et al., 2023), point-to-plane (Chen and Medioni, 1992; Zhang, 1994), or plane-to-plane (Segal et al., 2009) distances. In its original form, the residuals are minimized using least squares, which causes a strong susceptibility to outliers. To mitigate this sensitivity, many variants of the original ICP have been proposed that employ additional selection, matching, or weighting criteria for establishing correspondences or minimizing errors (Rusinkiewicz and Levoy, 2001; Fitzgibbon, 2003). Nevertheless, even if the optimal solution is certainly obtained in the minimization step, the ICP may still converge to a local optimum in case of incorrect correspondences. Thus, the algorithm relies on a ‘good enough’ initialization to converge towards the global optimum.

To reduce the impact of outliers, various methods utilizing robust loss functions have been proposed: In (Kaneko et al., 2003), a bi-weight function similar to Tukey’s loss (cf. Sec. 2.1.3) is utilized, which fully suppresses the influence of residuals larger than a specific threshold, yielding

a higher robustness towards outlying and non-overlapping data. In (Chetverikov et al., 2005), an ICP variant named Trimmed ICP (TrICP) is proposed, which is based on a least trimmed squares cost function. In (Vizzo et al., 2023), a modified point-to-point ICP is proposed, which represents the core of a LiDAR odometry algorithm. On the one hand, the outlier rejection criterion for correspondence matching is realized using an adaptive distance thresholding, and on the other hand, the robust Geman-McClure loss function (cf. Sec. 2.1.3) is employed in the optimization phase. Chebrolu et al. (2020) propose an enhanced formulation of the generalized robust kernel from Barron (2017), which enables to apply the adaptive kernel also in the presence of strong outliers. This is achieved by truncating the loss function, which allows to adapt the shape parameter over its full range during the optimization. In general, the approach is based on continuously adjusting the loss function to the distribution of the measurements, which avoids sticking to a fixed kernel. The adaptive kernel can be applied to various optimization problems. When integrated into an ICP, the results show better or on-par accuracy compared to fixed kernels, while avoiding the need for manual parameter tuning or additional outlier rejection schemes.

Granger and Pennec (2002) propose an ICP variant, which is able to avoid local minima while still being able to reach an optimal accuracy. Maier-Hein et al. (2012) tackle the assumption that the points are subject to isotropic Gaussian noise by introducing an ICP formulation which is able to account for anisotropic errors. Compared to the original ICP, it shows a higher robustness towards initial misalignment and a higher accuracy, especially for partial surface registration.

Myronenko and Song (2010) and Jian and Vemuri (2011) introduce probabilistic methods for both rigid and nonrigid point cloud registration, which provide robust and accurate results in the presence of noise and outliers. In (Myronenko and Song, 2010), one point cloud is represented as Gaussian Mixture Models (GMMs) and fitted to the other point cloud by maximizing the likelihood, and in (Jian and Vemuri, 2011), both point clouds are represented as GMMs and the point cloud registration problem is solved by aligning the two mixture distributions through minimization of the L_2 norm between both. Le et al. (2019b) propose a randomized approach with a new sampling strategy based on a graph matching formulation, which yields high quality hypotheses. These enable an efficient application to point cloud registration, where conventional randomized methods only obtain satisfactory results after evaluating a huge number of candidate models.

Similar to the presented local SPC methods, global approaches solve the registration problem without known point correspondences. However, in contrast to the local methods, they are able to provide globally optimal solutions and avoid to get stuck in a local optimum. Hence, they do not require any initial guesses. The main drawback is still the required computational cost. Many approaches are based on branch-and-bound strategies, whose runtime grows exponentially with the size of the point cloud (Parra Bustos and Chin, 2018). Various sophisticated approaches have been proposed to tighten the bounding and reduce the computational cost. Still, they only generate results in a reasonable time if large parts of the search space can be pruned, which is not the case in the presence of many local minima arising from high outlier ratios. In the following, various global SPC methods are presented.

Li and Hartley (2007) combine a Lipschitz error function with a branch-and-bound search over an octree data structure to provide globally optimal registration results without any initialization. However, the method relies on the assumption that there are no outliers or that those are removed beforehand.

Yang et al. (2016) propose a 3D point cloud registration approach named *Globally optimal ICP* (*Go-ICP*) that combines the local ICP with a global branch-and-bound scheme. It determines the

optimal solution under the L_2 -norm. To handle outliers, the approach can be extended using the LTS-based trimming strategy from Chetverikov et al. (2005), which leads to the *trimmed Go-ICP*.

Parra Bustos et al. (2014) introduce a novel branch-and-bound bounding function with respect to the robust maximum consensus criterion for fast global rotation search. It utilizes stereographic projections that enable efficient evaluation and an algorithm that is an order of magnitude faster than the matchlist-based branch-and-bound method from (Breuel, 2003).

Chin et al. (2014) introduce a fast rotation search based on branch-and-bound optimization that provides globally optimal results in real time. However, the algorithm only estimates one angular parameter. The method is embedded into a user-assisted point cloud registration tool.

Izatt et al. (2020) formulate the point cloud registration problem as mixed-integer convex program and solve it using a branch-and-bound solver. The method provides globally optimal results, however, it is only applicable to small point sets.

To complete the SPC methods, various deep learning-based methods are shortly presented. They have been proposed quite recently and open up a new direction of research. Aoki et al. (2019) combine the deep neural network *PointNet* (Qi et al., 2016), which has originally been proposed for 3D recognition tasks directly based on point clouds, with the Lucas & Kanade (LK) algorithm into a single deep neural network named *PointNetLK* for the task of point cloud registration. Wang and Solomon (2019a) introduce *Deep Closest Point (DCP)* with the goal to address difficulties of the ICP such as local optima, and they show that it is able to outperform the ICP, its variants such as (Yang et al., 2016), and *PointNetLK*. Wang and Solomon (2019b) introduce *Partial Registration Network (PRNet)*, which is well-suited for the registration of partially overlapping point clouds and outperforms *PointNetLK*, *DCP*, and conventional approaches. Avetisyan et al. (2019) propose a network to align clean and complete CAD models with noisy and incomplete 3D scans. Gojcic et al. (2019) introduce *3DSmoothNet*, which utilizes a siamese architecture to register 3D point clouds. Choy et al. (2020) present a framework named *Deep Global Registration* that uses Fully Convolutional Geometric Features (FCGF) (Choy et al., 2019) for robust and accurate registration of real-world 3D scans.

3.2. LiDAR-based localization

In this section, various related LiDAR-based localization and SLAM methods are presented. There are different ways to categorize the large number of existing approaches. Often, however, combinations these categories exist and their boundaries are blurred, resulting in an ambiguity with respect to a strict categorization. Nevertheless, to provide a structured overview, it is tried to sort the existing methods as best as possible. In the first place, it is differentiated between conventional and deep learning-based approaches, and in the second step, it is distinguished between localization approaches based on a pre-known map and LiDAR odometry or SLAM approaches without a prior map. Finally, the localization techniques with a prior map are categorized into local map matching and global place recognition methods. Additionally, one could distinguish between single pose estimation and filter techniques, or between feature-based and registration-based methods (Elhousni and Huang, 2020). However, to not over-complicate the presentation, these aspects will only be mentioned when the methods are briefly introduced below.

Conventional localization methods with a prior map:

In **map matching or pose tracking**, the task is to determine the vehicle pose from a known initial pose, which can be provided by a filter prediction or GNSS solution. Fox et al. (1999) and Dellaert et al. (1999) proposed the Monte-Carlo localization (MCL), which utilizes a particle filter, a framework that is still popular today. The robot’s belief is represented by a set of randomly drawn

samples, the so-called particles. In this way, arbitrary probability densities can be approximated, which makes it possible to deal with ambiguities. By initializing particles over the full area of operation, it can also be used for global localization, which, however, is accompanied by the corresponding computational costs. In (Levinson et al., 2007), first, a 2D road map containing ground reflectivity is generated using a graph SLAM approach that integrates GPS, IMU, wheel odometry, and LiDAR data, and second, a vehicle is localized within this map utilizing a particle filter and LiDAR measurements.

In (Wolcott and Eustice, 2014), to determine the 3D vehicle pose consisting of x, y, θ , the normalized mutual information between a measured camera image and a set of synthetic camera images rendered for positions in a regular grid with different heading angles is maximized using an exhaustive search. In (Wolcott and Eustice, 2015), a Gaussian mixture map, which is a 2D grid of one-dimensional GMMs modelling the distribution over the cell's z -height, is used together with a branch-and-bound method to find the guaranteed optimal localization solution in pre-defined search space.

Schlichting and Brenner (2014) detect poles and planes in LiDAR scans and define local feature patterns describing the spatial relationships. By matching these patterns to corresponding patterns in a reference map, the rate of wrong matches can be significantly reduced compared to a simple nearest neighbor feature assignment, especially in case of low initial accuracy. Steinke et al. (2021) follow a similar geometric fingerprinting approach, however, they generate a feature map from openly available data to avoid mapping the environment using LiDAR in the first place. Wei et al. (2020) and Zhang et al. (2022) propose localization approaches utilizing curb and intensity features.

Merfels and Stachniss (2016) register LiDAR scans to a globally georeferenced map point cloud and visual features to a globally georeferenced feature map. The obtained pose estimates are fused with information from GPS and wheel odometry in a sliding window pose graph. Wilbers et al. (2019) detect pole-like objects in LiDAR scans and match them to an existing map. The estimated pose information are combined with odometry and GNSS information in a sliding window factor graph.

In **global localization**, no initial pose is available and the task is to localize the vehicle within the full extent of the map. Chen et al. (2021) propose a Monte Carlo localization, where in the first phase particles are initialized over the entire map to determine the global pose, and in the second phase, after it is found, a classical particle filter is used for pose tracking. In both phases, with each new LiDAR scan, for every particle a synthetic LiDAR range image is rendered from a mesh-based map and every particle's resampling weight is calculated based on the mean absolute pixel-wise difference between its synthetic range image and the measured range image.

Brenner (2009) investigates the applicability of local pole patterns for global localization and proposes a descriptor for that purpose, which is the groundwork for (Schlichting and Brenner, 2014).

Lu et al. (2022) introduce a translation invariant descriptor named *Radon sinogram (RING)* and an orientation invariant similarity metric for global localization. The RING-based scan representation is derived by projecting the LiDAR scan points into Bird-Eye-View (BEV) and applying the radon transform.

Kim et al. (2021) present a descriptor for global localization named *Scan Context Descriptor (SCD)*, which encodes the structural appearance of a place by partitioning the LiDAR scan points into BEV bins using both a Cartesian and a polar grid and assigning the maximum height value to each bin.

Cop et al. (2018) propose a descriptor for global localization named *DELIGHT*, which encodes LiDAR intensities of the surroundings as a group of histograms. For the localization, first, the histograms of intensity are compared using chi-squared tests to find the position with the highest similarity, and second, a geometry-based verification is performed.

Conventional SLAM and LiDAR odometry methods:

SLAM and LiDAR odometry methods do not rely on a pre-known map. Instead, they generate a map simultaneously to the localization, where different types of map representations are used, such as voxel grids, NDT, surfel maps, and triangular meshes. Behley and Stachniss (2018) propose a SLAM approach based on a surfel map. The pose estimation relies on the projective association between the measured LiDAR scan and the surfel map, where the scan is represented as 2D scan image and a synthetic scan image is rendered from the map. The rendering of virtual views is not only used for the pose estimation, but also for the detection and verification of loop closures.

Chen et al. (2019) present an extension of the surfel-based SLAM approach from Behley and Stachniss (2018) that integrates semantic information obtained from a convolutional neural network. Based on point-wise labeled scans, a surfel map can be build that is semantically-enriched with labels. It allows to define semantic constraints in the projective data association, leading to an improved pose estimation accuracy in challenging situations with few static and many dynamic objects.

Nüchter et al. (2007) propose *6D SLAM* which considers both the three translational and the three rotational DoF for 3D mapping using 3D LiDAR scans. The method is based on an ICP to match 3D point clouds, where an octree-based coarse-to-fine search strategy is used to determine an initial guess for the registration of new scans and to detect loop closures. This heuristic works especially in diverse outdoor environments with few symmetries, since these result in a clear maximum of matched octree cubes. In the event of loop closure, a global relaxation is applied to minimize the overall registration error and thereby refine the generated model.

Zhang and Singh (2017) propose a SLAM method that performs the motion estimation from LiDAR odometry at an order of magnitude higher frequency than the LiDAR mapping. The LiDAR odometry is based on edge and planar feature points. To compensate for wrong associations and to obtain a robust behavior, Tukeys biweight function is utilized in the motion estimation.

Wen et al. (2018) investigate the performance of an NDT-based graph SLAM using LiDAR data and ascertain that it heavily depends on the traffic conditions. In dense traffic scenarios, and especially when large vehicles such as double-decker buses and trucks are present, the accuracy of the SLAM is reduced because these vehicles cause large occlusions of the LiDAR's FoV.

Vizzo et al. (2021) propose a SLAM approach that utilizes a triangular mesh as the map. The mesh is generated using Poisson surface reconstruction, which is able to represent more geometric details compared to other mapping methods based on TSDF or surfels.

Dellenbach et al. (2022) introduce a LiDAR-only odometry method named *Continuous-Time ICP (CT-ICP)*, which is implemented in a SLAM that incorporates loop closure constraints and uses a standard pose graph for optimization. The loops are detected by matching features computed in elevation images. The core contribution is that soft constraints are added in the pose estimation that allow for discontinuity between consecutive scans, yielding the ability to correct for errors. The proposed intra-scan continuity of poses is achieved by linear interpolation of both translation and rotation, and for the sake of robustness, a robust loss function is utilized in the ICP formulation.

Vizzo et al. (2023) introduce a LiDAR odometry method named *KISS-ICP*, which is based on a point-to-point ICP and a voxel map that simply stores the first N_{max} points in each voxel, with N_{max} being a fixed, user-defined number. To obtain a robust behavior, an outlier rejection based on an adaptive distance thresholding is used in the correspondence matching, and a robust Geman-McClure loss function is used in the optimization phase.

Deep learning-based methods:

The deep learning-based methods are categorized similar to the conventional approaches in local and global localization methods with a prior map, and LiDAR odometry methods without a prior map. Local localization methods with a prior map are proposed in (Bârsan et al., 2018; Lu et al., 2019; Ma et al., 2019). Bârsan et al. (2018) propose a deep localization network that determines a 3-DoF pose based on an intensity map and intensity information obtained from LiDAR scans. Lu et al. (2019) present a network that uses local feature descriptors generated by PointNet to determine the optimal pose in a bounded and discretized 3-DoF solution space. Ma et al. (2019) use networks for lane and traffic sign detection and exploit these information to localize with respect to a lightweight semantic map that only contains a lane graph and traffic sign locations.

Global localization methods with a prior map are presented in (Elbaz et al., 2017; Kim et al., 2019; Fallon et al., 2019; Yin et al., 2020; Sun et al., 2020). Elbaz et al. (2017) use a deep neural network auto-encoder to describe the local 3D geometric structure of a point in a low-dimensional descriptor instead of utilizing manually designed key-point descriptors. Based on these descriptors and a subset of points, a coarse registration is carried out using RANSAC, with a subsequent fine registration using ICP. Kim et al. (2019) formulate the place recognition problem as place classification problem and propose an end-to-end global localization method. A convolutional neural network learns a descriptor called Scan Context Image (SCI), which captures the structural information of a place in an image. SCI is based on scan context, a descriptor proposed in (Kim and Kim, 2018), which is also the groundwork of (Kim et al., 2021) described above. Although the data for training the network was collected on a single day, the localization performance is stable for over one year. Fallon et al. (2019) present a place recognition method that uses a novel learned descriptor based on segmented LiDAR scans. The approach does not require extensive preprocessing of the point clouds and is applicable to both urban and natural environments. The architecture is efficient and suitable for real-time applications, and the small model can run on a CPU, while loop closures are detected at around 1 Hz. Yin et al. (2020) use a siamese network that learns a semi-handcrafted representation of LiDAR point clouds to achieve global place recognition. Sun et al. (2020) combine a learning-based localization and a filtering-based localization. The learning-based module provides an optimized distribution for initializing the filtering-based module, which is a particle filter.

LiDAR odometry methods without a prior map are introduced in (Tang et al., 2018; Li et al., 2019; Wang et al., 2019; Cho et al., 2020). Tang et al. (2018) propose to train a regression model that predicts the bias between the solution of a classical state estimator and the ground truth, which is then applied as a correction. Li et al. (2019) present an end-to-end deep convolutional network, where an additional map is built that is used to refine the estimation result of each scan. Wang et al. (2019) present an end-to-end deep neural network that takes two consecutive LiDAR scans represented as depth images as input to estimate the LiDAR odometry. The network consists of two parallel sub-networks. One estimates the translation and the other the orientation, so that a 6-DoF pose is obtained. Cho et al. (2020) propose an unsupervised loss that captures the geometric consistency of point clouds and enables to train the presented LiDAR odometry network in an unsupervised manner.

3.3. Discussion

In this section, the requirements for the developed localization method are discussed along with the limitations of the presented related work. This reasoning leads to the research gap that is to be addressed in this thesis. Fundamentally, the focus of this work is on robust LiDAR-based localization, which means that the approach must be able to correctly estimate the position even under unfavorable conditions, such as dense traffic with very few static structure.

- The first requirement for the developed method is that it needs to utilize the measured points and not rely on features. If only a small number of correct correspondences, i.e. inliers, exist in a critical situation, the number of correct feature correspondences will be even smaller and eventually not sufficient anymore. Consequently, a correspondence-based registration method that relies on features cannot be the means of choice.
- The developed approach is not allowed to get stuck in local minima, thus it needs to be robust towards poor initialization. This condition excludes a local SPC strategy, which shows exactly this shortcoming.
- Another requirement is that the estimation needs to be highly robust towards outliers. In this regard, the maximum consensus criterion has shown promising results, however, it has not yet been investigated for the purpose of vehicle localization. Thus, the application of the maximum consensus criterion to LiDAR-based localization is determined as the first auspicious direction of research.
- Since autonomous driving is a safety-relevant task, the method must guarantee to find the globally optimal solution. This requirement excludes an randomized maximum consensus method such as RANSAC, which cannot provide these guarantees. Furthermore, in the presence of high outlier rates, randomized methods lose their sole advantage of faster execution. Hence, a global approach is to be pursued.
- The global SPC methods presented in most cases include some sophisticated branch-and-bound strategy. If a unique maximum exists, which can happen frequently in LiDAR-based localization depending on the scenario, such a strategy can provide acceleration. However, for a low-dimensional problem and a bounded parameter space, the implementation of an exhaustive search strategy with exponential complexity in the number of parameters is absolutely affordable. Moreover, an exhaustive approach has the benefit that it reveals knowledge about the entire search space, since each candidate pose is evaluated with respect to its consensus. Therefore, an exhaustive search is pursued instead of a branch-and-bound strategy. This immediately opens up the next direction of research, which is to investigate how the gained knowledge can be used to assess localization integrity.
- Some of the presented localization and SLAM approaches (Chen et al., 2021; Behley and Stachniss, 2018) render synthetic LiDAR range images to make use of projective data association. However, these approaches do not apply the robust maximum consensus criterion, but instead use an ICP scheme or weight particles based on the mean absolute deviation. Since the rendering of synthetic views and the projective data association can be performed both rapidly and for many candidate poses in parallel on a GPU, it appears to be a suitable approach to render candidate views and then evaluate them using the maximum consensus criterion. In this way, k d-tree queries for nearest neighbor data associations can be avoided.
- Many localization approaches rely on an extended Kalman filter or on a particle filter. Whereas the Kalman filter is limited to Gaussian distributions, the particle filter is prone to particle deprivation. Hence, both are not suitable for representing arbitrary distributions in a highly robust and reliable manner. This can be achieved by employing a histogram filter,

which is able to model arbitrary distributions while covering the entire space of potential solutions. To obtain more robust localization results than from a single pose estimation, another direction of research is to filter search spaces that have been evaluated with respect to the maximum consensus criterion or some other robust loss function, utilizing truncated probability distributions as introduced in (Barron, 2017; Chebrolu et al., 2020).

- Deep learning-based methods lack transparency and it is difficult to fully comprehend their behavior. This contradicts the high integrity idea, which is why deep learning-based approaches are not taken into consideration and instead conventional methods are pursued to tackle the problem.

To summarize, a generic method for the purpose of robust vehicle localization is to be developed that is based on the maximum consensus criterion and directly utilizes the measured points. It follows an exhaustive strategy and guarantees to provide globally optimal results, hence it cannot get stuck in local minima. The approach provides full knowledge of the search space, which is to be used in a histogram filter and in an integrity evaluation. To obtain comprehensibility and trustworthiness, a conventional method is to be pursued, and to obtain scalability and real-time capability, the utilization of projective data association is to be taken into consideration.

4. Methodological contributions

In this chapter, the methodological contributions of this work are presented. The sections of this chapter correspond to the structure of the robust and reliable localization framework, shown in Fig. 4.1. First, in Sec. 4.1, the basic concept and fundamental design criteria of maximum consensus localization are introduced, which also comprises the distinction between the *Euclidean space* and the so-called *observation space*, including their respective objective functions of the basic maximum consensus inlier count. Initial ideas for maximum consensus localization based on correspondences in *Euclidean space* and *observation space* were already published in (Axmann and Brenner, 2021) and (Axmann and Brenner, 2023), respectively. Afterwards, in Sec. 4.2, the optimization based on the *point-to-plane adjustment score*, a novel robust objective function based on Helmert’s point error, is proposed. The first concept of this approach was published in (Axmann et al., 2023b). In Sec. 4.3, the optimization based on various well-known loss functions is presented, and in Sec. 4.4, the extension of the single pose estimation into the maximum consensus filter is introduced. Finally, in Sec. 4.5, the protection level estimation and integrity evaluation are proposed, where first ideas were also published in (Axmann and Brenner, 2023).

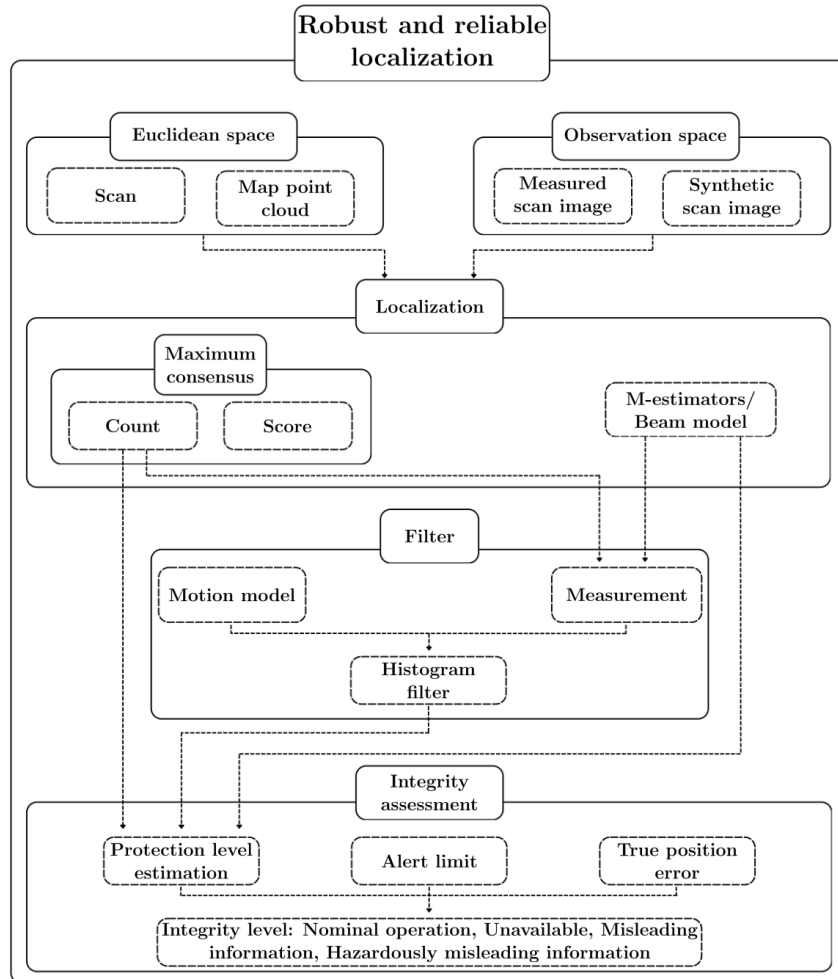


Figure 4.1.: Overview robust and reliable localization.

4.1. Methodology of maximum consensus localization

In the first part of this section, the basic concept and design criteria that apply to both approaches are explained. In the second part, the maximum consensus localization approach based on correspondences in the observation space is proposed, and in the third part, the maximum consensus localization approach based on correspondences in the Euclidean space is introduced.

4.1.1. Basic concept

In the maximum consensus localization, a scan from a low-cost automotive LiDAR, henceforth referred to as *car sensor scan* \mathcal{S} , is registered to a map \mathcal{M} , and thereby, the globally optimal vehicle pose \mathbf{x}^* is found in a bounded pre-defined search space. In the first place, the car sensor scan exists in the LiDAR coordinate system L (cf. Fig. 2.7), thus it is denoted as $^L\mathcal{S}$, and both the map and the optimal vehicle pose exist in the world frame W , thus they are denoted as $^W\mathcal{M}$ and $^W\mathbf{x}^*$, respectively. In the context of this work, the map is globally georeferenced with respect to a UTM coordinate system, where further information are provided in Sec. 5.2. However, in general, any reference system is applicable as world frame.

Maximum consensus localization is a local method, thus it relies on an initial pose $^W\mathbf{x}_0$, which can be provided by any source, for example by means of GNSS. Around the initial pose, a search space is defined that comprises three DoF, namely the 2D translation in the xy -plane \mathbf{t}_{xy} and the heading angle θ . It is presumed that the car is driving in a normal mode of operation on the road surface and that the variations in height as well as roll and pitch are limited and negligible. Accordingly, the goal of maximum consensus localization is to find the optimal three DoF rigid transformation $^C\mathbf{T}_V^*$ from the initial pose $^W\mathbf{x}_0$ to the optimal pose $^W\mathbf{x}^*$, which is defined as

$$\begin{aligned} ^C\mathbf{T}_V^* &= \mathbf{T}(t_x^*, t_y^*, 0, 0, 0, \theta^*) \\ &= \begin{pmatrix} \mathbf{R}_\theta^* & \mathbf{t}^* \\ \mathbf{0} & 1 \end{pmatrix}, \end{aligned} \quad (4.1)$$

where C denotes a candidate pose and V the vehicle coordinate system, $\mathbf{R}_\theta^* \in SO(3)$ is the rotation matrix of the optimal heading angle θ^* , and $\mathbf{t}^* \in \mathbb{R}^3$ is the translation vector of the optimal shift in the xy -plane, i.e. $\mathbf{t}^* = (\mathbf{t}_{xy}^*, 0)^\top$ with $\mathbf{t}_{xy}^* \in \mathbb{R}^2$. In the following, the notation for the transformation matrix of a candidate pose $^C\mathbf{T}_V = \mathbf{T}(t_x, t_y, 0, 0, 0, \theta)$ is abbreviated as $^C\mathbf{T}_V = \mathbf{T}(t_x, t_y, \theta)$. The rotation matrix $\mathbf{R}_\theta \in SO(3)$ is defined as

$$\mathbf{R}_\theta = \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (4.2)$$

The transformation matrix based on the initial pose $^W\mathbf{x}_0 = (x_0, y_0, z_0, \omega_0, \phi_0, \kappa_0)$, which provides the transformation from a candidate pose C into the world frame W , is defined as

$$\begin{aligned} ^W\mathbf{T}_C &= \mathbf{T}(x_0, y_0, z_0, \omega_0, \phi_0, \kappa_0) \\ &= \begin{pmatrix} & x_0 \\ \mathbf{R}_{\omega_0, \phi_0, \kappa_0} & y_0 \\ & z_0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \end{aligned} \quad (4.3)$$

where $\mathbf{R}_{\omega_0, \phi_0, \kappa_0}$ is defined as

$$\mathbf{R}_{\omega_0, \phi_0, \kappa_0} = R_{\kappa_0} R_{\phi_0} R_{\omega_0}, \quad (4.4)$$

with

$$R_{\omega_0} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \omega_0 & -\sin \omega_0 \\ 0 & \sin \omega_0 & \cos \omega_0 \end{pmatrix}, R_{\phi_0} = \begin{pmatrix} \cos \phi_0 & 0 & \sin \phi_0 & 0 \\ 0 & 1 & 0 & 0 \\ -\sin \phi_0 & 0 & \cos \phi_0 & 0 \end{pmatrix}, R_{\kappa_0} = \begin{pmatrix} \cos \kappa_0 & -\sin \kappa_0 & 0 \\ \sin \kappa_0 & \cos \kappa_0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

When transforming the car sensor scan into the world frame, the horizontal distortion, which actually exists in the UTM projection and for the location of the experiments used in this work is 2 cm at a distance of 50 m, is neglected. In theory, the size of the search space could span the entire map leading to a global localization approach that does not require any initialization. However, this is accompanied by large computational cost and does not represent a meaningful application. In this work, the case where an initialization is available is considered, and it is assumed that the optimal pose ${}^W\mathbf{x}^*$ is within a certain range around the initial pose ${}^W\mathbf{x}_0$. The size of the search space is selected according to these assumptions, where it is a fundamental condition of this localization approach to be successful that the true pose is certainly contained within the search space. If the true pose is not within the search space, it cannot be found by the method and the localization will inevitably fail. In practice, the size of the search space can be chosen according to uncertainty information of the initial pose, if those are available.

The task is to determine the pose within the search space, which maximizes an objective function with respect to all available measurements, for example the count of inliers. To this end, the search space is discretized and represented as a stack of regular grids, where each 2D grid corresponds to the candidate translations \mathbf{t}_{xy} for a candidate heading angle θ . The selection of the discretization is a trade-off between accuracy and computational cost. It can be chosen, for example, according to the measurement uncertainty of the automotive LiDAR or with respect to the accuracy requirements for autonomous driving. In this work, the discretization step size is chosen as 0.1 m and 0.17° for the translation and heading, respectively, according to the accuracy requirement for passenger vehicles on local roads (Reid et al., 2019). However, in one of the proposed maximum consensus localization approaches, the discretization step size of the heading angle is depending on the horizontal resolution $\Delta\alpha$ of the LiDAR sensor that is used for the localization. Therefore, the heading step size cannot be set exactly to 0.17° , and it is slightly larger instead, with more explanations provided in the following subsection. Additionally, a refinement based on the inliers can be conducted to achieve sub-cell accuracy.

Each combination of \mathbf{t}_{xy} and θ yields a candidate transformation ${}^C\mathbf{T}_V = \mathbf{T}(t_x, t_y, \theta)$ similar to Eq. 4.1, which is to be applied to the body frame of the vehicle. In this work, the RIEGL coordinate system serves as the body frame of the vehicle with respect to which the local corrections of the pose are determined. More information on the coordinate systems of the sensor setup are provided in Sec. 5.2. When a candidate transformation ${}^C\mathbf{T}_V$ is transformed into the world frame using the transformation ${}^W\mathbf{T}_C$ based on the initial pose ${}^W\mathbf{x}_0$, the corresponding candidate pose ${}^W\mathbf{x}$ is obtained. The relation is defined as follows:

$${}^W\mathbf{T}_V = {}^W\mathbf{T}_C {}^C\mathbf{T}_V. \quad (4.5)$$

The localization is carried out with respect to the highly precise prior map ${}^W\mathcal{M}$, which is acquired in a separate mapping campaign using a RIEGL Mobile Mapping System (MMS) (RIEGL Laser Measurement Systems GmbH, 2012). Hence, the map data initially exists as dense 3D LiDAR point cloud, which is denoted as ${}^W\mathbf{M} = {}^W\{\mathbf{m}_j\}_{j=1}^m$, with m points $\mathbf{m}_j \in \mathbb{R}^3$. However, depending on the maximum consensus approach, the map point cloud may also be converted into a triangular mesh ${}^W\mathcal{M}_{mesh}$. In the map point cloud, points from the ground, vegetation, and parked cars are removed as they either have no benefit for the pose estimation in the xy -plane, which is the case

for the ground, or even cause false correspondences and thus have a negative influence on the pose estimation, which is the case for the vegetation and parked cars. The removal is done using point-wise class labels obtained from a convolutional neural network based on KPConv (Thomas et al., 2019). Consequently, scan points acquired by the LiDAR of the vehicle that is to be localized, which are from the ground, vegetation and parked cars, cannot find a match in the map, do not contribute to the consensus, and are considered as outliers.

Each candidate transformation ${}^C\mathbf{T}_V$ of the discretized search space is evaluated with respect to the maximum consensus criterion, i.e. the inlier count, and a novel objective function based on Helmert’s point error, which is described in Sec. 4.2. The evaluation of all candidate transformations happens in an exhaustive manner, which is computationally affordable as the localization problem is limited to three DoF and the extent of the search space is bounded. Besides its guaranteed global optimality, the exhaustive approach has the advantage that it reveals the distribution of the objective function over the entire parameter space, which can be utilized in the integrity evaluation.

To establish correspondences between the car sensor scan and the map, two strategies are pursued. The first one is based on projective data associations, which is referred to as *maximum consensus in the observation space*, and the second one is based on point-to-point associations, which is referred to as *maximum consensus in the Euclidean space*. In the following two subsections, both approaches and their prerequisites are explained, and the objective functions of the basic maximum consensus inlier count are defined.

Finally, in Fig. 4.2, the basic concept of maximum consensus vehicle localization is illustrated when it is applied in practice. Each grid on the ground represents a fully evaluated search space of one epoch. However, from the entire stack of grids, only the 2D grid that corresponds to the heading angle yielding the highest consensus is shown. Furthermore, for the sake of better visibility, only the 2D grid of every third epoch is visualized. The consensus values are colored using temperature scale, from blue to red. Thus, the red area of the consensus distribution depicts the position of the highest consensus, which together with the heading represents the localization solution. The georeferenced map point cloud is shown in green, the 360° scan from the LiDAR mounted on the ego-vehicle is shown in red, and the inlier belonging to the pose with the highest consensus are shown in yellow.

4.1.2. Maximum consensus in the observation space

Maximum consensus localization in the observation space is based on projective data associations between the car sensor scan ${}^L\mathcal{S}$ and the map ${}^W\mathcal{M}$. To this end, the measured scan is represented as range image $\mathbf{I}_{R,s}$ and a synthetic range image $\mathbf{I}_{R,x}$ is generated for each candidate position within the xy -grid. The rendering of synthetic views requires knowledge about the LiDAR model and a triangular mesh as map representation. The concept of maximum consensus localization using projective associations is visualized in Fig. 4.3. In the following, the required prerequisites and depicted steps are introduced and the objective function of the maximum consensus inlier count is proposed.

In this work, the measured range images $\mathbf{I}_{R,s}$ are not generated directly from the raw measurements. The points are first transformed from the LiDAR coordinate system into the world coordinate system, using the exact vehicle pose at the time of measurement for each point. The point cloud in the world coordinate system is then projected back into the 2D image based on the vehicle pose when half the measurement has been taken, with the distance between the point and the LiDAR position being calculated. In this way, the ego motion of the vehicle during the 360° LiDAR sweep is taken into account and systematic differences between the measured ranges and

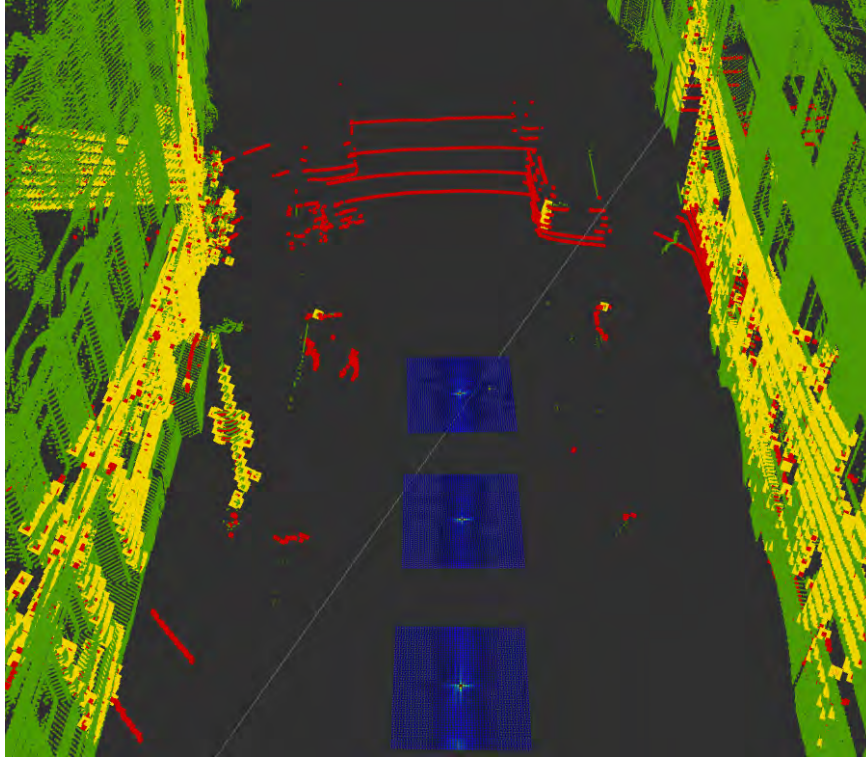


Figure 4.2.: Basic concept of maximum consensus vehicle localization when applied in practice: In every epoch, around an initial pose, a search space is defined, discretized, and evaluated with respect to an objective function. Each grid on the ground represents a fully evaluated search space of one epoch. However, from the entire stack of grids, only the 2D grid that corresponds to the heading angle yielding the highest consensus is shown. Furthermore, for the sake of better visibility, only the 2D grid of every third epoch is visualized. The consensus values are colored using temperature scale, from blue to red. Thus, the red area of the consensus distribution depicts the position of the highest consensus, which together with the heading represents the localization solution. The georeferenced map point cloud is shown in green, the 360° scan from the LiDAR mounted on the ego-vehicle is shown in red, and the inlier belonging to the pose with the highest consensus are shown in yellow.

the synthetic ranges are avoided. In the literature, this is often referred to as *scan deskewing*, for which a predicted trajectory is usually used. If the rendering of the synthetic ranges was not based on a single point of view, but also based on many points along a predicted trajectory, the measured range images $I_{R,s}$ could be generated directly from the raw measurements without introducing a systematic error.

In the context of this work, a Velodyne VLP-16 and a Hesai PandarXT-32 are used, which are 360° mechanically rotating, automotive LiDAR sensors with a horizontal resolution of $\Delta\alpha = 0.2^\circ$ and $\Delta\alpha = 0.18^\circ$ and $n_l = 16$ and $n_l = 32$ scan layers, respectively (Velodyne LiDAR, Inc., 2016; Hesai Technology Co., Ltd., 2021). In Fig. 4.5, the illustration at the top shows a measured range image from a Velodyne VLP-16, which acquires 1800 range measurements for each of its 16 layers in one 360° sweep, yielding an image of 16×1800 pixels. Similarly, the dimension of a range image from a Hesai PandarXT-32 is 32×2000 pixels.

As described in Sec. 2.3, LiDAR sensors are available with different operating principles, such as 360° rotating and solid-state, and with different FoVs and resolutions along both the horizontal and vertical axes. While the operating principle is irrelevant, both the FoV and the resolutions must be known for the rendering of synthetic scan images. For the Velodyne VLP-16 and the Hesai PandarXT-32, the relevant technical specifications are provided in Tab. 4.1.

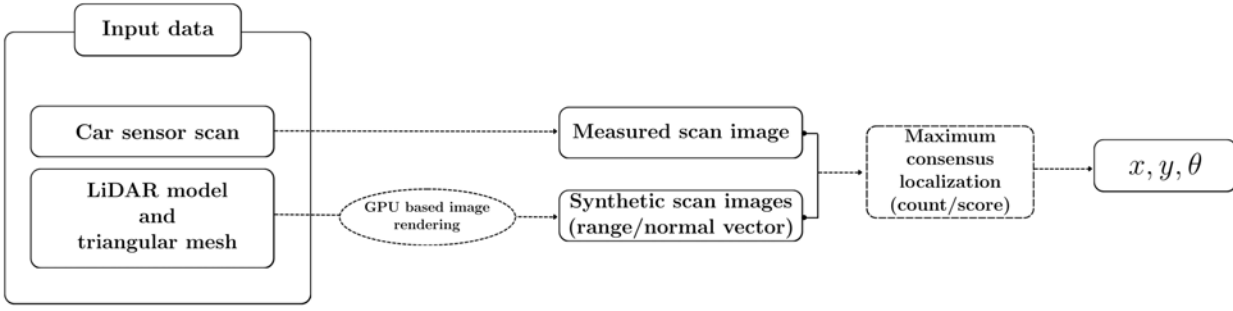


Figure 4.3.: Concept of the maximum consensus localization using projective associations.

Table 4.1.: Relevant technical LiDAR specifications for the rendering of synthetic scan images (V.: vertical; H.: horizontal).

LiDAR	Layer	FoV (V.)	FoV (H.)	Res. (V.)	Res. (H.)
VLP-16	16	30° (-15° to +15°)	360°	2°	0.2°
PandarXT-32	32	31° (-16° to +15°)	360°	1°	0.18°

To generate synthetic scan images, the rays of the LiDAR have to be traced and intersected with the map. In the context of this work, for this task triangular meshes are utilized as map representation, leveraging their capacity for rapid image rendering on GPUs. To generate a triangular mesh $^W\mathcal{M}_{mesh}$ from the initially existing map point cloud $^W\mathbf{M}$, the VDBFusion library (Vizzo et al., 2022) is used. Due to its ability to construct triangular meshes without assumptions about the size of the represented area, it is well suited for the large point clouds acquired by the RIEGL MMS. The VDBFusion library combines the TSDF (cf. Sec. 2.4.5) and the VDB data structure (Museth et al., 2013), which yields a fast and practical meshing method that is flexible regarding the size and type of the input point cloud. The adjustable parameters are the voxel size and truncation distance, which are chosen as 0.1 m and 0.25 m, respectively. As all points from the ground, vegetation and parked cars are removed from the 3D map point cloud, such parts do also not exist in the generated mesh. In Fig. 4.4, the meshes from two different places are shown. As can be seen, structures such as facades, street signs and tree trunks are represented in the mesh, whereas the ground, parked cars and vegetation, e.g. tree crowns, are not.

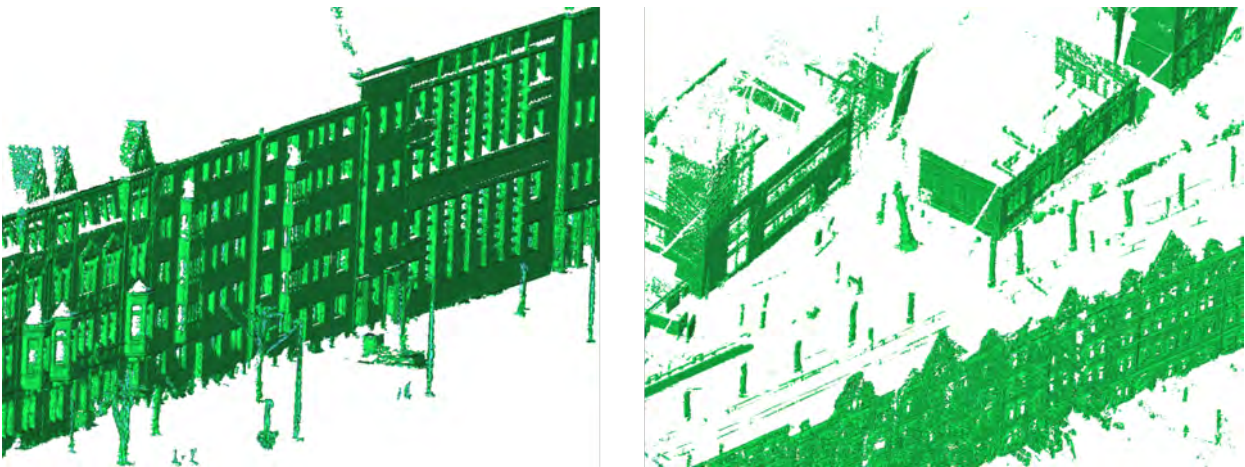


Figure 4.4.: Example meshes from two different places: Structures such as facades, street signs and tree trunks are represented in the mesh, whereas the ground, parked cars and vegetation, e.g. tree crowns, are not.

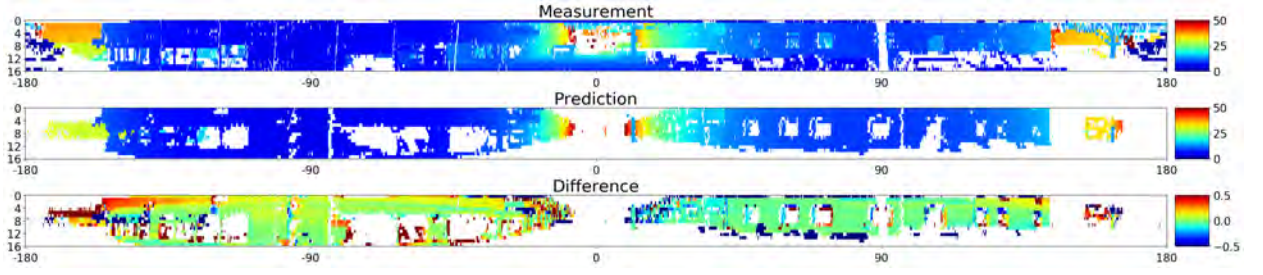


Figure 4.5.: Measured range image (top), synthetic range image (center), and image of range differences between rendered and measured ranges (bottom). The synthetic range image is generated using a triangular mesh. As the ground, vegetation, and parked cars are not represented in the mesh, such parts do also not exist in the synthetic range image, which partly causes areas of large deviations in the difference image.

