From Vehicle Setup to Dataset Generation: A Holistic Approach to Long-Range Automated Valet Parking Development

Amine Ben Arab*, Jing Gu†, Hassan Mohammadi†, Mia Book‡, Thomas Brandmeier‡, Alireza Ferdowsizadeh Naeeni †

Abstract: Sensor fusion is one of the trend topics in the automotive field, which aims to achieve robustness by relying on the information from different sensors. But to achieve further progress in this field, the availability or generation of multimodal data under different contexts is a vary important topic. In this paper we focus in the description of the necessary methodologies to generate such data considering Long-Range Automated Valet Driving scenarios, involving a combination of urban and indoor driving scenarios, we devised two distinct methods. For the initial approach, we utilized the installations present in the ISAFE Indoor Testing Facility at CARISSMA (Technische Hochschule Ingolstadt) to gather data for situations in severe environments and supply the necessary references. The second approach involves producing data from real-life situations using the prototyping vehicle of Expleo Germany GmbH. To accomplish this, we equipped the vehicle with various sensors and examined techniques from the latest technological developments to decrease the burden of dataset generation while meeting sensor fusion's demands. In this work, we present both approaches and our results for the methods employed to establish our data generation pipeline.

Schlüsselwörter: AVP, Sensors, Sensor-2-Sensor Calibration, Time Synchronization, Reference systems.

1 Introduction

Long-range Automated Valet Parking (LAVP)[1] is an extension of the AVP function. This service not only drives and parks autonomously in a parking facility like typical AVP but also considers driving the vehicle from a drop-off zone far from the parking facility's entry. Therefore, this function encompasses more diverse scenarios. On one hand, the system must cope with the challenges of urban environments, including dynamic conditions and various weather elements such as rain and fog. On the other hand, it must also manage the intricate maneuvering involved in enclosed parking situations.

Throughout the years, multiple datasets [2...13] have been made available to the public for AD use cases due to the significant time and cost required to generate a dataset.

*Expleo Germany GmbH, Wilhelm-Wagenfeld-Str.1-3, 80807 München (e-mail: amine.ben-arab@expleogroup.de).

†Expleo Germany GmbH, Salufer 8, 10587 Berlin (e-mail: name.lastname@expleogroup.com).

‡Technische Hochschule Ingolstadt, CARISSMA ISAFE, Esplanade 10, Ingolstadt (e-mail: name.nachname@carissma.eu).

These datasets have provided increasing sensor [2] and scenario [3] diversity, enabling the research community to rapidly prototype for scene understanding. However, to the best of our knowledge, there is currently no open dataset that accurately reflects the various scenarios that describe our particular use case.

In this paper, we present two approaches to data acquisition that focus on a diverse data set that can be collected in a time-efficient manner without extensive human intervention. The initial approach is achieved by capturing mock scenarios within a controllable indoor environment. By utilizing an indoor positioning system (IPS) to determine the location of individual objects, the labeling procedure can be partially automated, rendering it dependable even during undesirable weather conditions. The second approach involves gathering data in real-world environments. In this regard, we present the sensor setup we implemented in our prototyping vehicle, along with our sensor calibration and time synchronization methods. We also introduce our reference system for the vehicle's ego-localization task indoors.

2 Related Works

Over the last years, dataset generation has been a focus of research in AD, especially multi-modal datasets have become more and more important [3]. This is due to the rapid progress in sensor development in the LiDAR and radar fields. There are numerous techniques for creating a dataset to train and test driving-related algorithms, including collecting data in real-world settings, simulated environments, or performing mock scenarios. Real-world data acquisition provides realistic and authentic sensor data with accompanying noise but poses the risk of generating edge cases or critical scenarios. Additionally, annotating the data is a time-consuming and potentially incomplete process. For example, under rainy conditions, Lidar can't model the whole object of interest and therefore, it is challenging to generate a 3D bounding Box that limits this object. With simulated data, it is easy to generate critical and rare scenarios as well as annotations. However, the sensor data obtained from LiDAR or radar might not be comparable with sensors used in the actual vehicle. The Table 1 summarizes various real-world multi-modal datasets as well as datasets that include adverse weather conditions. It is evident that the majority of the datasets were collected during favorable conditions (cloudy or clear) or did not label all collected adverse conditions [4]. Only a few, such as Rain WCity, focus on collecting data during adverse weather conditions [5]. Additionally, parking scenarios are either not explicitly mentioned in the literature or are absent from most datasets. Different types of sensors are being researched to overcome the challenges of AD. However, to meet the automotive industry's standards, factors such as reliability, durability, and cost-effective manufacturing and maintenance must be considered. In this context, the sensors with the potential to solve high-level AD, found in the datasets listed in Table 1, are radar, camera, and LiDAR. These different sensors can be used as a redundancy solution, to enhance the reliability of the AD system, as well as a complementary solution to extend the continuity of the AD system.

To fuse the information from the sensors properly, a calibration is a necessary step. To ensure spatial consistency between sensor data, it is required to transform each sensor's readings into a common coordinate system, known as extrinsic calibration. It can be accomplished through different methodologies. For example, the sensor data can then be optimized through manual alignment, as seen in Radiate [4] and partially in Nuscenes [2] and EU long term [6], or the transformation between sensors can be estimated automatically using data from the sensors, which consists in target-based, and targetless sensor-2-sensor calibration [7]. Target-based calibration utilizes a specially designed target with known dimensions to extract features from each sensor. The chessboard target is frequently used to extract its corners as a 3D pattern. in Pixset [8], To estimate the transformation matrices between the cameras, the authors employed a closed-loop optimization method and used the perspective-n-point method for calibration between the camera and LiDAR. For chessboard-based calibration on the Kitti dataset, they utilized their later work [9]. Additionally, a unique target was created for the calibration of LiDAR, camera, and Radar at Tu-Delft [10]. A further investigation of the calibration method is presented in section 3.2. Targetless calibration consists of two sub-methods: Motion-based and feature-based calibration. Motion-based [11] calibration estimates the vehicle's trajectory from each sensor during various driving maneuvers and matches those trajectories to estimate the transformation between the sensors. This technique is commonly used for proprioceptive sensor calibration, including IMU or GNS. In the KITTI dataset, the hand-eye method was applied to calibrate IMU and LiDAR [9]. The feature-based calibration [12] extracts similar features from each sensor within the shared scene and estimates the transformation matrix by identifying correspondences between these features. In the A2D2 dataset [13], calibration between the camera and LiDAR is enhanced by utilizing edge correspondences.

Time synchronization of sensors is critical to the sensor fusion process. It is essential to maintain temporal consistency between sensor data. Simultaneous sensor triggering is the most accurate method for fusing sensors by simultaneously registering events on each sensor. In the KITTI dataset, cameras are triggered simultaneously with LiDAR by a reed contact in the LiDAR that activates the camera when the mechanical LiDAR's scanner faces forward. However, simultaneous sensor triggering is not feasible for the various automotive-grade sensors. The alternative is to synchronize the sensors to a common clock, and thus this method maintains the temporal correlation of events between sensors. this method can be achieved either at the software level or at the hardware level. In the EU long-term dataset, the LiDARs are synchronized at the hardware level by synchronizing the clocks of each LiDAR with the acquired GPS signal. Another method is to timestamp the sensor data with the time it arrives at the data logger. This method may be less accurate than others as it includes a non-deterministic delay in data transmission.

As most of the datasets are generated outside, an RTK GPS is usually used for generating reference positioning of the vehicle. It's an aviation-grade positioning system and is considered the most accurate positioning system for outdoor use cases.

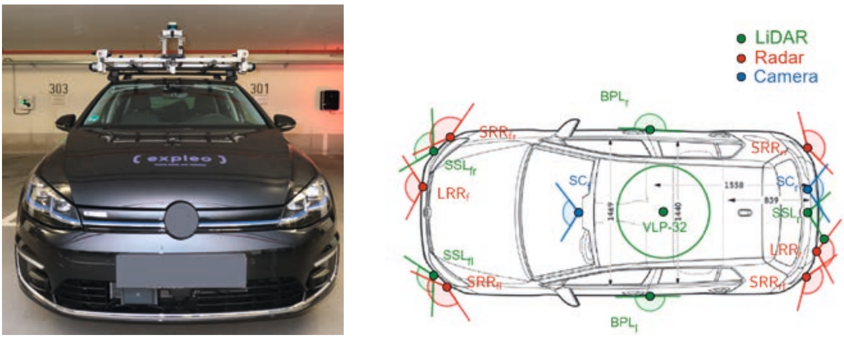
Data Annotation consists of labeling the different static and dynamic objects in the vehicle's surroundings, which are relevant to the AD function. It is done manually with the help of annotation tools [14][9] [3]. It's a rigorous task and hence very time-consuming as the precision of the labels has a big impact on the targeted task. In Cityscape [14], the semantic labels generation for each frame required 1.5 hours on average. however, there are some efforts to generate these labels automatically and reduce the labeling burden. Human supervision is still needed to validate or correct the proposed labels. For example, in the aiMotive dataset [15], a method was used to search for possible candidates in a sequence of data. the candidate's labels are further enhanced with an object tracking algorithm, where they are optimized recursively.