Similar to Chen et al. (2021), the synthetic scan images are rendered using OpenGL. Besides the LiDAR specifications and the triangular mesh, the 6-DoF LiDAR pose $\mathbf{x}_L = (x_L, y_L, z_L, \omega_L, \phi_L, \kappa_L)^T$ is required for the rendering. For an arbitrary vehicle pose ${}^W\mathbf{x}$ in the world frame, with the corresponding transformation matrix ${}^W\mathbf{T}_V$, the LiDAR pose ${}^W\mathbf{x}_L$ is calculated using the relation ${}^W\mathbf{T}_L = {}^W\mathbf{T}_V {}^V\mathbf{T}_L^{-1}$, where ${}^V\mathbf{T}_L$ denotes the transformation from the LiDAR frame into the body frame of the vehicle. Given both the LiDAR pose and the triangular mesh in the world frame, the rendering yields the synthetic range image $\mathbf{I}_{R,x}$ in the LiDAR frame L . This enables immediate comparison to the measured range image $\mathbf{I}_{R,s}$, which also exists in the LiDAR frame L . In Fig. 4.5, below the measured range image, the synthetic range image is visualized, which has been rendered for the same LiDAR pose. As the ground, vegetation, and parked cars are not represented in the mesh, such parts do also not exist in the synthetic range image.

By subtracting the ranges of the synthetic range image $\mathbf{I}_{R,x}$ from the ranges of the measured range image $\mathbf{I}_{R,s}$, the projective data associations can be utilized for maximum consensus localization. The pixel-wise subtractions yield an image of equal size that consists of the differences, which is denoted as \mathbf{D}_x . In Fig. 4.5, the bottom image illustrates the pixel-wise range differences of the synthetic and measured range images shown above. On the differences of the image \mathbf{D}_x , the maximum consensus criterion can be applied, which yields a consensus value for the candidate pose ${}^W\mathbf{x}$ with the rendered range image $\mathbf{I}_{R,x}$.

By rendering synthetic range images for a set of poses, the corresponding difference images can be computed with respect to the measured range image. In this work, this is done for all candidate positions within the xy -grid using the initial heading angle. Each difference image can then be evaluated with respect to the maximum consensus criterion, yielding an inlier count for each candidate position. In Fig. 4.6, difference images for positions in a 4-neighborhood around the ground truth position are visualized. No shift $\mathbf{t}_{xy} = (0\text{ m}, 0\text{ m})$ corresponds to the ground truth pose, thus the differences are close to zero for most pixels. The shifts in x -direction $\mathbf{t}_{xy} = (-0.3\text{ m}, 0\text{ m})$ and $\mathbf{t}_{xy} = (0.3\text{ m}, 0\text{ m})$ correspond to longitudinal shifts, which cause differences of the same size as the shift at perpendicular parts, especially visible at the arch-shaped structures on the right. Unsurprisingly, the two equally sized shifts in opposite directions cause equally sized differences of opposing sign. For the lateral shifts $\mathbf{t}_{xy} = (0\text{ m}, -0.3\text{ m})$ and $\mathbf{t}_{xy} = (0\text{ m}, 0.3\text{ m})$, the opposing behavior becomes even more visible as the differences do not occur at the small perpendicular structures, but at the main parts of the facades. Based on this visual impression, it becomes already clear that the inlier counts drop significantly for lateral deviations, whereas they are of similar magnitude for longitudinal deviations.

The real range measurements, and thus the measured range images, are taken from a continuously moving point of view, as the vehicle travels during a full 360° LiDAR rotation. However, the

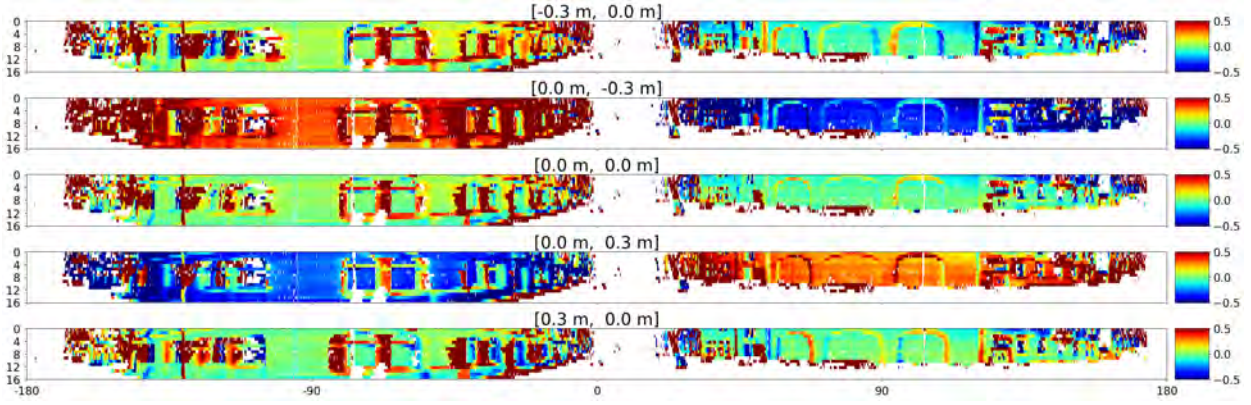


Figure 4.6.: Difference images for positions in a 4-neighborhood around the ground truth position: No shift $\mathbf{t}_{xy} = (0\text{ m}, 0\text{ m})$ corresponds to the ground truth pose, thus the differences are close to zero for most pixels. The shifts in x -direction $\mathbf{t}_{xy} = (-0.3\text{ m}, 0\text{ m})$ and $\mathbf{t}_{xy} = (0.3\text{ m}, 0\text{ m})$ correspond to longitudinal shifts along driving direction, which cause differences of the same size at perpendicular parts, especially visible at the arch-shaped structures on the right. It is visible that the two equally sized shifts in opposite directions cause differences of the same size, however, of opposing sign. For the lateral shifts $\mathbf{t}_{xy} = (0\text{ m}, -0.3\text{ m})$ and $\mathbf{t}_{xy} = (0\text{ m}, 0.3\text{ m})$, the opposing behavior becomes even more visible as the differences do not occur at the small perpendicular structures, but at the main parts of the facades.

synthetic range images are generated based on single poses and therefore static points of view. This difference causes systematic errors on the ranges as well as different occlusions between the measured and synthetic range images. As described above, by deskewing the measured point cloud and reprojecting the calculated ranges between the points and a single LiDAR position in the image, the systematic errors on the range are avoided. However, the difference in the occlusions, which is considered as the minor influence, remains for reasons of simplification.

To consider not only the xy -positions in the localization, but also the heading angle θ , the synthetic range images are shifted horizontally and the difference images are computed for each shift. Under the assumption that the scanner is exactly horizontally mounted, cyclically shifting the synthetic range image corresponds to a rotation around the up-axis. However, this approach restricts the discretization step size of the heading search range to be a multiple of the horizontal resolution of the LiDAR. Accordingly, the minimum step size, which is a shift by one column, equals the horizontal resolution.

To determine the localization solution, which is the pose with the highest consensus, for each difference image, the inlier count is determined, which is the number of pixels with an absolute difference less than the inlier threshold ϵ . Formally, the objective function for the maximum consensus localization in the observation space is defined as follows:

$$\theta^*, \mathbf{t}_{xy}^* = \arg \max_{\theta \in [-\pi, \pi), \mathbf{t}_{xy} \in \mathbb{R}^2} \Psi(\mathbf{R}_\theta, \mathbf{t}) \quad (4.6a)$$

$$\Psi(\mathbf{R}_\theta, \mathbf{t}) = \sum_{c=1}^n \sum_{r=1}^m \mathbb{I}(|I_{R,x}(\mathbf{R}_\theta, \mathbf{t}, c, r) - I_{R,s}(c, r)| \leq \epsilon), \quad (4.6b)$$

where \mathbf{t}_{xy}^* and θ^* are the optimal translation and heading angle, respectively, which are to be applied to the initial pose ${}^W\mathbf{x}_0$ to obtain the optimal pose ${}^W\mathbf{x}^*$, Ψ is the consensus function that is to be maximized, which is the count of inliers, $\mathbb{I}(c)$ is the indicator function, which returns +1 if the condition c is true, and 0 if not, ϵ is the inlier threshold, $\mathbf{R}_\theta \in SO(3)$ is the rotation matrix of the heading angle θ and $\mathbf{t} \in \mathbb{R}^3$, with $\mathbf{t} = (\mathbf{t}_{xy}, 0)^\top$ and $\mathbf{t}_{xy}^* \in \mathbb{R}^2$, is the translation vector

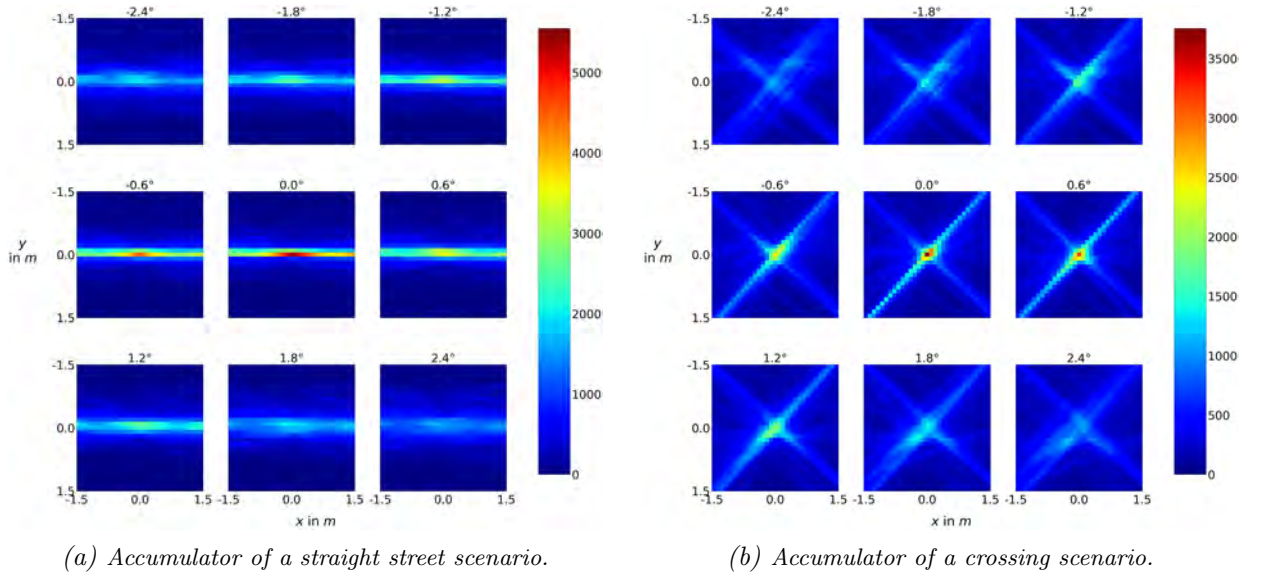


Figure 4.7.: The accumulators shown in (a) and (b) represent the maximum consensus localization solutions based on correspondences in the observation space for the straight street scenario visualized in Fig. 4.8 and the crossing scenario visualized in Fig. 4.9, respectively. The geometry of the straight street results in a single line within the accumulator as most matches occur at the parallel facades, whereas the geometry of the crossing scene yields multiple lines that arise from matches at facades with different orientations. The highest inlier count appears at $(x, y, \theta) = (0\text{ m}, 0\text{ m}, 0^\circ)$, which corresponds to the true pose. As the deviation from the true heading increases, the consensus values drop significantly and the peak disappears. The color scheme corresponds to temperature scale, from blue to red.

of the shift in the xy -plane, which together represent a candidate transformation, $\mathbf{I}_{R,x}$ and $\mathbf{I}_{R,s}$ are the synthetic and measured range image, respectively, m and n are the number of rows and columns, respectively, and (c, r) are the image coordinates. Since the inliers are summed up for each combination of θ and \mathbf{t}_{xy} , the evaluated 3D search space is also referred to as *accumulator*.

In Fig. 4.7, two accumulators are visualized each consisting of a 2D grid of candidate positions and nine different heading angles. For these accumulators and also the accumulators of the other approaches introduced during the course of this chapter, the search range is set to $\pm 1.5\text{ m}$ for both translational axes and $\pm 2.4^\circ$ for the heading angle. For the translation, a discretization step size of 10 cm is applied, which yields a grid of 31×31 positions, and for the heading, a discretization step size of 0.6° is implemented, which yields the nine angles. Given the horizontal resolution of 0.2° of the Velodyne VLP-16 that is used in this example, an angular step of 0.6° corresponds to a shift by three columns in the synthetic range image.

The accumulators shown in Fig. 4.7a and 4.7b represent the localization results for the straight street scenario visualized in Fig. 4.8 and the crossing scenario visualized in Fig. 4.9, respectively. The geometry of the straight street results in a single line within the accumulator as most matches occur at the parallel facades. Since only a few protruding facade elements restrict the position longitudinally, a reliable localization in driving direction is difficult. In contrast, the geometry of the crossing scene yields multiple lines that arise from matches at facades with different orientations. Such scenarios allow for a unique and reliable localization solution, which occurs in the intersection of the lines. The highest inlier count appears at $(x, y, \theta) = (0\text{ m}, 0\text{ m}, 0^\circ)$, which corresponds to the true pose. In this and also the following examples throughout this chapter, provided the maximum consensus localization works correctly, the highest inlier counts have to be at $(0\text{ m}, 0\text{ m})$, since the car sensor scans are initialized in the known ground truth pose. As the deviation from the true heading increases, the consensus values drop significantly and the peak disappears.

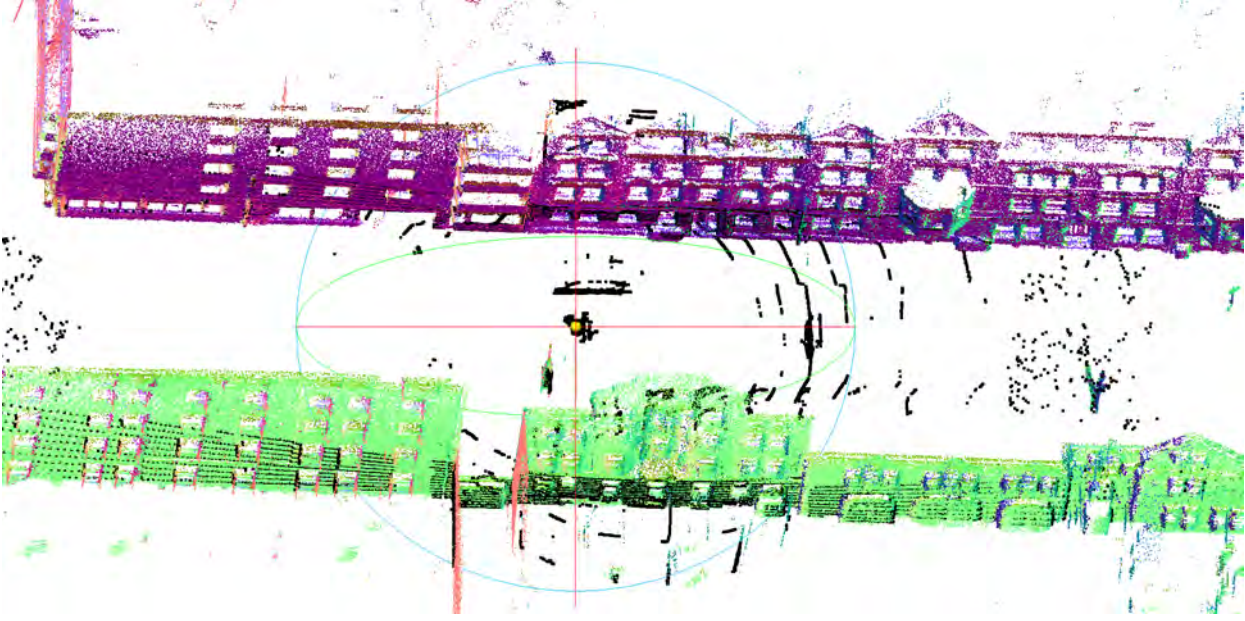


Figure 4.8.: Exemplary straight street scenario with the body frame of the vehicle illustrated in red, green, and blue, the car sensor scan shown in black, and the map point cloud colored according to its normal vectors. The geometry of the scene makes a reliable localization in driving direction difficult as only the few protruding facade elements restrict the position longitudinally.

4.1.3. Maximum consensus in the Euclidean space

As for maximum consensus in the observation space, the goal is to determine the consensus for each candidate transformation ${}^C\mathbf{T}_V$ of the discretized, 3-DoF search space, where the transformation ${}^C\mathbf{T}_V^*$, which yields the highest consensus, leads to the optimal pose ${}^W\mathbf{x}^*$ that is the localization solution. Both methods are based on an exhaustive search, which provides the globally optimal solution. However, whereas in the observation space the approach is based on projective data associations, in the Euclidean space it utilizes point-to-point correspondences. To this end, both the car sensor scan ${}^L\mathbf{S}$ and the globally georeferenced map ${}^W\mathbf{M}$ are represented as point clouds, where ${}^L\mathbf{S} = {}^L\{\mathbf{s}_i\}_{i=1}^n$, $\mathbf{s}_i \in \mathbb{R}^3$ denotes the car sensor scan point cloud with n points in the LiDAR coordinate system, and ${}^W\mathbf{M} = {}^W\{\mathbf{m}_j\}_{j=1}^m$, $\mathbf{m}_j \in \mathbb{R}^3$ denotes the map point cloud with m points in the world coordinate system. In the following, for the sake of readability, the car sensor scan point cloud is referred to as ‘scan point cloud’, a point from the scan point cloud is referred to as ‘scan point’, and a point from the map point cloud is referred to as ‘map point’.

In the maximum consensus approach in the Euclidean space, the consensus of a candidate transformation ${}^C\mathbf{T}_V$ within the search space is defined as the total number of scan points that match a map point for the corresponding pose ${}^W\mathbf{x}$, and a match exists if the l_∞ -distance between a scan and a map point is less than the inlier threshold ϵ . The primitive approach is to rotate and shift the scan point cloud ${}^L\mathbf{S}$ according to each candidate transformation ${}^C\mathbf{T}_V$, while counting the number of scan points matching a map point. Following this approach, the transformation of the scan point cloud ${}^L\mathbf{S}$ from the LiDAR frame into the world frame under the consideration of a candidate transformation ${}^C\mathbf{T}_V$ is defined as:

$${}^W\mathbf{S}_x = {}^W\mathbf{T}_C {}^C\mathbf{T}_V {}^V\mathbf{T}_L {}^L\mathbf{S}, \quad (4.7)$$

where ${}^V\mathbf{T}_L$ is the transformation from the LiDAR frame into the body frame of the vehicle, ${}^W\mathbf{T}_C$ is the transformation matrix based on the initial pose ${}^W\mathbf{x}_0$, and ${}^W\mathbf{S}_x$ is the resulting scan point

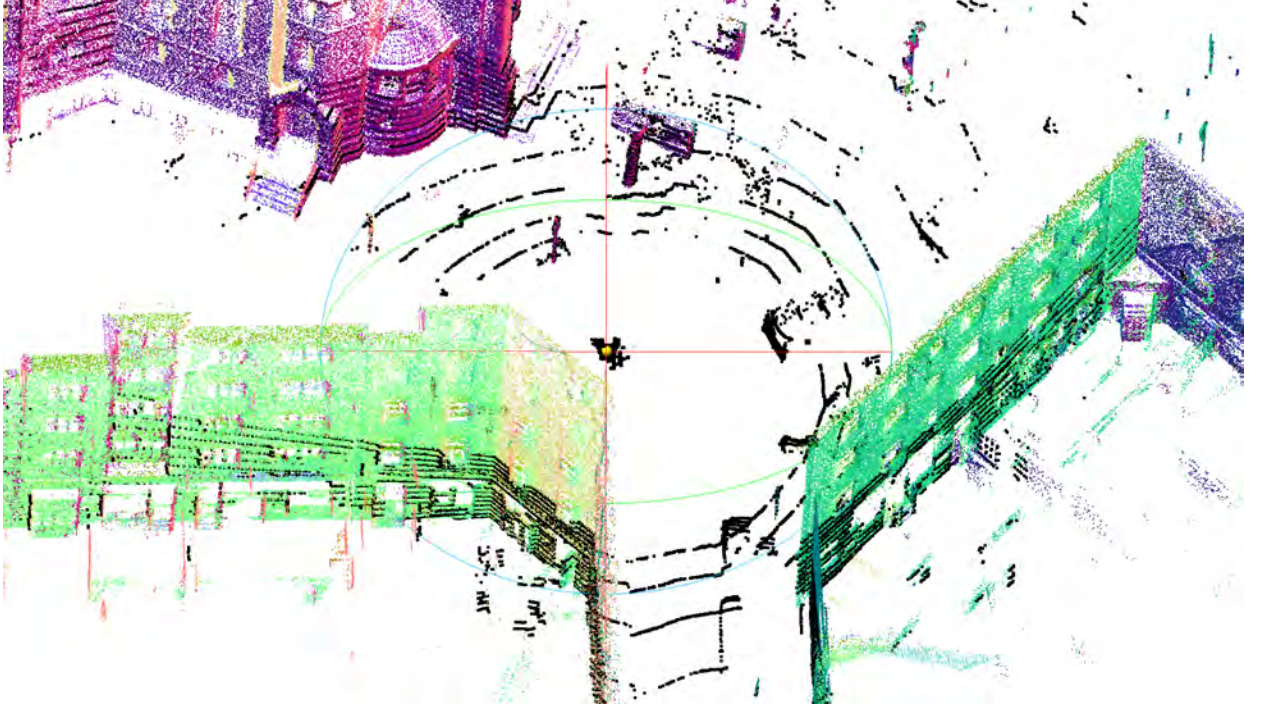


Figure 4.9.: Exemplary crossing scenario with the body frame of the vehicle illustrated in red, green, and blue, the car sensor scan shown in black, and the map point cloud colored according to its normal vectors. The geometry of the scene facilitates a unique and reliable localization solution.

cloud located in the candidate pose ${}^W\mathbf{x}$. For the transformation, ${}^L\mathbf{S}$ is represented as $4 \times n$ matrix with $\mathbf{s}_i = (x, y, z, 1)^T$ represented as homogeneous coordinates in the columns. Formally, the objective function for the maximum consensus localization in the Euclidean space is defined as:

$$\theta^*, \mathbf{t}_{xy}^* = \arg \max_{\theta \in [-\pi, \pi), \mathbf{t}_{xy} \in \mathbb{R}^2} \Psi(\mathbf{R}_\theta, \mathbf{t}) \quad (4.8a)$$

$$\Psi(\mathbf{R}_\theta, \mathbf{t}) = \sum_{i=1}^n \sum_{j=1}^m \mathbb{I}(\|\mathbf{R}_\theta \mathbf{s}_i + \mathbf{t} - \mathbf{m}_j\|_\infty \leq \epsilon), \quad (4.8b)$$

where \mathbf{t}_{xy}^* and θ^* are the optimal translation and heading angle, respectively, which maximize the consensus function Ψ , i.e., the total count of matches with respect to the point cloud registration. The symbols $\mathbf{s}_i \in \mathbb{R}^3$ and $\mathbf{m}_j \in \mathbb{R}^3$ denote a scan point and map point, respectively, $\mathbf{R}_\theta \in SO(3)$ is the rotation matrix of the heading angle θ and $\mathbf{t} \in \mathbb{R}^3$, with $\mathbf{t} = (\mathbf{t}_{xy}, 0)^T$ and $\mathbf{t}_{xy}^* \in \mathbb{R}^2$, is the translation vector of the shift in the xy -plane that together represent a candidate transformation, $\mathbb{I}(c)$ is the indicator function, which returns +1 if the condition c is true and 0 if not, ϵ is the inlier threshold, which is equal to half the cell size, and $\|\cdot\|_\infty$ denotes the l_∞ -norm, which is defined as $\|\mathbf{r}\|_\infty = \max(|r_1|, \dots, |r_N|)$ for a vector \mathbf{r} of length N . The l_∞ -norm leads to the quadratic shape and the quadratic grid cells of the search space of the translation \mathbf{t}_{xy} , allowing for the stencil operations described in the following.

To avoid iterating over all candidate poses and finding the matches for each pose individually, for each investigated heading angle of the search space, stencil operations are used to directly evaluate a full grid of 2D positions and fill an entire ‘slice’ of the accumulator. To this end, for a candidate heading angle θ , the scan point cloud is transformed into the world frame using the initial pose:

$${}^W\mathbf{S}_\theta = {}^W\mathbf{T}_C {}^C\mathbf{T}_{V,\theta} {}^V\mathbf{T}_L {}^L\mathbf{S}, \quad (4.9)$$

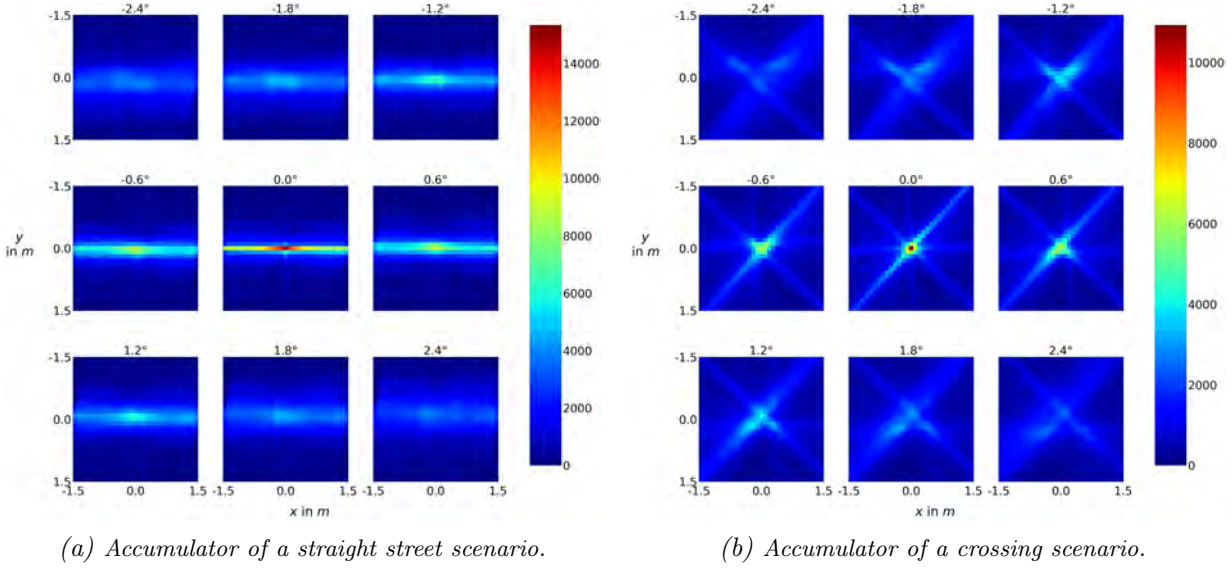


Figure 4.10.: The accumulators shown in (a) and (b) represent the maximum consensus localization solutions based on correspondences in the Euclidean space for the straight street scenario visualized in Fig. 4.8 and the crossing scenario visualized in Fig. 4.9, respectively. The geometry of the scene is reflected by lines of increased consensus values in the accumulator. Thus, for a straight street, a single line of high inlier counts appears representing an unreliable localization result, and for a crossing scene, multiple lines of high inlier counts appear, which yield a distinct peak at their intersection, representing a reliable localization solution. The highest inlier count appears at $(x, y, \theta) = (0\text{ m}, 0\text{ m}, 0^\circ)$, which corresponds to the true pose. As the deviation from the true heading increases, the consensus values drop significantly and the peak disappears. The color scheme corresponds to temperature scale, from blue to red.

where ${}^C T_{V, \theta}$ denotes the transformation matrix that only considers the rotation around the up-axis by θ , and ${}^W S_\theta$ denotes the corresponding scan point cloud after the transformation into the world frame. Then, the corresponding 2D ‘position’ slice of the accumulator is computed as follows: For each scan point ${}^W s_{\theta, i}$, the local neighborhood is discretized similarly to the grid of xy -positions, and the grid cells of the local neighborhood that contain at least one map point are incremented by +1 in the accumulator slice of that particular heading angle. Employing this technique, the inlier threshold ϵ equals half of the cell size of the position grid, and the candidate positions correspond to the center points of the 2D grid cells, with the initial position located in the center cell of the grid. By simply stacking the set of 2D accumulator slices the 3D accumulator ‘cube’ is obtained, which contains the inlier counts for the full 3-DoF search space. If a spatial data structure such as kd -tree is utilized for all map points, the map points inside a scan point’s local neighborhood can be efficiently selected. However, the strategy using stencil operations is merely another, more efficient way of implementation that does not change the results in any way.

The accumulators shown in Fig. 4.10a and 4.10b represent the maximum consensus localization solutions based on correspondences in the Euclidean space for the straight street scenario visualized in Fig. 4.8 and the crossing scenario visualized in Fig. 4.9, respectively. The chosen parameters are the same as for the observation space, namely a search range from -1.5 m to $+1.5\text{ m}$ for both translational axes and from -2.4° to $+2.4^\circ$ for the heading, with a discretization of 0.1 m and 0.6° , respectively, resulting in accumulators of 31×31 candidate positions and nine different heading angles. Similar to the accumulators determined in the observation space (cf. Fig. 4.7), the geometry of the scene is reflected by lines of increased consensus values in the accumulator. Thus, for a straight street, a single line of high inlier counts appears representing an unreliable localization result, and for a crossing scene, multiple lines of high inlier counts appear, which yield a distinct peak at their intersection, representing a reliable localization solution.

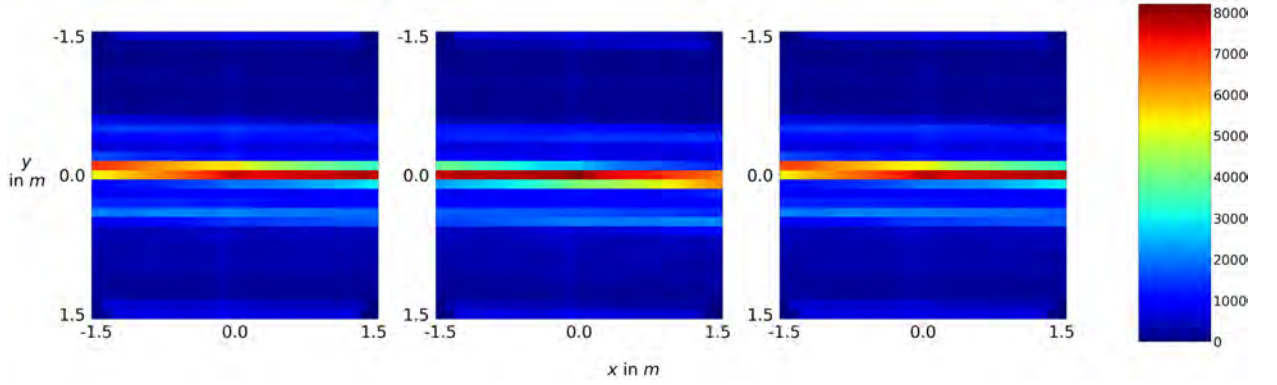


Figure 4.11.: Epoch where grid shifts prevent a localization failure caused by the discretization of the search space: The accumulator slice of 0° is shown without any additional shift (left) and with a shift by half the size of a grid cell along the y -axis (center) and along the x -axis (right). Without the additional shifts, the highest consensus would have occurred far away from the true position at $(x, y) = (0\text{ m}, 0\text{ m})$, but with the shift, in this case along the y -axis (center), the highest consensus occurs at the correct position and a localization failure can be avoided.

To prevent localization failures caused by the discretization of the search space, shifts by half a grid cell along each axis are conducted. In Fig. 4.11, for an epoch where the shifts prevent a localization failure, the accumulator slice of 0° is shown without any additional shift (left) and with a shift by half the size of a grid cell along the y -axis (center) and along the x -axis (right). In this epoch, without the additional shifts, the highest consensus would have occurred far away from the true position at $(x, y) = (0\text{ m}, 0\text{ m})$, but with the shift, in this case along the y -axis (center), the highest consensus occurs at the correct position and a localization failure can be avoided.

4.2. Optimization based on point-to-plane adjustment score

In this section, a novel robust objective function based on Helmert’s point error is proposed, which can be used for point cloud registration and thus also for robust and reliable vehicle localization. Instead of the count of inliers from the original maximum consensus criterion, the novel objective function provides the so-called *point-to-plane adjustment score*, which can be utilized one-to-one in the existing maximum consensus localization framework. In this way, the shortcomings of the maximum consensus localization based on the simple inlier count can be successfully tackled. This especially concerns the localization in straight streets with few protruding structures that fix the position in driving direction. The novel objective function has the advantage that it only requires very few correspondences to successfully and robustly localize, enabling a reliable positioning also in such challenging scenarios.

In general, counting is a simple, fast, and at the same time robust operation, however, it is not able to model the constraints generated by each correspondence with respect to the unknown pose. Typically, the majority of LiDAR points are located on surfaces such as facades, walls, and roads, which represent point-to-plane associations. However, if only the number of those ‘hits’ is considered, as it applies to the maximum consensus inlier count, the imposed constraint from the point-to-plane correspondence on the pose, which acts orthogonal to the plane, is not taken into account and thus a significant part of the information is thrown away. In straight streets, the vast majority of correspondences exist at parallel building facades to the left and right of the vehicle, and only few correspondences stem from protruding parts, for example balconies, which constrain the position in longitudinal direction. In such scenarios, a localization relying on the plain count of matches is unreliable and prone to failure, since the few matches actually constraining the

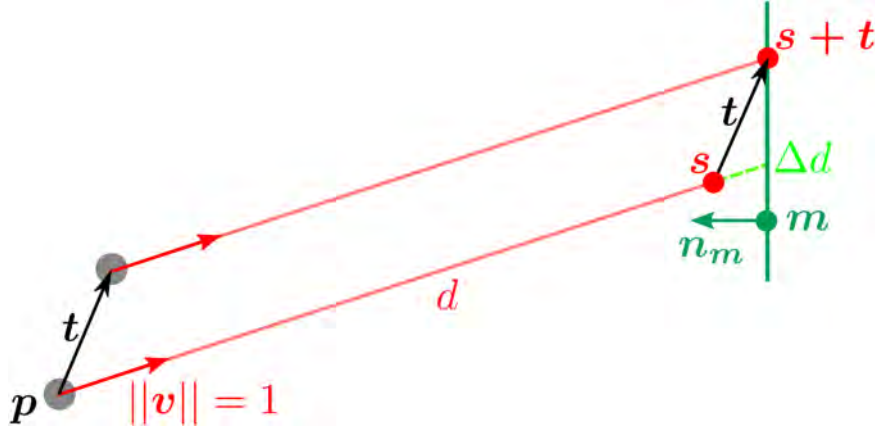


Figure 4.12.: Fundamental concept of the point-to-plane adjustment illustrated based on one out of many measurements. Given a scan point s with the measured distance d in the direction of v , the goal is to estimate the translation t that s is located on the local plane of the map point m , which is defined by the normal vector n_m .

pose in the longitudinal direction can be easily outvoted, which may happen, for example, in the presence of varying point density along the facade or unfavorable outlier occurrences. To avoid such registration failures and to increase the localization reliability, the imposed point-to-plane constraint of each correspondence is to be modeled and utilized in an objective function.

The presented novel objective function for point cloud registration is based on a point-to-plane adjustment taking into account all matches from the original consensus set and is applicable to both Euclidean and projective data associations. The point-to-plane correspondences are modeled using the normal vectors of the matched map points or mesh vertices, since they are more stable than those from the car sensor scan because of the higher and more regular point density of the map point cloud. The covariance matrix of the adjustment states the errors for the estimated pose parameters, and if these errors are small and of similar size for all components, the trace of the covariance matrix will be small as well. For the translational part, the square root of the trace is also referred to as Helmert's point error, and the inverse of the trace, i.e. the inverse of the square of Helmert's point error, is defined as the novel robust point-to-plane adjustment score that is to be maximized. This is reasonable because for n i.i.d. measurements of unit variance, the variance of the mean estimator will be $1/n$, thus the score can be considered as a generalized form of the inlier count.

In the following, the point-to-plane adjustment is modeled and the score is derived for the approach in both the observation space and the Euclidean space. In Fig. 4.12, the fundamental concept of the adjustment is illustrated based on one out of many measurements. Given a scan point s with the measured distance d in the direction of v , the goal is to estimate the translation t that s is located on the local plane of the map point m , which is defined by the normal vector n_m .

Observation space

From a location p , along the ray given by the unit vector v , the point s is measured. The distance Δd to a locally planar surface, given by map point m and normal vector n_m , measured along the ray, is obtained from

$$\langle s + \Delta d \cdot v - m, n_m \rangle = 0, \quad (4.10)$$

where $\langle \cdot, \cdot \rangle$ denotes the scalar product. If the location \mathbf{p} , or equivalently \mathbf{s} , is shifted by translation \mathbf{t} , $\mathbf{s} + \mathbf{t}$ is located on the local plane if

$$\langle \mathbf{s} + \mathbf{t} - \mathbf{m}, \mathbf{n}_m \rangle = 0, \quad (4.11)$$

and by subtracting Eq. 4.11 from Eq. 4.10 one obtains

$$\langle \Delta d \cdot \mathbf{v} - \mathbf{t}, \mathbf{n}_m \rangle = 0, \quad (4.12)$$

or

$$\Delta d = \frac{\mathbf{n}_m^T}{\langle \mathbf{n}_m, \mathbf{v} \rangle} \mathbf{t}, \quad (4.13)$$

which is the functional model, with observed Δd and unknown translation \mathbf{t} . Thus, the design matrix \mathbf{A} contains the normal vector \mathbf{n}_m^T of the assigned surface element in each row, weighted by $\frac{1}{\langle \mathbf{n}_m, \mathbf{v} \rangle}$.

For the point-to-plane adjustment, the observation of every match, i.e. of every correspondence with an absolute pixel value less than the inlier threshold ϵ in the difference image \mathbf{D}_x , is taken into consideration. In the first place, however, not the translation itself, but the error of the unknown translation is sought, which is described by the covariance matrix Σ_{tt} . For the comparison of different candidate poses based on the point-to-plane adjustment score, it is sufficient to determine the cofactor matrix \mathbf{Q}_{tt} , as it corresponds to the covariance matrix Σ_{tt} up to the constant factor σ_0^2 , which is the variance of the observations:

$$\Sigma_{tt} = \sigma_0^2 \mathbf{Q}_{tt} \quad (4.14)$$

The cofactor matrix \mathbf{Q}_{tt} is calculated as follows:

$$\mathbf{Q}_{tt} = \mathbf{N}^{-1} = (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} = \begin{bmatrix} \sum \left(\frac{n_{x,i}}{\langle \mathbf{n}_{m_i}, \mathbf{v} \rangle} \right)^2 & \sum \frac{n_{x,i}}{\langle \mathbf{n}_{m_i}, \mathbf{v} \rangle} \frac{n_{y,i}}{\langle \mathbf{n}_{m_i}, \mathbf{v} \rangle} \\ \sum \frac{n_{x,i}}{\langle \mathbf{n}_{m_i}, \mathbf{v} \rangle} \frac{n_{y,i}}{\langle \mathbf{n}_{m_i}, \mathbf{v} \rangle} & \sum \left(\frac{n_{y,i}}{\langle \mathbf{n}_{m_i}, \mathbf{v} \rangle} \right)^2 \end{bmatrix}^{-1}, \quad (4.15)$$

with $\mathbf{A} = \begin{bmatrix} \frac{n_{x,1}}{\langle \mathbf{n}_{m_1}, \mathbf{v} \rangle} & \frac{n_{y,1}}{\langle \mathbf{n}_{m_1}, \mathbf{v} \rangle} \\ \frac{n_{x,2}}{\langle \mathbf{n}_{m_2}, \mathbf{v} \rangle} & \frac{n_{y,2}}{\langle \mathbf{n}_{m_2}, \mathbf{v} \rangle} \\ \vdots & \vdots \\ \frac{n_{x,k}}{\langle \mathbf{n}_{m_k}, \mathbf{v} \rangle} & \frac{n_{y,k}}{\langle \mathbf{n}_{m_k}, \mathbf{v} \rangle} \end{bmatrix}$ and $\mathbf{P} = \mathbf{I}_k$,

where \mathbf{A} denotes the design matrix, and \mathbf{P} denotes the weight matrix, which is chosen as the identity matrix because all observations are assumed to be equally accurate. The elements of the design matrix correspond to the x - and y -parts of the normal vectors of the matched mesh faces \mathbf{n}_{m_i} divided by the scalar product of the same normal vector \mathbf{n}_{m_i} with the direction vector of the LiDAR ray \mathbf{v} . The division by the scalar product has the effect that correspondences with a sharp incidence angle, and thus a scalar product close to zero, obtain a very large value, which leads to overly small eigenvalues of \mathbf{Q}_{tt} . To avoid this effect, the scalar product considering the incidence angle is neglected and the denominator is always set to one in the context of this work.

To obtain the normal vectors of the matched mesh faces \mathbf{n}_{m_i} using projective associations, for each grid pose, in addition to the range image, a scan image containing the normal vectors is rendered, each with an x , y , and z channel. The normal vectors of matching ranges are then included in the point-to-plane adjustment and thus in the score estimation. In Fig. 4.13, a rendered scan image is shown with three channels representing the x , y , and z components of the normal vector.

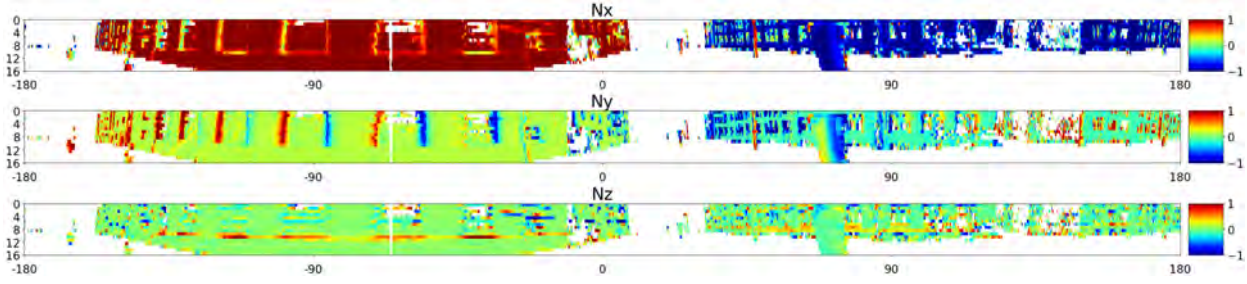


Figure 4.13.: Synthetic LiDAR scan image with three channels representing the x , y , and z components of the normal vector.

Euclidean space

Similar to the observation space, the goal of the adjustment is that the scan point \mathbf{s} matches the local planes of the map point \mathbf{m} based on an unknown translation \mathbf{t} (cf. Fig. 4.12). However, instead of considering the distance Δd along the direction vector \mathbf{v} , the adjustment is modeled based on the Euclidean scan point \mathbf{s} , and the functional model is defined as:

$$\begin{aligned} 0 &= \langle \mathbf{n}_m, \mathbf{s} + \mathbf{t} - \mathbf{m} \rangle \\ \langle \mathbf{n}_m, \mathbf{m} - \mathbf{s} \rangle &= \langle \mathbf{n}_m, \mathbf{t} \rangle = \mathbf{n}_m^\top \cdot \mathbf{t}, \end{aligned} \quad (4.16)$$

where the adjustment aims to minimize the point-to-plane distance $\langle \mathbf{n}_m, \mathbf{m} - \mathbf{s} \rangle$. Similar to the approach in the observation space, the cofactor matrix \mathbf{Q}_{tt} is determined using the design matrix \mathbf{A} and the weight matrix \mathbf{P} :

$$\begin{aligned} \mathbf{Q}_{tt} = \mathbf{N}^{-1} &= (\mathbf{A}^\top \mathbf{P} \mathbf{A})^{-1} = \begin{bmatrix} \sum w_i n_{x,i}^2 & \sum w_i n_{x,i} n_{y,i} \\ \sum w_i n_{x,i} n_{y,i} & \sum w_i n_{y,i}^2 \end{bmatrix}^{-1} \\ \text{with } \mathbf{A} &= \begin{bmatrix} n_{x,1} & n_{y,1} \\ n_{x,2} & n_{y,2} \\ \vdots & \vdots \\ n_{x,k} & n_{y,k} \end{bmatrix} \text{ and } \mathbf{P} = \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & w_m \end{bmatrix}. \end{aligned} \quad (4.17)$$

Compared to the design matrix \mathbf{A} in Eq. 4.15, the elements are also the normal vector components $n_{x,i}$ and $n_{y,i}$ of the matched map element (here: map point), however, without the quotient considering the scalar product of \mathbf{n}_m and \mathbf{v} . It was decided to neglect this quotient as it yields excessively low uncertainties, which is why the calculation of the design matrices is the same. In the Euclidean space, a matched map point is determined based on the spatial vicinity to a scan point, with a distance less than the inlier threshold ϵ , similar to the localization approach using the inlier count explained in Sec. 4.1. While in Eq. 4.15 the weight matrix \mathbf{P} is set to the identity for reasons of simplicity, in the Euclidean space the cosine of the angle between the normal vector of the scan point \mathbf{n}_{s_i} and the normal vector of the map point \mathbf{n}_{m_i} is used as weight in a diagonal matrix. However, the weight is only used if it is positive, and if it is not, it is set to zero, which effectively excludes the correspondence. Following this, the weights are calculated as $w_i = \max(0, \langle \mathbf{n}_{s_i}, \mathbf{n}_{m_i} \rangle)$.