Table 1: State-of-the-art datasets. Abbreviations scenarios: (U)urban, (SU)suburban, (H)highway and (P)parking; conditions: (G)good, (N)night, (R)rain, (F)fog and (S)snow; Sensors: (M)mono/(S)stereo (C)camera, (M)mechanical/(SS)solid-state (L)LiDAR and (R)radar; synchronization (Sync.): (H)hardware and (S)software; calibration: (M)manual, (TL)target-based and (TB)target-less; (-)no information available or not available in the dataset.

Dataset	Scenarios				Conditions					Sensors			GPS/IMU	Annotation			Classes	Sync.	Calib.
	U	SU	H	P	G	N	R	F	S	C	L	R		2D	3D	Sem.			
KITTI [9]	✓	✓	✓	-	✓	-	-	-	-	1 SC	1 ML	-	✓	-	✓	✓	28	H/S	TB
nuScene [2]	✓	-	-	-	✓	✓	✓	✓	-	6 MC	1 ML	5 R	✓	-	✓	✓	23	S	TB/TL
View of Delft [10]	✓	-	-	-	✓	-	-	-	-	1 SC	1 ML	1 R	✓	-	✓	-	13	-	TB
Waymo Open [16]	✓	-	-	-	✓	-	✓	-	-	5 MC	5 ML	-	-	✓	✓	✓	4	S	-
A2D2 [13]	✓	✓	✓	-	✓	-	✓	-	-	6 MC	5 ML	-	-	-	✓	✓	38	-	TB/TL
Cityscape (and 3D) [14]	✓	-	-	-	✓	-	✓	-	-	1 MC	-	-	-	-	✓	✓	30	-	-
Radiate [4]	✓	-	✓	✓	✓	✓	✓	✓	✓	1 SC	1 ML	1 R (360°)	✓	✓	(✓)	-	8	-	TB
Pixset [8]	✓	✓	-	✓	✓	-	✓ (10%)	-	-	3 MC	1 ML/ 1 SSL	1 R	✓	-	✓	-	20	PTP	TB
aiMotive [15]	✓	-	✓	-	✓	-	✓ (4%)	-	-	5 MC	1 ML	2 R	✓ (GNSS)	✓	✓	-	14	-	-
Rain Wcity [5]	✓	-	-	-	✓	-	1 MC	-	-	1 MC	-	-	-	-	-	-	-	-	-
EU Long-term [6]	✓	✓	-	-	✓	-	-	-	✓	2 SC/ 2 MC	3 ML/ 1 SSL	1 R	✓ (GNSS)	-	-	-	-	H/S	TB
Zenseact [3]	✓	✓	✓	-	✓ (80%)	✓ (19%)	✓ (16%)	✓ (2%)	✓ (2%)	1 MC	3 ML	-	✓ (GNSS)	✓	✓	✓	15	-	-

3 Fulfilling the Requirements of Sensor Fusion

3.1 Sensors and Sensors Placement



(a) Prototyping vehicle

(b) Sensor Placement Architecture

Figure 1: Perception System

Expleo Germany GmbH has been developing a prototyping vehicle and corresponding functionalities for highly automated and autonomous driving functions like AVP and

Autonomous Valet Driving System (AVDS) in the last few years. The newly enhanced sensor setup integrated into Expleo's vehicle collects data in real-world scenarios. It consists of two sub-systems: the AD setup and the reference setup. The autonomous driving (AD) setup includes a set of sensors that can potentially be integrated into a production vehicle, and it's composed of three solid-state LiDARs (SSL), Robosense M1, two stereo cameras (SC), four short-range radars (SRR), and two long-range radars (LRR). Those sensors have been integrated into the vehicle chassis as shown in Figure 1b. The reference sensor setup consists of three mechanical LiDARs (one Velodyne VLP-32 and two Robosense BlackPearls(BP)) positioned atop the vehicle to model the prototyping vehicle's entire surroundings, as depicted in Figure 1a.

3.2 Sensor Calibration

For our prototype vehicle sensor setup, manual calibration can prove challenging due to some sensors being embedded in the vehicle chassis, making their centers invisible. To overcome this, we have explored the possibility of estimating the transformation between sensors using the data they generate. In this study, our focus is on target-based multi-sensor calibration. It is a challenging task to extract environmental features with radars that are comparable to LiDARs or cameras as well as obtaining an accurate trajectory with radars. In [17], a multi-sensor calibrator (MSCT) was proposed to calibrate sensors similar to our setup, including LiDAR, camera, and radar. The Target comprises of four rings in the center of a rectangular board alongside a radar reflector positioned in the middle behind the board. The 3D positions of the four rings' centers are estimated as a pattern from the LiDAR and camera data in addition to the corner reflector from the radar data. The optimization method is further refined by incorporating a loop closure constraint between the sensors.

However, we discovered while testing this calibration procedure on our sensor setup, the patterns were not detectable by the BP LiDAR. This is because the implemented detection procedure is based on the horizontal line scans that reflect the rings on the board. Consequently, we chose an alternative solution that is not dependent on the LiDAR's point cloud structure. We replaced the rings board (RB) with a chessboard (CB), which is a widely used target for camera-2-LiDAR calibration [18]. For the detection of chessboard edges using cameras, we implemented the commonly-used pattern detection method [19]. To detect chessboard edges with LiDAR, we first estimated the point clouds that belong to the chessboard using the RANSAC algorithm. We then used the convex hull algorithm to extract the border of the chessboard, and based on that, used a minimum rectangle fitting algorithm to extract the edges of the chessboard.

For comparing and validating the various calibration solutions tested, we constructed a test bench as illustrated in Figure 2. To accurately estimate the sensor center, we designed customized mountings to digitally measure the distance between the sensor center and a reference point on the mountings. We manually measured the distances between these designated reference points for each sensor to calculate the transformation matrix between the sensors.

Figure 2: Test bench for testing different sensor calibration tools

Table 2: Results of multiple calibration tools on different sensor-to-sensor constellations

Sensor-to-Sensor	Calibration tool	Rotation Error (deg)	Translation error (cm)
VLP-32 to camera	MSCT Tool with CB	[0.83 0.52 0.74]	[2.08 1.31 3.94]
	MSCT Tool with RB	[0.3 0.72 1.33]	[2.09 1.38 3.9]
	Matlab tool	[0.15 0.21 2.1]	[2.02 3.17 0.19]
VLP-32 to BP	MSCT Tool	[1.51 0.25 1.164]	[2.62 0.67 1.39]
M1 to LRR	MSCT Tool	[0.074 1.49 1.41]	[1.3 3.08 5.39]

Table 2 displays the outcomes of our calibration method for various sensor-to-sensor configurations examined on the constructed test bench. We also evaluated the VLP-32-to-camera calibration with alternate methodologies like the MSCT Tool with RB, and the Lidar and Camera Calibration tool from Matlab. We observed similar results for this sensor constellation. However, in general, the calibration results differ with an average 5.5 cm translational and 2-degree rotational error from manual calibration. This can have a negative impact on the accuracy of sensor fusion applications. It is also worth mentioning that the positioning of the sensors on the test bench differs from the prototyping vehicle setup and can lead to diverse calibration results.

3.3 Time Synchronisation

As mentioned in section 2, time synchronization plays a crucial role in sensor fusion, particularly for autonomous driving applications. A time offset between different sensors can lead to significant positional errors when observing the same object, especially in highly dynamic conditions. As outlined in [20], the delay in message delivery can be divided into three components: the delay from the sender (sensors), the delay from propagation, and the delay from the receiver (data logger).

Conversely, the time delay from the receiver is non-deterministic due to potential operating system overheads. When the data logger approaches its data processing capacity limit, a conventional solution like that described in [21], which uses the data logger to assign timestamps to different sensors, struggles to assign timestamps reliably. This challenge prompted the development of our time synchronization framework, specifically designed to address time delays in message delivery (see figure 3).

The propagation delay is deterministic and relies on the distance between the sender and the receiver. In our use case, the delay from the sender (the sensors) is also deterministic because it is implemented on a hardware level. On the other hand, the time delay from the receiver is non-deterministic due to possible overheads from the operating

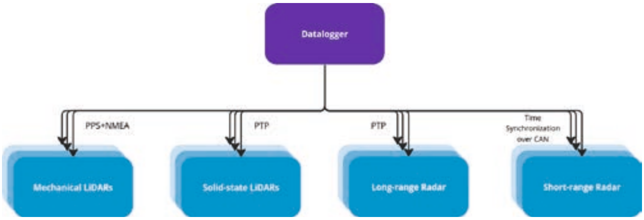


Figure 3: Overview of the created time synchronization framework

system. When the data logger nears its data processing limit, a conventional solution like the one described in [21], which uses the data logger to assign timestamps to different sensors, struggles to assign timestamps reliably. To address this issue, we developed a time synchronization framework (refer to figure 3) that is specifically tailored to handle message delivery delays.