Definition of the point-to-plane adjustment score based on correspondences in both the observation and the Euclidean space

The eigenvalues λ_1 and λ_2 of the cofactor matrix \mathbf{Q}_{tt} state up to the factor σ_0^2 the minimum and maximum variance of the estimated translation in the directions of the respective eigenvectors

(cf. Eq. 4.14). They can be determined by an eigendecomposition, with $\mathbf{Q}_{tt} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$ and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2)$, where small values for both λ_1 and λ_2 indicate that the observations enable a precise estimation of \mathbf{t} . Thus, $l = \max(\lambda_1, \lambda_2)$ or $l = \lambda_1 + \lambda_2$ are appropriate loss definitions, where the latter one corresponds to the square of Helmert's point error (neglecting the variance factor), which is commonly used as a measure of accuracy for 2D point estimates and is defined as the square root of the sum of variances $l_H = \sqrt{\sigma_x^2 + \sigma_y^2}$. To obtain an objective function that is to be maximized, the inverse of the square of Helmert's point error is defined as the point-to-plane adjustment score:

$$\text{score} = \frac{1}{\lambda_1 + \lambda_2} = \frac{1}{\text{tr}(\mathbf{Q}_{tt})}. \quad (4.18)$$

To motivate that the defined point-to-plane adjustment score is an appropriate generalization of the maximum consensus inlier count, n_1 observations are defined in one normal vector direction and n_2 observations are defined orthogonal to n_1 , where all observations have unit variance and unit weight, then the score increases linearly with n , since $\text{score} = 1/(1/n_1 + 1/n_2) = n/2$ for $n_1 = n_2 = n$. In the case that there is only one observation $n_1 = 1$ in one of the directions, then $\text{score} = n_2/(n_2 + 1)$, i.e. the score is always less than one, indicating a translation with high uncertainty, even if n_2 is infinitely large. The score can be determined without computing a matrix inversion or eigendecomposition by utilizing the relation $\text{tr}(\mathbf{Q}_{tt}) = \text{tr}(\mathbf{N}^{-1}) = \text{tr}(\mathbf{N})/\det(\mathbf{N})$.

Finally, to obtain an accuracy below the discretization step size of the search space, the adjusted translation \mathbf{t}^* is estimated for the grid cell or candidate pose of the highest score:

$$\begin{aligned} \mathbf{t}^* &= (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \mathbf{l} \\ \text{with } \mathbf{l} &= \langle \mathbf{n}_m, \mathbf{m} - \mathbf{s} \rangle \text{ in Euclidean space or} \\ \text{with } \mathbf{l} &= \langle \mathbf{n}_m, \mathbf{m} - d\mathbf{v} \rangle \text{ in observation space.} \end{aligned} \quad (4.19)$$

In Fig. 4.14, evaluated search spaces using point-to-plane adjustment score based on correspondences in the observation space and Euclidean space are illustrated. The scenarios are the same as for the maximum consensus inlier count in Sec. 4.1, where in Fig. 4.14a and 4.14c, the search spaces are based on the straight street scenario visualized in Fig. 4.8, and in Fig. 4.14b and 4.14d, the search spaces are based on the crossing scenario visualized in Fig. 4.9. For the sake of comparability, the chosen parameters are also the same as for the inlier count in Sec. 4.1, with a search range from -1.5 m to $+1.5\text{ m}$ for both translational axes and from -2.4° to $+2.4^\circ$ for the heading, with a discretization of 0.1 m and 0.6° , respectively, resulting in accumulators of 31×31 candidate positions and nine different heading angles. In addition, the ground truth pose is again used for initialization, thus the highest scores are expected to be at $(0\text{ m}, 0\text{ m})$.

Using the inlier count, the accumulators reflect the scene geometry, which leads to one strong line of multiple high consensus values for a straight street (cf. Fig. 4.7a and 4.10a) and multiple lines intersecting in a sharp and unique peak for a crossing (cf. Fig. 4.7b and 4.10b). In a straight road with no side streets, the parallel facades to the left and right of the vehicle are the dominant structures that render trees or protruding facade parts invisible. In a crossing, however, the facades with different orientations unambiguously constrain the vehicle pose, resulting in a distinct solution. A plane inherently restricts the vehicle pose to a parallel shift along it, and if multiple planes exist, their corresponding lines intersect in one point, which corresponds to the localization solution. However, in a straight street scenario, where almost exclusively the parallel planes exist, multiple candidate poses in longitudinal direction along the corresponding line generate high consensus values, yielding a localization solution of low certainty.

In contrast to the inlier count, the point-to-plane adjustment score enables to provide a distinct localization solution also for straight streets. While the score is low for consensus sets that rely only on matches with normal vectors pointing in the same direction, it is high for consensus sets that consist of matches with normal vectors pointing in different directions. Thus, as can be seen in Fig. 4.14a and 4.14c, for a straight street, a high consensus score only remains at the true pose $(x, y, \theta) = (0\text{ m}, 0\text{ m}, 0^\circ)$, where also the few protruding facade parts and other perpendicular elements align, while the consensus values of the candidate poses with matches only at the parallel facades are significantly reduced. The same applies to the crossing scene, as can be seen in Fig. 4.14b and 4.14d. The single lines of high inlier counts, which reflect the scene geometry in the accumulator, vanish, and only the peak at their intersection remains, eventually even more distinct. The robust characteristic of the novel objective function reduces the susceptibility to registration and thus localization failures, which may occur when using the inlier count in conjunction with an irregular point density within the map, especially in a straight street situation (cf. Sec. 5.3). The objective function of the simple inlier count aims to maximize the number of matches, thus the method inherently overlaps parts with high point densities. As the car sensor scan principally yields an irregular point density (cf. Fig. 2.8), the original maximum consensus localization based on the count is always prone to fail when the point density in the map is irregular as well.

4.3. Optimization based on common loss functions

So far, the inlier threshold ϵ is applied to the residuals to determine inliers and matches, respectively, both for the count and score objective function. In the observation space, the residuals are the differences of the ranges of the synthetic range image $\mathbf{I}_{R,x}$ and the measured range image $\mathbf{I}_{R,s}$, thus $r(\mathbf{R}_\theta, \mathbf{t}) = |\mathbf{I}_{R,x}(\mathbf{R}_\theta, \mathbf{t}, c, r) - \mathbf{I}_{R,s}(c, r)|$, and in the Euclidean space, the residuals are the xy -distances between the scan points \mathbf{s}_i and map points \mathbf{m}_j , thus $r(\mathbf{R}_\theta, \mathbf{t}) = \|\mathbf{R}_\theta \mathbf{s}_i + \mathbf{t} - \mathbf{m}_j\|$. Instead of applying a hard threshold to the residuals, with $r \leq \epsilon$, any loss function can be used to estimate a loss for a certain candidate pose. In general, the loss function has to adapt to the distribution of the measurements, which however varies strongly during the course of a real world driving scenario. This is why there have been approaches to constantly estimate and adapt the shape of the loss function (Chebrolu et al., 2020). In this work, in addition to the maximum consensus inlier count, the loss functions introduced in Sec. 2.1.3 (cf. Fig. 2.2) are applied to the residuals and thereby investigated for the localization task. The goal of this investigation is to determine the varying sensitivity with respect to outliers. To this end, the pose of the global minimum, which represents the localization solution with respect to the utilized loss function, is compared to the ground truth pose. Furthermore, to assess the reliability of the localization solution, the distribution within the evaluated search space is analyzed.

In the observation space, the loss function can be directly applied to the residual value of a pixel in the difference image \mathbf{D}_x . Combining Eq. 2.10 with Eq. 4.6, the following general formulation of the estimation problem is defined for the observation space:

$$\theta^*, \mathbf{t}_{xy}^* = \arg \min_{\theta \in [-\pi, \pi], \mathbf{t}_{xy} \in \mathbb{R}^2} \sum_{c=1}^n \sum_{r=1}^m \rho(|\mathbf{I}_{R,x}(\mathbf{R}_\theta, \mathbf{t}, c, r) - \mathbf{I}_{R,s}(c, r)|). \quad (4.20)$$

In the Euclidean space, the car sensor scan is transformed into each candidate pose. Then, for each scan point, the nearest neighbor is determined and the loss function is applied to the xy -distance.

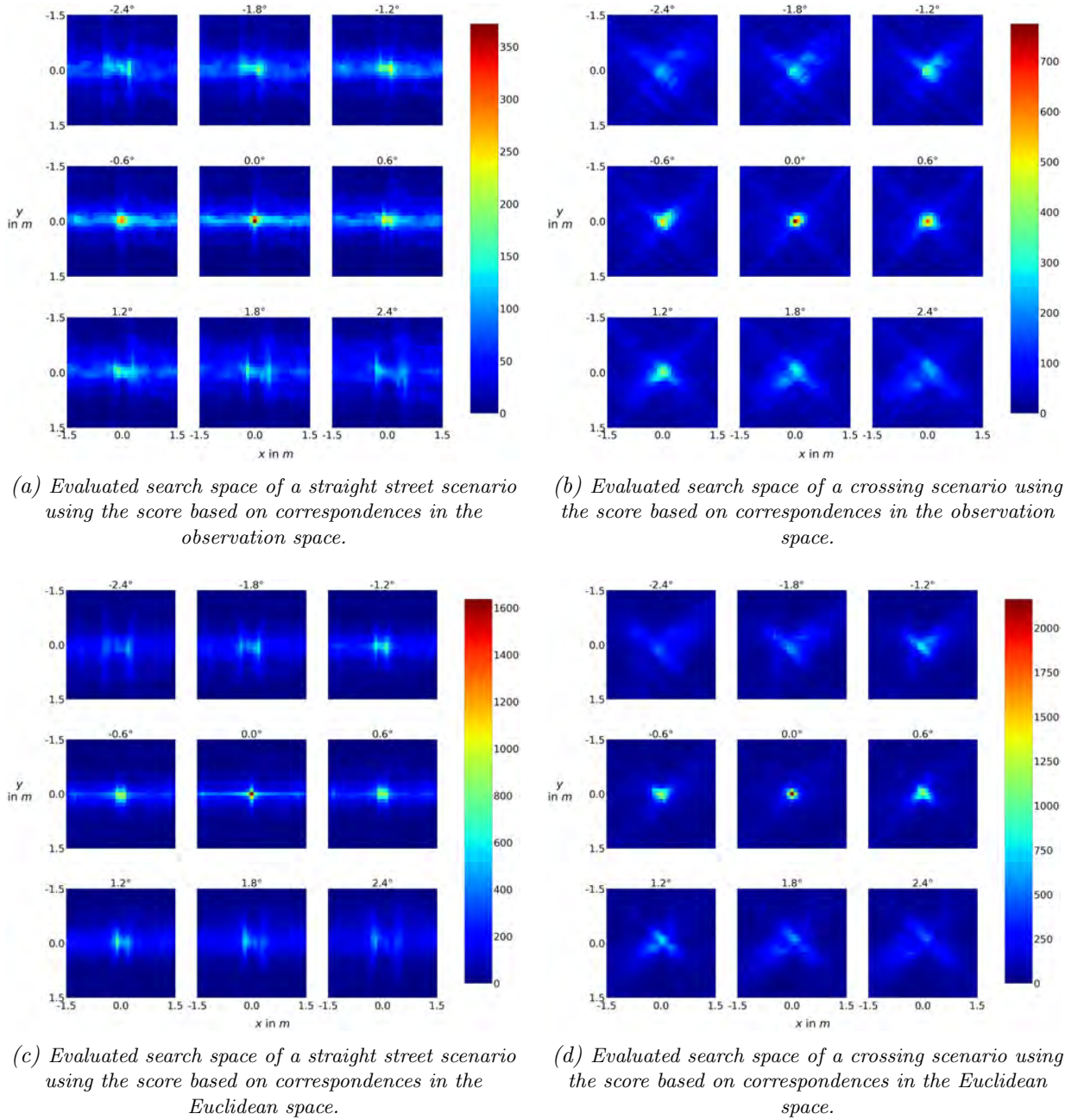


Figure 4.14.: The evaluated search spaces represent the localization solutions using the point-to-plane adjustment score for the straight street scenario visualized in Fig. 4.8 and the crossing scenario visualized in Fig. 4.9, respectively. In (a) and (b), the results are based on correspondences in the observation space, and in (c) and (d), the results are based on correspondences in the Euclidean space. In contrast to the inlier count, the point-to-plane adjustment score enables to provide a distinct localization solution also for straight streets. While the score is low for consensus sets that rely only on matches with normal vectors pointing in the same direction, it is high for consensus sets that consist of matches with normal vectors pointing in different directions. Thus, as can be seen in (a) and (c), for a straight street, a high consensus score only remains at the true pose $(x, y, \theta) = (0 \text{ m}, 0 \text{ m}, 0^\circ)$, where also the few protruding facade parts and other perpendicular elements align, while the consensus values of the candidate poses with matches only at the parallel facades are significantly reduced. The same applies to the crossing scene, as can be seen in (b) and (d). The single lines of high inlier counts, which reflect the scene geometry in the accumulator, vanish, and only the peak at their intersection remains, eventually even more distinct. For all four cases, the score values drop significantly and the peaks disappear as the deviation from the true heading increases. The color scheme corresponds to temperature scale, from blue to red.

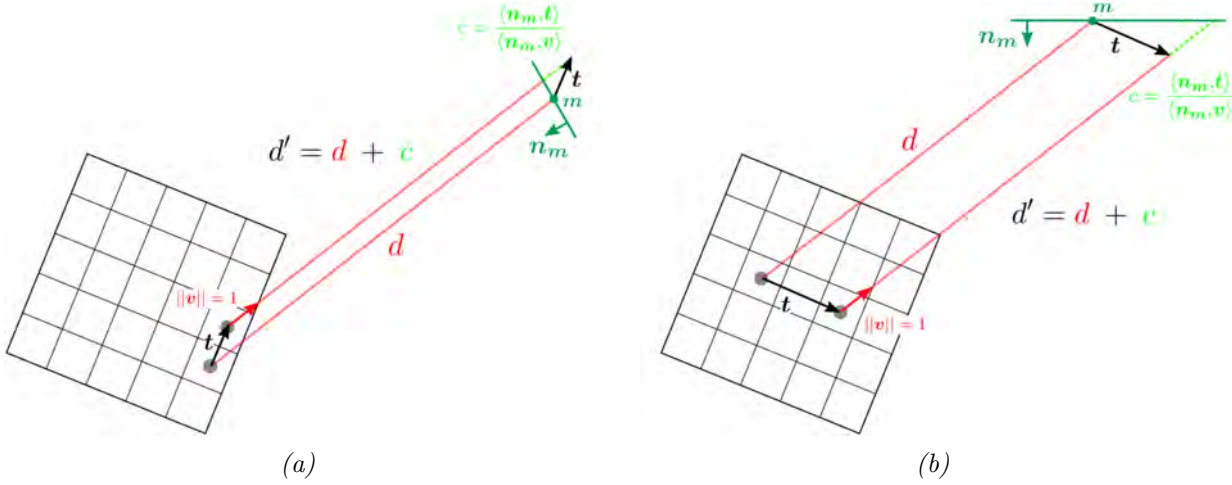


Figure 4.15.: Approach to determine missing correspondences using the normal vector of the hit map object shown for two examples. The additional range c can be negative as shown in (a) and positive as shown in (b). It is depending on the normal vector of the hit map object \mathbf{n}_m , the direction of the LiDAR ray \mathbf{v} , and the shift between the two candidate poses \mathbf{t} .

Combining Eq. 2.10 with Eq. 4.8, the following general formulation of the estimation problem is defined for the Euclidean space:

$$\theta^*, \mathbf{t}_{xy}^* = \arg \min_{\theta \in [-\pi, \pi), \mathbf{t}_{xy} \in \mathbb{R}^2} \sum_{i=1}^n \sum_{j=1}^m \rho(\|\mathbf{R}_\theta \mathbf{s}_i + \mathbf{t} - \mathbf{m}_j\|). \quad (4.21)$$

To determine the overall loss of a specific candidate pose, the loss values of all pixels and scan points, respectively, are summed up. After evaluating each candidate pose, a loss accumulator is obtained for each loss function. Since there is no maximum threshold applied to the residuals, and a correspondence is taken into account for each scan point or measured pixel, for loss functions such as L2, where the residuals have an unbounded influence on the estimate, the estimates can be heavily biased by outliers.

In the Euclidean space, as for each scan point the nearest neighbor is selected, for each candidate pose always the same number of correspondences exist. However, in the observation space, each candidate pose leads to a different scan image regarding both the number and the location of filled pixel. Consequently, the number of correspondences varies for each candidate pose. A varying number of correspondences has the effect that poses with a high number of correspondences receive a greater loss than poses with a small number of correspondences. To obtain the same number of correspondences for each candidate pose, missing correspondences need to be established. Given a certain pixel in the measured scan image, for which a range value exists, the same pixel is considered for all poses and the corresponding synthetic scan images, respectively, and if there exists a rendered range value and thus a correspondence to the measured range in at least one synthetic range image, all missing synthetic ranges are calculated as follows:

$$d' = d + c = d + \frac{\langle \mathbf{n}_m, \mathbf{t} \rangle}{\langle \mathbf{n}_m, \mathbf{v} \rangle}, \quad (4.22)$$

where d' is the missing synthetic range that is to be determined, d is the rendered range from another candidate pose, and c is the range correction. In Fig. 4.15, the approach to determine missing correspondences is visualized for two examples. The additional range c can be negative as shown in Fig. 4.15a and positive as shown in Fig. 4.15b. It is depending on the normal vector of

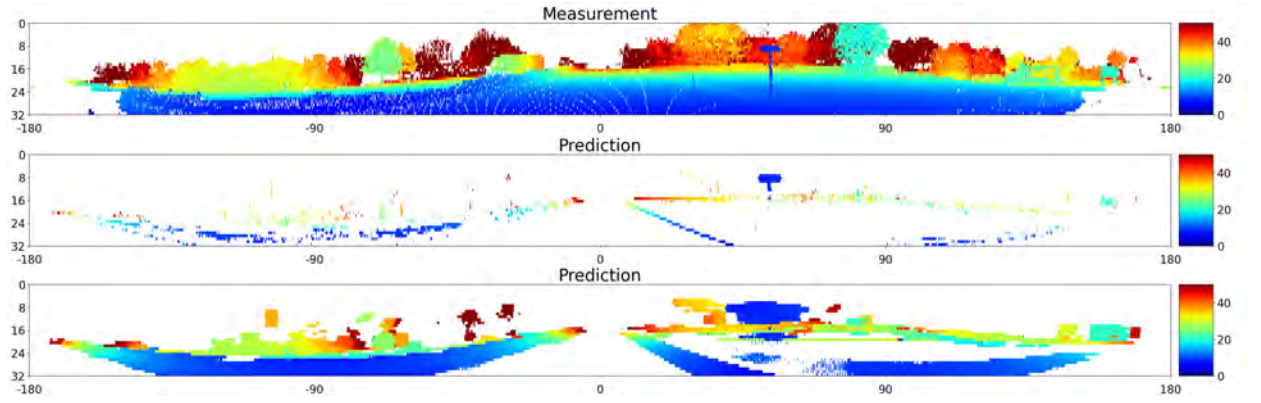


Figure 4.16.: From top to bottom: Measured range image, and a synthetic range image both before and after the determination of missing correspondences. In general, the synthetic range image without the missing correspondences appears sparse compared to the measured range image because the ground and vegetation are removed from the map and only a few static parts remain. Comparing both synthetic range images, the one after adding the missing correspondences appears to be more complete, however, it becomes clear that establishing missing correspondences also is accompanied by the drawback of false correspondences. This is especially visible at the street sign, which becomes overly large after adding missing correspondences in the synthetic scan image.

the hit map object \mathbf{n}_m , the direction of the LiDAR ray \mathbf{v} , and the shift between the two candidate poses \mathbf{t} . As the normal vector of the map is utilized, ideally, every candidate pose with a rendered range d should provide a correction c that leads to the same d' . Therefore, one pose with a rendered range d is randomly chosen to determine all missing synthetic ranges of a specific pixel.

In Fig. 4.16, a measured range image, and a synthetic range image both before and after the determination of missing correspondences are visualized from top to bottom. In general, the synthetic range image without the missing correspondences appears sparse compared to the measured range image because the ground and vegetation are removed from the map and only a few static parts remain. Comparing both synthetic range images, the one after adding the missing correspondences appears to be more complete, however, it becomes clear that establishing missing correspondences is also accompanied by the drawback of false correspondences. This is especially visible at the street sign, which becomes overly large after adding missing correspondences in the synthetic scan image.

Overall, however, the advantage outweighs, as the added correspondences improve the estimation since they yield more consistent and smoother loss distributions, which can be seen by comparing Fig. 4.17 and Fig. 4.18, which show the loss accumulators of various loss functions for the translation in the xy -plane without and after adding the missing correspondences, respectively. In general, comparing the different loss functions among each other, it becomes clear that with increasing robustness of the loss function the minimum in the loss distribution gradually moves towards the correct position at $(x, y) = (0\text{ m}, 0\text{ m})$.

4.4. State estimation filtering

In unfavorable scenarios in conjunction with an imperfect map representation, the maximum consensus localization may lead to the wrong pose estimate. Normally, such erroneous state estimates only occur sporadically or in very short time intervals, as occlusions caused by other road users disappear very quickly or the number and arrangement of static objects used for localization improve again. In order to withstand these faulty epochs and thus make localization even more

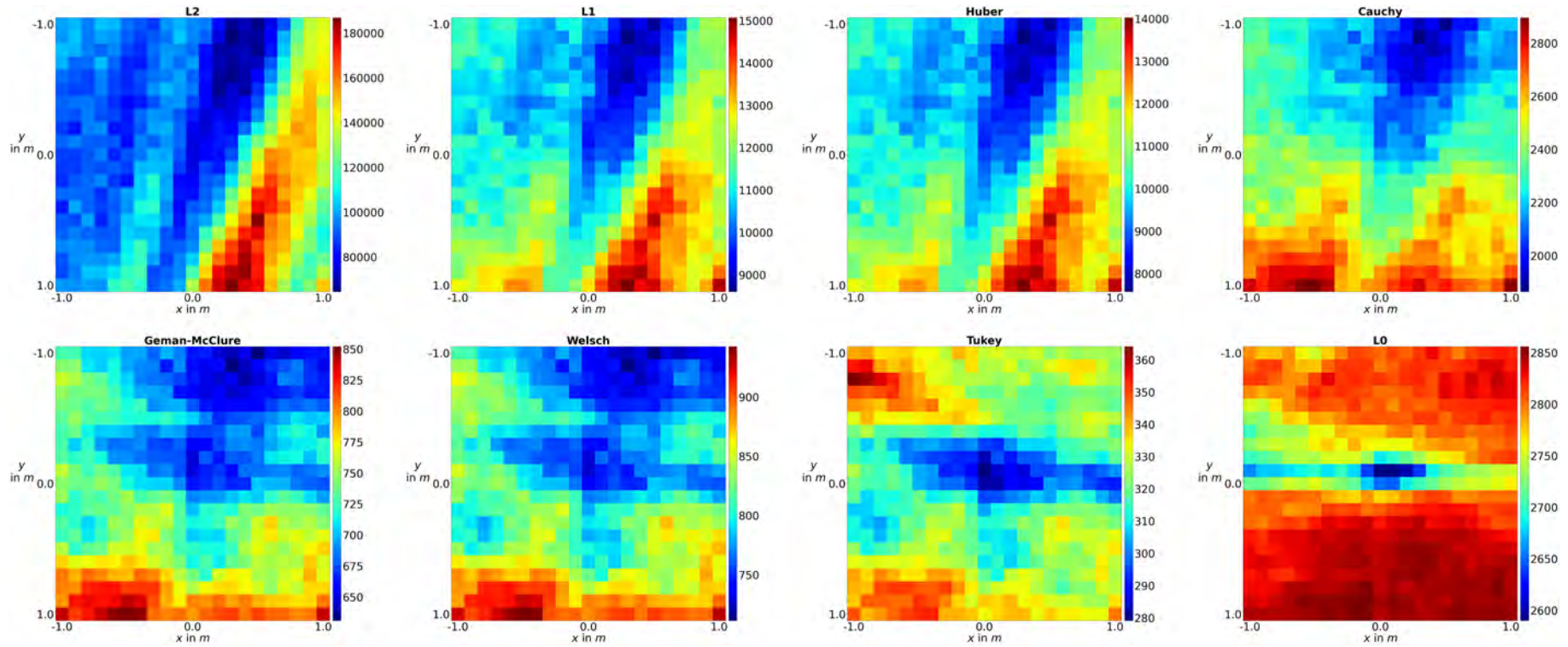


Figure 4.17.: Loss accumulators of various loss functions showing the loss distribution for the translation in the xy -plane without adding the missing correspondences. Due to the different number of correspondences that contribute to the respective loss of each candidate pose the loss distributions appear inconsistent and noisy.

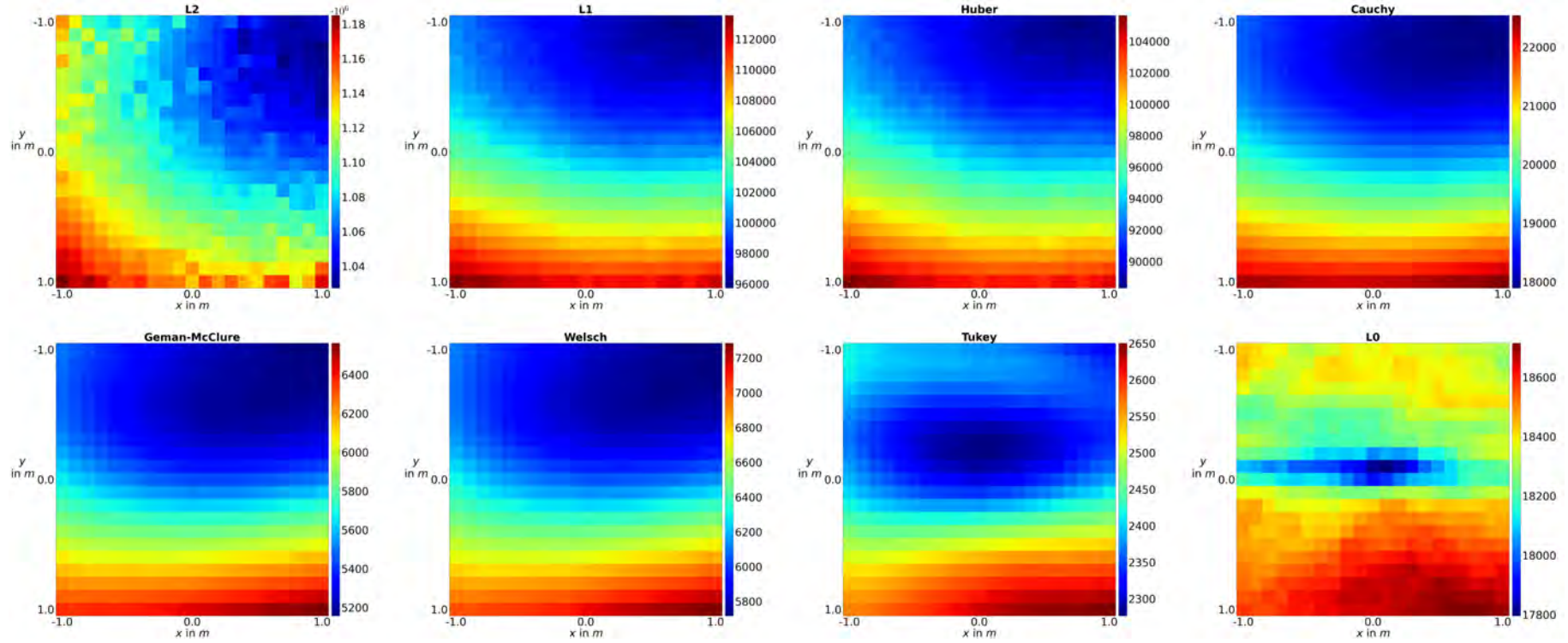


Figure 4.18.: Loss accumulators of various loss functions showing the loss distribution for the translation in the xy -plane after adding the missing correspondences. The added correspondences improve the estimation since they yield more consistent and smoother loss distributions. By comparing the different loss functions among each other, it becomes clear that with increasing robustness of the loss function the minimum in the loss distribution gradually moves towards the correct position at $(x, y) = (0 \text{ m}, 0 \text{ m})$.

robust, the independent single position estimation is transferred into a filter. The goal is to use the information from the previous epoch, which is the accumulator containing the inlier values of the full search space, and fuse it with the new inlier values of the current measurement to obtain a more robust localization result. The question is, how to do the prediction based on a state space, which is described by a set of inlier counts, while guaranteeing the adherence to the Bayesian filter rules.

In this chapter, a strategy is proposed utilizing the mathematical formulations from Barron (2017) and Chebrolu et al. (2020), based on which the maximum consensus accumulator is propagated to the next epoch, which enables the formulation of a maximum consensus filter. To this end, the maximum consensus problem is to be considered as an outlier minimization problem. Based on Eq. 2.12, the corresponding loss accumulator is determined. This accumulator contains the number of outliers for each candidate pose and is also referred to as L0 loss throughout this thesis. The fundamental idea is to transfer the loss accumulator into a discrete probability distribution, where a probability is assigned to each candidate pose, and then to make use of a standard histogram filter.

To this end, first, the work from Barron (2017) is considered, who proposes a generalized loss $\rho(r, \alpha, c)$:

$$\rho(r, \alpha, c) = \frac{|\alpha - 2|}{\alpha} \left(\left(\frac{(r/c)^2}{|\alpha - 2|} + 1 \right)^{\alpha/2} - 1 \right), \quad (4.23)$$

where α and c are the shape and the scale parameter, respectively. By changing the parameter α , different kernels can be realized, e.g. L2 loss for $\alpha = 2$, Cauchy for $\alpha = 0$, Geman-McClure for $\alpha = -2$, and Welsch for $\alpha = -\infty$, where for $\alpha = 2$, $\alpha = 0$, and $\alpha = -\infty$, the loss functions are derived by taking the limit of $\rho(r, \alpha, c)$. Based on this generalized loss, the following general probability distribution is constructed:

$$\begin{aligned} P(r, \alpha, c) &= \frac{1}{cZ(\alpha)} e^{-\rho(r, \alpha, c)}, \\ Z(\alpha) &= \int_{-\infty}^{\infty} e^{-\rho(r, \alpha, 1)} dr, \end{aligned} \quad (4.24)$$

with $cZ(\alpha)$ being the normalization term. For the probability distribution $P(r, \alpha, c)$ the corresponding loss $\rho_p(r, \alpha, c)$ is formulated as the negative log-likelihood (Chebrolu et al., 2020):

$$\begin{aligned} \rho_p(r, \alpha, c) &= -\log P(r, \alpha, c) \\ &= \rho(r, \alpha, c) + \log cZ(\alpha), \end{aligned} \quad (4.25)$$

where the loss $\rho_p(r, \alpha, c)$ corresponds to the generalized loss $\rho(r, \alpha, c)$ shifted by the log partition. Since $Z(\alpha)$ goes to infinity for $\alpha < 0$, the probability distribution $P(r, \alpha, c)$ is only defined for $\alpha \geq 0$. To realize a probability distribution, which is also defined for values of $\alpha < 0$, Chebrolu et al. (2020) propose to apply a truncation and to compute the integral only within the limits $[-\tau, \tau]$, which leads to the finite partition function $\tilde{Z}(\alpha)$:

$$\tilde{Z}(\alpha) = \int_{-\tau}^{\tau} e^{-\rho(r, \alpha, 1)} dr, \quad (4.26)$$

with τ being the truncation limit. The truncated probability distribution $\tilde{P}(r, \alpha, c)$ and the truncated loss $\tilde{\rho}_p(r, \alpha, c)$ are then defined accordingly:

$$\tilde{P}(r, \alpha, c) = \frac{1}{c\tilde{Z}(\alpha)} e^{-\rho(r, \alpha, c)}, \quad (4.27a)$$

$$\begin{aligned} \tilde{\rho}_p(r, \alpha, c) &= -\log \tilde{P}(r, \alpha, c) \\ &= \rho(r, \alpha, c) + \log c\tilde{Z}(\alpha). \end{aligned} \quad (4.27b)$$

In (Chebrolu et al., 2020), the truncation is introduced to be able to use not only positive but also negative α values for the adaption of the shape of the loss function. In this work, the truncated probability distribution and truncated loss function are the key to represent the loss accumulator containing the count of outliers as discrete probability distribution and thereby create the basis for the maximum consensus histogram filter. The histogram filter has the advantages that it considers all potential states within the pre-defined state space and it is able to represent arbitrary distributions as well as multiple hypotheses.

For the maximum consensus filter, in the first step, the loss ρ , which corresponds to the grid of outlier counts, is represented as discrete truncated probability distribution \tilde{P} according to Eq. 4.27a. Basically, the utilization of the truncated probability distribution allows to use any loss in the filter. Thus, the different loss functions considered in the robust kernel comparison (cf. Sec. 4.3) can also be used in the filter and compared with respect to their robustness and accuracy.

In this work, instead of utilizing additional sensors for the prediction of the pose, a constant velocity model is applied, which has been shown to be sufficiently accurate (Vizzo et al., 2023). To model the uncertainty of the prediction and to obtain the predicted belief $\overline{bel}(\mathbf{x}_t)$, the Bayes filter prediction step is applied using the prior belief $bel(\mathbf{x}_{t-1})$ and the state transition probability $p(\mathbf{x}_t|\mathbf{u}_t, \mathbf{x}_{t-1})$:

$$\overline{bel}(\mathbf{x}_t) = \sum_{\mathbf{x}_{t-1}} p(\mathbf{x}_t|\mathbf{u}_t, \mathbf{x}_{t-1}) \cdot bel(\mathbf{x}_{t-1}), \quad (4.28)$$

where $\overline{bel}(\mathbf{x}_t)$ and $bel(\mathbf{x}_{t-1})$ are discrete probability distributions and $p(\mathbf{x}_t|\mathbf{u}_t, \mathbf{x}_{t-1})$ is modeled using a Gaussian kernel with $\sigma = 5 \text{ cm}$.

The update step can either be conducted in the probability space by multiplying the predicted truncated probability distribution with the newly measured truncated probability distribution or in the loss space by adding the predicted truncated loss to the newly measured truncated loss. In both cases, the newly measured scan is transformed into the predicted pose to determine the truncated probability distribution or the truncated loss needed for the update. The transition from the probability space into the loss space is done according to Eq. 4.27b. In general, the update is performed in the probability space, since there is no benefit of transitioning back into the loss space. Certainly, the multiplication of the probabilities yields exactly the same results as the addition of the losses.

In Fig. 4.19, the methodology of the maximum consensus filter is visualized. The purpose of the chart is merely to illustrate the relationship between the truncated loss and the truncated probability space, and it is not to be understood as a graph of a dynamic Bayes network as shown in Fig. 2.6. Using the relations $\tilde{P} = e^{-\tilde{\rho}}$ and $\tilde{\rho} = -\log \tilde{P}$, a truncated loss accumulator $\tilde{\rho}$ can be represented as a truncated probability distribution \tilde{P} , and vice versa. When a new measurement, and thus a new loss accumulator ρ , is acquired, for the update step it must be transferred either into the truncated probability space using the relationship given in Eq. 4.27a or into the truncated loss space using the relationship given in Eq. 4.27b, which are illustrated in simplified terms by $\frac{1}{\tilde{Z}}e^{-\rho}$ and $\rho + \log \tilde{Z}$ in Fig. 4.19.

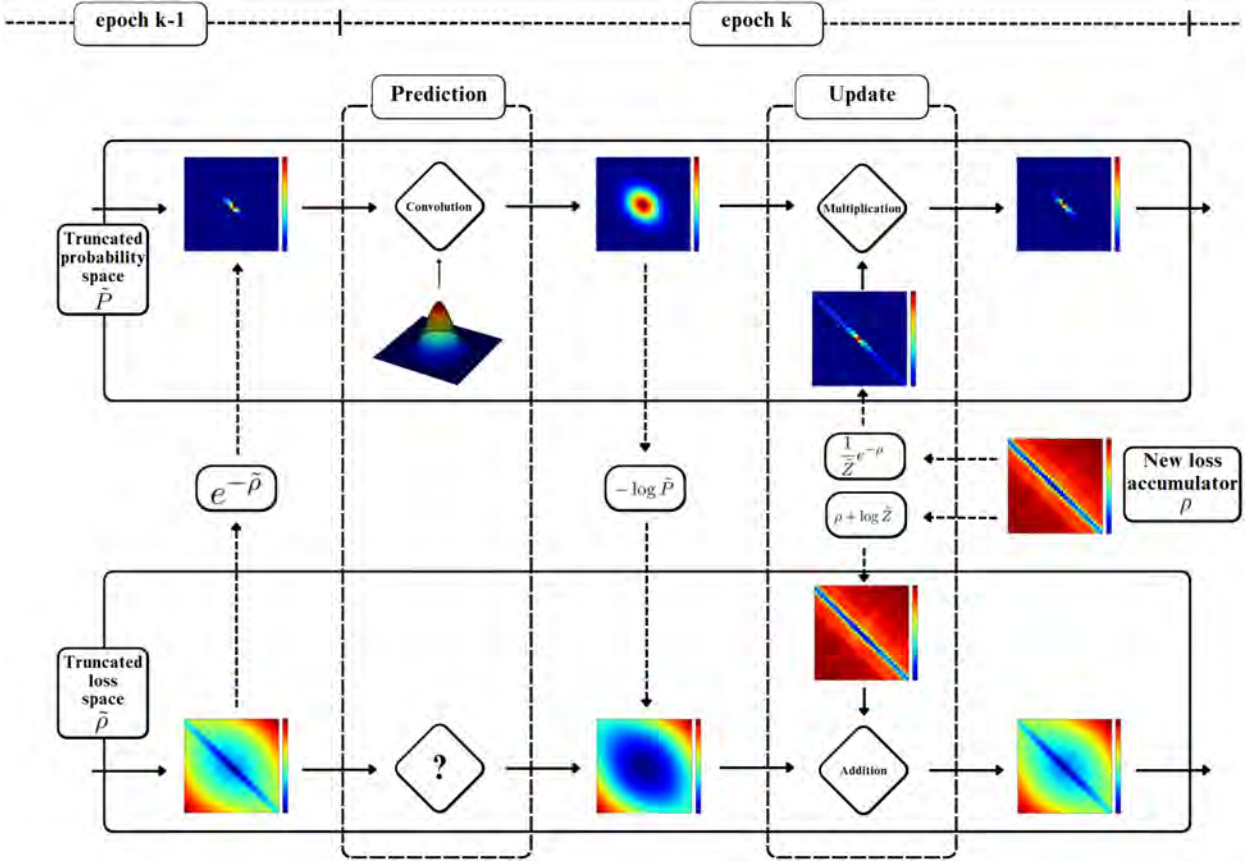


Figure 4.19.: Methodology of the maximum consensus filter: Using the relations $\tilde{P} = e^{-\tilde{\rho}}$ and $\tilde{\rho} = -\log \tilde{P}$, a truncated loss accumulator $\tilde{\rho}$ can be represented as a truncated probability distribution \tilde{P} , and vice versa. When a new measurement, and thus a new loss accumulator ρ , is acquired, for the update step it must be transferred either into the truncated probability space using the relationship given in Eq. 4.27a or into the truncated loss space using the relationship given in Eq. 4.27b, which are illustrated in simplified terms by $\frac{1}{\tilde{Z}}e^{-\rho}$ and $\rho + \log \tilde{Z}$.

To prevent numerical instability the *exp-normalize trick* is applied, which is typically used when dealing with softmax probabilities in machine learning. It is defined as:

$$p_i = \frac{\exp(x_i)}{\sum_{j=1}^n \exp(x_j)} = \frac{\exp(x_i - b)}{\sum_{j=1}^n \exp(x_j - b)}, \quad (4.29)$$

where b is typically selected as $b = \max_{i=1}^n x_i$. Subtracting the maximum value from each element before exponentiating does not affect the relative relationships between the calculated probabilities and allows to handle a vector of large negative values, such as the negated loss accumulators in the context of this work. Applying the exp-normalize trick to Eq. 4.26 and Eq. 4.27a, which means that $x = -\rho(r, \alpha, c)$, results in the following:

$$\begin{aligned} \tilde{P}(r, \alpha, c) &= \frac{1}{c\tilde{Z}(\alpha)} e^{(-\rho(r, \alpha, c) - \max(-\rho(r, \alpha, c)))}, \\ \tilde{Z}(\alpha) &= \int_{-\tau}^{\tau} e^{-\rho(r, \alpha, 1) - \max(-\rho(r, \alpha, 1))} dr. \end{aligned} \quad (4.30)$$

So far, the standard deviation σ of the Gaussian kernel applied in the prediction is the only parameter of the filter. Since the correlations between the individual LiDAR measurements are still

neglected and not taken into account, although they have a significant influence on the probability distribution, an additional parameter is introduced. The probability distribution from the measurement depends on the absolute differences in the number of outliers. As a LiDAR sensor often acquires many thousand scan points in one 360° sweep, the absolute differences between the loss values of the candidate poses are usually large, which leads to very sharp probability distributions. Even between the smallest and second smallest loss, the difference can be so large that, in the extreme case, one candidate pose has a probability of 1 while all other candidate poses have a probability of 0. Considering the fact that simply changing the LiDAR model from 32 to 16 or from 32 to 64 scan layers would half or double the absolute differences and therefore significantly influence the probabilities, even though the relative loss values did not change, shows that there is an additional parameter required. Actually, the correlations between the measurements must be determined, where two fully correlated measurements would act as one measurement, however, as this is not trivial, the problem is simplified by introducing an additional parameter and testing its influence for different values. The new parameter is referred to as the *correlation quotient*, and it defines the ratio between the real measurements and the so-called *effective measurements*. For example, with 10000 real measurements and a correlation quotient of 10, the number of effective measurements is 1000. In the context of this work, the correlation quotient is applied to the loss accumulator of the measurement, i.e. all loss values are divided by the correlation quotient, which roughly has the same effect as uniformly reducing the number of measurements. The influence of the correlation quotient on the performance of the filter is shown in the experiments in Sec. 5.3.3. As the parameter significantly affects the shape of the probability distribution, it will also be important for the protection level estimation, which is also based on the discrete probability distributions and introduced in the next section.

4.5. Protection level estimation and integrity evaluation

In this section, a novel formulation of protection levels is proposed with respect to the concept of integrity (cf. Sec. 2.5). The goal is to utilize the full knowledge of the search space to define a parameter subspace, where it is guaranteed up to a certain probability that the vehicle pose is within these parameter bounds. Reid et al. (2019) define an integrity risk of 10^{-8} failures per hour of operation. This is taken as a reference point and a probability of $p > 1 - 10^{-8}$ is used to determine the protection level in this work, where it is defined by the set of candidate poses with a probability $p > 1 - 10^{-8}$.

Since the protection levels cannot be determined directly based on the accumulator of inlier counts, the first step to estimate a protection level is to calculate a probability for each candidate pose. There are two strategies to obtain such a discrete probability distribution, either by utilizing the truncated probability distribution $\tilde{P}(r, \alpha, c)$ (cf. Eq. 4.27a) or by utilizing the beam model from Thrun et al. (2005) (cf. Sec. 2.1.4), both of which are investigated in this work. Using the truncated probability distribution $\tilde{P}(r, \alpha, c)$ enables to define a protection level based on any loss function, however, as the focus is on maximum consensus in this work, the L0 loss is considered here (cf. Eq. 2.12). Since the beam model is applied to the range measurements in the observation space, the protection levels based on the L0 loss are also calculated based on correspondences in the observation space for the purpose of comparability. Furthermore, as will become clear from the experiments in Sec. 5.3, the localization errors are larger for the localization based on correspondences in the observation space, which offers a wider range of cases for the evaluation of the protection level estimation.

In Fig. 4.20, for a search space from -1 m to $+1\text{ m}$ for the translation in x - and y -directions and from -0.72° to $+0.72^\circ$ for the heading θ , a truncated probability distribution based on the L0 loss

is visualized using a logarithmic representation. In the estimation of the probability distribution, a correlation quotient of 10 was applied, where more information regarding the influence of the correlation quotient on the protection level estimation are provided at the end of this section. The white borders contain all candidate poses with a probability of $p > 1 - 10^{-8}$, i.e. in this example seven candidate poses with a heading from -0.18° to $+0.18^\circ$. As the integrity concept from Reid et al. (2019) only considers one alert limit for each axis, where it does not differentiate between positive and negative directions, for each axis only one protection level value is to be determined, and as the true position is unknown in a real application, the size of the protection is defined as the maximum extent of candidate poses with a probability of $p > 1 - 10^{-8}$ in longitudinal, lateral, and heading directions starting from the estimated localization solution, which is the candidate pose with the highest probability. In Fig. 4.20, this is the candidate pose at $(x, y, \theta) = (0.2\text{ m}, 0.0\text{ m}, 0.0^\circ)$. Consequently, in longitudinal direction, which corresponds to the x -axis, the size of the protection level, which is later compared to the alert limit, is 0.2 m . In lateral direction, the size of the protection level is 0, which means that along this axis only the localization solution is contained in the protection level, and in heading direction, the size of the protection level is 0.18° . In Fig. 4.20, the green borders show the protection level, which is the maximum extent of 0.2 m and 0.18° along the longitudinal and heading axis, respectively, each applied in both directions starting from the localization solution. If the alert limit were defined in a more sophisticated way, e.g. by distinguishing between positive and negative directions along each axis or by defining a polygon taking into account information from the map or other road users, then the obtained polygon containing all candidate poses with a probability of $p > 1 - 10^{-8}$ would be of greater benefit. However, for now only one alert limit value is considered along each axis, against which the protection level is compared.

By using the beam model from Thrun et al. (2005), basically a data-adapted loss function is employed. In this work, the intrinsic beam model parameters Θ , which are σ_{hit} , λ_{short} , ω_{hit} , ω_{short} , ω_{random} , and ω_{max} , are estimated from the measured scan data (cf. Sec. 2.1.4). To this end, besides the set of measured ranges $\mathbf{Z} = \{z_i\}$, the set of reference poses $\mathbf{X} = \{\mathbf{x}_i\}$, in which the measurements were carried out, and the map \mathcal{M} , based on which the expected ranges are rendered, are required. The reference poses and the map are obtained from the ground truth trajectory and point clouds both acquired by the RIEGL MMS, where the final map representation for the rendering of the expected ranges is a triangle mesh. To account for varying environmental conditions, for each driving scenario an individual set of beam model parameters is estimated, however, strong short term deviations from the final beam model, e.g. caused by dynamic traffic participants, still remain. Detailed information on the used data and on the estimated intrinsic parameters are given in Sec. 5.2 and Sec. 5.3.2, respectively.

To calculate the probability of each candidate pose, the first step is to derive the probabilities for the actual measurements given the expected measurements, denoted as $p(z_i|\mathbf{x}, \mathcal{M})$, from the beam model. Here, z_i represents a measured range, while \mathbf{x} and \mathcal{M} represent the candidate pose and map, respectively, which are required to determine the expected range through ray casting. Then, the discrete probability distribution over all candidate poses $p(\mathbf{x}|\mathbf{Z}, \mathcal{M})$ is computed as follows:

$$p(\mathbf{x}|\mathbf{Z}, \mathcal{M}) = \frac{p(\mathbf{Z}|\mathbf{x}, \mathcal{M}) \cdot p(\mathbf{x}|\mathcal{M})}{\sum_{\tilde{\mathbf{x}}} p(\mathbf{Z}|\tilde{\mathbf{x}}, \mathcal{M}) p(\tilde{\mathbf{x}}|\mathcal{M})}, \quad (4.31)$$

with $p(\mathbf{Z}|\mathbf{x}, \mathcal{M}) = \prod_i p(z_i|\mathbf{x}, \mathcal{M})$,

where $p(\mathbf{Z}|\mathbf{x}, \mathcal{M})$ denotes the probability of all real measurements, under the assumption of independence, given a particular candidate pose \mathbf{x} , and $p(\mathbf{x})$ denotes the prior, where in this work $p(\mathbf{x}) \neq 1$ only if the approach is implemented in a filter and $p(\mathbf{x}) = 1$ if the approach considers a single pose independently.

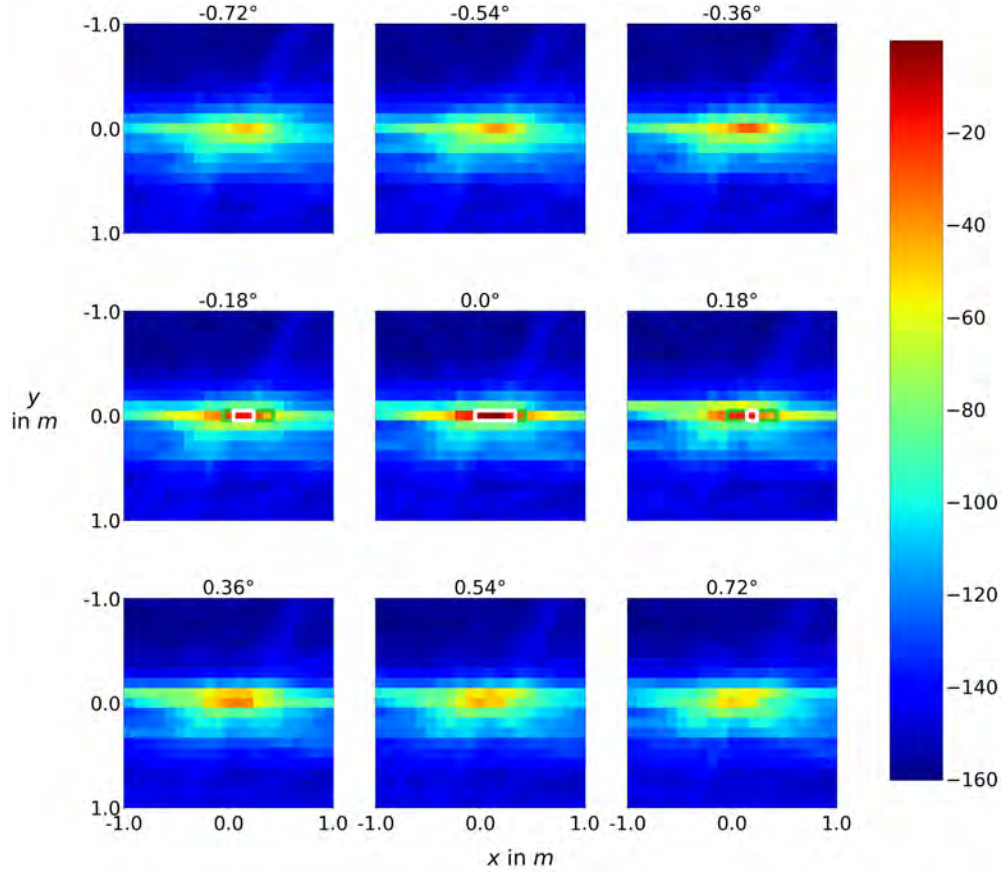


Figure 4.20.: Truncated probability distribution based on the L0 loss using a logarithmic representation for a search space from -1 m to $+1\text{ m}$ for the translation in x - and y -directions and from -0.72° to $+0.72^\circ$ for the heading θ . The white borders contain all candidate poses with a probability of $p > 1 - 10^{-8}$, i.e. in this example seven candidate poses with a heading from -0.18° to $+0.18^\circ$. The green borders show the protection level being the maximum extent of candidate poses with a probability of $p > 1 - 10^{-8}$ applied in both directions along each axis starting from the localization solution.

For the maximum consensus accumulator shown in Fig. 4.21a, the corresponding discrete probability distribution based on a beam model is shown using logarithmic representation in Fig. 4.21b. The used beam model parameters are stated in Tab. 5.13, where the example is from the data set *Haltenhoffstraße*. Similar to Fig. 4.20, the search space reaches from -1 m to $+1\text{ m}$ for the translation in x - and y -directions and from -0.72° to $+0.72^\circ$ for the heading θ . The white borders contain all candidate poses with a probability of $p > 1 - 10^{-8}$, i.e. in this example eight candidate poses with a heading from -0.18° to $+0.18^\circ$. The green borders show the protection level, which is the maximum extent of 0.1 m , 0.1 m , and 0.36° along the longitudinal, lateral, and heading axis, respectively, each applied in both directions starting from the localization solution, which occurs at $(x, y, \theta) = (0.1\text{ m}, 0\text{ m}, 0.18^\circ)$. As can be seen, the pose with the highest probability based on the beam model deviates from the pose with the highest consensus and therefore also from the pose with the highest probability based on the L0 loss shown in Fig. 4.21a and Fig. 4.20, respectively, which occur at $(x, y, \theta) = (0.2\text{ m}, 0\text{ m}, 0^\circ)$. This illustrates the fact that both probability distributions are based on different assumptions about the distribution of the measurements, however, for a valid probability distribution the underlying model has to be in accordance with the distribution of the measurements. The shown white borders claim to contain the true pose for a probability of $p > 1 - 10^{-8}$, however, in Fig. 4.20 under the assumption that the measurements follow a distribution for which the negative log-likelihood corresponds to the L0 loss function, and in Fig. 4.21b

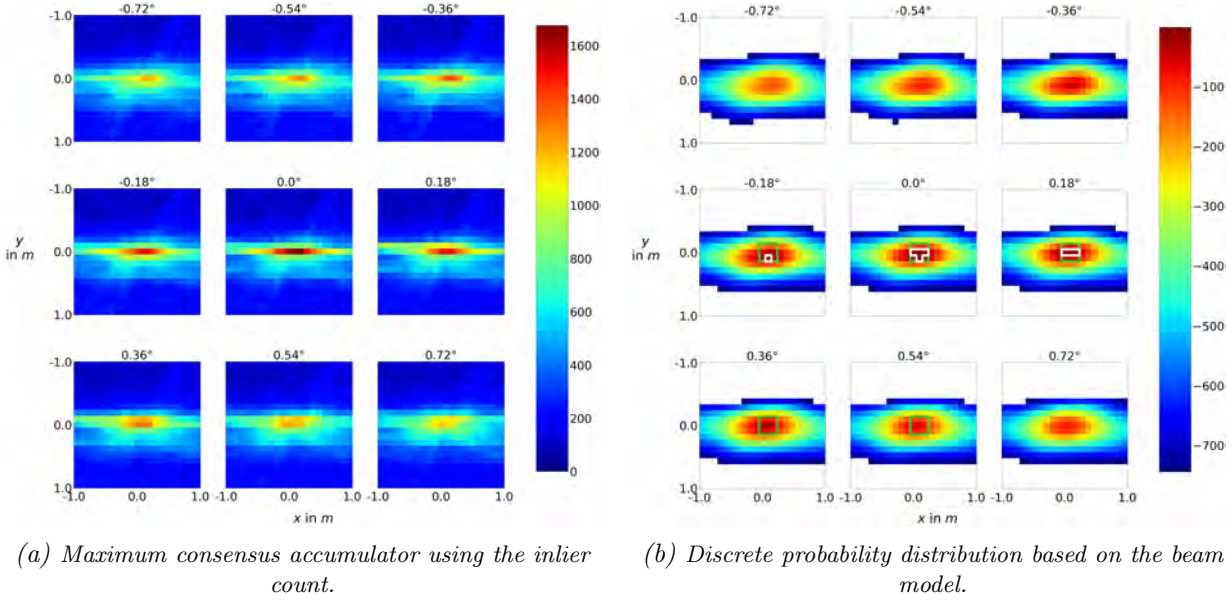


Figure 4.21.: For the maximum consensus accumulator shown in (a), the corresponding discrete probability distribution based on the beam model is shown using logarithmic representation in (b). The search space reaches from -1 m to $+1\text{ m}$ for the translation in x - and y -directions and from -0.72° to $+0.72^\circ$ for the heading θ . The white borders contain all candidate poses with a probability of $p > 1 - 10^{-8}$, i.e. in this example eight candidate poses with a heading from -0.18° to $+0.18^\circ$. The green borders show the protection level being the maximum extent of candidate poses with a probability of $p > 1 - 10^{-8}$ applied in both directions along each axis starting from the localization solution.

under the assumption that the measurements follow the beam model. In general, from Fig. 4.21 it can be seen that taking the maximum extent of candidate poses with a probability of $p > 1 - 10^{-8}$ along each axis as protection level is overly pessimistic. Hence, a more differentiated definition of the alert limit is desirable, against which the exact, white polygon of the protection level can be compared.

In general, instead of the truncated probability distribution based on any loss function, the probability distribution based on a beam model can also be used in the histogram filter proposed in Sec. 4.4. Moreover, the beam model likelihood distribution $p(z|\mathbf{x}, \mathcal{M})$ can be interpreted as a data-adapted loss function, with $\rho(z|\mathbf{x}, \mathcal{M}) = -\log p(z|\mathbf{x}, \mathcal{M})$, and be included in the comparison of loss functions in Sec. 4.3. The loss accumulator based on the beam model can be assessed and compared to the ones from the other loss functions regarding the localization accuracy and robustness, which are determined based on the pose with the minimum loss, and the localization reliability, which is based on the distribution within the accumulator.

Given the probability distributions and the derived protection levels, the overall objective is to assess the integrity of the localization system using the four operational states *Nominal Operation* (NO), *Unavailable* (UA), *Misleading Information* (MI), and *Hazardously Misleading Information* (HMI) from the Stanford diagram. The navigation algorithm should be designed to provide as little misleading information as possible, and preferably no hazardously misleading information, while at the same time entering the unavailable state as rarely as possible. As was discussed in Sec. 4.4, the probability distributions are generally calculated without taking correlations into account, which makes them overly optimistic. To compensate for that, the correlation quotient has been introduced, which is a simplification compared to the complex determination of correlations and which leads to broader and flatter, and therefore more conservative, probability distributions. Therefore, the correlation quotient directly affects the protection level estimation and the size of

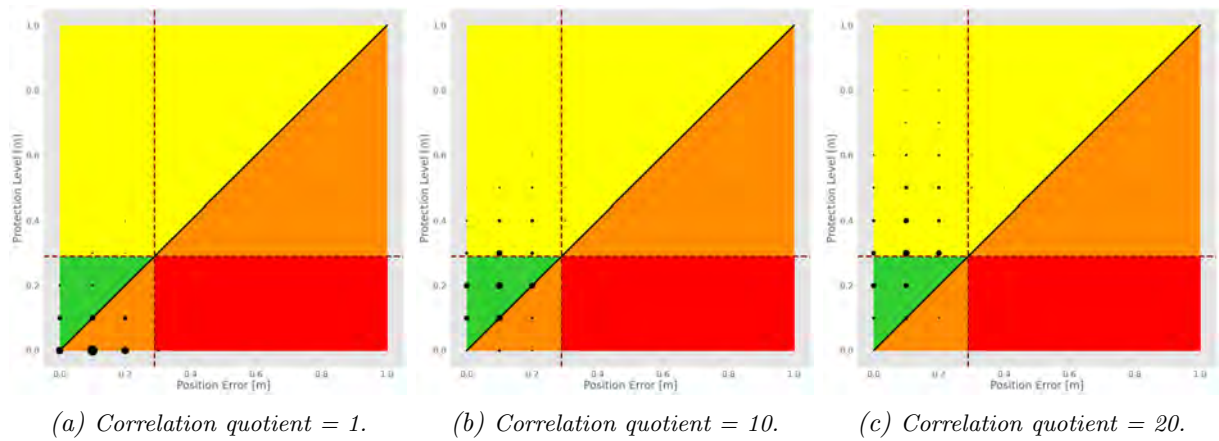


Figure 4.22.: Influence of the correlation quotient on the distribution between the four operational states in the Stanford diagram. Due to the discrete nature of the results, the point size is used to encode the number of epochs belonging to each protection level - position error - pair. The larger the correlation quotient, the larger are the estimated protection levels and the more conservative is the integrity system. This leads to more epochs of the state *UA*, but also to less epochs of the states *MI* and *HMI*.

the protection level, which is why it can be considered as a tuning parameter of the integrity system. Consequently, it defines the ratios between the four operational states *NO*, *UA*, *MI*, and *HMI* in the Stanford diagram. For example, if the correlation quotient is chosen to be small, the probability distribution will be highly optimistic and therefore the protection level will be small, resulting in an aggressive integrity system with fewer epochs in the operational state *UA* and more epochs in the states *MI* and *HMI*. In contrast, if the correlation quotient is chosen to be large, the probability distribution will be highly pessimistic and therefore the protection level will be large, resulting in a conservative integrity system with more epochs in the operational state *UA* and fewer epochs in the states *MI* and *HMI*. This relationship is illustrated in Fig. 4.22, which shows the Stanford diagrams of the same test scenario using three different correlation quotients. Whereas the sizes of the position errors do not change, the sizes of the protection levels change and so do the distributions within the Stanford diagrams.

5. Experiments

In this chapter, the proposed methods of the robust and reliable localization pipeline are evaluated using data from real world driving scenarios. To this end, at first, the objectives and corresponding evaluation metrics are presented in Sec. 5.1, and the used data sets are introduced in Sec. 5.2. Afterwards, in Sec. 5.3, for each method the examination results are presented, which forms the main part of this chapter. Finally, in Sec. 5.4, the results are summarized and discussed.

5.1. Objectives and evaluation metrics

In this section, the objectives of the experiments are introduced, and the corresponding evaluation metrics, which serve to fulfill the objectives, are explained. The objectives of the evaluation are defined for the individual parts of the localization pipeline and set the structure for the results presented in Sec. 5.3. In general, the goal is to determine both strengths and weaknesses or limitations of the localization framework. The evaluations are conducted using three different data sets, which have been acquired in an urban, a suburban, and a highway environment. Based on their amount of map objects for the localization, the driving scenarios take place in a structured, semi-structured, and unstructured environment, respectively, and the three data sets enable to assess the localization performance with respect to different levels of difficulty. In Sec. 5.2, the data sets are introduced in detail.

The objectives comprise the following aspects:

1. Evaluate and compare the inlier count and point-to-plane adjustment score of the maximum consensus localization with respect to accuracy and robustness as well as reliability, while using other state-of-the-art localization techniques for comparison. Show whether and if so where the point-to-plane adjustment score yields a better localization performance than the inlier count.
2. Evaluate and compare the different loss functions (cf. Sec. 4.3) for the purpose of xy -position estimation. Similar to the count and point-to-plane adjustment score, do the examination with respect to accuracy and robustness as well as reliability. Besides the well-known M-estimators, evaluate if a data-adapted loss function based on the beam model from Thrun et al. (2005) is suitable for vehicle localization.
3. Evaluate the maximum consensus filter based on the inlier count and compare it to the single position estimation with respect to accuracy and robustness. Analyze how the number of effective measurements affects the localization performance of the filter.
4. Estimate the protection level using the beam model from Thrun et al. (2005) and the truncated probability distribution based on the L0 loss, and evaluate the localization integrity with respect to the four operational states from the Stanford diagram. Show how the size of the protection level can be adjusted and thereby the ratio of the four operational states be shifted by changing the number of effective measurements.

Whereas the maximum consensus localization is evaluated considering both the xy -position and the heading, the different loss functions and also the maximum consensus filter are investigated

only considering the xy -position. For the sake of simplicity, the heading is excluded, which does not impair the methodological findings of those investigations, as they still become very clear.

Since the maximum consensus localization is an exhaustive search which provides the globally optimal solution with respect to the objective function, it is independent on the initialization and only requires that the true pose is contained within the search space. Therefore, the ground truth pose can be utilized as initial pose, which is used to transform the car sensor scan into the global coordinate system and which also represents the center of the search space. Then, the correct pose is found if the localization solution is at $(x, y, \theta) = (0\text{ m}, 0\text{ m}, 0^\circ)$. This is done for the single pose estimation using the inlier count, the point-to-plane adjustment score, and the different loss functions, but not for the maximum consensus filter where the initial position corresponds to the predicted position.

For the purpose of comparison, the following three approaches are used: a point-to-plane ICP, the KISS-ICP proposed by Vizzo et al. (2023), and the Range-MCL proposed by Chen et al. (2021). While the point-to-plane ICP and KISS-ICP are based on 3D points, in the Range-MCL algorithm, for a fixed number of particles synthetic range images are rendered in a mesh and compared with the range image of the measurement to determine the resampling weights of the particles. Accordingly, the point-to-plane ICP and KISS-ICP can be considered as algorithms operating in Euclidean space, while the Range-MCL can be considered as an algorithm operating in observation space. The point-to-plane ICP uses a quadratic loss function and minimizes point-to-plane errors, the KISS-ICP uses the robust Geman-McClure loss function and minimizes point-to-point errors, and the Range-MCL uses the mean of the absolute pixel-wise differences between the synthetic and the measured range image to determine the particle weight. For the point-to-plane ICP, the crucial parameter to be set is the inlier threshold in the phase of establishing correspondences. It heavily affects the registration result, which is why it is set to 0.5 m and 1 m in this work in order to generate different results depending on this parameter.

The KISS-ICP method, referred to by the author as a LiDAR odometry approach, originally does not require a pre-known map and instead builds a map by accumulating the aligned LiDAR scans in a voxel grid. To obtain a localization method, the algorithm is adapted that it uses a provided, pre-known map point cloud instead of creating one. However, the voxelization strategy for the map is maintained, with the first N_{max} points stored in each voxel, where N_{max} is a fixed number set to the default value of 20. A second modification is that the algorithm takes an initial pose as input instead of predicting it using a constant velocity model, and the third change is that the adaptive distance threshold for the outlier rejection is replaced by a fixed threshold. Thus, the robust registration part based on the Geman-McClure kernel is extracted from the original LiDAR odometry framework and used to register car sensor scans to the pre-known map point cloud and thereby determine the vehicle pose. The only parameter to be set is the σ , which defines the outlier rejection threshold $\tau = 3\sigma$ as well as the scale parameter $\kappa = \sigma/3$ of the Geman-McClure kernel and is set to $\sigma = 0.5\text{ m}$. For both the point-to-plane ICP and KISS-ICP, to investigate the accuracy of the alignment the ground truth pose is used as initial pose, and to investigate the robustness in terms of the ability for convergence in case of bad initialization the ground truth pose with an added initialization error is used as initial pose. Since a large initial error inevitably leads to the failure of the ICP, two tests are carried out in which the initialization error is set to $+0.5\text{ m}$ and $+1\text{ m}$, each time along both the x - and y -axes.

The Range-MCL is a Monte Carlo localization approach that originally first does a global localization by initializing particles over the entire map and then carries out a pose tracking using a classical particle filter. For the comparison in this work, it is modified to discard the particle filter and instead newly initialize particles in each epoch around the ground truth pose with an error for the xy -position and heading up to the size of the search space of the maximum consensus

localization. The number of particles is set to the number of candidate poses from the maximum consensus localization, e.g. $21 \cdot 21 \cdot 9 = 3969$ for an xy -grid of 21×21 candidate positions and nine heading angles, and the particle with the highest weight is considered as the localization solution. Overall, the three algorithms selected for comparison consider different types of correspondences, different types of loss functions, and different types of localization strategies.

So far, the measures *accuracy*, *robustness*, and *reliability* have been used to formulate the evaluation objectives. In the following, the evaluation metrics are defined which are used to assess and quantify such measures in this work:

1. The accuracy is evaluated using the Root Mean Square Error (RMSE), which is determined with respect to the RIEGL ground truth trajectory while considering all epochs of a driving scenario. For the translation and heading, it is calculated as follows:

$$RMSE_{xy} = \sqrt{\frac{\sum_{i=1}^N \|\mathbf{t}_{xy,i}^* - \mathbf{t}_{xy,i}\|_2^2}{N}}, \quad RMSE_{\theta} = \sqrt{\frac{\sum_{i=1}^N |\theta_i^* - \theta_i|^2}{N}}, \quad (5.1)$$

where $\mathbf{t}_{xy,i}^*$ and θ_i^* are the ground truth translation and heading, respectively, $\mathbf{t}_{xy,i}$ and θ_i are the estimated translation and heading, respectively, and N is the number of epochs.

2. The robustness is evaluated by analysing the frequency of localization failures over all epochs. To define a failure, the alert limit for passenger vehicles on US local roads proposed by Reid et al. (2019) is used, which is 0.29 m in the longitudinal as well as lateral directions and 0.5° for the heading. In this work, the alert limit for the translation is simplified to an overall error of 0.29 m , without separating along longitudinal and lateral direction. The failure rates fr_{xy} and fr_{θ} for the translation and heading, respectively, are then defined as:

$$fr_{xy} = \frac{\sum_{i=1}^N (\mathbb{I}(\|\mathbf{t}_{xy,i}^* - \mathbf{t}_{xy,i}\|_2 > AL_{xy}))}{N}, \quad fr_{\theta} = \frac{\sum_{i=1}^N (\mathbb{I}(|\theta_i^* - \theta_i| > AL_{\theta}))}{N}, \quad (5.2)$$

where AL_{xy} and AL_{θ} are the alert limits for the xy -position and heading, respectively. In this work, the goal is to reduce the significant localization errors by using the robust maximum consensus localization, thus a special focus is on the robustness.

3. The reliability of an estimate is determined by the observations. In this work, the 2D accumulators reflect the observations, and therefore the reliability is evaluated based on the distribution within the accumulators. The fundamental idea is that a wide and flat distribution indicates low reliability, while a narrow and distinct peak indicates high reliability. To analyze the distribution within the accumulator and evaluate the reliability of the localization, two measures are used. The first measure is the kurtosis, which describes the tailedness of a distribution. There exist two definitions from Pearson and Fisher, where the latter one is also referred to as *excess kurtosis*. They only differ by a constant offset of three and are defined as follows:

$$\text{Pearson's definition: } \beta_2 = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma} \right)^4 \quad (5.3)$$

$$\text{Fisher's definition: } \gamma = \beta_2 - 3$$

The parameter x_i denotes the inlier count, point-to-plane adjustment score, or loss value of a candidate pose. In this work, Fisher's definition is used where the normal distribution has a kurtosis of zero, which is referred to as 'mesokurtic' distribution. A flat, rather uniform distribution is called 'platykurtic' and has a negative kurtosis value, while a sharp distribution with a distinct maximum is called 'leptokurtic' and has a positive kurtosis value. In Fig. 5.1,

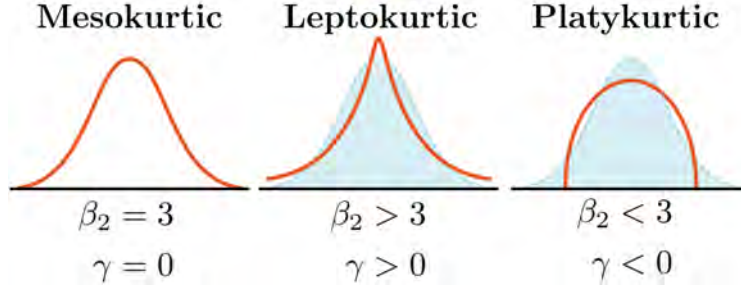


Figure 5.1.: Kurtosis values according to Pearson’s (β_2) and Fisher’s (γ) definition for mesokurtic, leptokurtic and platykurtic distributions. The illustration is adapted from Hemmerich (2020).

mesokurtic, leptokurtic and platykurtic distributions are illustrated, and the corresponding kurtosis values according to Pearson’s (β_2) and Fisher’s (γ) definition are given. To calculate the kurtosis for a loss accumulator, the loss accumulator is multiplied by -1 that the minimum loss becomes the maximum value.

The second measure is the value of the second largest consensus Ψ_2 relative to the largest consensus Ψ_{max} , which is referred to as *second peak ratio* $r_{secondpeak}$ and defined as follows:

$$r_{secondpeak} = \frac{\Psi_2}{\Psi_{max}} \quad (5.4)$$

It can be similarly applied to a loss accumulator by taking the smallest and second smallest loss values. The idea is that the smaller the second peak ratio, the more unique is the registration solution and thus the more reliable is the localization.

Since both metrics are based on the distribution within the accumulator, the reliability is only evaluated for the maximum consensus localization using count and score and the localization using the various loss functions, but not for the comparison algorithms ICP, Kiss-ICP, and Range-MCL.

5.2. Used sensors and data

In this section, the three data sets used for the experiments are introduced. They all have the same structure as they all contain a map point cloud, a reference trajectory, and a set of car sensor scans. However, the environments where they were acquired significantly differ from each other. The first driving scenario took place in the *Nordstadt* of Hannover, Germany, where the environment is characterized by a dense building structure and many pole-like objects, the second driving scenario was conducted in the *Haltenhoffstraße* in Hannover, Germany, where the environment is characterized by many trees along the street and comparatively few buildings, and the third driving scenario happened on the *Autobahn 2* and *Bundesstraße 6* (interchange *Hannover-Herrenhausen*) in Hannover, Germany, where the environment is characterized by guardrails, few traffic signs, and many other traffic participants, such as cars and trucks. According to the amount of static objects that can be used for vehicle localization, the three environments are categorized as structured, semi-structured, and unstructured, respectively. In Fig. 5.2, from each data set a part of the trajectory is visualized on top of an orthophoto to give an impression about the appearance of the environment and the type of driving scenario.

For the data set *Nordstadt*, the car sensor scans were acquired using a Velodyne VLP-16, and for the data sets *Haltenhoffstraße* and *Highway A2/B6*, they were acquired using a Hesai PandarXT-



(a) Nordstadt



(b) Haltenhoffstraße



(c) Autobahn 2 and Bundesstraße 6 (interchange Hannover-Herrenhausen)

Figure 5.2.: Orthophotos showing the environments of the three data sets along with parts of the driven trajectories (Geoinformation Hannover, 2022).

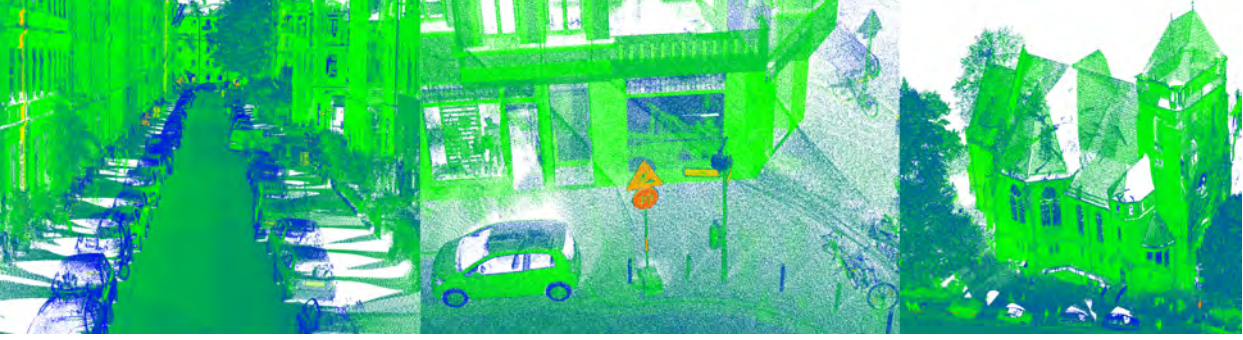


Figure 5.3.: Sections of the map point cloud from the Nordstadt data set. The points are colored according to the measured reflectivity value.

32. The VLP-16 and PandarXT-32 were both operated with 10 Hz and provide a range accuracy of $\pm 3\text{ cm}$ and $\pm 1\text{ cm}$, respectively. Further technical specifications are stated in Tab. 4.1.

The map point clouds are acquired using the two RIEGL VQ-250 LiDARs, which are part of the RIEGL VMX-250 MMS, and aligned using the adjustment strategy from Brenner (2016). Detailed information is provided in Sec. 2.4.2. In Fig. 5.3, sections of the map point cloud from the Nordstadt data set are visualized. The points are colored according to the measured reflectivity value.

By means of KPConv (Thomas et al., 2019) point-wise labels are generated, which are used to remove points from the ground, vegetation, parked cars, and other traffic participants. In Fig. 5.4, a section of the map point cloud from the Nordstadt data set is visualized after the removal has been conducted. In this example, only facades, tree trunks, street lights, and street signs remain.

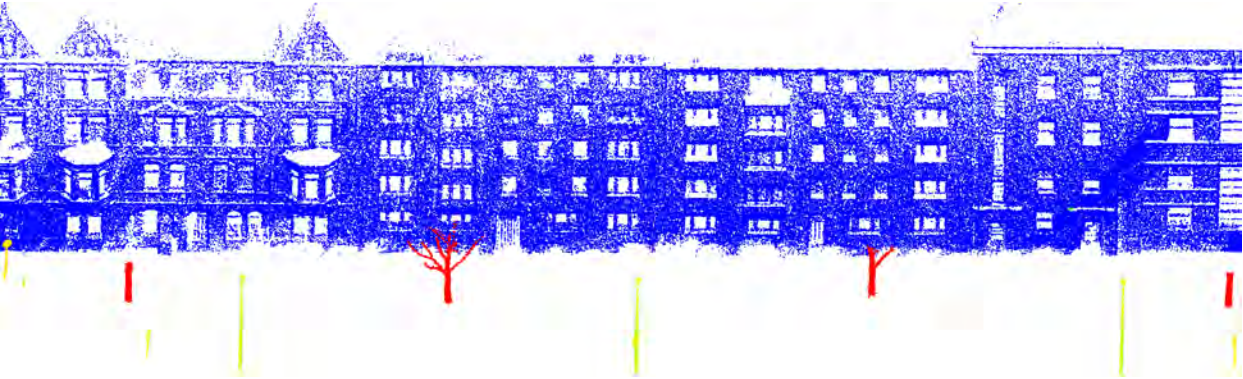


Figure 5.4.: Section of the map point cloud from the Nordstadt data set where points from the ground, vegetation, parked cars, and other traffic participants are removed. The points are colored according to the assigned class label. Only facades (blue), tree trunks (red), street lights (green), and street signs (yellow) remain.

The reference trajectories provide the ground truth poses for the localization. They are acquired using the Applanix POS LV GNSS/IMU system, which is also a part of the RIEGL VMX-250 MMS, and refined using both RTK correction data from SAPOS® (GeoBasis-DE, SAPOS, 2023) and local shifts from the point cloud adjustment (cf. Sec. 2.4.2).

In Tab. 5.1, the used LiDAR sensor, the number of car sensor scans (referred to as *epochs*), and the length of the trajectory are summarized for the three data sets. The Nordstadt data set is a part of the LUCOOP: Leibniz University Cooperative Perception and Urban Navigation Dataset (Axmann

Table 5.1.: Overview of data sets used in the experiments.

Environment	Area	LiDAR	Epochs	Length [m]
Structured	Nordstadt	Velodyne VLP-16	7094	2089.89
Semi-structured	Haltenhoffstr.	PandarXT-32	1694	1754.23
Unstructured	Highway A2/B6	PandarXT-32	7851	12299.75

et al., 2023a), which is publicly available. For the investigations based on correspondences in the observation space, only 2600 epochs from the Highway data set were considered.

Finally, the transformations between the different coordinate systems of the sensors are explained. In Fig. 5.5, a photo of the basic sensor setup is shown, which has been taken before the measurement drive for the *Highway A2/B6* data set. The Hesai PandarXT-32 is mounted on a platform at the front of the roof, and the RIEGL MMS is mounted at the end of the roof.



Figure 5.5.: Photo of the basic sensor setup: The Hesai PandarXT-32 is mounted on a platform at the front of the roof, and the RIEGL MMS is mounted at the end of the roof.

In Fig. 5.6, a sketch of the vehicle with the sensor setup is shown, which includes the different coordinate systems and the transformations between those. The green circle represents the PandarXT-32, the white rectangle the platform on which it is mounted, and the grey rectangle the RIEGL MMS. The transformation from the LiDAR coordinate system L into the platform coordinate system P is denoted as ${}^P\mathbf{T}_L$, the transformation from the RIEGL coordinate system R into the platform coordinate system P is denoted as ${}^P\mathbf{T}_R$, and the transformation from the RIEGL coordinate system R into the world coordinate system W is denoted as ${}^W\mathbf{T}_R$. In Fig. 5.6, the world coordinate system is visualized next to the vehicle. In this work, a UTM projection (Easting, Northing, ellipsoidal height) with GRS80 ellipsoid and ETRS89 (epsg:25832) datum and reference frame is used as world coordinate system. The data referenced in the world coordinate system are the map point clouds and reference trajectories.

The calibration to determine the transformations was done in two steps. In the first step, the PandarXT-32 was calibrated with respect to the platform coordinate system to determine ${}^P\mathbf{T}_L$, and in the second step, the platform was mounted on the roof of the vehicle and the RIEGL MMS

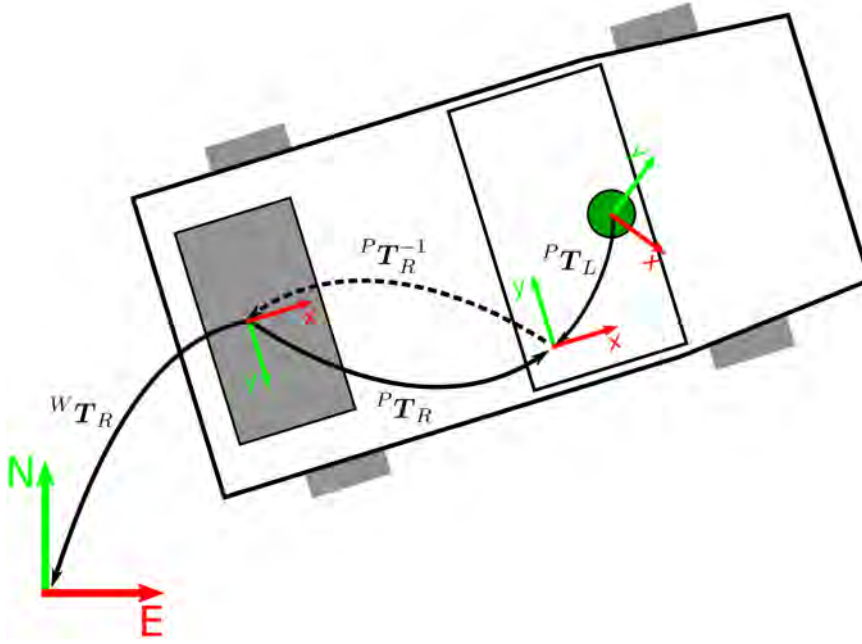


Figure 5.6.: Sketch of the vehicle with the sensor setup, the different coordinate systems, and the transformations between those coordinate systems. The green circle represents the PandarXT-32, the white rectangle the platform on which it is mounted, and the grey rectangle the RIEGL MMS.

was calibrated with respect to it to determine $^P T_R$. Additional sensors could be easily added to the sensor setup by mounting them on the platform and calibrating them with respect to the platform coordinate system in the first step of the calibration.

To transform a car sensor scan point s_i , which initially exists in the coordinate system L , into the world coordinate system W , the following the chain of transformations has to be applied:

$$^W s_i = ^W T_R (^P T_R)^{-1} ^P T_L ^L s_i, \quad (5.5)$$

where $^W T_R$ is defined by the pose of the vehicle in the world coordinate system at the time of the measurement of s_i , which is given by the RIEGL trajectory.

5.3. Results

In this section, the results of the experiments are presented. The section is structured according to the evaluation objectives. In Sec. 5.3.1, the results of the maximum consensus localization based on both the inlier count and the point-to-plane adjustment score are shown, in Sec. 5.3.2 the localization results using the different loss functions, in Sec. 5.3.3 the results of the maximum consensus filter, and in Sec. 5.3.4 the results of the protection level estimation and integrity evaluation.

For the experiments, the search range for the maximum consensus localization based on correspondences in the observation space is set to $2\text{ m} \times 2\text{ m}$ for the translation t_{xy} (from -1 m to $+1\text{ m}$ for both axes), which, with a discretization of 0.1 m , leads to a grid of 21×21 candidate positions. The search range for the maximum consensus localization based on correspondences in the Euclidean space is set to $4\text{ m} \times 4\text{ m}$ for the translation t_{xy} (from -2 m to $+2\text{ m}$ for both axes), which, with the same discretization of 0.1 m , leads to a grid of 41×41 candidate positions. For both approaches, the search range of the heading angle θ is set to -0.8° to $+0.8^\circ$ and -0.72° to



Figure 5.7.: Application of the maximum consensus localization: For a section of the Nordstadt trajectory, the 2D maximum consensus accumulator slices of the optimal heading angle are visualized on top of an orthophoto. Every 20th epoch is illustrated, the size of the 2D xy -grids is $4\text{ m} \times 4\text{ m}$, and the red dots indicate the localization solutions.

$+0.72^\circ$ with a discretization of 0.2° and 0.18° for the Velodyne VLP-16 and Hesai PandarXT-32, respectively, which, for both sensors, leads to nine different heading angles.

In Fig. 5.7, to show the application of the maximum consensus localization, for a section of the Nordstadt trajectory, the 2D maximum consensus accumulator slices of the optimal heading angle are visualized on top of an orthophoto. In this image, every 20th epoch is illustrated, the size of the 2D grids to determine the translation \mathbf{t}_{xy} is $4\text{ m} \times 4\text{ m}$, and the red dots indicate the localization solutions.

5.3.1. Maximum consensus localization based on the inlier count and point-to-plane adjustment score

To improve the structure, the description of the results is divided into small sections, which cover the accuracy, robustness, and reliability of position and heading, as well as additional investigations.

Position accuracy:

For all three data sets, the maximum consensus localization is performed based on the inlier count and the score, using both correspondences in the observation space and in the Euclidean space, which results in four different localization solutions for each data set. In Tab. 5.2, the RMSE values with respect to the xy -position are stated for the three methods considered for comparison point-to-plane ICP, KISS-ICP, and Range-MCL as well as for the four maximum consensus localization approaches. As the point-to-plane ICP with an inlier radius of 1 m shows significantly poorer results than with an inlier radius of 0.5 m , the tables and discussions in this section refer to the point-to-plane ICP with an inlier radius of 0.5 m . Only in Fig. 5.8 and Fig. 5.9, later in this section, are the results of the point-to-plane ICP shown with both 0.5 m and 1 m inlier radii. The maximum consensus localization solution corresponds to the candidate pose with the highest consensus for the approach using correspondences in the observation space and to the center point of the grid cell with the highest consensus for the approach using correspondences in the Euclidean space.

Data set	ICP	KISS	R-MCL	Count (o.)	Score (o.)	Count (E.)	Score (E.)
Nordstadt	0.047	0.051	0.22	0.07	0.097	0.006	0.005
Haltenhoffstr.	0.055	0.033	0.217	0.151	0.218	0.028	0.032
Highway	0.204	0.056	0.755	0.381	0.431	0.178	0.152

Table 5.2.: RMSE [m] with respect to the xy -positioning for the three methods considered for comparison point-to-plane ICP, KISS-ICP, and Range-MCL as well as for the maximum consensus localization approaches using the inlier count and the point-to-plane adjustment score. The results for the maximum consensus localization are determined based on correspondences both in the observation space (o.) and in the Euclidean space (E.).

Considering the results in Tab. 5.2, except for the KISS-ICP, the RMSE values for the *Nordstadt* data set are the smallest, and they increase for the *Haltenhoffstraße* and increase even more for the *Highway*, which is conclusive as the amount of static objects for the localization decreases between those data sets. Comparing the localization results from the observation space and Euclidean space, the RMSE values for the localization based on correspondences in the observation space are significantly larger, which is also conclusive as there are multiple reasons for that. First, the synthetic scan images are rendered based on static candidate poses while the actual measurement is performed from a continuously moving LiDAR pose, which leads to different occlusions and deviations between the measured and rendered ranges. Second, the approach in the Euclidean space uses directly the very precise map point cloud while the approach in the observation space uses a triangle mesh, which first had to be generated from the map point cloud. This meshing step induces additional inaccuracies. Third, to determine the scan image column of a measured scan point, its azimuth angle is calculated and rounded, which also induces deviations between the measured and synthetic scan image.

Comparing the count and the score, in the observation space, the count performs better than the score, which could question the benefit of the new score. However, due to the mentioned error sources and additional inaccuracies from using the normal vectors, these results should not be overrated. Instead, the focus is to be put on the results from the Euclidean space, as these are free from systematic effects and therefore much more suitable for a methodical analysis as well as comparison of the approaches. For the *Nordstadt*, the RMSE value of the count is slightly larger than the one of the score, but for the *Haltenhoffstraße*, the RMSE of the count is smaller than the one of the score, so that there is still no recognizable advantage of using the score. However, for the *Highway*, which represents a challenging scenario for LiDAR-based localization, the benefit of using the point-to-plane adjustment score becomes visible as the RMSE values are significantly smaller compared to the count.

For the comparison to the point-to-plane ICP, KISS-ICP, and Range-MCL, it should first be noted that for the maximum consensus localization the error is exactly zero if the ground pose is found, and the error is immediately at least the size of the discretization, which is with 0.1 m quite large in this case, if the localization solution is not the ground truth pose. For the *Nordstadt*, where the ground truth pose is often found, the RMSE value is therefore overly optimistic, while it is overly pessimistic for the *Highway*, where also candidate poses around the ground pose are often determined as localization solutions. This effect does not occur for the methods considered for comparison, which is why their RMSE values tend to be higher for the *Nordstadt* and lower for the *Highway* compared to the RMSE values of maximum consensus localization. For the point-to-plane ICP, the RMSE values are low at around 5 cm for *Nordstadt* and *Haltenhoffstraße*, while they are about four times as high at 20.4 cm for the *Highway*, which indicates that the ICP enables accurate localization for *Nordstadt* and *Haltenhoffstraße*, while it leads to incorrect localization at least partially on the *Highway*. For the KISS-ICP, the RMSE values are low at around 5 cm for all three

Data set	ICP	KISS	R-MCL	Count (o.)	Score (o.)	Count (E.)	Score (E.)
Nordstadt	0.0	0.0	0.054	0.005	0.022	0.0	0.0
Haltenhoffstr.	0.0	0.0	0.05	0.034	0.11	0.0	0.0
Highway	0.051	0.002	0.62	0.252	0.307	0.032	0.028

Table 5.3.: Shares of epochs with an xy -error larger than the position alert threshold for the three methods considered for comparison point-to-plane ICP, KISS-ICP, and Range-MCL as well as for the maximum consensus localization approaches using the inlier count and the point-to-plane adjustment score. The results for the maximum consensus localization are determined based on correspondences both in the observation space (o.) and in the Euclidean space (E.).

data sets, indicating that the method allows accurate localization in each case, even on the *Highway*. Comparing the point-to-plane ICP and KISS-ICP, which are both based on correspondences in the Euclidean space, to both maximum consensus approaches based on correspondences in the Euclidean space, the point-to-plane ICP performs worse for the *Highway*, while the KISS-ICP performs much better for the *Highway*. For the *Nordstadt* and *Haltenhoffstraße*, the exact RMSE values of ICP, KISS-ICP and both maximum consensus approaches are hard to compare due to the described bias from the initialization, however, as all RMSE values are low, it can be concluded that for these two data sets all the approaches allow an accurate localization. Looking at the RMSE values of the Range-MCL, they are significantly larger compared to all other localization approaches, even compared to the maximum consensus localization based on correspondences in the observation space similar to the Range-MCL itself. This indicates that the objective function of Range-MCL, which is the mean of the absolute pixel-wise differences between the synthetic and the measured range image, leads to significant localization errors at least in some parts of the trajectories.

Position robustness:

In Tab. 5.3, for the four maximum consensus localization approaches and the three methods of comparison the shares of epochs with an xy -error larger than the position alert threshold are stated. These results help to assess the robustness of the methods, which is a particular focus of this work. In general, the frequencies of positioning failures are in line with the RMSE values, i.e. the higher the RMSE value, the higher is the frequency of failures. While both the point-to-plane ICP and KISS-ICP only lead to positioning failures on the *Highway* and only with small shares of failures, the Range-MCL leads to positioning failures in all three data sets, with small shares in the *Nordstadt* and *Haltenhoffstraße*, but an enormous share on the *Highway*. Considering the maximum consensus localization, similar to the RMSE values, the shares of positioning failures in the observation space are higher when using the score than when using the count, and they increase with less objects in the environment. The shares of positioning failures in the Euclidean space are smaller than in the observation space, and for *Nordstadt* and *Haltenhoffstraße* they are even zero for both the count and the score. For the highway, the share of positioning failures for the score is smaller, which emphasizes its advantage over the count.

When comparing the point-to-plane ICP and KISS-ICP with both maximum consensus localization approaches based on correspondences in the Euclidean space in terms of robustness, all approaches perform optimally for *Nordstadt* and *Haltenhoffstraße* and the point-to-plane ICP performs worse and the KISS-ICP performs better than the maximum consensus approaches on the *Highway*. However, it is important to note that the point-to-plane ICP and KISS-ICP have so far been initialized in the ground truth pose, which means that positioning failures only occur if the registration significantly diverges from the true pose, and that the failures do not give any indication about the robustness towards initialization errors. Therefore, two additional tests have been carried out where the point-to-plane ICP and KISS-ICP have been initialized in the ground

Data set	Count CP	Count refined	Score CP	Score refined
Nordstadt	0.006	0.006	0.005	0.008
Haltenhoffstr.	0.028	0.028	0.032	0.029
Highway	0.178	0.177	0.152	0.15

Table 5.4.: RMSE values [m] of the xy -positions corresponding to the center points (CP) of the grid cells with the highest consensus and of the refined xy -positions for the maximum consensus localization based on both the count and the score using correspondences in the Euclidean space. The refined xy -positions are determined as the mean of the residuals from all matches contributing to the highest consensus and according to Eq. 4.19 for the count and score, respectively.

truth position plus an offset of 0.5 m and 1 m , each time along both the x - and y -axes. With an initialization error of 0.5 m , for the point-to-plane ICP the share of positioning failures is 0, 0, and 0.073, and for the KISS-ICP the share of positioning failures is 0.032, 0, and 0.091 for the *Nordstadt*, *Haltenhoffstraße*, and *Highway*, respectively. While this only means an increase of 0.022 on the *Highway* for the point-to-plane ICP, for the KISS-ICP, in contrast to the perfect initialization, positioning failures now also occur in the *Nordstadt* and the share of failures increases by 0.089 on the *Highway*. Even though the failure rates are not very high, the results show that an error of 0.5 m can already prevent convergence to the correct solution. However, if the initialization error is now increased to 1 m , the shares of positioning failures tremendously rise to 0.108, 0.255, and 0.325 for the point-to-plane ICP and 0.396, 0.424, and 0.362 for the KISS-ICP, which shows that both methods strongly depend on a good initialization. Considering that such comparatively small initialization errors already lead to such high error rates, the share of positioning failures of 0.032 and 0.028 for the maximum consensus localization using count and score, respectively, can be considered good results, especially since the search area was $4\text{ m} \times 4\text{ m}$. Furthermore, it applies to both ICP-approaches, the worse the initialization, the greater the threshold value for outlier rejection must be in order to somehow still achieve convergence to the true solution, which, however, is accompanied by poorer accuracy, especially in case of the point-to-plane ICP with the quadratic loss function. This becomes clear when looking at Fig. 5.8, which shows, among others, the positioning error of the point-to-plane ICP with an inlier radius of 0.5 m and 1 m .

Position refinement:

As already shortly discussed, the candidate pose and the center point of the grid cell with the highest consensus have so far been used as localization solutions for the maximum consensus approaches using correspondences in the observation space and Euclidean space, respectively. On the one hand, this leads to overly optimistic RMSE values as the error of an epoch is always zero if the candidate pose or grid cell corresponding to the ground truth pose yields the highest consensus and therefore is the localization solution. On the other hand, if the localization solution is not the ground truth, the error is immediately at least the size of the discretization, which may be overly pessimistic. Therefore, the goal is to refine the localization solution and determine a more precise position. For the count, the refined position is calculated as the mean of the residuals from all matches contributing to the highest consensus, and for the score, the refined position is calculated according to Eq. 4.19. In Tab. 5.4, for the maximum consensus localization using correspondences in the Euclidean space the RMSE values of the xy -positions corresponding to the center points of the grid cells with the highest consensus (similar to Tab. 5.2) and of the refined xy -positions are stated.

Apart from the score for the *Nordstadt* data set, the refined xy -positions lead to the same or smaller RMSE values than the center points of the grid cells with the highest consensus. To make the benefit of the refinement even clearer, in Tab. 5.5 the RMSE values are stated for the conventional positioning using the center point of the grid cell with the highest consensus and for

Data set	Count CP	Count refined	Score CP	Score refined
Nordstadt	0.052	0.042	0.052	0.043
Haltenhoffstr.	0.077	0.073	0.067	0.06
Highway	0.224	0.223	0.184	0.181

Table 5.5.: RMSE [m] with respect to the xy -positioning for the maximum consensus localization using both the inlier count and the point-to-plane adjustment score based on correspondences in the Euclidean space. The RMSE values are stated for the conventional positioning using the center point of the grid cell with the highest consensus and for the refined positioning, however, taking into account only those epochs where the count or score localization error is not equal to zero.

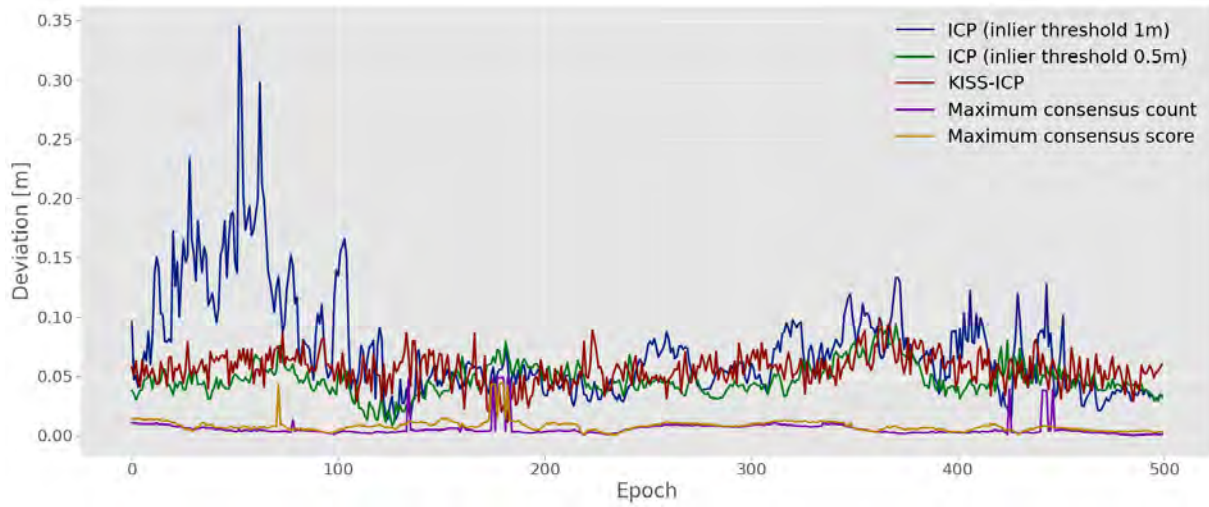
the refined positioning, however, taking into account only those epochs where the count or score localization error is not equal to zero. In each case, the RMSE based on the refined position is smaller than the one based on the center point of the grid cell with the highest consensus.

In Fig. 5.8, the errors with respect to the xy -position are visualized for the point-to-plane ICP with an inlier threshold of 0.5 m and 1 m, the KISS-ICP, and the maximum consensus localization using both the inlier count and the point-to-plane adjustment score based on correspondences in the Euclidean space. For reasons of clarity and comprehensibility, only a section of 500 epochs is shown for each data set. In Fig. 5.8a, the deviations are plotted for the *Nordstadt* data set, in Fig. 5.8b for the *Haltenhoffstraße* data set, and in Fig. 5.8c for the *Highway* data set. Since the number of map objects gradually decreases between these data sets, the localization errors of all five methods increase (except for KISS-ICP between *Nordstadt* and *Haltenhoffstraße*). Comparing the two ICP results, it becomes clear that the accuracy of the ICP strongly depends on the correspondence threshold, and that lowering this threshold reduces the localization error, provided that the initialization is sufficiently good. Overall, the ICP with the inlier threshold of 1 m performs worst and is not able to reliably provide localization solutions with errors below the alert threshold of 0.29 m. The ICP with the inlier threshold of 0.5 m performs already much better, but its lack of robustness also becomes clear when considering the *Highway*. The KISS-ICP performs best, as it is able to provide precise localization results for all three data sets, even in challenging situations on the *Highway*, where it rarely produces localization errors close to or above the alert threshold. However, it is important to be aware that for the results shown here, the two point-to-plane ICPs and the KISS-ICP are based on an initialization in the ground truth pose and that their results would be significantly worse in the event of initialization errors. The two maximum consensus approaches provide precise localization results for the *Nordstadt* and *Haltenhoffstraße*, however, they show significant localization errors in some parts of the *Highway*. When comparing the count and the score, the benefit of the score becomes clear when looking at the *Highway* data set, as the score leads to a lower number of large localization errors.

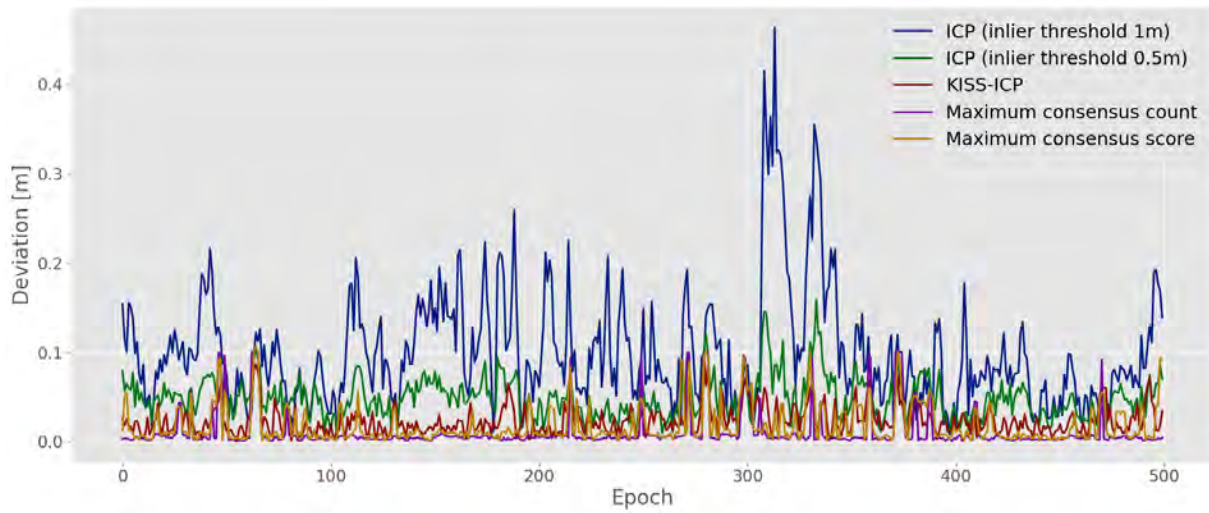
Heading accuracy and robustness:

So far, only the xy -position has been considered, however, the localization also comprises the heading angle θ . In Tab. 5.6, for the three methods considered for comparison point-to-plane ICP, KISS-ICP, and Range-MCL as well as for the maximum consensus localization approaches using the count and the score, each in the observation and Euclidean space, the RMSE values with respect to the heading are stated.

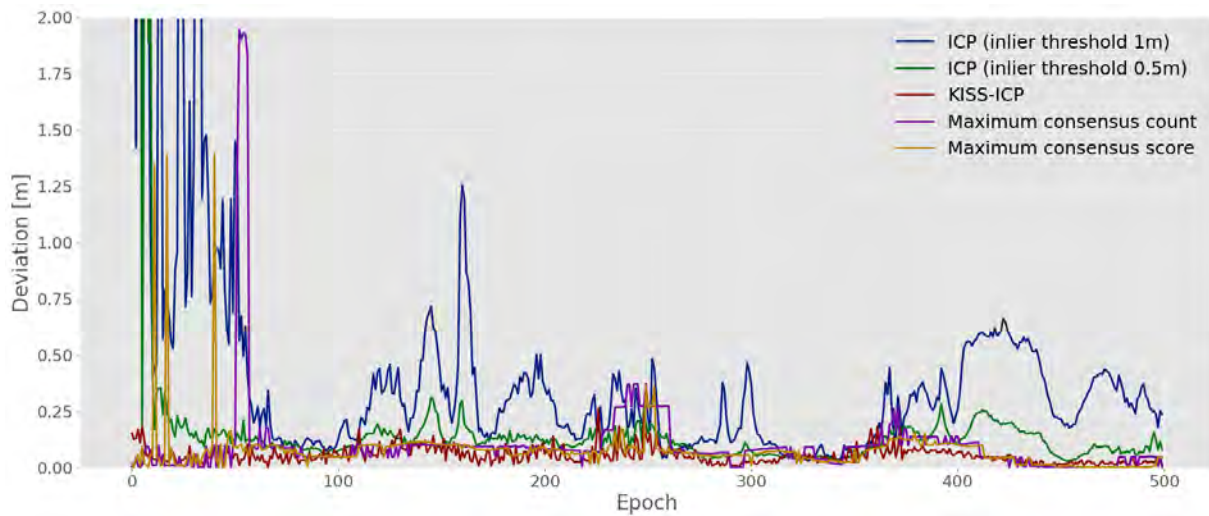
For the maximum consensus approaches, similar to the xy -position, the RMSE values increase from *Nordstadt* to *Haltenhoffstraße* and from *Haltenhoffstraße* to *Highway*, and they are significantly larger for the localization based on correspondences in the observation space than for the localization based on correspondences in the Euclidean space. However, when comparing the count and the score, not only in the observation space, but also in the Euclidean space, the count shows



(a) Nordstadt.



(b) Haltenhoffstraße.



(c) Highway.

Figure 5.8.: Errors [m] with respect to the xy -position for the point-to-plane ICP with an inlier threshold of 0.5m and 1m, the KISS-ICP, and the maximum consensus localization using both the inlier count and the point-to-plane adjustment score based on correspondences in the Euclidean space.

Data set	ICP	KISS	R-MCL	Count (o.)	Score (o.)	Count (E.)	Score (E.)
Nordstadt	0.042	0.054	0.023	0.171	0.203	0.032	0.038
Haltenhoffstr.	0.045	0.045	0.021	0.154	0.264	0.038	0.044
Highway	0.077	0.04	0.021	0.269	0.321	0.058	0.1

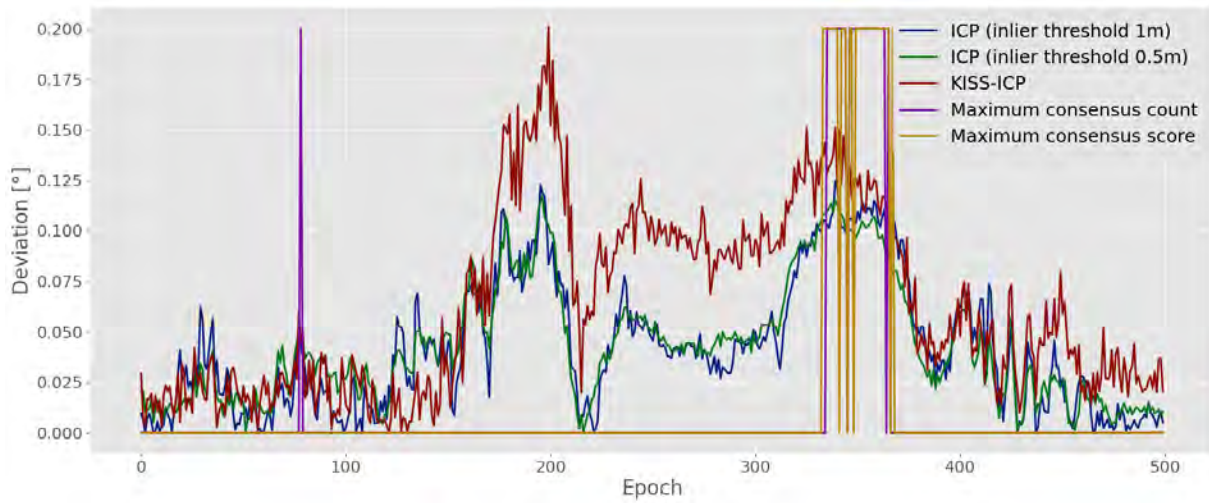
Table 5.6.: RMSE [°] with respect to the heading for the three methods considered for comparison point-to-plane ICP, KISS-ICP, and Range-MCL as well as for the maximum consensus localization approaches using the inlier count and the point-to-plane adjustment score. The results for the maximum consensus localization are determined based on correspondences both in the observation space (o.) and in the Euclidean space (E.).

Data set	ICP	KISS	R-MCL	Count (o.)	Score (o.)	Count (E.)	Score (E.)
Nordstadt	0.0	0.0	0.0	0.001	0.018	0.0	0.0
Haltenhoffstr.	0.0	0.0	0.0	0.008	0.06	0.0	0.0
Highway	0.002	0.0	0.0	0.053	0.052	0.0	0.002

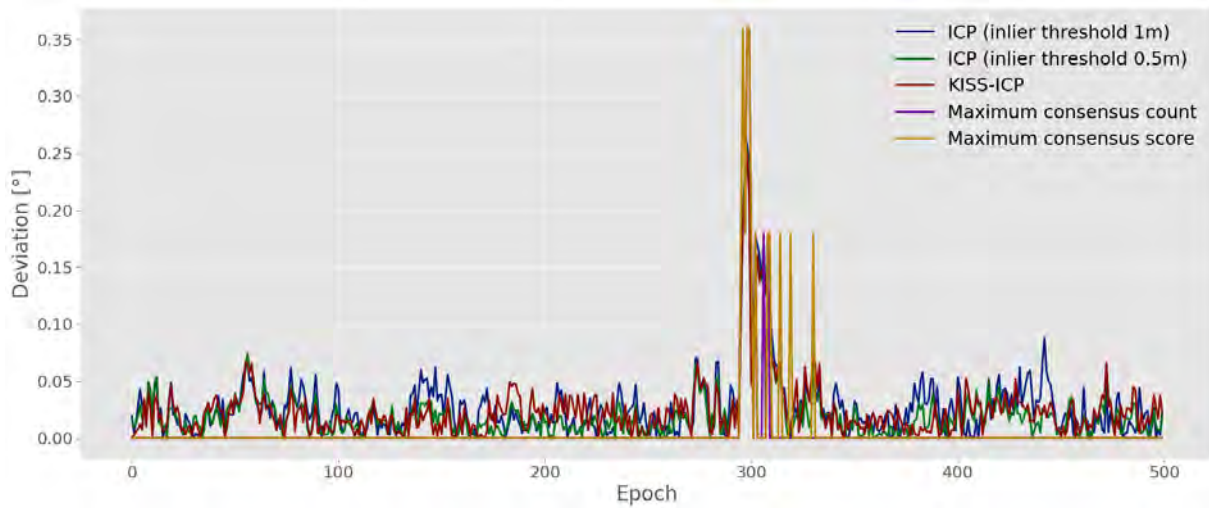
Table 5.7.: Share of epochs with a heading error larger than the heading alert threshold for the three methods considered for comparison point-to-plane ICP, KISS-ICP, and Range-MCL as well as for the maximum consensus localization approaches using the inlier count and the point-to-plane adjustment score. The results for the maximum consensus localization are determined based on correspondences both in the observation space (o.) and in the Euclidean space (E.).

smaller RMSE values than the score. When focusing on the RMSE values in the Euclidean space, which are considered to be more reliable, it becomes clear that the overall error magnitudes are very small though, and that both the count and score do not show a weakness in determining the heading sufficiently accurately. This also applies to the three methods considered for comparison, which all show RMSE values that are in the same order of magnitude as the RMSE values of the maximum consensus approaches in the Euclidean space. Whereas for the point-to-plane ICP the RMSE values increase from *Nordstadt* to *Haltenhoffstraße* and from *Haltenhoffstraße* to *Highway*, for the KISS-ICP they decrease underlining its robustness. The Range-MCL shows the smallest RMSE values of all approaches, which are almost similar for all three data sets. Thus, both maximum consensus approaches based on correspondences in the Euclidean space and the three methods of comparison can be used to reliably determine the heading, which is confirmed by the shares of positioning failures, where a failure corresponds to a heading error greater than the heading alert threshold of 0.5° . As stated in Tab. 5.7, for all three data sets, the three methods considered for comparison and both the count and the score based on correspondences in the Euclidean space almost always result in a failure rate of zero. The only exceptions are the point-to-plane ICP and the score on the *Highway*, for which the failure rate is close to zero at 0.002. The maximum consensus localization based on correspondences in the observation space cannot reach such results as for all three data sets and for both the count and the score comparatively high failure rates occur.

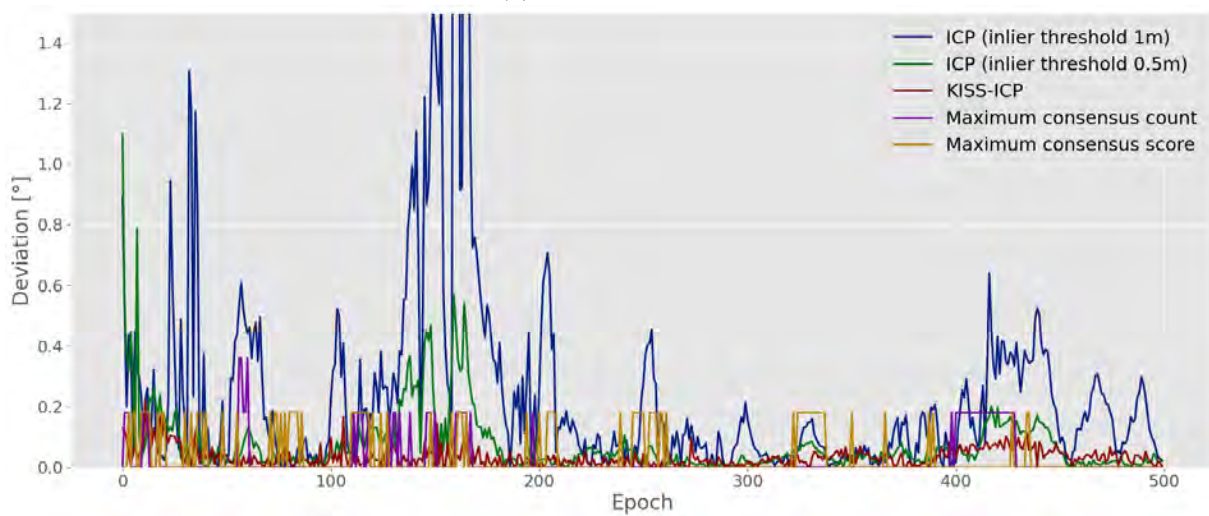
In Fig. 5.9, the errors with respect to the heading are visualized for the point-to-plane ICP with an inlier threshold of 0.5 m and 1 m , the KISS-ICP, and the maximum consensus localization using both the inlier count and the point-to-plane adjustment score based on correspondences in the Euclidean space. For reasons of clarity and comprehensibility, only a section of 500 epochs is shown for each data set, which are the same epochs as in Fig. 5.8. In Fig. 5.9a, the deviations are plotted for the *Nordstadt* data set, in Fig. 5.9b for the *Haltenhoffstraße* data set, and in Fig. 5.9c for the *Highway* data set. As the ground truth pose is used as initial pose, the error of both maximum consensus localization approaches is zero, if the correct heading is found, and otherwise a multiple of the heading discretization of 0.2° and 0.18° for the VLP-16 and PandarXT-32, respectively. For the *Nordstadt* and *Haltenhoffstraße*, all five localization approaches show only small deviations



(a) Nordstadt.



(b) Haltenhoffstraße.



(c) Highway.

Figure 5.9.: Errors [°] with respect to the heading for the point-to-plane ICP with an inlier threshold of 0.5 m and 1 m, the KISS-ICP, and the maximum consensus localization using both the inlier count and the point-to-plane adjustment score based on correspondences in the Euclidean space.

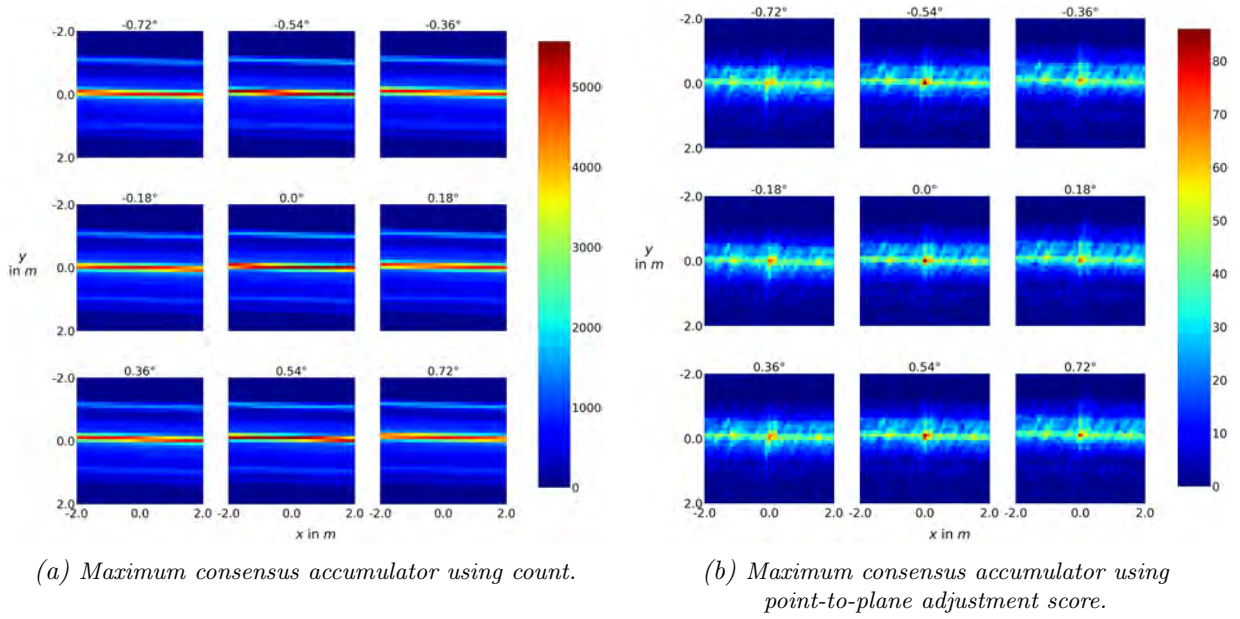


Figure 5.10.: Maximum consensus accumulator based on the count (a) and based on the score (b) for a challenging epoch from the Highway scenario. Whereas the count yields many high consensus values along driving direction, the score generates a clear peak, which vanishes with increasing deviation from the true heading. Furthermore, the count leads to a localization error larger the alert threshold, as it determines $(x, y, \theta) = (1.6\text{ m}, 0\text{ m}, -0.54^\circ)$ as the localization solution, while the score yields $(x, y, \theta) = (0\text{ m}, 0\text{ m}, 0.0^\circ)$, which is the true pose.

from the true heading, below the heading alert threshold of 0.5° . However, for the *Highway*, the heading error of the point-to-plane ICP with an inlier threshold of 1 m increases significantly, frequently surpassing the heading alert threshold, and also the heading error of the point-to-plane ICP with an inlier threshold of 0.5 m increases, although not as strong as for the inlier threshold of 1 m , but it still exceeds the alert threshold in a few epochs. The maximum consensus approaches using both the count and the score only show heading errors of up to 0.36° in these epochs, which do not exceed the heading alert threshold. This also applies to the KISS-ICP, which shows the most robust behavior of all approaches. However, it is important to be aware that for the results shown here, the two point-to-plane ICPs and the KISS-ICP are initialized in the ground truth pose and that the deviations in the event of an initialization error are not considered here.

Example accumulator using count and point-to-plane adjustment score:

To give an example for the benefit of the point-to-plane adjustment score, for a challenging epoch from the *Highway* scenario, both a maximum consensus accumulator based on the count and a maximum consensus accumulator based on the score are visualized in Fig. 5.10. Whereas the count yields many high consensus values along driving direction, the score generates a clear peak, which vanishes with increasing deviation from the true heading. Furthermore, the count leads to a localization error larger the alert threshold, as it determines $(x, y, \theta) = (1.6\text{ m}, 0\text{ m}, -0.54^\circ)$ as the localization solution, while the score yields $(x, y, \theta) = (0\text{ m}, 0\text{ m}, 0.0^\circ)$, which is the true pose.

Application of grid shifts:

As described in Sec. 4.1.3, to prevent localization failures caused by the discretization of the search space, for the maximum consensus localization based on correspondences in the Euclidean space, shifts by half a grid cell along each axis are conducted. To see the effect of this measure, for the maximum consensus localization using both the count and score, the RMSE values and shares of positioning failures with and without the application of the grid shifts are stated in Tab. 5.8. The

Translation				
	Count RMSE [m]	Score RMSE [m]	Count failures	Score failures
Without grid shifts	0.291	0.167	0.068	0.04
With grid shifts	0.178	0.152	0.032	0.028

Heading				
	Count RMSE [°]	Score RMSE [°]	Count failures	Score failures
Without grid shifts	0.075	0.12	0.001	0.004
With grid shifts	0.058	0.1	0.0	0.002

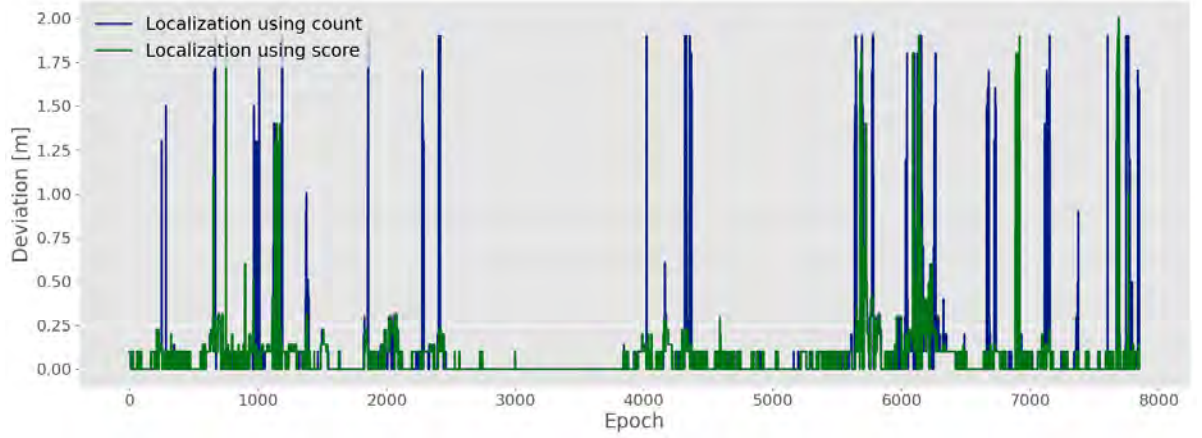
Table 5.8.: Comparison of maximum consensus localization without and with grid shifts for the highway scenario: RMSE and share of positioning failures with respect to the xy -positioning and heading for the localization using both the inlier count and the point-to-plane adjustment score. The results are based on correspondences in the Euclidean space and the localization solution for the xy -position corresponds to the center point of the grid cell with the highest consensus value.

values are based on the localization using the *Highway* data set. For both the count and the score, the RMSE values and shares of positioning failures reduce for translation as well as heading.

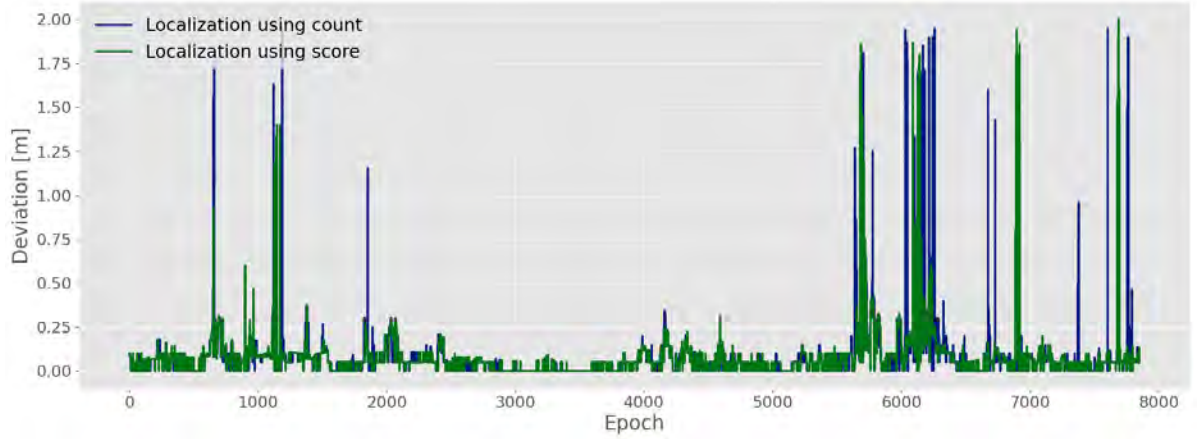
In Fig. 5.11, the effect of the grid shifts on the localization performance is visualized for the *Highway* data set, where Fig. 5.11a shows the localization error without applying the grid shifts, and Fig. 5.11b shows the localization error with the grid shifts applied. It has already been expressed in Tab. 5.8 and it is made visually clear in Fig. 5.11 that the grid shifts significantly reduce the number of large localization errors for both the count and the score.

Localization performance in the presence of a systematic heading error:

So far, the results of the maximum consensus localization based on the count have been almost as good as the results based on the score. Even though from the examples shown in Fig. 4.10, Fig. 4.14, and Fig. 5.10 it became visible that the score yields a more distinct localization solution, a crystal clear benefit of the score was not apparent in the RMSE values and shares of positioning failures. The reasons for that are the almost perfect circumstances for the LiDAR-based localization, which comprise the following aspects: First, the map point clouds are complete, up-to-date, and provide a very regular point density, especially for the *Nordstadt* and *Haltenhoffstraße*. Second, the transformations between the sensors were determined highly accurately in a calibration procedure. Third, it is corrected for the motion distortion during the 0.1 s of one full LiDAR rotation. Fourth, the initialization using the ground truth trajectory results in basically no error in roll and pitch. Finally, as the true heading is known and as it is a methodological investigation, the heading search range as well as the discretization are chosen very small in this work. If the discretization would be selected much larger, the determined heading would always be the initial and true heading, respectively, and an influence from the heading would not be visible. However, in a real application where the initial heading is highly uncertain, the heading search range can be much larger and the discretization much coarser due to limitations in computation time. Depending on the discretization of the xy -grid, a certain discretization error in the heading can lead to a significant error in the determined xy -position. The finer the xy -discretization, the smaller the heading error that is sufficient to cause an xy -error. This does not only apply to discretization errors in the heading, but also applies to errors in roll and pitch or systematic errors in the transformations between the sensors, which are not existent in this work, however might be present in real world. To analyze the behavior of the count and score in the presence of such an error, the maximum consensus localization was carried out with a systematic heading error of two degrees. In Fig. 5.12, for the *Nordstadt* data set, the deviation from the ground truth is shown using the count and the score. Due to the systematic heading error, for both approaches a basic error of



(a) Localization error without grid shifts.



(b) Localization error with grid shifts.

Figure 5.11.: Effect of the grid shifts on the localization performance tested on 7851 epochs of the Highway data set: The grid shifts significantly reduce the number of large localization errors for both the count and the score.

around 20 cm occurs in most epochs. However, the difference between both approaches is that the score is much more resistant towards significant errors. Whereas the count leads to errors larger than 1 m in multiple epochs, the score is able to contain the error. This advantage in robustness is even more important when using a smaller xy -discretization, since then already smaller errors in the angles or transformations can cause such significant positioning failures.

Reliability:

Finally, the count and the score are compared with respect to the reliability of the registration, which directly determines the reliability of the localization. In this work, the reliability is assessed by the kurtosis and the second peak ratio (cf. Sec. 5.1). The kurtosis describes the distribution within the 2D grid of xy -positions, and the higher the kurtosis, the more distinct is the peak and therefore the localization solution. The second peak ratio is the ratio between the second highest and highest consensus values within the 2D grid of xy -positions, and the smaller the second peak ratio, the more distinct is the localization solution. Additionally, if there are one or multiple grid cells in the accumulator that yield at least 90% of the highest consensus, the largest distance between the ground truth position and these grid cells is taken into consideration. It is referred to as *largest peak distance*, and the aim is to define a measure that provides information on whether other high consensus values occur near or far from the ground truth position. The larger the

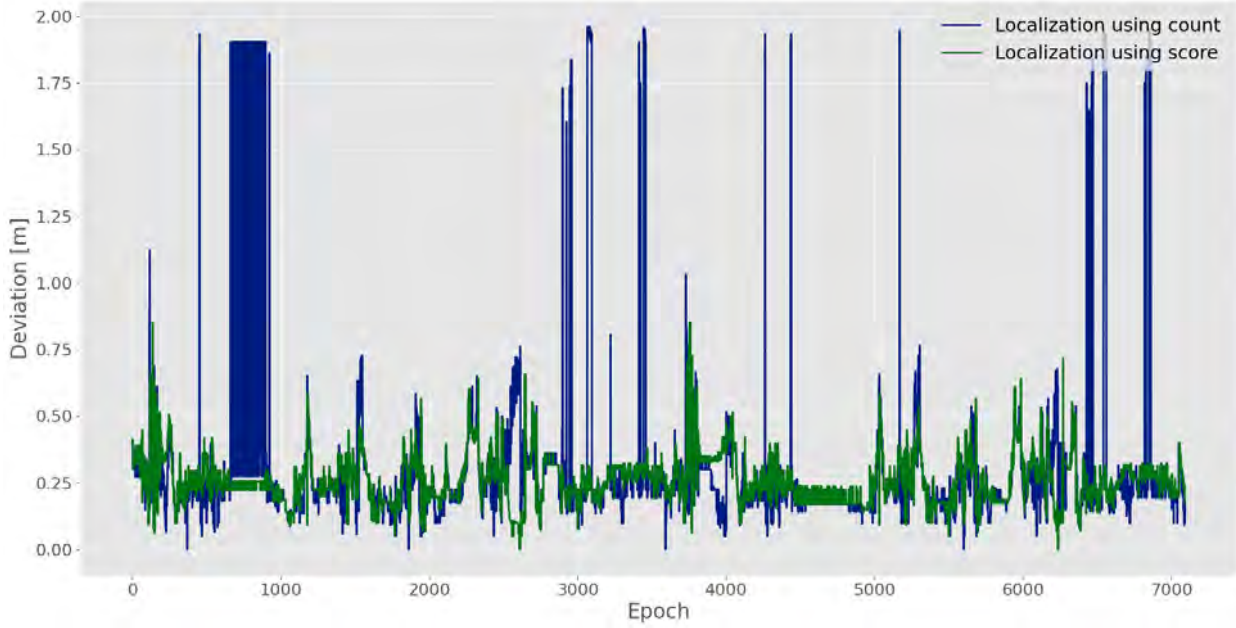


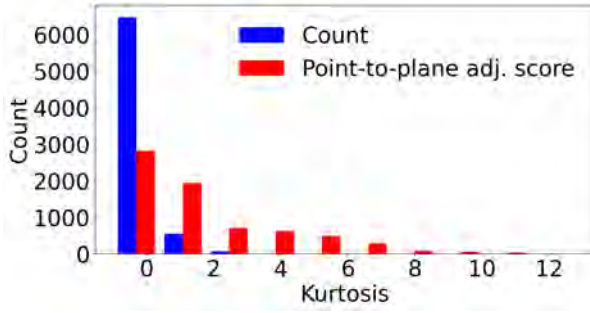
Figure 5.12.: Localization error [m] of the count and the score in the presence of a systematic heading error of two degrees shown for the Nordstadt data set: For both approaches a basic error of around 20 cm occurs in most epochs. However, the score is much more resistant towards significant errors, as it is able to contain the error, whereas the count leads to errors larger than 1 m in multiple epochs.

Data set	Kurtosis		Second peak ratio		Largest peak distance	
	Count	Score	Count	Score	Count	Score
Nordstadt	-0.336	1.64	0.834	0.579	0.1	0.1
Haltenhoffstr.	0.046	3.53	0.851	0.652	0.086	0.064
Highway	-0.804	0.819	0.948	0.809	0.629	0.216

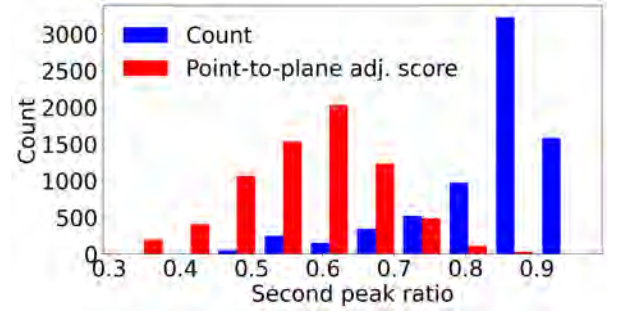
Table 5.9.: Mean of kurtosis values, second peak ratios, and largest peak distances over all epochs for the localization using count and score based on correspondences in the Euclidean space. The point-to-plane adjustment score yields higher kurtosis values and smaller second peak ratios, and thus more distinct localization solutions. Furthermore, compared to the count, the point-to-plane adjustment score yields similar or smaller distances between the ground truth position and a grid cell yielding at least 90% of the highest consensus.

largest peak distance, the larger and more critical is a potential localization error. In Tab. 5.9, for the localization using count and score based on correspondences in the Euclidean space, the mean values of the kurtosis values, second peak ratios, and largest peak distances over all epochs are stated. For the mean of the largest peak distances only those epochs are taken into consideration where at least one grid cell exists that yields minimum 90% of the highest consensus.

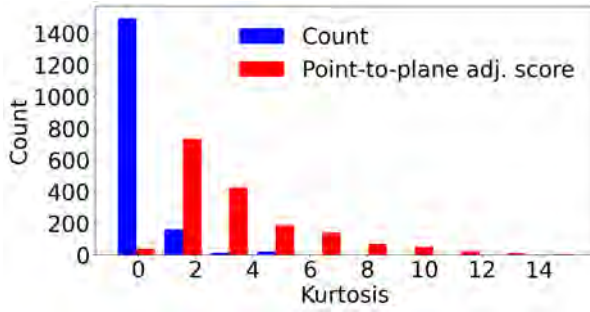
For all three data sets, the point-to-plane adjustment score yields higher mean values of the kurtosis and smaller mean values of the second peak ratio, and thus more distinct localization solutions. Furthermore, compared to the count, the score yields similar or smaller distances between the ground truth position and a grid cell yielding at least 90% of the highest consensus. For the *Nordstadt* and *Haltenhoffstraße*, both for the count and score the mean values of the largest peak distance indicate that other high consensus values occur near the true position. In some epochs, they even occur at the true position, which explains the mean values of the *Haltenhoffstraße* that are smaller than the grid cell size of 0.1 m. For the *Highway*, the mean of the largest peak distances slightly increases to 0.216 m for the score, however, it significantly increases to 0.629 m



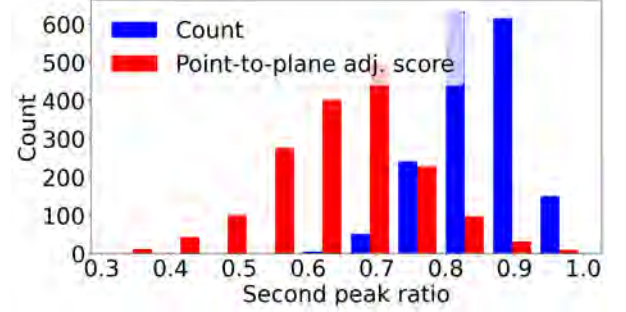
(a) Kurtosis values using Nordstadt data set.



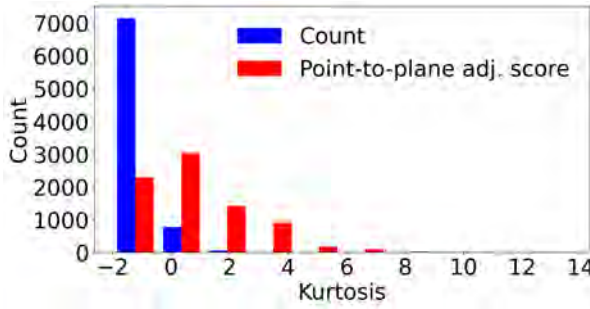
(b) Second peak ratios using Nordstadt data set.



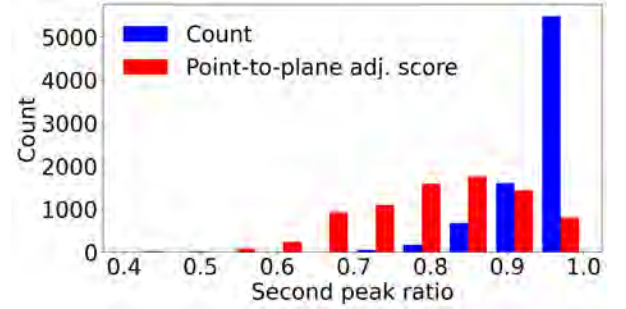
(c) Kurtosis values using Haltenhoffstraße data set.



(d) Second peak ratios using Haltenhoffstraße data set.



(e) Kurtosis values using Highway data set.



(f) Second peak ratios using Highway data set.

Figure 5.13.: Histograms for each of the three data sets containing the kurtosis values and second peak ratios for the count and score. The histograms illustrate that the point-to-plane adjustment score yields higher kurtosis values and smaller second peak ratios, both of which indicate a more distinct registration and therefore more reliable localization.

for the count, which shows that the count leads to high consensus values also far away from the ground truth position. In general, the number of epochs that are used to calculate the mean of the largest peak distance is much larger for the count than for the score because using the score there are significantly fewer epochs where a second consensus value exists that is larger than 90% of the highest consensus. This is consistent with the mean values of the second peak ratio, which are much smaller than 0.9 for the score.

In Fig. 5.13, for each of the three data sets, histograms are shown containing the kurtosis values and second peak ratios for the count and score that were used to calculate the mean values listed in Tab. 5.9. The histograms illustrate what has already become clear by the mean values, i.e., the point-to-plane adjustment score yields higher kurtosis values and smaller second peak ratios, both of which indicate a more distinct registration and therefore more reliable localization.

Data set	L2	L1	Huber	Cauchy	G.-McC.	Welsch	Tukey	L0
Nordstadt	0.89	0.026	0.243	0.065	0.019	0.021	0.01	0.003
Haltenhoffstr.	1.119	0.3	0.576	0.198	0.055	0.054	0.038	0.035
Highway	1.315	1.238	1.263	1.074	0.474	0.355	0.124	0.128

Table 5.10.: RMSE [m] with respect to the xy -positioning for the localization using various loss functions. The results are based on correspondences in the Euclidean space. The localization solution corresponds to the center point of the grid cell with the smallest loss.

5.3.2. Localization using different loss functions

In Sec. 4.3, it is described that the localization framework can be based not only on the inlier count and the point-to-plane adjustment score, but also on different loss functions. The loss functions shown in Fig. 2.2 were evaluated for the purpose of vehicle localization and compared with respect to accuracy, robustness, and reliability. For the purpose of comparison, the L0 loss was also considered (cf. Eq. 2.12). This section presents the results of the evaluations.

The loss functions are applied to the residuals in both observation and Euclidean space. Since the results of both approaches are similar, only the results based on correspondences in the Euclidean space are presented and discussed in this section, and the results based on correspondences in the observation space are stated in the appendix in Sec. A.1. Similar to the comparison of the count and score, the accuracy is evaluated based on the RMSE value of the localization solution with respect to the ground truth, the robustness is evaluated based on the share of epochs with an error greater than the alert limit, and the reliability is evaluated based on the kurtosis, second peak ratio, and largest peak distance. However, in contrast to the experiments in Sec. 5.3.1, only the xy -position is considered in this investigation and the heading is excluded for simplicity. The xy -position of the center point of the grid cell yielding the smallest loss represents the localization solution with respect to the utilized loss function and is compared to the ground truth position.

For the experiments, the size of the xy -grid is set to $2m \times 2m$ (from $-1m$ to $+1m$ for both axes), which, with a discretization of $0.1m$, leads to a grid of 21×21 candidate positions. For computational reasons, only every second epoch is considered in the experiments. For the Huber, Cauchy, Geman-McClure, Welsch, and Tukey loss functions, the shape parameter c is set to $c = 1$.

Position accuracy:

In Tab. 5.10, for the different loss functions, the RMSE values [m] with respect to the xy -positioning are presented. Similar to the count and score, for each loss function, the RMSE values are smallest for the *Nordstadt* data set, increasing for the *Haltenhoffstraße* data set, and increasing even more for the *Highway* data set. The reason is that between these data sets, the amount of static objects that facilitate localization decreases and the amount of dynamic objects that cause outliers increases. When comparing the different loss functions, it becomes clear that (apart from the Cauchy loss in the *Nordstadt* data set and L0 loss on the *Highway* data set) the RMSE values decrease for each data set the more robust the loss function is, i.e. the smaller the influence of large residuals is. With regard to the L1 and Huber loss, it should be noted that the Huber loss performs worse than the L1 loss due to the quadratic part for residuals of up to $1m$. Furthermore, it is noteworthy that Geman-McClure and Welsch as well as Tukey and L0 loss show very similar RMSE values.

Position robustness:

Considering the shares of epochs with an xy -error greater than the position alert threshold of $0.29m$, which are stated in Tab. 5.11, it can be seen that these are in line with the RMSE values. Between the three data sets, the frequency of positioning failures increases from *Nordstadt* to *Haltenhoffstraße* and from *Haltenhoffstraße* to *Highway*. While in *Nordstadt*, apart from L2 and

Data set	L2	L1	Huber	Cauchy	G.-McC.	Welsch	Tukey	L0
Nordstadt	0.831	0.0	0.181	0.0	0.0	0.0	0.0	0.0
Haltenhoffstr.	0.976	0.168	0.693	0.12	0.0	0.0	0.0	0.0
Highway	0.998	0.959	0.993	0.908	0.317	0.242	0.034	0.044

Table 5.11.: Share of epochs with an xy -error greater than the position alert threshold for various loss functions. The results are based on correspondences in the Euclidean space. The localization solution corresponds to the center point of the grid cell with the smallest loss.

Huber, all loss functions are able to localize without causing a positioning failure, only Geman-McClure, Welsch, Tukey, and L0 achieve this in *Haltenhoffstrasse*, and no loss function is capable of localizing failure-free on the *Highway*. The frequency of positioning failures generally decreases the more robust the loss function is. The L2 loss clearly causes the highest rate of positioning failures, followed by Huber, while Tukey and L0 lead to the lowest failure rates. Even on the *Highway* with a failure rate of 3.4% and 4.4%, respectively, they lead to correct localization solutions in most epochs. Overall, both the RMSE values and the shares of positioning failures are consistent with the shape and robustness of the different loss functions, and they confirm the different sensitivities to outliers.

For the sake of completeness, the RMSE values and shares of positioning failures for localization using the different loss functions based on correspondences in the observation space are stated in the appendix in Tab. A.1 and Tab. A.2, respectively.

In Fig. 5.14, for a section of 50 epochs from the *Highway* data set, the localization errors using the different loss functions are shown. The graph reflects the results stated in Tab. 5.10 and Tab. 5.11. The L2, L1, Huber, and Cauchy loss do not lead to correct localization results in case of a localization scenario as challenging as the *Highway* data set. Using the Geman-McClure and Welsch loss, the performance is already significantly better, however, the estimated positions still deviate from the correct positions in most epochs, sometimes significantly. Only Tukey and L0 loss lead to localization solutions that match or are at least very close to the ground truth.

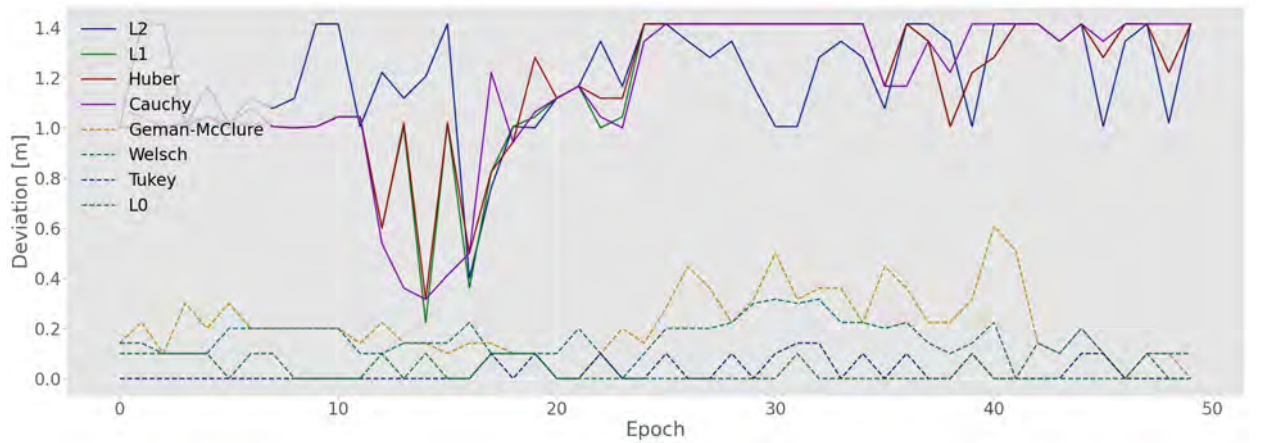


Figure 5.14.: Localization errors using the different loss functions for a section of 50 epochs from the *Highway* data set. The L2, L1, Huber, and Cauchy loss show significant localization errors and the estimated solutions are not even close to the ground truth. Using the Geman-McClure and Welsch loss, the performance is already significantly better, however, the estimated positions still deviate from the correct positions in most epochs, sometimes significantly. Only Tukey and L0 loss lead to localization solutions that match or are at least very close to the ground truth.

Kurtosis								
Data set	L2	L1	Huber	Cauchy	G.-McC.	Welsch	Tukey	L0
Nordstadt	-1.072	-0.67	-0.877	-0.798	-0.736	-0.778	-0.588	-0.07
Haltenhoffstr.	-1.133	-0.85	-1.002	-0.846	-0.777	-0.801	-0.655	0.032
Highway	-1.203	-1.173	-1.18	-1.115	-0.891	-0.86	-0.711	-0.482

Second loss ratio								
Data set	L2	L1	Huber	Cauchy	G.-McC.	Welsch	Tukey	L0
Nordstadt	1.001	1.006	1.001	1.002	1.005	1.005	1.012	1.114
Haltenhoffstr.	1.001	1.001	1.0	1.0	1.001	1.001	1.002	1.024
Highway	1.001	1.0	1.0	1.0	1.0	1.0	1.0	1.003

Largest peak distance								
Data set	L2	L1	Huber	Cauchy	G.-McC.	Welsch	Tukey	L0
Nordstadt	1.376	0.821	1.152	0.867	0.444	0.448	0.251	0.101
Haltenhoffstr.	1.414	1.387	1.413	1.368	1.141	1.089	0.809	0.658
Highway	1.413	1.412	1.413	1.411	1.391	1.384	1.339	1.366

Table 5.12.: Mean of kurtosis values, second loss ratios, and largest peak distances for the localization using different loss functions. The L0 loss yields the highest kurtosis values and the largest second loss ratios, and thus the most distinct localization solutions. Furthermore, the L0 loss yields the smallest distances between the ground truth position and a grid cell that yields a loss that is at most 10% greater than the smallest loss, but the distances are still significantly larger than the distances of the point-to-plane adjustment score (cf. Tab. 5.9).

Reliability:

Similar to the count and score, to assess the reliability of the localization solution, the distribution within the evaluated search space is analyzed using the kurtosis, the second peak ratio or, in this case, the second loss ratio, and the largest peak distance. To calculate the kurtosis of the loss accumulators, they are first multiplied by -1 , so that the minimum loss becomes the maximum value. The second loss ratio is calculated as the ratio between the smallest loss and second smallest loss. Consequently, it is greater than one and, in contrast to the second peak ratio, the greater the value, the more reliable the localization solution. The largest peak distance is calculated between the ground truth position and a grid cell that yields a loss that is at most 10% greater than the smallest loss.

In Tab. 5.12, the mean values of the kurtosis values, second loss ratios, and largest peak distances for the localization using the different loss functions are shown. Comparing the loss functions with regard to the kurtosis, it becomes clear that the kurtosis values increase in the overall trend with increasing robustness of the loss function, with the L0 loss leading to the highest kurtosis values and thus the most distinct localization solutions. Comparing the second loss ratios, it is noticeable that the values are all very similar and close to one. No clear trend can be identified, which is why this measure appears unsuitable in this case. Only the L0 loss shows second loss ratios that are somewhat more clearly greater than one, which is why one could conclude that it achieves the most distinct solutions. In principle, however, the second loss ratios should not be over-weighted and are mentioned more for the sake of completeness. Comparing the largest peak distances, the trend can again be seen that the largest peak distances decrease with increasing robustness of the loss function and that the L0 loss leads to the smallest distances, although these are still significantly larger than the distances of the point-to-plane adjustment score (cf. Tab. 5.9).

Data set	σ_{hit}	λ_{short}	ω_{hit}	ω_{short}	ω_{rand}	ω_{max}
Nordstadt	0.171	0.095	0.588	0.132	0.152	0.128
Haltenhoffstr.	0.231	0.071	0.483	0.166	0.253	0.098
Highway	0.551	0.083	0.488	0.191	0.191	0.13

Table 5.13.: Estimated intrinsic parameters for the three data sets.

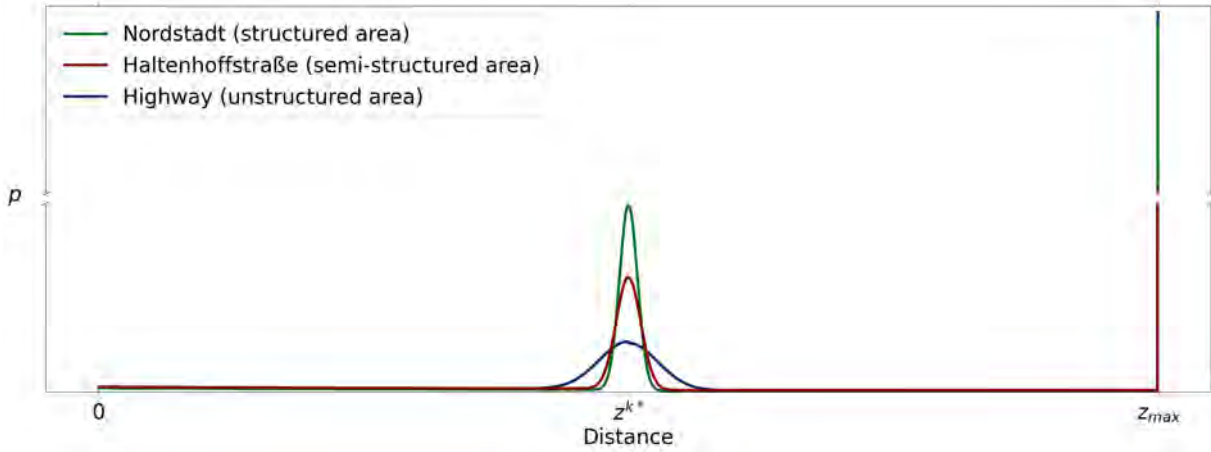
Beam model-based localization

As described in Sec. 4.5, the likelihood distribution $p(z|\mathbf{x}, \mathcal{M})$ from the beam model, which was originally intended for estimating the protection level in this work, can also be used as data-adapted loss function $\rho(z|\mathbf{x}, \mathcal{M}) = -\log p(z|\mathbf{x}, \mathcal{M})$ for the localization. As the beam model is applied to range measurements, the corresponding localization is based on correspondences in the observation space. Thus, the car sensor scan is represented as a 2D scan image and a synthetic range image is rendered for each candidate pose. The prerequisite for the localization is a set of beam model parameters Θ , which is determined individually for each data set to account for the different environments. For each epoch, a synthetic range image is rendered based on the ground truth pose from the RIEGL MMS and the triangle mesh, which is used together with the measured range image to estimate the six intrinsic parameters. Finally, using all epochs from a data set, the mean value for each parameter is computed, which represents the final intrinsic parameter. For the three data sets, the final intrinsic parameters are stated in Tab. 5.13. The corresponding beam models, underlying histograms for the beam model parameter estimation, and loss functions based on the beam models are visualized in Fig. 5.15. The histograms shown in Fig. 5.15b illustrate the frequency of the measured distance z^k relative to the expected distance z^{k*} . For visualization purposes, the bin size of the histograms in Fig. 5.15b is ten times larger than the sampling of the beam model distributions in Fig. 5.15a, which is why the peak of the maximum range response appears larger in Fig. 5.15a and smaller in Fig. 5.15b relative to the normally distributed part.

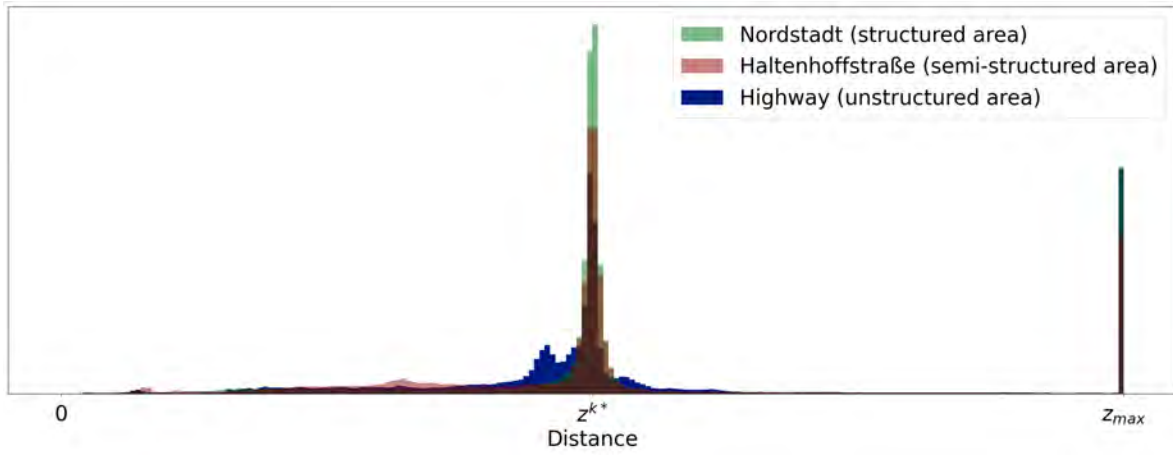
Looking at the individual intrinsic parameter values, the standard deviation σ_{hit} increases from *Nordstadt* to *Haltenhoffstraße* and from *Haltenhoffstraße* to *Highway*, while the weight of the normal distribution ω_{hit} is lower for *Haltenhoffstraße* and *Highway* compared to *Nordstadt*. Accordingly, the proportion of the normal distribution between the three data sets becomes smaller and the appearance of the bell curve becomes flatter and broader, as can be seen in Fig. 5.15a. Furthermore, the weight of the exponential distribution ω_{short} increases between the three data sets, which can be seen in Fig. 5.15c.

The reasons are that in the *Nordstadt* the environment is primarily characterized by flat building facades, which can be mapped more accurately by the mesh compared to the signs and tree trunks in *Haltenhoffstraße* and *Highway*, and that in the *Nordstadt* only very few occlusions by other road users occur, so that the majority of the LiDAR beams hit the map objects, while especially on the *Highway* many beams scan other road users.

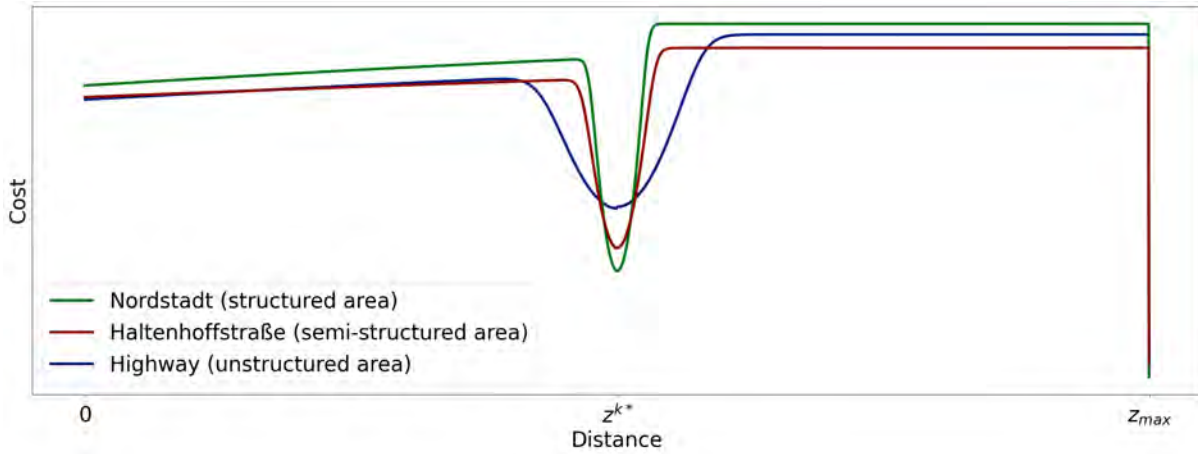
In the histogram and also in the beam model estimation, only correspondences, or more specifically scan image pixels, are taken into consideration where an expected measurement exists. If there is a real measurement in addition to the expected measurement, the correspondence contributes to the normally distributed part, the random part, and the exponential part, but only to the latter if the measured range is smaller than the expected range. If there is only an expected but no real measurement, the ‘correspondence’ contributes to the maximum range response. The case where there is a real measurement but no expected measurement is excluded. This applies to measurements of parts that are removed from the map, such as the ground, vegetation, and parked cars, but also to measurements of other dynamic traffic participants. The inclusion of this



(a) Estimated beam models.



(b) Underlying histograms for the beam model parameter estimation.



(c) Loss functions based on the beam models.

Figure 5.15.: Visualization of (a) the estimated beam models of the three data sets, (b) the underlying histograms for the beam model parameter estimation, and (c) the corresponding loss functions based on the beam models. The intrinsic parameters of the beam models are stated in Tab. 5.13.

case would lead to a significantly higher weighting of the exponential component, but since this does not reflect the distribution of the residuals, the inclusion is incorrect and is not carried out. The case where neither an expected nor a real measurement exists is also neglected.

Data set	RMSE		Failures	
	Position [m]	Heading [°]	Position	Heading
Nordstadt	0.084	0.103	0.002	0.0
Haltenhoffstr.	0.097	0.098	0.0	0.001
Highway	0.536	0.349	0.543	0.167

Table 5.14.: RMSE values with respect to the xy -positioning [m] and the heading [°] and shares of epochs with an xy -error greater than the position alert threshold and a heading error greater than the heading alert threshold for the localization using the beam model for range sensors.

Looking at Fig. 5.15c, it becomes clear that the condition for an M-estimator that there is only one unique minimum is not fulfilled. Nevertheless, the almost constant parts outside the normally distributed range result in a robust characteristic of the loss function, which is why it could be suitable for accurate and robust localization of the vehicle. Therefore, the beam model-based loss function is applied to the localization problem considering the xy -position and heading. Similar to the other loss functions, the pose with the minimum loss within the loss accumulator is considered as localization solution, and the accuracy and robustness are evaluated using the RMSE and share of positioning failures, respectively. In Tab. 5.14, for all three data sets the RMSE values and shares of positioning failures for the position and heading are stated.

Considering the position, compared to the maximum consensus count based on correspondences in the observation space, the RMSE value is slightly worse for the *Nordstadt*, slightly better for the *Haltenhoffstraße*, and significantly worse for the *Highway* (cf. Tab. 5.2), and the share of positioning failures is slightly better for the *Nordstadt* and *Haltenhoffstraße*, and significantly worse for the *Highway* (cf. Tab. 5.3). Considering the heading, compared to the maximum consensus count based on correspondences in the observation space, the RMSE value and share of positioning failures are better for the *Nordstadt* and *Haltenhoffstraße*, but worse for the *Highway* (cf. Tab. 5.6 and Tab. 5.7). Thus, overall, there is no advantage of using the beam model over the inlier count. The problem of the beam model-based localization, as it is applied here, is that it relies on only a single set of intrinsic parameters for each scenario, while the actual distribution of measurements varies considerably throughout a scenario due to changes in the static environment and the presence of other road users. Thus, as the beam model only represents an average distribution, it often does not represent the actual distribution of measurements, and instead strongly deviates from it. Given the results in Tab. 5.14, this leads to a poor localization performance especially for the *Highway* data set. To improve the performance, a localization approach based on a beam model with updated parameters can be employed. This ensures that changes in the environment are taken into account, and the assumed distribution of measurements matches the true distribution more closely.

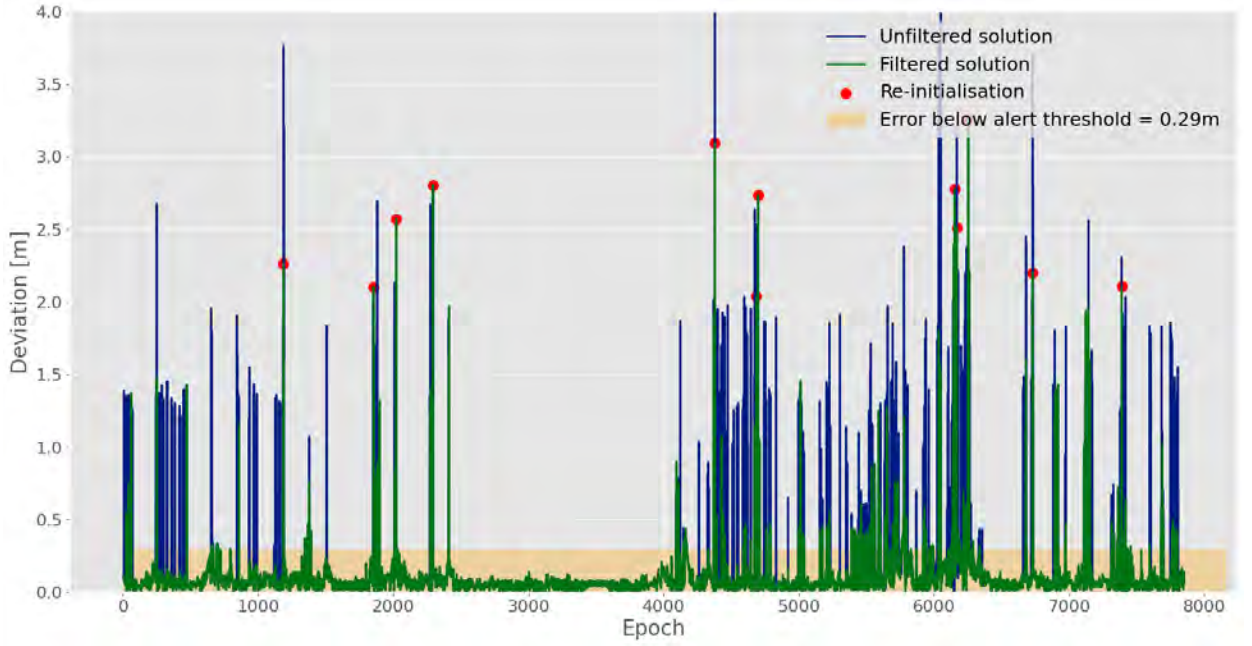
5.3.3. Maximum consensus filter

In this section, the evaluation results of the maximum consensus filter are presented, which was introduced in Sec. 4.4. The maximum consensus filter is a histogram filter that has the objective to make the localization even more robust by utilizing the knowledge from previous epochs. For the prediction, a constant velocity model is applied and the induced uncertainty is modeled using a Gaussian kernel with $\sigma = 5 \text{ cm}$. To avoid jumps in the estimated velocity due to incorrect localization solutions in individual epochs, the velocity is determined over the last 10 epochs. The key to transfer the loss accumulator containing the outlier counts into a discrete probability distribution is the truncation proposed by Chebrolu et al. (2020), which leads to the truncated probability distribution $\tilde{P}(r, \alpha, c)$ on which the maximum consensus filter is based. In order to compensate

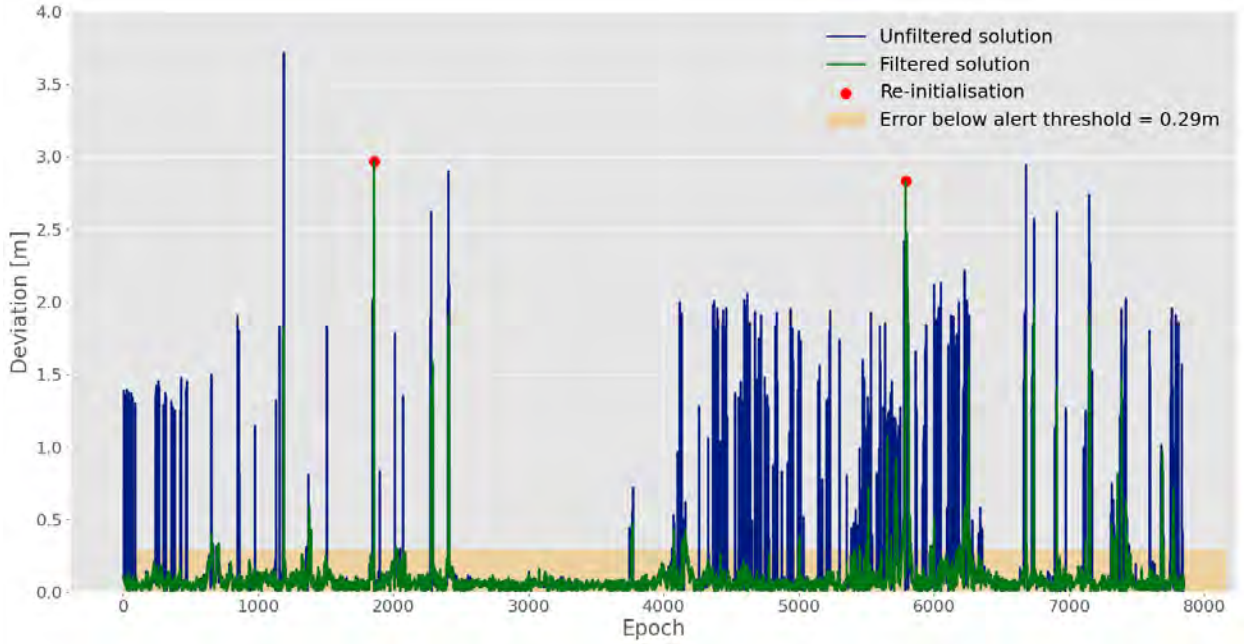
for neglecting the correlations and thus avoid overly optimistic probability distributions, the correlation quotient was introduced as additional parameter. The higher the correlation quotient is selected, the smaller is the effective number of measurements and the broader and flatter is the probability distribution. Since this is a plain simplification compared to the exact determination of the correlations, the aim is to show how the parameter influences the localization performance and whether it can improve it, which is why different values are tested in the experiments. Thus, the maximum consensus filter has two tuning parameters: the standard deviation σ of the prediction and the correlation quotient.

As the maximum consensus localization only leads to positioning errors greater than the alert threshold for the *Highway* data set (cf. Tab. 5.3, Euclidean space), this comparatively challenging scenario is selected to test the performance of the maximum consensus filter. As already mentioned, only the xy -position is considered for the filter, and to allow a methodological investigation without external error sources, the filter is applied to the localization based on correspondences in the Euclidean space. In Fig. 5.16, the deviations of the xy -positions with respect to the ground truth positions are shown for the unfiltered solution, which is based on the accumulator using only the measurement, and the filtered solution, which also takes the prediction into account. The red dots mark the epochs in which the deviation from the ground truth position becomes so large that the search space no longer includes it, which is why a re-initialization is necessary and carried out in such epochs. The orange area illustrates the size of the error up to the alert threshold of 0.29 m . In Fig. 5.16a, the results are shown for a correlation quotient of 1, which means that all measurements are taken into account, and in Fig. 5.16b, the results are shown for a correlation quotient of 20, which means that the number of effective measurements taken into account is $1/20$ of the number of actual measurements. The presented results exemplify the application of the effective measurements. A correlation quotient of 20 is used because it leads to a significant improvement compared to the correlation quotients of 5 and 10, and a higher correlation quotient of 30 results in a deterioration again. In Tab. 5.15, the corresponding RMSE values and failure rates are stated for the filter that takes into account all real measurements and for the filter that takes into account only the effective measurements. In both cases, the unfiltered solution leads to a significant number of localization failures, which can be seen in Fig. 5.16 and is confirmed by the RMSE values of 0.39 m and 0.359 m and failure rates of 0.088 and 0.083 stated in Tab. 5.15. Looking at the filtered solutions, both filters already lead to an enormous improvement compared to the unfiltered solutions, which is confirmed by the RMSE values of 0.225 m and 0.188 m and failure rates of 0.055 and 0.037, even though neither is yet able to complete the journey completely failure-free. If the number of re-initializations is considered, a positive effect of applying the correlation quotient becomes apparent, as the number reduces from 11 when taking all measurements into account to 2 when taking only the effective number of measurements into account. The benefit also becomes clear when looking at the RMSE values and failure rates in Tab. 5.15, which improve from 0.225 m to 0.188 m and from 0.055 to 0.037, respectively. In any case, the advantage over the unfiltered solution is even greater and represents the main advantage and contribution of the maximum consensus filter, regardless of whether and to what extent the correlation quotient is applied.

Finally, with regard to the RMSE values and positioning failures, it should be noted that these are larger for both the filtered and unfiltered solutions compared to the single pose estimation in Sec. 5.3.1 (cf. Tab. 5.2 and Tab. 5.3). This is due to the fact that for the single pose estimation, in contrast to here, the initialization was carried out in the ground truth pose in each epoch and the grid shift was applied, both of which have a positive effect on the susceptibility to errors. In addition, the initialization bias has a favorable effect, since the error is zero if the ground truth pose is determined as localization solution. Therefore, the values in Tab. 5.15 should be used to compare the filtered and unfiltered solutions with each other.



(a) Correlation quotient = 1.



(b) Correlation quotient = 20.

Figure 5.16.: Localization results of the maximum consensus filter using a correlation quotient of 1 shown in (a), which means that all measurements are taken into account, and of 20 shown in (b), which means that the number of effective measurements taken into account is 1/20 of the number of actual measurements. The deviations of the xy -positions with respect to the ground truth positions are shown for the unfiltered solution, which is based on the accumulator using only the measurement, and the filtered solution, which also takes the prediction into account. The red dots mark the epochs in which the deviation from the ground truth position becomes so large that the search space no longer includes it, which is why a re-initialization is necessary and carried out in such epochs. The orange area illustrates the size of the error up to the alert threshold of 0.29 m.

To illustrate the benefit of the maximum consensus filter, the probability grids of the prediction, measurement, and update are visualized for an example epoch in Fig. 5.17. The probability distri-

Data set	Real measurements		Effective measurements	
	RMSE [m]	Failures	RMSE [m]	Failures
Unfiltered position	0.39	0.088	0.359	0.083
Filtered position	0.225	0.055	0.188	0.037

Table 5.15.: RMSE values [m] and failure rates with respect to the xy -positioning for the maximum consensus localization (unfiltered position) and the maximum consensus filter (filtered position). The columns of ‘Real measurements’ refer to the filter that takes all actual measurements into account and the columns of ‘Effective measurements’ refer to the filter that only takes the effective measurements into account. The localization solution corresponds to the center point of the grid cell with the highest consensus value or probability.

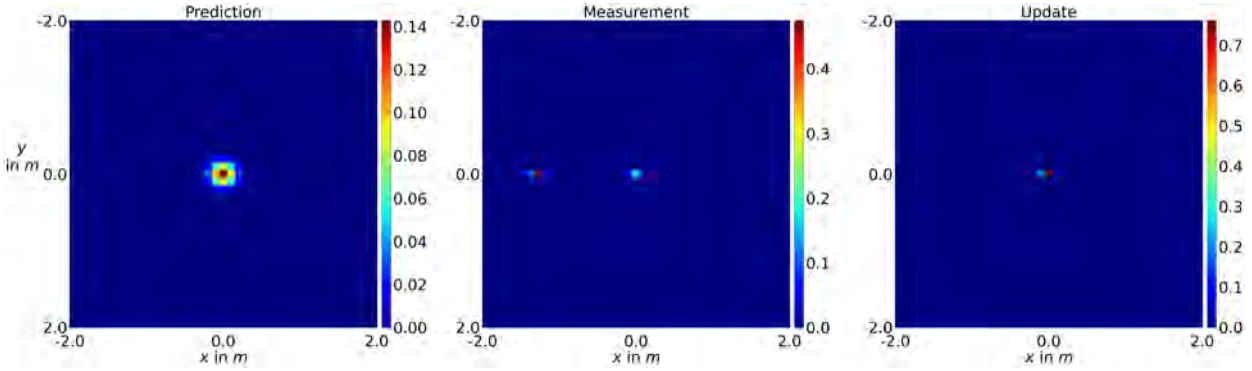
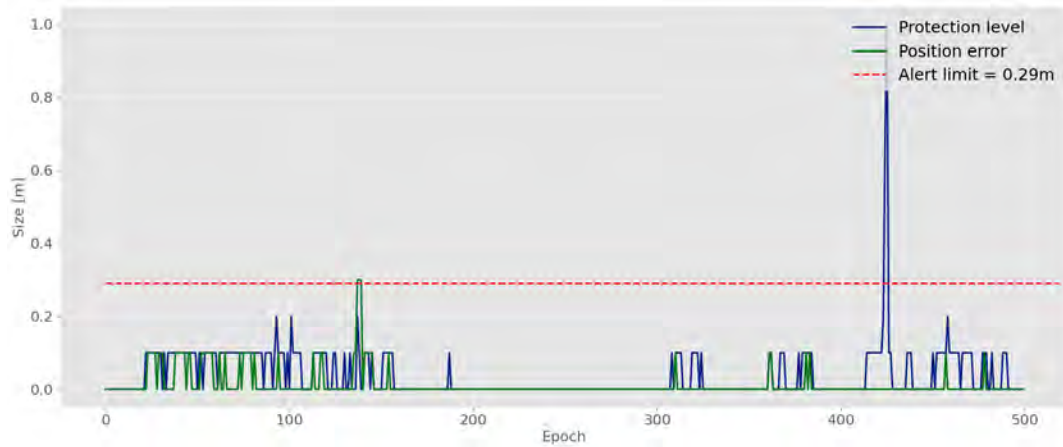


Figure 5.17.: Probability grids of the prediction, measurement, and update for an example epoch to illustrate the benefit of the maximum consensus filter. The probability distribution of the prediction is rather broad and flat, with a maximum probability of approximately 0.14 located at the ground truth position. The probability distribution of the measurement shows two hypotheses, where the one with the higher probability is far away from the ground truth position and the one with the smaller probability is at the ground truth position, which is why the maximum consensus localization, which only considers the measurement, would lead to a large error in this case. However, the combination of prediction and measurement, which results in the probability distribution of the update, leads to a localization solution of the filter at the ground truth position.

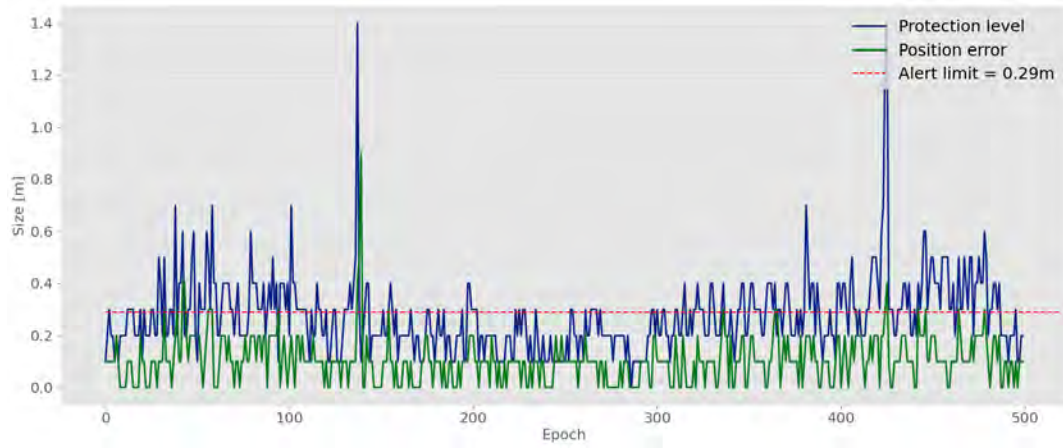
bution of the prediction is rather broad and flat, with a maximum probability of approximately 0.14 located at the ground truth position. The probability distribution of the measurement shows two hypotheses, where the one with the higher probability is far away from the ground truth position and the one with the smaller probability is at the ground truth position, which is why the maximum consensus localization, which only considers the measurement, would lead to a large error in this case. However, the combination of prediction and measurement by multiplying the probabilities and subsequent normalization, which results in the probability distribution of the update, leads to a localization solution of the maximum consensus filter at the ground truth position.

5.3.4. Protection level estimation and integrity evaluation

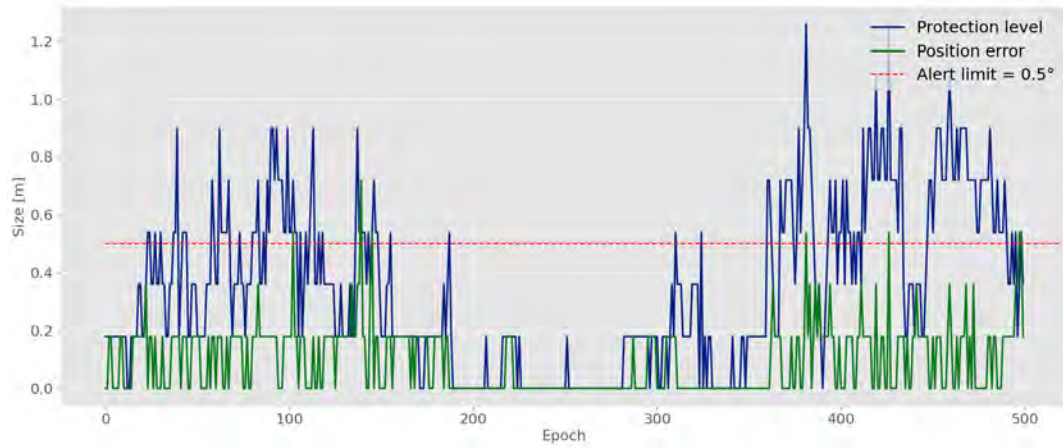
In this section, the results of the protection level estimation and integrity evaluation are presented, which were introduced in Sec. 4.5. The protection levels are defined as the maximum extent of candidate poses with a probability of $p > 1 - 10^{-8}$ starting from the estimated localization solution in longitudinal, lateral, and heading directions. The probabilities assigned to the candidate poses are either based on the truncated probability distribution $\tilde{P}(r, \alpha, c)$ using the L0 loss, which is the outlier count, or based on the beam model from Thrun et al. (2005), which is shown in Fig. 5.15 and where the intrinsic parameters are stated in Tab. 5.13. Based on the relations between the true position error, the estimated protection level and the pre-defined alert limit, the integrity levels of the system can be derived, which correspond to the four operational states *Nominal Operation*



(a) Lateral direction.



(b) Longitudinal direction.



(c) Heading.

Figure 5.18.: The protection level, position error, and alert limit are shown for the lateral direction, longitudinal direction, and heading for a section of 500 epochs. In general, the protection level is larger than the position error in most epochs, however, not in all epochs, which shows that besides the main proportion of Nominal Operation also a proportion of Misleading Information exists. Whereas the protection level is smaller than the alert limit in almost all epochs in lateral direction, which indicates a small share of the state Unavailable, the protection level exceeds the alert limit frequently in longitudinal direction, which indicates a comparatively high share of the state Unavailable. Concerning the state of Hazardously Misleading Information, it exists, though not often, in all three directions, since the position error exceeds the alert limit, while the protection level does not.

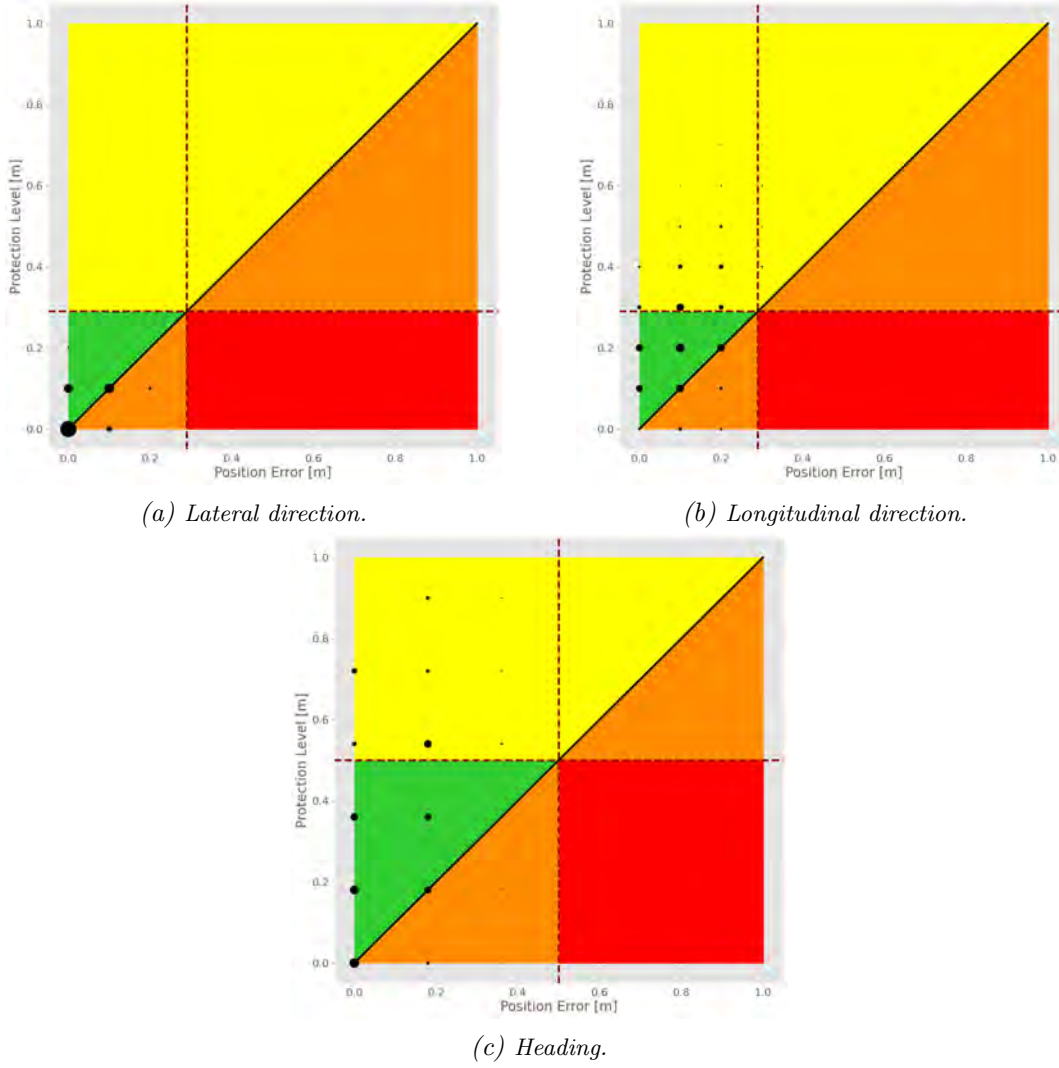


Figure 5.19.: Stanford-ESA Integrity Diagrams for the Haltenhoffstraße for the lateral direction, longitudinal direction, and heading (axes cut at 1 m). The protection levels are based on the L0 loss and a correlation quotient of 10 was applied in the probability estimation. The diagrams are based on the protection levels and position errors presented in Fig. 5.18. Due to the discrete nature of the results, the point size is used to encode the number of epochs belonging to each protection level - position error - pair.

(NO), Unavailable (UA), Misleading Information (MI), and Hazardously Misleading Information (HMI) that can be represented using the Stanford-ESA Integrity Diagram. In the probability estimation, a correlation quotient can be applied, which can be considered as a tuning parameter of the integrity system, since it adjusts the size of the protection level and thereby the ratios between the four operational states.

To illustrate how it is differentiated between the four states and how the final Stanford diagrams are determined, for the lateral direction, longitudinal direction, and heading, the protection level, position error, and alert limit are plotted for 500 epochs in Fig. 5.18. The shown results are based on the data set *Haltenhoffstraße*, the probabilities assigned to the candidate poses are based on the truncated probability distribution $\tilde{P}(r, \alpha, c)$ using the L0 loss, and in the probability estimation a correlation quotient of 10 was applied, where the influence of this parameter is evaluated in more detail later in this section. In general, the protection level is larger than the position error in most epochs, however, not in all epochs, which shows that besides the main proportion of *Nominal*

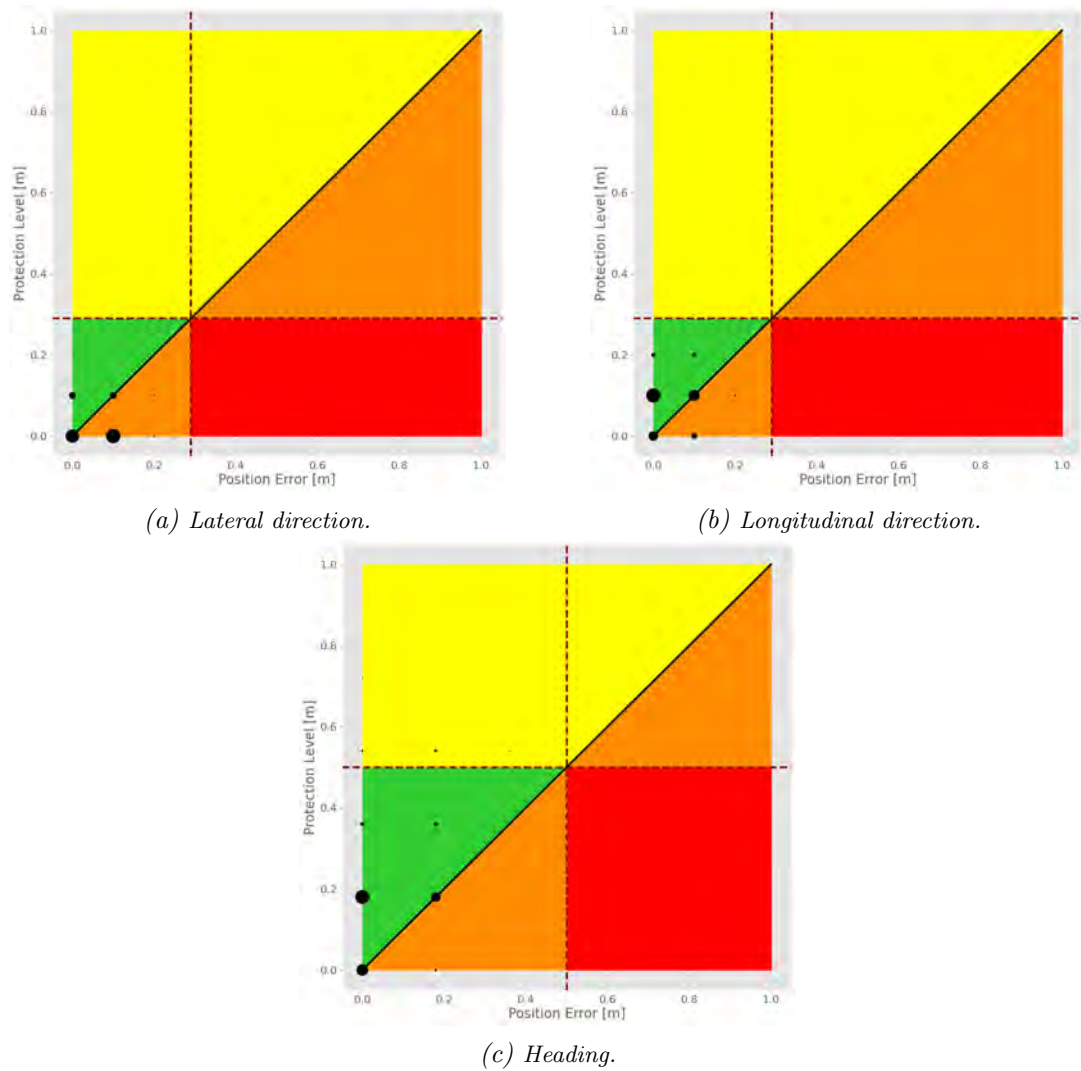


Figure 5.20.: Stanford-ESA Integrity Diagrams for the Haltenhoffstraße for the lateral direction, longitudinal direction, and heading (axes cut at 1 m). The protection levels are based on the beam model and a correlation quotient of 10 was applied in the probability estimation. Due to the discrete nature of the results, the point size is used to encode the number of epochs belonging to each protection level - position error - pair.

Operation also a proportion of *Misleading Information* exists. Whereas the protection level is smaller than the alert limit in almost all epochs in lateral direction, which indicates a small share of the state *Unavailable*, the protection level exceeds the alert limit frequently in longitudinal direction, which indicates a comparatively high share of the state *Unavailable*. Concerning the state of *Hazardously Misleading Information*, it exists, though not often, in all three directions, since the position error exceeds the alert limit, while the protection level does not.

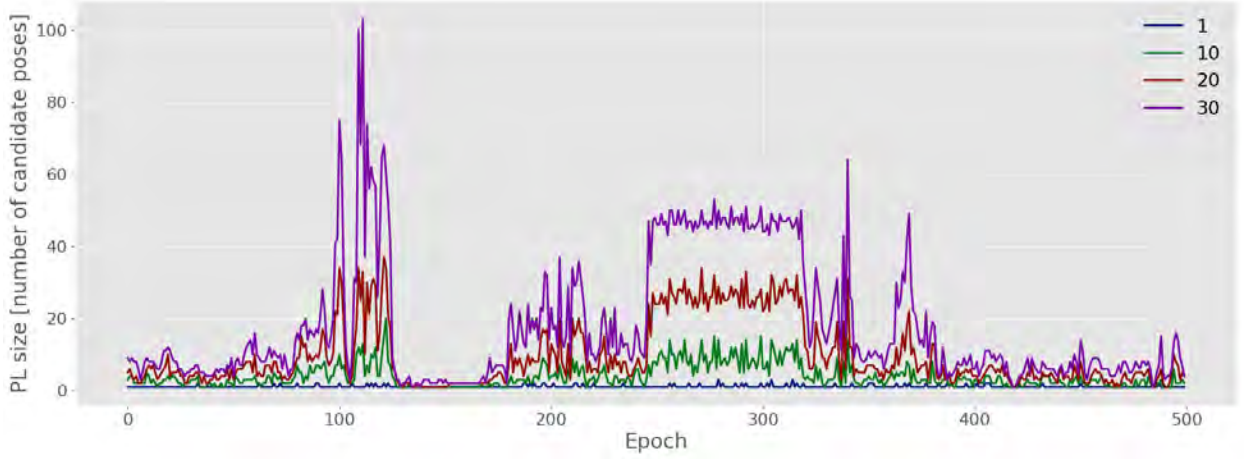
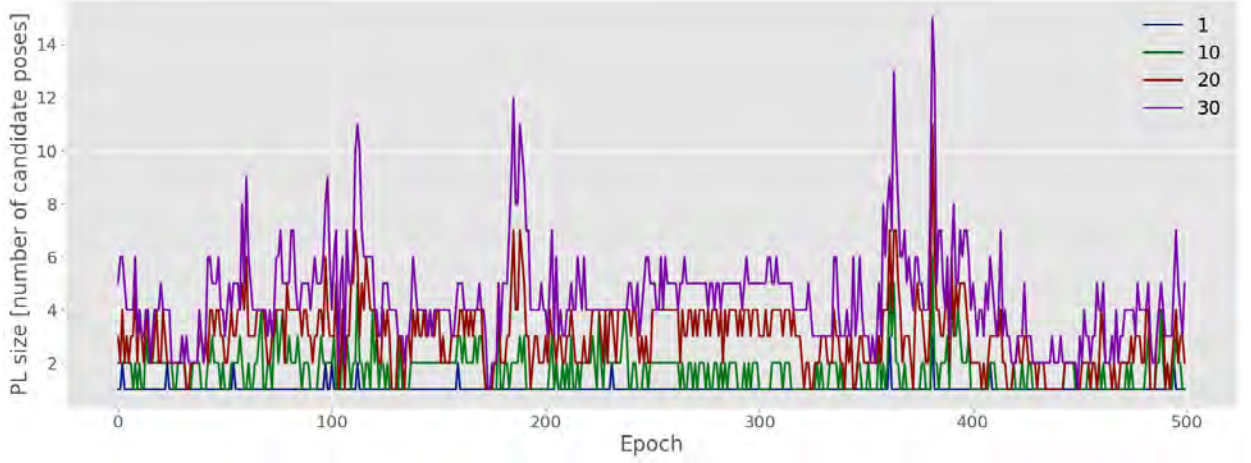
In Fig. 5.19, the Stanford-ESA Integrity Diagrams are shown for each axis, which are based on the protection levels and position errors presented in Fig. 5.18. The green area represents the state *Nominal Operation*, the yellow area the state *Unavailable*, the orange area the state *Misleading Information*, the red area the state *Hazardously Misleading Information*, and the red dotted line the alert limit. Due to the discrete nature of the results, the point size is used to encode the number of epochs belonging to each protection level - position error - pair to be able to see the proportions in the results. The epochs on the main diagonal belong to the states *Nominal Operation* or *Unavailable*, which is also important for the quantitative analysis in the following.

Underlying model		L0 loss				Beam model			
Data set	Direction	NO	UA	MI	HMI	NO	UA	MI	HMI
Nordstadt	Lateral	0.87	0.0	0.13	0.0	0.665	0.0	0.335	0.0
	Longitudinal	0.954	0.02	0.024	0.002	0.883	0.0	0.115	0.002
	Heading	0.74	0.11	0.15	0.0	0.937	0.0	0.063	0.0
Haltenhoffstr.	Lateral	0.924	0.003	0.069	0.004	0.557	0.001	0.442	0.0
	Longitudinal	0.597	0.342	0.05	0.011	0.924	0.002	0.074	0.0
	Heading	0.666	0.306	0.024	0.004	0.958	0.03	0.011	0.001
Highway	Lateral	0.851	0.092	0.046	0.011	0.672	0.136	0.101	0.091
	Longitudinal	0.324	0.623	0.041	0.012	0.233	0.55	0.191	0.026
	Heading	0.439	0.53	0.017	0.014	0.328	0.636	0.015	0.021

Table 5.16.: Results of the integrity evaluation using a correlation quotient of 10 in the probability estimation. For the three data sets, the proportions of the four operational states for the lateral direction, longitudinal direction, and heading are shown, on the one hand based on the L0 loss and on the other hand based on the beam model.

In Tab. 5.16, the proportions of the four operational states are stated for the three data sets, on the one hand based on the L0 loss and on the other hand based on the beam model. The proportions of the *Haltenhoffstraße* based on the L0 loss are the ones which are illustrated in Fig. 5.19. For illustration and comparison purposes, the proportions of the *Haltenhoffstraße* based on the beam model are visualized in Fig. 5.20, also using Stanford-ESA Integrity Diagrams. When comparing the integrity levels based on the L0 loss and beam model for the *Haltenhoffstraße*, the first aspect which becomes clear is that the protection levels are significantly smaller when using the beam model, which results in much smaller proportions for the state *Unavailable*. However, as the position errors are not necessarily smaller using the beam model, this comes at the cost of higher proportions of the state *Misleading information*. Nevertheless, using the beam model, the large position errors disappear, which has the positive effect that almost no *Hazardously Misleading Information* occur. For the Nordstadt, the results also show that when using the beam model, the proportions of the state *Unavailable* become smaller and the proportions of the state *Misleading information* in total larger. Overall, for the *Nordstadt* and *Haltenhoffstraße*, the beam model leads to more optimistic probability distributions due to its strongly pronounced normally distributed part (cf. Fig. 5.15). For the Highway, the first finding is that the proportions of the state *Nominal Operation* are much smaller and the proportions of the state *Unavailable* much larger than those of the *Haltenhoffstraße* and *Nordstadt*, both based on the L0 loss and based on the beam model. This is in line with the much worse localization results stated in Tab. 5.2 and Tab. 5.14 and underlines that the localization is much more difficult and uncertain in this scenario. The failure rates for the localization based on the inlier count and beam model stated in Tab. 5.3 and Tab. 5.14, respectively, have shown that the position error exceeds the alert limit in at least 25.2% and 54.2% of the epochs (only with respect to the position). While these results are completely unacceptable in terms of robust localization, it is positive to note that, especially based on the L0 loss, most of these epochs are correctly recognized as unavailable, leading to the correspondingly high proportions of the state *Unavailable* and the significantly lower proportions of the state *Hazardously Misleading Information*, at about 1% for each of the three axes. Furthermore, the proportions of the state *Misleading Information* are also comparatively small and of a similar order of magnitude to the other two data sets.

In general, the beam model-based localization has the disadvantage that it is based on a single set of intrinsic parameters, while the actual distribution in the measurements changes considerably between individual epochs due to the changing static environment, but also due to the rapidly

(a) Protection level sizes based on $L0$ loss.

(b) Protection level sizes based on beam model.

Figure 5.21.: Size of the protection level, which is represented as the number of candidate poses with a probability of $p > 1 - 10^{-8}$, shown for correlation quotients of 1, 10, 20, and 30 for 500 epochs of the Nordstadt data set. A correlation quotient of 1 corresponds to the case that all real measurements are taken into account. The protection level sizes shown in (a) are based on the $L0$ loss and the protection level sizes shown in (b) are based on the beam model. For both approaches, the size of the protection level increases with a higher correlation quotient.

changing appearance of other road users. Since the beam model parameters do not reflect such dynamic changes, which occur especially for the *Highway* data set in this work, the underlying distribution often strongly deviates from the actual distribution. On the one hand, this leads to a localization that is no longer robust, resulting in significantly poorer localization performance. On the other hand, it leads to protection levels that are not correctly estimated, since they are based on a wrong distribution. Consequently, the proportion of the state of *Nominal Operation* is smaller for the *Highway* when using the beam model compared to the $L0$ loss and, due to the incorrect protection levels, the proportions of the states *Misleading Information* and *Hazardously Misleading Information* are larger. As already suggested for the beam model-based localization, one possible way to improve not only the localization performance but also the protection level estimation could be to constantly update the parameters of the beam model instead of using a fixed set, so that changes in the environment are taken into account, making the beam model more consistent with the actual distribution of the measurements and the protection levels more accurate.

Underlying model		L0 loss				Beam model			
Data set	Direction	NO	UA	MI	HMI	NO	UA	MI	HMI
Nordstadt	Lateral	0.774	0.0	0.226	0.0	0.636	0.0	0.364	0.0
	Longitudinal	0.846	0.001	0.148	0.005	0.766	0.0	0.232	0.002
	Heading	0.506	0.005	0.485	0.004	0.776	0.0	0.224	0.0
Haltenhoffstr.	Lateral	0.745	0.001	0.249	0.005	0.479	0.0	0.521	0.0
	Longitudinal	0.381	0.027	0.564	0.028	0.688	0.0	0.312	0.0
	Heading	0.684	0.013	0.289	0.014	0.794	0.0	0.205	0.001
Highway	Lateral	0.788	0.008	0.166	0.038	0.485	0.001	0.339	0.175
	Longitudinal	0.453	0.098	0.278	0.171	0.311	0.087	0.238	0.364
	Heading	0.481	0.057	0.383	0.079	0.494	0.108	0.23	0.168

Table 5.17.: Results of the integrity evaluation using a correlation quotient of 1 in the probability estimation. For the three data sets, the proportions of the four operational states for the lateral direction, longitudinal direction, and heading are shown, on the one hand based on the L0 loss and on the other hand based on the beam model.

Finally, the influence of the correlation quotient is considered, which acts as a tuning parameter for the integrity systems as it directly influences the size of the protection level. In Fig. 5.21, for 500 epochs of the *Nordstadt* data set, the size of the protection level, which is represented as the number of candidate poses with a probability of $p > 1 - 10^{-8}$ in this case, is shown for correlation quotients of 1, 10, 20, and 30, where a correlation quotient of 1 corresponds to the case that all real measurements are taken into account. In Fig. 5.21a, the protection level sizes are based on the L0 loss, and in Fig. 5.21b, they are based on the beam model. For both approaches, the size of the protection level increases with a higher correlation quotient. Apart from that, the difference between both approaches becomes clear, which is that for the *Nordstadt*, the protection levels based on the beam model are overall much smaller than those based on the L0 loss, which leads to the more optimistic behavior that can be seen when looking at the proportions of the operational states stated in Tab. 5.16.

In Tab. 5.17, the proportions of the four operational states are given, using a correlation quotient of 1 in the probability estimation. Similar to Tab. 5.16, the proportions are stated for all three data sets, on the one hand based on the L0 loss and on the other hand based on the beam model. A comparison of the results in Tab. 5.16 and Tab. 5.17 reveals clear trends. With a correlation quotient of 1, the proportion of the state *Unavailable* is smaller, and in some cases significantly smaller, than with a correlation quotient of 10, the proportion of *Misleading Information* is significantly larger, and the proportion of *Hazardously Misleading Information* is larger or the same. The share of the state *Nominal Operation* is overall smaller for *Nordstadt* and *Haltenhoffstraße*, but there is no clear trend for the *Highway*.

When tuning the integrity system, on the one hand, the goal is to reduce the proportion of the state *Unavailable* as the aim is to have a system that is available as often as possible, which is in favor of the smallest possible correlation quotient. However, on the other hand, the goal is also to reduce the proportion of misleading and, in particular, hazardously misleading states in order to ensure safe operation, which is in favor of a high correlation quotient. Eventually, a trade-off has to be found and the integrity system tuned according to the pre-defined integrity risk. The step of finding an appropriate correlation quotient based on data that was recorded in advance can be considered as a learning step, as already done by Al Hage et al. (2022), who first gathered data to adjust a Student's *t*-distribution in order to fulfill a pre-defined integrity risk.

Correlation quotient	Direction	NO	UA	MI	HMI
1	Lateral	0.944	0.008	0.048	0.0
	Longitudinal	0.753	0.092	0.141	0.014
4	Lateral	0.957	0.029	0.014	0.0
	Longitudinal	0.657	0.335	0.007	0.001
6	Lateral	0.955	0.039	0.006	0.0
	Longitudinal	0.506	0.493	0.001	0.0

Table 5.18.: Example of parameter tuning of the integrity system: Proportions of the four operational states for the lateral and longitudinal directions using different correlation quotients in the probability estimation. The example is based on the Highway data set and the probability distributions are based on the L0 loss using correspondences in the Euclidean space. With an increasing correlation quotient, the proportions of the states *Misleading Information* and *Hazardously Misleading Information* decrease, while the state *Unavailable* increases.

Finally, an example of the parameter tuning of the integrity system is presented based on the *Highway* data set. Since the localization based on the correspondences in the observation space is affected by additional influences, which weaken the localization itself, the tuning of the integrity system is shown for the localization based on correspondences in Euclidean space. In Tab. 5.18, the proportions of the four operational states for the lateral and longitudinal directions using different correlation quotients in the probability estimation are stated, where the probability distributions are based on the L0 loss. It can be seen that with an increasing correlation quotient, the proportions of the states *Misleading Information* and *Hazardously Misleading Information* decrease, while the proportion of the state *Unavailable* increases. Consequently, reducing the integrity risk comes at the cost of fewer availability, while increasing it allows for higher availability. For an integrity risk of 0%, a nominal operation of only 50.6% is achieved (based on the lower value of the longitudinal axis), which is not surprising considering the difficult circumstances for the localization.

If the circumstances are more favorable, such as in the *Nordstadt* or *Haltenhoffstraße*, significantly higher rates of nominal operation are achieved. For a correlation quotient of 1, in the *Nordstadt*, for 100% of the epochs the state *Nominal Operation* occurs both for lateral and longitudinal direction, and in the *Haltenhoffstraße*, for 99.8% and 91% the state *Nominal Operation* and for 0.2% and 9% the state *Misleading Information* occurs for lateral and longitudinal direction, respectively, so that both states *Unavailable* and *Hazardously Misleading Information* do not exist. By increasing the correlation quotient, the proportions of the state *Misleading Information* can be reduced, however, at some point also the state *Unavailable* occurs.

5.4. Summary and discussion

The most important findings are now summarized and discussed against the background of the objectives defined in Sec. 5.1. While doing so, strengths, weaknesses, and limitations of the localization framework are highlighted.

In Sec. 5.3.1, the results of the experiments using real-world data from different kinds of environments show that the maximum consensus localization both based on the inlier count and the point-to-plane adjustment score is able to provide accurate and robust localization solutions. In fact, it achieves outstanding results in terms of robustness, which especially becomes clear when considering the failure rates for the methods considered for comparison in the presence of inaccurate initialization, where ‘inaccurate’ already refers to an initialization error of 0.5 m. The results show that an exhaustive approach such as the maximum consensus localization is a suitable strategy for

overcoming such problems. In general, the inlier count performs unexpectedly well, and while it is equivalent to the point-to-plane adjustment score in Euclidean space, it performs even better than the score in observation space. However, it should be noted that the count benefits enormously from the error-free initialization and the almost perfect data, i.e. the regular point density and the completeness of the map, which is particularly true for *Nordstadt* and *Haltenhoffstraße*. The experiments based on the *Highway* data set and those including the artificial heading error ultimately show that the point-to-plane adjustment score has its advantages in terms of robustness. This is also demonstrated by the examination of the reliability based on the kurtosis, second peak ratio, and largest peak distance, all of which emphasize the higher robustness of the score. It can be assumed that if, for example, the count and score were integrated in a SLAM approach to register a new scan to an incomplete and irregular map, the score would be able to exploit its robustness advantages. However, the robustness of the score is only assured if the normal vectors can be accurately determined. If the normal vectors cannot be determined accurately, the simple inlier count may outperform the score, as is frequently the case in the observation space.

In general, the localization results based on correspondences in the observation space are significantly worse than in the Euclidean space, which is related to additional simplifications and error sources, namely, firstly, that the synthetic scan images are rendered based on static candidate poses, while the actual measurement is performed from a continuously moving LiDAR pose, and secondly, that the approach in the observation space uses a triangle mesh, where additional inaccuracies are induced in the step of generating the mesh from the map point cloud, while the approach in the Euclidean space works directly on the very precise map point cloud. The above simplifications lead directly to potential improvements of the approach. Firstly, the ranges could be rendered not from a single point of view, but based on many points along a predicted trajectory. The prediction of the trajectory for one scan rotation of 0.1 s is comparatively accurate, and differences in the field of view are avoided in this way. Secondly, the rendering could be performed directly based on the map points to avoid introducing additional inaccuracies from the mesh. Finally, integrating the azimuth angle in the determination of correspondences could be explored, as it is currently completely neglected and correspondences are only established when the azimuth angles exactly match. To summarize, the approach based on correspondences in the observation space still has its justification because the implementation is comparatively efficient, since it avoids searching for correspondences using spatial data structures such as *kd*-trees, and because it also provides knowledge about the expected measurements for different poses, which allows, for example, the application of the beam model.

Despite the problems of the maximum consensus localization based on correspondences in the observation space, both the count and the score are able to outperform the method considered for comparison, which is also based on correspondences in the observation space, namely the Range-MCL. However, the Range-MCL also exhibits the poorest results among all three methods considered for comparison, and it seems that its objective function, namely the mean of the absolute pixel-wise differences, would be unsuitable. The other two methods considered for comparison, which are based on correspondences in Euclidean space, namely the point-to-plane ICP with a quadratic loss function and the KISS-ICP with the Geman-McClure loss function, show a significantly better localization performance than the Range-MCL. Both achieve, just like the maximum consensus localization, a failure-free localization for the *Nordstadt* and *Haltenhoffstraße*. In the case of the *Highway*, the advantage of the robust loss function of the KISS-ICP is evident, as it achieves the best results, even compared to the maximum consensus localization, however, based on an initialization in the ground truth. The investigation with an additional error in the initialization shows that both approaches require an accurate initialization, as already half a meter leads to a significant rate of positioning failures.

In Sec. 5.3.2, the localization results based on different loss functions show that without additional outlier rejection, only those loss functions where large residuals or outliers have no influence lead to precise and robust localization results. This is the case for the Geman-McClure, Welsch, Tukey, and L0 loss functions, all of which provide failure-free results for *Nordstadt* and *Haltenhoffstraße*. For the *Highway*, there is an additional difference in performance among these four loss functions, with Tukey’s loss performing best, followed by the L0 loss. The results of the localization based on the beam model indicate that the use of a data-adapted loss function is only suitable to a limited extent. While the results for *Nordstadt* and *Haltenhoffstraße* are equivalent to those of the count based on correspondences in the observation space, with almost no positioning failures, the performance for the *Highway* data set is very poor. The shortcoming of the beam model-based localization is that there is only one fixed beam model that is estimated and used for each scenario, while within each scenario the actual distribution of measurements varies considerably. Consequently, the beam model often deviates strongly from the actual distribution of the measurements, and due to the discrepancy between the assumed and the actual distribution, the estimation leads to an incorrect solution. Therefore, a possibility to improve the localization approach could be to constantly update the beam model parameters so that changes in the environment are captured and the assumed distribution of measurements represented by the beam model better matches the actual distribution.

In Sec. 5.3.3, the results of the maximum consensus filter show a significant improvement in localization performance compared to the single position estimation, as it notably reduces the number of large positioning errors. Due to its histogram filter property, it is generally capable of recovering from large positioning errors. However, a total of 11 re-initializations are still required over the entire trajectory of the challenging *Highway* scenario because the error has become so large that the true position is located outside the search range. One way to alleviate this problem is to expand the search space, but it could also be addressed by applying the correlation quotient. The correlation quotient serves as a tuning parameter for the probability distribution calculated on the basis of the loss accumulator. Applying a correlation quotient of 20, for example, leads to an additional significant improvement in performance in terms of RMSE and the rate of positioning failures, and only two re-initializations are required over the entire trajectory of the *Highway* data set.

Nevertheless, there is still room for improvement for the filter. Currently, only the center of the grid cell with the highest probability is used as the localization solution. On the one hand, this inaccuracy affects the calculated motion for the constant velocity model, making it less accurate, and on the other hand, the localization solution serves as the starting point for the prediction, carrying the inaccuracy directly into the next epoch. Potential ways to achieve a higher accuracy could include a sub-pixel interpolation or a finer discretization of the xy -search space. In addition, the influence of the correlation quotient must be investigated in more detail in order to select it optimally. In principle, it should also be possible to calculate the probability distributions based on the beam model from Thrun et al. (2005), although this should then be re-estimated in each epoch in order to avoid the problems seen with beam model-based localization. Finally, the good performance of the filter must be qualified by the fact that the heading was completely excluded, eliminating a potential source of error. Extending the filter to include the heading remains one of the future tasks.

In Sec. 5.3.4, the results of the protection level estimation demonstrate that the idea of defining confidence intervals based on the discrete probability distribution and thus utilizing the knowledge about the entire search space is beneficial. However, the results have also shown that the selection of the underlying loss function is crucial as it directly affects the estimated protection levels. In general, the loss function must reflect the distribution of the measurements, which is why

a protection level estimation based on the beam model from Thrun et al. (2005) is generally suitable. However, as can be seen from the results and as explained above, its parameters need to be constantly updated so that changes in the environment are captured and the protection levels are reliable.

Furthermore, the correlation quotient, which defines the number of effective measurements, serves as a tuning parameter for the integrity system, as it directly influences the size of the protection level and thereby defines the proportions between the four operational states as well as the integrity risk that is achieved. The selection of the correlation quotient must be carried out in a learning step using pre-recorded data, where the challenge is to tune the integrity system in such a way that the occurrence of the state *Unavailable* is minimized while the state *Hazardously Misleading Information* is avoided up to the integrity risk. In general, reducing the integrity risk comes at the cost of fewer availability, while increasing it allows for higher availability. For the *Highway* data set, an example for tuning the integrity system is presented, which shows that with an increasing correlation quotient, the proportions of the states *Nominal Operation*, *Misleading Information*, and *Hazardously Misleading Information* decrease, while the proportion of the state *Unavailable* increases.

Finally, it should be noted that the definition of the protection level based on the discrete probability distribution enables a more differentiated definition of the alert limit, in contrast to the current approach where only one alert limit value along each axis is considered against which the protection level is compared. Both a distinction between positive and negative directions along each axis and a location- and situation-dependent definition of the size, e.g. by defining a polygon that takes into account information from the map or from other road users, are possible as the protection level also allows for this. This improvement would lead to higher availability rates at constant integrity risk.

6. Conclusion and outlook

Finally, a conclusion on the main contributions and findings is given, followed by an outlook on future work.

6.1. Conclusion

In this work, a robust and reliable localization framework was proposed with the aim to provide highly robust and high-integrity localization solutions. At the core of this localization framework is an exhaustive search strategy that allows the application of non-convex loss functions, such as the maximum consensus criterion, and provides knowledge of a certain parameter space enabling integrity assessments. In order to draw a conclusion on the most important findings, the research questions formulated at the beginning of this thesis in Sec. 1.2 are revisited, where the first research question is as follows:

How to apply the maximum consensus criterion to the task of vehicle localization? How can maximum consensus techniques improve autonomous vehicle localization, and what are the advantages and limitations of a maximum consensus-based localization?

The maximum consensus localization determines the relative pose to a given initial pose by the registration of a sparse ‘car sensor’ point cloud from a LiDAR mounted on the ego-vehicle to a dense, high resolution ‘map’ point cloud using the maximum consensus criterion. To this end, around the initial pose, a three dimensional search space comprising the 2D position in the xy -plane and the heading angle θ is defined and discretized to candidate poses. These candidate poses are then evaluated exhaustively. In the basic approach, the number of measurements matching the map is determined for each candidate pose, and the candidate pose with the highest consensus set is considered as the localization solution. The experiments have shown that the exhaustive search strategy yields outstanding results in terms of robustness, especially in the presence of inaccurate initialization, and that the simple inlier count performs unexpectedly well. Therefore, it can be concluded that the simple maximum consensus localization using the inlier count is an excellent new localization method. To reduce the computational effort, which is the biggest limitation, implementations using large matrix operations or the rendering of synthetic views can be used, both of which can be computed quickly using modern GPUs. In principle, however, a limitation of the search space is unavoidable, where the size is to be set according to uncertainty information in a practical application. With regard to the inlier count, it should be noted that it benefits enormously from the error-free initialization of the three angles and the complete and regular map data, and that it shows weaknesses when these ideal conditions are not present, which leads directly to the next research question:

Which measures can be taken to cope with existing limitations and improve the maximum consensus-based localization? What alternative objective functions can increase the localization reliability?

To increase the robustness of the localization and reduce the susceptibility to errors caused by inaccuracies in orientation or imperfect map data, a new objective function based on Helmert’s point error has been proposed. Instead of simply counting inliers, the constraints generated by each correspondence with respect to the unknown pose are modelled. In the approach, the uncertainty

of a point-to-plane adjustment is considered, and the corresponding point-to-plane adjustment score is defined as the inverse of the square of Helmert's point error. The score is therefore inversely related to the uncertainty of the point-to-plane adjustment, which has the effect that candidate poses that have constraints only in one direction and thus are undetermined in the other direction obtain a significantly smaller score than candidate poses that are constrained in all directions. Based on the results, it can be concluded that the score outperforms the count in the presence of imperfect conditions. Recently, robust loss functions have been increasingly integrated into localization algorithms, and the adaptation of the loss function to the distribution of measurements has also been introduced, which leads to the next question:

How do other known and data-adapted loss functions perform in comparison to the maximum consensus criterion?

The localization framework is designed to be general, allowing for the use of any loss function, including common M-estimators and data-adjusted loss functions such as Thrun's beam model, whose parameters are estimated from the data. From the results, it can be concluded that without additional outlier rejection, only those loss functions lead to precise and robust localization results where large residuals or outliers have no influence. However, in case of inaccurate initialization, also these loss functions can lead to localization failures, as shown by both the evaluation using the *Highway* data set and the KISS-ICP that is based on the Geman-McClure loss function. It should be emphasized that when using other loss functions, a suitable selection of the scale parameter must still be made, which is difficult in advance and difficult in general due to the strongly changing distribution of the measurements, which is why an adapted loss appears suitable. However, it should be constantly updated, which can be concluded from the results of the beam model-based localization, which have shown that the use of a data-adapted loss function is only applicable, if it does not strongly deviate from the actual distribution of measurements, and this is only the case, if it is constantly updated so that changes in the environment are taken into account. So far, single positions were estimated independently, however, valuable information can be obtained from taking estimates from previous epochs into account, which refers to a filter and leads to the next question:

How to extent the localization based on the maximum consensus criterion or other loss functions into a filter solution?

To obtain a filtering method capable of using the accumulator of inlier counts from the previous epoch, it has been proposed to consider the loss accumulator containing the outlier counts and to derive a discrete probability distribution that can then be used in a classical histogram filter, preserving the full knowledge of the search space and avoiding the shortcomings of Monte Carlo-based filtering techniques. Two conclusions can be drawn from the results of the maximum consensus filter. The first is that the filter is generally able to improve localization reliability compared to single position estimation, as it notably reduces the number of large position errors and thus increases robustness, while also being able to recover from large position errors due to its histogram filtering property. The second is that the probability distributions estimated by taking into account all the measurements are overly optimistic and that by reducing this optimism, achieved by introducing the correlation quotient, an even better performance of the filter can be obtained. However, in highly challenging scenarios, the filter is still not able to always determine the correct solution, which is why such situations should be detected to warn the driver. This leads to the final research question:

How to bridge the gap between maximum consensus localization and an integrity system based on confidence or protection levels, which guarantee up to a certain probability to contain the true pose?

In this work, a protection level definition was proposed that makes use of the full knowledge of

the investigated search space. Using the probability distributions based on the loss accumulators or the beam model, the protection levels are defined as the maximum extent of candidate poses with a probability of $p > 1 - 10^{-8}$ starting from the estimated localization solution in longitudinal, lateral, and heading direction. Based on the relations between the true position error, the estimated protection level, and the alert limit defined by Reid et al. (2019), the integrity levels of the system can be derived and represented using the well-known Stanford-ESA Integrity Diagram. Due to the overly optimistic probability distributions, the application of the correlation quotient in the probability estimation is proposed, which can be considered as a tuning parameter of the integrity system. From the results of the protection level estimation, it can be concluded that defining confidence intervals based on the discrete probability distribution and utilizing the knowledge about the entire search space is sensible. However, the results have also shown that the selection of the underlying loss function and the correlation quotient are crucial. They directly affect the estimated protection levels and thus define the proportions between the four operational states and the integrity risk achieved. The underlying loss function should match the distribution of the measurements and therefore be constantly adapted, and the correlation quotient should be determined in a learning step using pre-recorded data.

6.2. Outlook

Finally, this section provides an outlook on future work. It identifies potential areas for improvement in existing methods and suggests new research directions. Specifically, there is room for improvement in maximum consensus localization based on correspondences in the observation space. Systematic deviations caused by rendering from a single point of view must be eliminated. This can be achieved by generating synthetic views based on several points along a predicted trajectory. Additionally, inaccuracies introduced during meshing can be avoided by rendering directly on the points or, more precisely, on their local planes based on the normal vector. Finally, the approach could be enhanced by allowing a slight deviation in the azimuth angle when determining correspondences. This would eliminate the need for them to match up to the resolution of the LiDAR, resulting in a greater number of correspondences.

One of the primary research objectives is to explain as many measurements as possible, rather than simply distinguishing between inliers and outliers. Asymmetric and changing measurement distributions, which occur in dynamic road traffic, cannot be represented by a single loss function with a fixed scale parameter. Therefore, a modelling approach such as Thrun's beam model seems promising. As previously mentioned in this thesis, continuously adjusting the beam model parameters is a way to accurately represent the measurement distribution. An alternative approach to accurately represent the actual distribution of measurements, including potential dynamic objects, is to expand the map into a voxel grid, where each voxel contains information about its occupancy, including potential dynamic occupancy from parked cars or vegetation. This would allow for the rendering of not just an expected range, but a distribution along the beam. Similar to the beam model approach, the probability for each candidate pose can be determined based on the expected measurement distribution for each beam and the actual measured range. This voxel grid approach models the environment more accurately by taking into account not only static objects but also potential dynamic objects in the foreground. Overall, a more comprehensive and accurate mapping results in a more accurate representation of the expected measurement distribution. This, in turn, leads to improved localization and more accurate discrete probability distributions. These distributions can then form the basis for the maximum consensus filter and the estimation of the protection level.

In principle, maps can store more information than just the arrangement of static objects. For instance, synthetic scan images can be stored for each position in the map along the streets, which avoids the need for repeated rendering of synthetic views. Additionally, beam model parameters could be stored and updated in the map instead of constantly estimating them while driving. Another suggestion is to store the expected accumulators along with the expected number of inliers for each position, which can then be used for validation. This idea is discussed as another perspective in the next paragraph.

A further research direction is to validate a maximum consensus accumulator against a theoretical accumulator. By generating a simulated scan, an artificial ideal accumulator can be produced, which represents the best possible localization result for the corresponding position in terms of the shape and uniqueness of the peak. The ideal accumulator can also be stored in the map. If the geometry of the static objects in the environment is unfavourable, reliable localization may not be possible, which can be predicted by the artificial accumulator. Additionally, it is useful to predict the expected number of inliers to assess the actual number. For instance, having 100 measured inliers may seem insufficient if 1000 were expected, but adequate if only 120 were expected. Accordingly, the comparison of the expected appearance and number of inliers to the obtained appearance and number of inliers can be used to validate the accumulator. However, the appropriate metrics for this validation still need to be examined, where the Kullback-Leibler divergence may be a useful metric to consider. Another point is that the loss definition has so far been based on the number of measurements, which is not really meaningful, as the number of measurements is strongly influenced by other road users. In addition, the exp-normalize trick is applied so that the loss definition is ultimately based on the highest number of inliers, which, however, completely neglects the expected number of inliers. Therefore, a loss definition based on the expected number of inliers could also be sensible, although it should be noted that the expected number of inliers is different for each candidate position.

Several interesting ideas for improving the integrity system remain, along with open questions. The loss function should correspond to the distribution of the measurements. Thrun's beam model is capable of achieving this, making it suitable for accurately estimating protection levels. The first improvement is to continuously adjust the beam model parameters to minimize the difference between the beam model and the actual measurement distribution, which is crucial, as demonstrated by the results. Another way to obtain appropriate beam model parameters is to predict them based on the classes present in the scan using deep learning. Both approaches will allow changes in the environment to be taken into account, which should lead to more accurate protection levels. The second aspect is a more comprehensive evaluation of the integrity approach. This concerns the accuracy of the protection levels, which could be compared with protection levels based on other approaches and could be checked in a long-term test phase to determine whether the protection level really does contain the true pose up to a certain probability. It should also be investigated whether the correlation quotient can be determined differently. The correlation quotient could also be determined in the long-term test phase and it could be investigated whether it varies for different environments. The last aspect regarding future work for the integrity system is to implement the vision of defining the alert limit based on the location and situation. This can be achieved by considering map and lane information, as well as other road users detected in the environment. The alert limit can be defined as an arbitrary polygon with different extent in the positive and negative directions along each axis. This enables a more efficient utilization of the precise shape of the protection level proposed in this thesis, resulting in increased availability rates while maintaining a constant level of integrity risk.

The final aspect to consider is whether a classifier can be trained to predict the correctness of a localization on the basis of an accumulator. The accumulator and the information about the

correctness of the localization could be used as training data, collected over many epochs in a learning phase. The classifier could be adjusted to increase its sensitivity to localization failures, ensuring that mislocalizations are not missed and potential errors are detected. The aim of such a localization assessment is to further enhance the trustworthiness of the positioning system. In initial investigations, this has already been done using a random forest classifier with hand-crafted features derived from the accumulator. However, the overall results still need improvement as not all mislocalizations could be detected. Despite this, the approach could still be promising to assess the localization result, as it uses the full knowledge of the search space, which is provided by the exhaustive search strategy. The idea should therefore be pursued further. However, in the next step, deep learning approaches should be tested to avoid the selection of hand-crafted features, which may lead to better classification results.

Appendices

A. Experiments

A.1. Localization using different loss functions

Accuracy position (obs.):

Data set	L2	L1	Huber	Cauchy	G.-McC.	Welsch	Tukey	L0	BM
Nordstadt	0.742	0.19	0.26	0.133	0.097	0.099	0.094	0.068	0.099
Haltenhoffstr.	0.723	0.259	0.308	0.156	0.107	0.108	0.098	0.139	0.107
Highway	0.87	0.707	0.751	0.63	0.518	0.517	0.46	0.331	0.517

Table A.1.: RMSE [m] with respect to the xy -positioning for the localization using various loss functions. The results are based on correspondences in the observation space. The localization solution corresponds to the center point of the grid cell with the smallest loss.

Robustness position (obs.):

Data set	L2	L1	Huber	Cauchy	G.-McC.	Welsch	Tukey	L0	BM
Nordstadt	0.696	0.049	0.128	0.011	0.0	0.001	0.001	0.005	0.007
Haltenhoffstr.	0.521	0.056	0.102	0.017	0.001	0.001	0.0	0.021	0.005
Highway	0.931	0.767	0.851	0.657	0.502	0.515	0.35	0.184	0.525

Table A.2.: Share of epochs with an xy -error larger than the position alert threshold for various loss functions. The results are based on correspondences in the observation space. The localization solution corresponds to the center point of the grid cell with the smallest loss.

List of Figures

1.1. Common reasons for incorrect associations.	2
1.2. Various loss functions and the corresponding estimation results for a 1D estimation problem in the presence of outliers.	3
1.3. Evaluated search space represented as stack and unwrapped.	5
1.4. Overview of the robust and reliable localization framework.	6
2.1. Illustration of a line fitting problem in the presence of outliers.	11
2.2. Plots and formulas of various well-known loss functions.	13
2.3. 1D estimation problem and the corresponding estimation results and loss curves.	14
2.4. Components of the beam model for range sensors.	16
2.5. Physical beam model of range sensors represented by the mixture distribution $p(z^k z^{k*})$	17
2.6. The dynamic Bayes network.	19
2.7. Structure of a mechanically rotating LiDAR.	23
2.8. 3D LiDAR point cloud colored according to the measured intensity value.	23
2.9. Basic principle how a range image is created.	24
2.10. Illustration of a measured range image.	25
2.11. RIEGL VMX-250 mobile mapping system.	28
2.12. Map point cloud acquired by the RIEGL VMX-250 MMS and colored according to manually annotated labels.	29
2.13. Graphical representation of a mapping problem.	30
2.14. Illustration of a simple inverse sensor model $l(\mathbf{m}_i \mathbf{z}_t, \mathbf{x}_t)$	31
2.15. Illustration of the NDT map generation process.	33
2.16. Shortcoming of the NDT regarding the representation of planar structures.	33
2.17. NDT map and surfel map.	34
2.18. Principle of the TSDF.	36
2.19. Stanford-ESA Integrity Diagram.	38
4.1. Overview robust and reliable localization.	49
4.2. Application of maximum consensus to vehicle localization.	53
4.3. Concept of the maximum consensus localization using projective associations.	54
4.4. Example meshes	54
4.5. Measured range image, synthetic range image, and image of range differences between rendered and measured ranges.	55
4.6. Difference images for positions in a 4-neighborhood.	56
4.7. Maximum consensus count accumulators for a straight street and crossing scenario based on correspondences in the observation space.	57
4.8. Car sensor scan and map point cloud for a straight street scenario.	58
4.9. Car sensor scan and map point cloud for a crossing scenario.	59
4.10. Maximum consensus count accumulators for a straight street and crossing scenario based on correspondences in the Euclidean space.	60

4.11. Epoch where grid shifts prevent a localization failure caused by the discretization of the search space.	61
4.12. Fundamental concept of the point-to-plane adjustment.	62
4.13. Synthetic LiDAR scan image with three channels representing the x , y , and z components of the normal vector.	64
4.14. Maximum consensus score accumulators for a straight street and crossing scenario based on correspondences both in the observation and Euclidean space.	67
4.15. Approach to determine missing correspondences.	68
4.16. Measured range image, and a synthetic range image both before and after the determination of missing correspondences.	69
4.17. Loss accumulators of various loss functions without adding the missing correspondences. . .	70
4.18. Loss accumulators of various loss functions after adding the missing correspondences. . . .	71
4.19. Methodology of the maximum consensus filter.	74
4.20. Truncated probability distribution based on the L0 loss and the estimated protection level. . .	77
4.21. Maximum consensus accumulator and the corresponding discrete probability distribution based on the beam model including the estimated protection level.	78
4.22. Influence of the correlation quotient on the distribution between the four operational states in the Stanford diagram.	79
5.1. Kurtosis values according to Pearson's (β_2) and Fisher's (γ) definition for mesokurtic, leptokurtic and platykurtic distributions. The illustration is adapted from Hemmerich (2020).	84
5.2. Orthophotos showing the environments of the three data sets.	85
5.3. Map point cloud from the Nordstadt data set.	86
5.4. Map point cloud from the Nordstadt data set where points from the ground, vegetation, parked cars, and other traffic participants are removed.	86
5.5. Photo of the sensor setup.	87
5.6. Sketch of the vehicle with the sensor setup, the different coordinate systems, and the transformations between those coordinate systems.	88
5.7. Application of the maximum consensus localization.	89
5.8. Errors [m] with respect to the xy -position for the point-to-plane ICP, the KISS-ICP, and the maximum consensus localization based on correspondences in the Euclidean space.	94
5.9. Errors [$^\circ$] with respect to the heading for the point-to-plane ICP, the KISS-ICP, and the maximum consensus localization based on correspondences in the Euclidean space.	96
5.10. Maximum consensus accumulator based on the count and based on the score for a challenging epoch from the <i>Highway</i> scenario.	97
5.11. Effect of the grid shifts on the localization performance.	99
5.12. Localization error [m] of the count and the score in the presence of a systematic heading error of two degrees.	100
5.13. Histograms for each of the three data sets containing the kurtosis values and second peak ratios for the count and score.	101
5.14. Localization errors using the different loss functions for a section of 50 epochs from the <i>Highway</i> data set.	103
5.15. Visualization of the estimated beam models, the underlying histograms, and the corresponding loss functions of the three data sets.	106
5.16. Localization results of the maximum consensus filter.	109

5.17. Probability grids of the prediction, measurement, and update for an example epoch to illustrate the benefit of the maximum consensus filter.	110
5.18. The protection level, position error, and alert limit are shown for the lateral direction, longitudinal direction, and heading for a section of 500 epochs.	111
5.19. Stanford-ESA Integrity Diagrams for the <i>Haltenhoffstraße</i> for the lateral direction, longitudinal direction, and heading using protection levels based on the L0 loss.	112
5.20. Stanford-ESA Integrity Diagrams for the <i>Haltenhoffstraße</i> for the lateral direction, longitudinal direction, and heading using protection levels based on the beam model.	113
5.21. Sizes of the protection level for correlation quotients of 1, 10, 20, and 30.	115

List of Tables

4.1. Relevant technical LiDAR specifications for the rendering of synthetic scan images.	54
5.1. Overview of data sets used in the experiments.	87
5.2. RMSE [m] with respect to the xy -positioning for the methods considered for comparison and the maximum consensus localization.	90
5.3. Shares of epochs with an xy -error larger than the position alert threshold for the methods considered for comparison and the maximum consensus localization.	91
5.4. RMSE values [m] of the xy -positions corresponding to the center points of the grid cells with the highest consensus and of the refined xy -positions.	92
5.5. RMSE values [m] for the conventional positioning using the center point of the grid cell with the highest consensus and for the refined positioning taking into account only those epochs where the count or score localization error is not equal to zero.	93
5.6. RMSE [$^\circ$] with respect to the heading for the methods considered for comparison and the maximum consensus localization.	95
5.7. Share of epochs with a heading error larger than the heading alert threshold for the methods considered for comparison and the maximum consensus localization.	95
5.8. Comparison of maximum consensus localization without and with grid shifts for the highway scenario.	98
5.9. Mean of kurtosis values, second peak ratios, and largest peak distances over all epochs for the localization using count and score.	100
5.10. RMSE [m] with respect to the xy -positioning for the localization using various loss functions. The results are based on correspondences in the Euclidean space.	102
5.11. Share of epochs with an xy -error greater than the position alert threshold for various loss functions. The results are based on correspondences in the Euclidean space.	103
5.12. Mean of kurtosis values, second loss ratios, and largest peak distances for the localization using different loss functions.	104
5.13. Estimated intrinsic parameters for the three data sets.	105
5.14. RMSE values and shares of positioning failures for the localization using the beam model for range sensors.	107
5.15. RMSE values [m] and failure rates with respect to the xy -positioning for the maximum consensus localization and the maximum consensus filter.	110
5.16. Results of the integrity evaluation using a correlation quotient of 10 in the probability estimation.	114
5.17. Results of the integrity evaluation using a correlation quotient of 1 in the probability estimation.	116
5.18. Example of parameter tuning of the integrity system: Proportions of the four operational states using different correlation quotients in the probability estimation.	117
A.1. RMSE [m] with respect to the xy -positioning for the localization using various loss functions. The results are based on correspondences in the observation space.	129

A.2. Share of epochs with an xy -error larger than the position alert threshold for various loss functions. The results are based on correspondences in the observation space.	129
--	-----

List of Acronyms

LiDAR	Light Detection And Ranging
GNSS	Global Navigation Satellite System
IMU	Inertial Measurement Unit
ICP	Iterative Closest Point
RANSAC	Random Sample Consensus
PL	Protection Level
TPE	True Position Error
AL	Alert Limit
ToF	Time-of-Flight
FoV	Field of View
SLAM	Simultaneous Localization and Mapping
MMS	Mobile Mapping System
RTK	Real-Time Kinematic positioning
NLOS	Non-Line-Of-Sight
NDT	Normal Distributions Transform
PCA	Principal Component Analysis
TSDF	Truncated Signed Distance Function
GMM	Gaussian Mixture Model
BEV	Bird-Eye-View
DoF	Degrees of Freedom
RMSE	Root Mean Square Error

Nomenclature

General notation

$\mathbf{A}, \mathbf{B}, \mathbf{C}$	Matrices: Bold upper case
$\mathbf{a}, \mathbf{b}, \mathbf{c}$	Vectors: Bold lower case
a, b, c	Scalars: Regular lower case
\mathbb{R}	Real numbers

Operators

\mathbf{M}^\top	Transpose of matrix \mathbf{M}
$\langle \cdot, \cdot \rangle$	Scalar product
$\max(\cdot, \cdot)$	Max operator
\det	Determinant
tr	Trace
l_∞	l_∞ -norm

Fundamental

$\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2)$	Diagonal matrix of eigenvalues
$\boldsymbol{\mu}, \boldsymbol{\Sigma}$	Mean, covariance matrix
\mathbf{A}	Design matrix
\mathbf{I}	Identity matrix
\mathbf{N}	Normal equation matrix
\mathbf{n}	Normal vector
\mathbf{P}	Precision matrix
\mathbf{Q}_{tt}	Cofactor matrix of translation \mathbf{t}
\mathbf{R}	Rotation matrix
\mathbf{t}	Translation vector
$\mathbf{T} = \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{pmatrix}$	Transformation matrix
λ	Eigenvalue
\mathcal{L}	Likelihood
\mathcal{N}	Normal distribution

ρ	Loss
σ, σ^2	Standard deviation, variance
θ	Heading angle
d	Distance
d_M	Mahalanobis distance
n_x, n_y	Normal vector part in x , normal vector part in y
r	Residual
w	Weight

LiDAR

$\Delta\alpha$	Horizontal resolution of a LiDAR
$\{\mathbf{p}_i\}_{i=1}^n$	Point cloud with n points, $\mathbf{p}_i \in \mathbb{R}^3$
n_l	Number of scan layer of a LiDAR
c	Speed of light

Maximum Consensus

$\mathbf{R}_\theta, \mathbf{t}_{xy}$	Rotation matrix only considering θ , translation in the xy -plane
$\mathbf{t}_{xy}^*, \theta^*$	Optimal translation in the xy -plane, optimal heading angle
$\mathbf{x}, \mathbf{x}_0, \mathbf{x}^*$	Candidate pose, initial pose, optimal pose
ϵ	Inlier threshold
$\mathbb{I}(\cdot)$	Indicator function
\mathcal{M}	Map
\mathcal{S}	Car sensor scan
$\Psi(\cdot)$	Consensus function
${}^C\mathbf{T}_V^*$	Optimal transformation based on \mathbf{t}_{xy}^* and θ^*
${}^C\mathbf{T}_V$	Candidate transformation $\mathbf{T}(t_x, t_y, \theta)$
${}^W\mathbf{T}_C$	Transformation based on initial pose \mathbf{x}_0

Observation space

$\mathbf{I}_{R,s}, \mathbf{I}_{R,x}$	Measured range image, synthetic range image of a candidate pose
\mathbf{v}	Direction vector of a LiDAR ray
$\{\mathbf{I}_{R,x}\}$	Set of synthetic range images
c, r	Image coordinate column, image coordinate row
n, m	Number of columns, number of rows

Euclidean space

$M = \{\mathbf{m}_j\}_{j=1}^m$	Map point cloud with m points, $\mathbf{m}_j \in \mathbb{R}^3$
$\mathbf{m}_j = (m_{x,j}, m_{y,j}, m_{z,j})^T$	Map point with index j
\mathbf{n}_m	Normal vector of a map point \mathbf{m}
$S = \{\mathbf{s}_i\}_{i=1}^n$	Car sensor scan point cloud with n points, $\mathbf{s}_i \in \mathbb{R}^3$
$\mathbf{s}_i = (s_{x,i}, s_{y,i}, s_{z,i})^T$	Car sensor scan point with index i

Beam model for range sensors

λ_{short}	Parameter of the exponential part
σ_{hit}	Standard deviation of the Gaussian part
Θ	Intrinsic parameters
k	Index of observation
$p_{hit}, p_{short}, p_{max}, p_{rand}$	Likelihood of Gaussian part, exponential part, point mass, and uniform part
$w_{hit}, w_{short}, w_{max}, w_{rand}$	Weights of Gaussian part, exponential part, point mass, and uniform part
z^k, z^{k*}	Measured range, expected range

State estimation filtering

α, c	Shape parameter, scale parameter
$\mathbf{x}, \mathbf{u}, \mathbf{Z}$	State vector, control, set of observations
η	Normalization factor
$\rho(r, \alpha, c), \tilde{\rho}_p(r, \alpha, c)$	Generalized loss, truncated loss
τ	Truncation limit
bel, \overline{bel}	Belief, predicted belief
$P(r, \alpha, c), \tilde{P}(r, \alpha, c)$	General probability distribution, truncated probability distribution
$Z(\alpha), \tilde{Z}(\alpha)$	Partition function, finite partition function

Coordinate systems

${}^A\mathbf{p}_i = ({}^Ap_{x,i}, {}^Ap_{y,i}, {}^Ap_{z,i})^T$	Point with index i in the coordinate system A
${}^A\{\mathbf{p}\}$	Point cloud in the coordinate system A
${}^B\mathbf{R}_A$	Rotation matrix from coordinate system A to coordinate system B
${}^B\mathbf{T}_A$	Transformation matrix from coordinate system A to coordinate system B
${}^B\mathbf{t}_A$	Translation vector from coordinate system A to coordinate system B
L, V, W	LiDAR coordinate system, Vehicle coordinate system, World coordinate system

Bibliography

- Al Hage, J., Xu, P., Bonnifait, P., Ibañez-Guzmán, J., 2022. Localization Integrity for Intelligent Vehicles through Fault Detection and Position Error Characterization. *IEEE Transactions on Intelligent Transportation Systems* 23 (4), pp. 2978–2990.
- Aoki, Y., Goforth, H., Arun Srivatsan, R., Lucey, S., 2019. PointNetLK: Robust & Efficient Point Cloud Registration Using PointNet. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Arun, K. S., Huang, T. S., Blostein, S. D., 1987. Least-Squares Fitting of Two 3-D Point Sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-9 (5), pp. 698–700.
- Ask, E., Enqvist, O., Kahl, F., 2013. Optimal Geometric Fitting Under the Truncated L-2-Norm. In: *2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE - Institute of Electrical and Electronics Engineers Inc., United States, pp. 1722–1729, 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013 ; Conference date: 23-06-2013 Through 28-06-2013.
- Avetisyan, A., Dai, A., Niessner, M., 2019. End-to-End CAD Model Retrieval and 9DoF Alignment in 3D Scans. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.
- Axmann, J., Brenner, C., 2021. MAXIMUM CONSENSUS LOCALIZATION USING LIDAR SENSORS. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* V-2-2021, pp. 9–16.
- Axmann, J., Brenner, C., 2023. Maximum Consensus based Localization and Protection Level Estimation using Synthetic LiDAR Range Images. In: *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*. pp. 5917–5924.
- Axmann, J., Moftizadeh, R., Su, J., Tennstedt, B., Zou, Q., Yuan, Y., Ernst, D., Alkhatib, H., Brenner, C., Schön, S., 2023a. LUCOOP: Leibniz University Cooperative Perception and Urban Navigation Dataset. In: *2023 IEEE Intelligent Vehicles Symposium (IV)*. pp. 1–8.
- Axmann, J., Zhang, Y., Brenner, C., 2023b. Maximum Consensus Localization Using an Objective Function Based on Helmert’s Point Error. In: *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*. pp. 2302–2309.
- Barron, J. T., 2017. A General and Adaptive Robust Loss Function. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4326–4334.
- Bársan, I. A., Wang, S., Pokrovsky, A., Urtasun, R., 2018. Learning to Localize Using a LiDAR Intensity Map. *ArXiv* abs/2012.10902.
- Bazin, J.-C., Seo, Y., Pollefeys, M., 2013. Globally Optimal Consensus Set Maximization through Rotation Search. In: Lee, K. M., Matsushita, Y., Rehg, J. M., Hu, Z. (Eds.), *Computer Vision – ACCV 2012*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 539–551.
- Behley, J., Stachniss, C., 2018. Efficient Surfel-Based SLAM using 3D Laser Range Data in Urban Environments. In: *Proc. of Robotics: Science and Systems (RSS)*.
- Bentley, J. L., 1975. Multidimensional binary search trees used for associative searching. *Commun. ACM* 18 (9), pp. 509–517.
- Beraldin, J.-A., Blais, F., Lohr, U., 2010. Laser Scanning Technology. In: Vosselman, G., Maas, H. (Eds.), *Airborne and Terrestrial Laser Scanning*. Whittles Publishing.

- Besl, P. J., McKay, N. D., 1992. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14 (2), pp. 239–256.
- Biber, P., Strasser, W., 2003. The normal distributions transform: a new approach to laser scan matching. In: *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453)*. vol. 3. pp. 2743–2748 vol.3.
- Bishop, C. M., 2006. Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag, Berlin, Heidelberg.
- Björck, Å., 1996. Numerical Methods for Least Squares Problems. Society for Industrial and Applied Mathematics.
- Black, M. J., Rangarajan, A., 1996. On the Unification of Line Processes, Outlier Rejection, and Robust Statistics with Applications in Early Vision. *International Journal of Computer Vision* 19 (1), pp. 57–91.
- Boyd, S., Vandenberghe, L., 2004. Convex Optimization. Cambridge University Press.
- Brenner, C., 2009. Global Localization of Vehicles Using Local Pole Patterns. In: Denzler, J., Notni, G., Süße, H. (Eds.), *Pattern Recognition*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 61–70.
- Brenner, C., 2016. Scalable Estimation of Precision Maps in a MapReduce Framework. In: *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. SIGSPACIAL '16. Association for Computing Machinery, New York, NY, USA.
- Brenner, C., 2023. Internal report about integration of scans.
- Breuel, T. M., 1992. Fast recognition using adaptive subdivisions of transformation space. In: *Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. pp. 445–451.
- Breuel, T. M., 2003. Implementation techniques for geometric branch-and-bound matching methods. *Computer Vision and Image Understanding* 90 (3), pp. 258–294.
- Briales, J., Gonzalez-Jimenez, J., 2017. Convex Global 3D Registration With Lagrangian Duality. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Cai, Z., Chin, T.-J., Koltun, V., 2019. Consensus Maximization Tree Search Revisited. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.
- Campbell, D., Petersson, L., Kneip, L., Li, H., 2017. Globally-Optimal Inlier Set Maximisation for Simultaneous Camera Pose and Feature Correspondence. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.
- Chebrolu, N., Läbe, T., Vysotska, O., Behley, J., Stachniss, C., 2020. Adaptive Robust Kernels for Non-Linear Least Squares Problems. *IEEE Robotics and Automation Letters* 6, pp. 2240–2247.
- Chen, X., Milioto, A., Palazzolo, E., Giguère, P., Behley, J., Stachniss, C., 2019. SuMa++: Efficient LiDAR-based Semantic SLAM. In: *Proceedings of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*. pp. 4530–4537.
- Chen, X., Vizzo, I., Läbe, T., Behley, J., Stachniss, C., 2021. Range Image-based LiDAR Localization for Autonomous Vehicles. In: *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*.
- Chen, Y., Medioni, G. G., 1992. Object modelling by registration of multiple range images. *Image Vis. Comput.* 10 (3), pp. 145–155.
- Chetverikov, D., Stepanov, D., Krsek, P., 2005. Robust Euclidean alignment of 3D point sets: the trimmed iterative closest point algorithm. *Image and Vision Computing* 23 (3), pp. 299–309.
- Chin, T., Purkait, P., Eriksson, A., Suter, D., 2017. Efficient Globally Optimal Consensus Maximisation with Tree Search. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (4), pp. 758–772.

- Chin, T.-J., Cai, Z., Neumann, F., 2018. Robust fitting in computer vision: easy or hard? In: *Proceedings of the European Conference on Computer Vision (ECCV)*.
- Chin, T.-J., Parra Bustos, Á., Brown, M. S., Suter, D., 2014. Fast Rotation Search for Real-Time Interactive Point Cloud Registration. In: *Proceedings of the 18th Meeting of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*. I3D '14. Association for Computing Machinery, New York, NY, USA, p. 55–62.
- Chin, T.-J., Suter, D., 2017. The Maximum Consensus Problem: Recent Algorithmic Advances. *Synthesis Lectures on Computer Vision* 7, pp. 1–194.
- Cho, Y., Kim, G., Kim, A., 2020. Unsupervised Geometry-Aware Deep LiDAR Odometry. In: *2020 IEEE International Conference on Robotics and Automation (ICRA)*. pp. 2145–2152.
- Choy, C., Dong, W., Koltun, V., 2020. Deep Global Registration. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Choy, C., Park, J., Koltun, V., 2019. Fully Convolutional Geometric Features. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.
- Cop, K. P., Borges, P. V. K., Dubé, R., 2018. Delight: An Efficient Descriptor for Global Localisation Using LiDAR Intensities. *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3653–3660.
- Curless, B., Levoy, M., 1996. A Volumetric Method for Building Complex Models from Range Images. In: *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '96. Association for Computing Machinery, New York, NY, USA, p. 303–312.
- Dellaert, F., Fox, D., Burgard, W., Thrun, S., 1999. Monte Carlo localization for mobile robots. In: *Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No.99CH36288C)*. vol. 2. pp. 1322–1328 vol.2.
- Dellenbach, P., Deschaud, J.-E., Jacquet, B., Goulette, F., 2022. CT-ICP: Real-Time Elastic LiDAR Odometry with Loop Closure. In: *2022 International Conference on Robotics and Automation (ICRA)*. IEEE Press, p. 5580–5586.
- Drost, B., Ulrich, M., Navab, N., Ilic, S., 2010. Model globally, match locally: Efficient and robust 3D object recognition. In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. Institute of Electrical and Electronics Engineers (IEEE), p. 998–1005.
- Elbaz, G., Avraham, T., Fischer, A., 2017. 3D Point Cloud Registration for Localization Using a Deep Neural Network Auto-Encoder. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2472–2481.
- Elhousni, M., Huang, X., 2020. A Survey on 3D LiDAR Localization for Autonomous Vehicles. In: *2020 IEEE Intelligent Vehicles Symposium (IV)*. pp. 1879–1884.
- Enqvist, O., Ask, E., Kahl, F., Åström, K., 2012. Robust Fitting for Multiple View Geometry. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (Eds.), *Computer Vision – ECCV 2012*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 738–751.
- Enqvist, O., Josephson, K., Kahl, F., 2009. Optimal Correspondences from Pairwise Constraints. In: *IEEE International Conference on Computer Vision*. IEEE - Institute of Electrical and Electronics Engineers Inc., United States, pp. 1295–1302, IEEE International Conference on Computer Vision (ICCV), 2009 ; Conference date: 27-09-2009 Through 04-10-2009.
- Esri, 2023. Grundlagen der Topologie. <https://pro.arcgis.com/de/pro-app/latest/help/data/topologies/topology-basics.htm>, [Online; accessed 30-October-2023].
- Fallon, M., Tinchev, G., Penate-Sanchez, A., 2019. Learning to see the wood for the trees: Deep laser localization in urban and natural environments on a CPU. vol. 4. Institute of Electrical and Electronics Engineers, pp. 1327–1334.

- Fan, L., Xiong, X., Wang, F., Wang, N., Zhang, Z., 2021. RangeDet: In Defense of Range View for LiDAR-Based 3D Object Detection. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.
- Fischler, M. A., Bolles, R. C., 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Commun. ACM* 24 (6), pp. 381–395.
- Fitzgibbon, A. W., 2003. Robust Registration of 2D and 3D Point Sets. *Image Vis. Comput.* 21, pp. 1145–1153.
- Fox, D., Burgard, W., Dellaert, F., Thrun, S., 1999. Monte Carlo Localization: Efficient Position Estimation for Mobile Robots. In: *AAAI/IAAI*.
- GeoBasis-DE, SAPOS, 2023. <https://zentrale-stelle-sapos.de/>, [Online; accessed 31-October-2023].
- Geoinformation Hannover, 2022. Digitale Orthophotos 2021. Publisher: Landeshauptstadt Hannover, FB Planen und Stadtentwicklung, Bereich Geoinformation, Geoservice@hannover-stadt.de.
- Gojcic, Z., Zhou, C., Wegner, J. D., Andreas, W., 2019. The Perfect Match: 3D Point Cloud Matching with Smoothed Densities. In: *International conference on computer vision and pattern recognition (CVPR)*.
- Granger, S., Pennec, X., 2002. Multi-scale EM-ICP: A Fast and Robust Approach for Surface Registration. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (Eds.), *Computer Vision — ECCV 2002*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 418–432.
- Hartley, R., Zisserman, A., 2004. Multiple View Geometry in Computer Vision, 2nd ed. Cambridge University Press.
- Hartley, R. I., Kahl, F., 2009. Global Optimization through Rotation Space Search. *International Journal of Computer Vision* 82 (1), pp. 64–79.
- Hemmerich, W. A., 2020. StatistikGuru: Kurtosis, Wölbung, Exzess. <https://statistikguru.de/lexikon/kurtosis-woelbung.html>, [Online; accessed 04-March-2024].
- Hesai Technology Co., Ltd., 2021. PandarXT-32 32-Channel Medium-Range Mechanical LiDAR User Manual.
- Horn, B. K. P., 1987. Closed-form solution of absolute orientation using unit quaternions. *J. Opt. Soc. Am. A* 4 (4), pp. 629–642.
- Huang, Q.-X., Anguelov, D., 2010. High quality pose estimation by aligning multiple scans to a latent map. In: *2010 IEEE International Conference on Robotics and Automation*. pp. 1353–1360.
- International Civil Aviation Organization, 2006. Aeronautical Telecommunications: International Standards and Recommended Practices: Annex 10 to the Convention on International Civil Aviation. Vol. I. Radio Navigation Aids, 6th ed. International Civil Aviation Organization, Montreal, Quebec, Canada.
- Izatt, G., Dai, H., Tedrake, R., 2020. Globally Optimal Object Pose Estimation in Point Clouds with Mixed-Integer Programming. In: Amato, N. M., Hager, G., Thomas, S., Torres-Torriti, M. (Eds.), *Robotics Research*. Springer International Publishing, Cham, pp. 695–710.
- Jian, B., Vemuri, B. C., 2011. Robust Point Set Registration Using Gaussian Mixture Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33 (8), pp. 1633–1645.
- Kalman, R., 1960. A New Approach To Linear Filtering and Prediction Problems. *Journal of Basic Engineering (ASME)* 82D, pp. 35–45.
- Kaneko, S., Kondo, T., Miyamoto, A., 2003. Robust matching of 3D contours using iterative closest point algorithm improved by M-estimation. *Pattern Recognition* 36 (9), pp. 2041–2047, kernel and Subspace Methods for Computer Vision.

- Kim, G., Choi, S., Kim, A., 2021. Scan Context++: Structural Place Recognition Robust to Rotation and Lateral Variations in Urban Environments. *IEEE Transactions on Robotics* 38, pp. 1856–1874.
- Kim, G., Kim, A., 2018. Scan Context: Egocentric Spatial Descriptor for Place Recognition Within 3D Point Cloud Map. In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. pp. 4802–4809.
- Kim, G., Park, B., Kim, A., 2019. 1-Day Learning, 1-Year Localization: Long-Term LiDAR Localization Using Scan Context Image. *IEEE Robotics and Automation Letters* 4, pp. 1948–1955.
- Kutterer, H., 2010. Mobile Mapping. In: Vosselman, G., Maas, H. (Eds.), *Airborne and Terrestrial Laser Scanning*. Whittles Publishing.
- Langerwisch, M., 2014. Kartierung und Lokalisation durch mobile Serviceroboter unter der Annahme unbekannter aber begrenzter Sensorfehler. Dissertation, Leibniz Universität Hannover, Hannover.
- Le, H. M., Chin, T.-J., Eriksson, A., Do, T.-T., Suter, D., 2019a. Deterministic Approximate Methods for Maximum Consensus Robust Fitting. *IEEE Trans. Pattern Anal. Mach. Intell. (PAMI)*.
- Le, H. M., Do, T.-T., Hoang, T., Cheung, N.-M., 2019b. SDRSAC: Semidefinite-based Randomized Approach for Robust Point Cloud Registration Without Correspondences. In: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* pp. 124–133.
- Levinson, J., Montemerlo, M., Thrun, S., 2007. Map-Based Precision Vehicle Localization in Urban Environments. In: *Robotics: Science and Systems*.
- Li, H., 2009. Consensus set maximization with guaranteed global optimality for robust geometry estimation. In: *2009 IEEE 12th International Conference on Computer Vision*. pp. 1074–1080.
- Li, H., Hartley, R., 2007. The 3D-3D Registration Problem Revisited. In: *Proc. IEEE Int. Conf. Comput. Vis.* pp. 1–8.
- Li, N., Ho, C. P., Xue, J., Lim, L. W., Chen, G., Fu, Y. H., Lee, L. Y. T., 2022. A Progress Review on Solid-State LiDAR and Nanophotonics-Based LiDAR Sensors. *Laser & Photonics Reviews* 16 (11), pp. 2100511.
- Li, Q., Chen, S., Wang, C., Li, X., Wen, C., Cheng, M., Li, J., 2019. LO-Net: Deep Real-Time Lidar Odometry. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8465–8474.
- Lorensen, W. E., Cline, H. E., 1987. Marching Cubes: A High Resolution 3D Surface Construction Algorithm. *SIGGRAPH Comput. Graph.* 21 (4), pp. 163–169.
- Lu, S., Xu, X., Yin, H., Xiong, R., Wang, Y., 2022. One RING to Rule Them All: Radon Sinogram for Place Recognition, Orientation and Translation Estimation. *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2778–2785.
- Lu, W., Zhou, Y., Wan, G., Hou, S., Song, S., 2019. L3-Net: Towards Learning Based LiDAR Localization for Autonomous Driving. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 6382–6391.
- Ma, W.-C., Tartavull, I., Bârsan, I. A., Wang, S., Bai, M., Mátyus, G., Homayounfar, N., Lakshmikanth, S. K., Pokrovsky, A., Urtasun, R., 2019. Exploiting Sparse Semantic HD Maps for Self-Driving Vehicle Localization. *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5304–5311.
- MacTavish, K., Barfoot, T. D., 2015. At all Costs: A Comparison of Robust Cost Functions for Camera Correspondence Outliers. In: *2015 12th Conference on Computer and Robot Vision*. pp. 62–69.

- Maier-Hein, L., Franz, A. M., dos Santos, T. R., Schmidt, M., Fangerau, M., Meinzer, H.-P., Fitzpatrick, J. M., 2012. Convergent Iterative Closest-Point Algorithm to Accomodate Anisotropic and Inhomogenous Localization Error. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34 (8), pp. 1520–1532.
- Meer, P., Mintz, D., Rosenfeld, A., Kim, D. Y., 1991. Robust Regression Methods for Computer Vision: A Review. *International Journal of Computer Vision* 6 (1), pp. 59–70.
- Merfels, C., Stachniss, C., 2016. Pose fusion with chain pose graphs for automated driving. In: *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. pp. 3116–3123.
- Montani, C., Scateni, R., Scopigno, R., 1994. A modified look-up table for implicit disambiguation of Marching Cubes. *The Visual Computer* 10 (6), pp. 353–355.
- Museth, K., Lait, J., Johanson, J., Budsberg, J., Henderson, R., Aldén, M., Cucka, P., Hill, D. R., Pearce, A., 2013. OpenVDB: an open-source data structure and toolkit for high-resolution volumes. In: *International Conference on Computer Graphics and Interactive Techniques*.
- Myronenko, A., Song, X., 2010. Point Set Registration: Coherent Point Drift. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32 (12), pp. 2262–2275.
- Nüchter, A., Lingemann, K., Hertzberg, J., Surmann, H., 2007. 6D SLAM—3D mapping outdoor environments. *Journal of Field Robotics* 24 (8-9), pp. 699–722.
- Olsson, C., Enqvist, O., Kahl, F., 2008. A polynomial-time bound for matching and registration with outliers. In: *2008 IEEE Conference on Computer Vision and Pattern Recognition*. pp. 1–8.
- Olsson, C., Kahl, F., Oskarsson, M., 2009. Branch-and-Bound Methods for Euclidean Registration Problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31 (5), pp. 783–794.
- Parra Bustos, Á., Chin, T.-J., 2018. Guaranteed Outlier Removal for Point Cloud Registration with Correspondences. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40 (12), pp. 2868–2882.
- Parra Bustos, Á., Chin, T.-J., Neumann, F., Friedrich, T., Katzmann, M., 2019. A Practical Maximum Clique Algorithm for Matching with Pairwise Constraints. *Computing Research Repository (CoRR)* abs/1902.01534.
- Parra Bustos, Á., Chin, T.-J., Suter, D., 2014. Fast Rotation Search with Stereographic Projections for 3D Registration. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Qi, C. R., Su, H., Mo, K., Guibas, L. J., 2016. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. *arXiv preprint arXiv:1612.00593*.
- Reid, T. G., Houts, S. E., Cammarata, R., Mills, G., Agarwal, S., Vora, A., Pandey, G., 2019. Localization Requirements for Autonomous Vehicles. *SAE International Journal of Connected and Automated Vehicles* 2 (3), pp. 173–190.
- RIEGL Laser Measurement Systems GmbH, 2012. RIEGL VMX-250.
- Rusinkiewicz, S., Levoy, M., 2001. Efficient variants of the ICP algorithm. In: *Proceedings Third International Conference on 3-D Digital Imaging and Modeling*. pp. 145–152.
- Rusu, R. B., Blodow, N., Beetz, M., 2009. Fast Point Feature Histograms (FPFH) for 3D registration. In: *2009 IEEE International Conference on Robotics and Automation*. pp. 3212–3217.
- Schlichting, A., Brenner, C., 2014. Localization using automotive laser scanners and local pattern matching. In: *2014 IEEE Intelligent Vehicles Symposium Proceedings*. pp. 414–419.
- Segal, A., Hähnel, D., Thrun, S., 2009. Generalized-ICP. In: Trinkle, J., Matsuoka, Y., Castellanos, J. A. (Eds.), *Robotics: Science and Systems*. The MIT Press.

- Statistisches Bundesamt, 2023. Statistischer Bericht - Verkehrsunfälle Zeitreihen - 2013-2022. <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Verkehrsunfaelle/Publikationen/Downloads-Verkehrsunfaelle/statistischer-bericht-verkehrsunfaelle-zeitreihen-5462403227005.xlsx>, [Online; accessed 10-October-2023].
- Steinke, N., Ritter, C.-N., Goehring, D., Rojas, R., 2021. Robust LiDAR Feature Localization for Autonomous Vehicles Using Geometric Fingerprinting on Open Datasets. *IEEE Robotics and Automation Letters* 6 (2), pp. 2761–2767.
- Stückler, J., Behnke, S., 2014. Multi-resolution surfel maps for efficient dense 3D modeling and tracking. *Journal of Visual Communication and Image Representation* 25 (1), pp. 137–147, visual Understanding and Applications with RGB-D Cameras.
- Sun, L., Adolfsson, D., Magnusson, M., Andreasson, H., Posner, I., Duckett, T., 2020. Localising Faster: Efficient and precise lidar-based robot localisation in large-scale environments. In: *2020 IEEE International Conference on Robotics and Automation (ICRA)*. pp. 4386–4392.
- Tang, T. Y., Yoon, D. J., Pomerleau, F., Barfoot, T. D., 2018. Learning a Bias Correction for Lidar-Only Motion Estimation. In: *2018 15th Conference on Computer and Robot Vision (CRV)*. pp. 166–173.
- Thomas, H., Qi, C. R., Deschaut, J.-E., Marcotegui, B., Goulette, F., Guibas, L. J., 2019. KPConv: Flexible and Deformable Convolution for Point Clouds. *Proceedings of the IEEE International Conference on Computer Vision*.
- Thrun, S., Burgard, W., Fox, D., 2005. Probabilistic Robotics. The MIT Press.
- Tombari, F., Salti, S., Di Stefano, L., 2013. Performance Evaluation of 3D Keypoint Detectors. *Int. J. Comput. Vision* 102 (1–3), pp. 198–220.
- Tossaint, M., Samson, J., Toran, F., Ventura-Traveset, J., Hernandez-Pajares, M., Juan, J., Sanz, J., Ramos-Bosch, P., 2007. The Stanford - ESA Integrity Diagram: A New Tool for The User Domain SBAS Integrity Assessment. *NAVIGATION: Journal of The Institute of Navigation* 54 (2), pp. 153–162.
- Tzoumas, V., Antonante, P., Carlone, L., 2019. Outlier-Robust Spatial Perception: Hardness, General-Purpose Algorithms, and Guarantees. In: *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. pp. 5383–5390.
- Velodyne LiDAR, Inc., 2016. VLP-16 Velodyne LiDAR Puck.
- Verma, A., Ajit, S., Karanki, D., 2015. Reliability and Safety Engineering. Springer Series in Reliability Engineering. Springer London.
- Vizzo, I., Chen, X., Chebrolu, N., Behley, J., Stachniss, C., 2021. Poisson Surface Reconstruction for LiDAR Odometry and Mapping. In: *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*.
- Vizzo, I., Guadagnino, T., Behley, J., Stachniss, C., 2022. VDBFusion: Flexible and Efficient TSDF Integration of Range Sensor Data. *Sensors* 22 (3).
- Vizzo, I., Guadagnino, T., Mersch, B., Wiesmann, L., Behley, J., Stachniss, C., 2023. KISS-ICP: In Defense of Point-to-Point ICP – Simple, Accurate, and Robust Registration If Done the Right Way. *IEEE Robotics and Automation Letters (RA-L)* 8 (2), pp. 1029–1036.
- Wang, W., Saputra, M. R. U., Zhao, P., Gusmao, P., Yang, B., Chen, C., Markham, A., Trigoni, N., 2019. DeepPCO: End-to-End Point Cloud Odometry through Deep Parallel Neural Network. In: *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. pp. 3248–3254.
- Wang, Y., Solomon, J. M., 2019a. Deep Closest Point: Learning Representations for Point Cloud Registration. In: *The IEEE International Conference on Computer Vision (ICCV)*.
- Wang, Y., Solomon, J. M., 2019b. PRNet: Self-Supervised Learning for Partial-to-Partial Registration. In: Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R. (Eds.), *Advances in Neural Information Processing Systems*. vol. 32. Curran Associates, Inc.

- Wei, P., Wang, X., Guo, Y., 2020. 3D-LIDAR Feature Based Localization for Autonomous Vehicles. In: *2020 IEEE 16th International Conference on Automation Science and Engineering (CASE)*. pp. 288–293.
- Wen, W., Hsu, L.-T., Zhang, G., 2018. Performance Analysis of NDT-based Graph SLAM for Autonomous Vehicle in Diverse Typical Driving Scenarios of Hong Kong. *Sensors* 18 (11).
- Wilbers, D., Merfels, C., Stachniss, C., 2019. Localization with Sliding Window Factor Graphs on Third-Party Maps for Automated Driving. In: *2019 International Conference on Robotics and Automation (ICRA)*. pp. 5951–5957.
- Williamson, D. P., Shmoys, D. B., 2011. The Design of Approximation Algorithms. Cambridge University Press.
- Wolcott, R. W., Eustice, R. M., 2014. Visual localization within LIDAR maps for automated urban driving. In: *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. pp. 176–183.
- Wolcott, R. W., Eustice, R. M., 2015. Fast LIDAR localization using multiresolution Gaussian mixture maps. *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2814–2821.
- Wörner, M., Schuster, F., Dölitzscher, F., Keller, C. G., Haueis, M., Dietmayer, K., 2016. Integrity for autonomous driving: A survey. In: *2016 IEEE/ION Position, Location and Navigation Symposium (PLANS)*. pp. 666–671.
- Yang, H., Carlone, L., 2019. A Polynomial-Time Solution for Robust Registration with Extreme Outlier Rates. In: *Proc. Robot.: Sci. Syst.* [Online]. Available: http://rss2019.informatik.uni-freiburg.de/papers/0013_FI.pdf (pdf), http://rss2019.informatik.uni-freiburg.de/videos/0013_VI.fi.mp4 (video), <http://news.mit.edu/2019/spotting-objects-cars-robots-0620> (media), <https://www.sciencedaily.com/releases/2019/06/190620121444.htm> (media), http://www.ansa.it/canale_scienza_tecnica/notizie/tecnologie/2019/06/21/i-robot-imparano-a-vedere-nella-nebbia-_9e59485c-ff17-4d62-8224-1d42f44111b9.html (media).
- Yang, H., Shi, J., Carlone, L., 2021. TEASER: Fast and Certifiable Point Cloud Registration. *IEEE Transactions on Robotics* 37 (2), pp. 314–333.
- Yang, J., Li, H., Campbell, D., Jia, Y., 2016. Go-ICP: A Globally Optimal Solution to 3D ICP Point-Set Registration. *IEEE Trans. Pattern Anal. Mach. Intell.* 38 (11), pp. 2241–2254.
- Yin, H., Wang, Y., Ding, X., Tang, L., Huang, S., Xiong, R., 2020. 3D LiDAR-Based Global Localization Using Siamese Neural Network. *IEEE Transactions on Intelligent Transportation Systems* 21, pp. 1380–1392.
- Zhang, J., Singh, S., 2017. Low-drift and Real-time Lidar Odometry and Mapping. *Autonomous Robots* 41 (2), pp. 401 – 416.
- Zhang, Y., Wang, L., Jiang, X., Zeng, Y., Dai, Y., 2022. An efficient LiDAR-based localization method for self-driving cars in dynamic environments. *Robotica* 40 (1), pp. 38–55.
- Zhang, Z., 1994. Iterative point matching for registration of free-form curves and surfaces. *International journal of computer vision* 13, pp. 119–152.
- Zhang, Z., 1997. Parameter estimation techniques: A tutorial with application to conic fitting. vol. 15. Elsevier.
- Zheng, Y., Sugimoto, S., Okutomi, M., 2011. Deterministically maximizing feasible subsystem for robust model fitting with unit norm constraint. In: *CVPR 2011*. pp. 1825–1832.
- Zhou, Q.-Y., Park, J., Koltun, V., 2016. Fast Global Registration. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (Eds.), *Computer Vision – ECCV 2016*. Springer International Publishing, Cham, pp. 766–782.
- Zhu, N., Marais, J., Bétaille, D., Berbineau, M., 2018. GNSS Position Integrity in Urban Environments: A Review of Literature. *IEEE Transactions on Intelligent Transportation Systems* 19 (9), pp. 2762–2778.

Acknowledgements

This dissertation was written as part of my work as a research associate at the Institute of Cartography and Geoinformatics at the Leibniz University Hannover and as a member of the Research Training Group 2159 ‘Integrity and Collaboration in Dynamic Sensor Networks’ (i.c.sens), funded by the Deutsche Forschungsgemeinschaft (DFG).

First of all, I would like to express my deepest gratitude to my supervisor, Prof. Dr.-Ing. Claus Brenner, for his constant support and guidance throughout the time of my doctoral research. Your deep knowledge and enthusiasm about laser scanning, LiDAR-based localization, and robust estimation have always motivated and inspired me for my own work. I am especially grateful for the long discussions, where new ideas emerged and I learned a lot.

I would also like to thank Prof. Dr.-Ing. Steffen Schön and Prof. Dr. rer. nat. Andreas Nüchter for their participation in the examination committee of this dissertation.

Furthermore, I would like to thank Prof. Dr.-Ing. habil. Monika Sester for giving me the possibility to pursue my Ph.D. at the Institute of Cartography and Geoinformatics. I am especially thankful for the great opportunities I have had, the highlights being the research stay in Hong Kong and attending various conferences.

A big thank you also to all colleagues at IKG for the great working atmosphere that made work actually not feel like work. I am grateful for the collaborations and discussion where I could learn and get new insights. I really enjoyed my time here.

Moreover, I want to thank all i.c.sens members for the scientific exchange during the various types of gatherings, such as Jour-Fixes, colloquia and retreats. I am thankful for the participation in the joint projects such as the hackathon, the measurement campaigns, and the data publication, which were great collaborations where I could gain a lot of experience.

Finally, I would like to thank my family for their unwavering support and encouragement, not only during my doctoral studies, but also during school and university.

Curriculum Vitae

Personal information

Name	Jeldrik Axmann
Date of birth	June 9th, 1995
Place of birth	Neustadt am Rübenberge, Germany

Education

2013	Abitur (Gymnasium Stolzenau)
2013 - 2017	Dual study program consisting of: <ul style="list-style-type: none">- Apprenticeship as a toolmaker (Volkswagen AG)- B.Eng. Automotive Engineering (Ostfalia University of Applied Sciences)
2017 - 2020	M.Sc. Navigation and Field Robotics (Leibniz University Hannover)

Work experience

Since June 2020	Research associate at Institute of Cartography and Geoinformatics Hannover (Leibniz University Hannover) and member of the Research Training Group 2159 'Integrity and Collaboration in Dynamic Sensor Networks' (i.c.sens)
-----------------	---

Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Leibniz Universität Hannover

(Eine vollständige Liste der Wiss. Arb. ist beim Geodätischen Institut, Nienburger Str. 1, 30167 Hannover erhältlich.)

Nr. 374	REN, Le:	GPS-based Precise Absolute and Relative Kinematic Orbit Determination of Swarm Satellites under Challenging Ionospheric Conditions (Diss. 2021)
Nr. 375	XU, Wei:	Automatic Calibration of Finite Element Analysis Based on Geometric Boundary Models from Terrestrial Laser Scanning (Diss. 2021)
Nr. 376	FENG, Yu:	Extraction of Flood and Precipitation Observations from opportunistic Volunteered Geographic Information (Diss. 2021)
Nr. 377	YANG, Chun:	A hierarchical deep learning framework for the verification of geospatial databases (Diss. 2021)
Nr. 378	MEHLTRETTER, Max:	Uncertainty Estimation for Dense Stereo Matching using Bayesian Deep Learning (Diss. 2021)
Nr. 379	KAZIMI, Bashir:	Self Supervised Learning for Detection of Archaeological Monuments in LiDAR Data (Diss. 2021)
Nr. 380	PETERS, Torben:	Learning Multi-View 2D to 3D Label Transfer for Semi-Supervised Semantic Segmentation of Point Clouds (Diss. 2022)
Nr. 381	WASSINK, Martin:	Kommunal- und Regionalentwicklung durch Kooperation und Teilung von Verantwortung in ländlichen Räumen - eine multiperspektivische Untersuchung an Beispielen aus dem Raum Steinwald/Fichtelgebirge (Diss. 2022)
Nr. 382	GOLDSCHMIDT, Jürgen:	Die Berücksichtigung künftiger Entwicklungen bei der Verkehrswertermittlung (Diss. 2022)
Nr. 383	KRUSE, Christian:	Impact maps from bomb craters detected in aerial wartime images using marked point processes (Diss. 2023)
Nr. 384	ZOURLIDOU, Stefania:	Traffic Regulation Recognition from GPS Data (Diss. 2023)
Nr. 385	SLEDZ, Artuom:	Thermal anomaly detection based on information fusion from optical and infrared images (Diss. 2023)
Nr. 386	WITTICH, Dennis:	Deep Domain Adaptation for the Pixel-wise Classification of Aerial and Satellite Images (Diss. 2023)
Nr. 387	SINGH, Vishwa Vijay:	Lunar Laser Ranging - Improved Modelling and Parameter Estimation (Diss. 2023)
Nr. 388	HARTMANN, Jens:	Hochgenaue 3D-Erfassung von Großstrukturen durch kinematisches terrestrisches Laserscanning (Diss. 2023)
Nr. 389	ZHUGE, Xia:	Characterizing slope instability kinematics by integrating multi-sensor satellite remote sensing observations (Diss. 2023)
Nr. 390	DOROZYNSKI, Mareike Marianne:	Image Classification and Retrieval in the Context of Silk Heritage using Deep Learning (Diss. 2023)
Nr. 391	KNABE, Annike:	New Concepts for Gravity Field Recovery using Satellites (Diss. 2023)
Nr. 392	KALIA, Andre:	Landslide activity detection based on nationwide Sentinel-1 PSI datasets (Diss. 2023)
Nr. 393	BROCKMEYER, Marco:	Modellierung von Bodenbewegungen anhand heterogener Messverfahren am Beispiel der niedersächsischen Landesfläche (Diss. 2023)
Nr. 394	ZHANG, Mingyue:	Characteristics and Benefits of Differential Lunar Laser Ranging (Diss. 2023)
Nr. 395	DENNIG, Dirk:	Entwicklung eines kinematischen Profilvermessungssystems am Beispiel Kranbahnvermessung (Diss. 2024)
Nr. 396	FUEST, Stefan:	Nudging travelers to societally favorable routes by means of cartographic symbolization (Diss. 2024)
Nr. 397	MOFTIZADEH, Rozhin:	Advanced Particle Filtering for Vehicle Navigation based on Collaborative Information (Diss. 2024)
Nr. 398	VASSILEVA, Magdalena Stefanova:	Satellite Radar Interferometry for Geohazards: from ground deformation to processes understanding (Diss. 2024)
Nr. 399	MALINOVSKAYA, Anna:	Statistical Process Monitoring of Networks (Diss. 2024)
Nr. 400	BANNERT, Jörn:	Der Einfluss von Straßenverkehrslärm und Umgehungsstraßen auf Grundstückswerte in Ortslagen - Bestimmung mittels Expertenbefragung nach der Delphi-Methode (Diss. 2024)
Nr. 401	AXMANN, Jeldrik:	Maximum consensus localization using LiDAR (Diss. 2024)