In our framework, mechanical LiDARs (e.g. VLP-32C and BP) use the GPS time synchronization method [22, 23]. However, in underground parking garages, there is no GPS signal. As a solution, we propose an approach that simulates GPS signals via our data logger's clock. We simulate the GPS signal for mechanical LiDARs by producing the Pulse Per Second (PPS) and National Marine Electronics Association (NMEA) sentence, following a previous study [24].

Precision Time Protocol (PTP) is used for time synchronization of solid-state LiDARs (M1) and LRRs. PTP, introduced in 2002 as a method for synchronizing clocks in distributed systems (Eidson, 2002), is used in conjunction with the data logger as the master clock and the solid-state LiDARs and long-range radar as the slave clock.

The SRRs exclusively support time synchronization over the Controller Area Network (CAN). We adhere to the specifications outlined in the Automotive Open System Architecture (AUTOSAR) [25] to implement this time synchronization protocol.

Table 3: Time Synchronization Framework comparison

Sensors	Typical Solution STD(ms)	Our STD(ms)
VLP-32C	0.15	**0.02**
BlackPearl	0.4	**0.05**
M1	0.1	**0.03**
SRR	4.4	**1.2**
LRR	3.6	**0.6**

We compare our framework with the typical solution, which assigns the arrival time to the message when the message arrives at the data logger [21]. We use the standard deviation of the sensor clock with the time of arrivals, which is an evaluation metric used in [24], to evaluate our time synchronization protocol.

As shown in table 3, we collected 3000 messages and measured the stability of the packet-to-packet period by calculating the standard deviation. Our framework shows more stability because the operating system overheads will not affect the time synchronization performance.

4 Methodology

4.1 Longterm AVP Data Generation in Controlled Environment

Currently, available open datasets for AD's perception module are reliant on data collected in the real world where critical situations and adverse weather conditions are infrequent. Additionally, research is confined to using sensor setups that may not satisfy the demands of new sensor arrangements, and where the collected data needs precise annotations. In most modern datasets, LiDAR sensors are frequently utilized as a reference owing to their high accuracy. However, during adverse weather conditions like foggy weather, LiDAR signal performance can be significantly compromised, which negatively impacts the quality of the annotation process. To resolve these concerns, we present a methodology for collecting and annotating data for vehicle ego localization as well as object detection and prediction in an LAVP context, specifically in regard to pedestrian-related scenarios, which is the most common Vulnerable Road User present in LAVP scenarios. Considering the ISAFE indoor testing facility, it is feasible to generate various scenarios involving harsh weather conditions, such as varying rain and fog intensities, as well as controlled lighting conditions. In the available test area, one can simulate different real-world scenarios. Parking situations are also taken into account for data collection. A brief outline of the potential weather conditions that could be incorporated into the data is presented in figure 4.

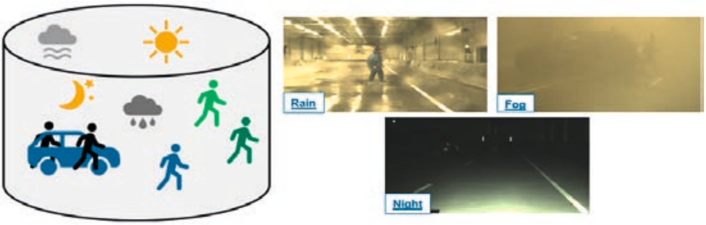


Figure 4: Scenario variety for pedestrian detection at ISAFE.

Each condition presents distinct challenges for various sensors. For example, cameras face visibility issues while the radar point cloud may have ghost objects when multiple objects are in close proximity. With different combinations collected with different objects, it is possible to either improve or develop new sensor fusion algorithms to achieve a robust performance under different conditions. Additionally, it can serve as a complementary source to the existing state-of-the-art datasets.

4.2 Reference Systems

Reference systems are essential in order to evaluate the accuracy of the developed perception functions such as Object Detection and Classification or Vehicle ego-localization.

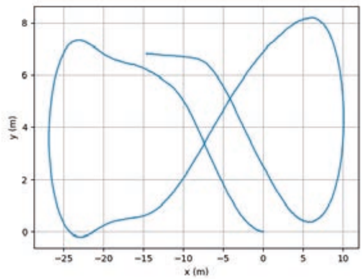
4.2.1 LiDAR-based Reference Localization System

For testing in GPS-deprived environments, utilizing localization systems that are integrated into the infrastructure yields an accurate solution, such as the IPS system (discussed in section 4.2.2) or MOCAP systems. Nevertheless, conducting real-world testing in various locations incurs significant overhead for integrating and calibrating these systems into the infrastructure. In our use case, we chose to implement a LiDAR-based localization system, as it provides an accurate and robust solution.

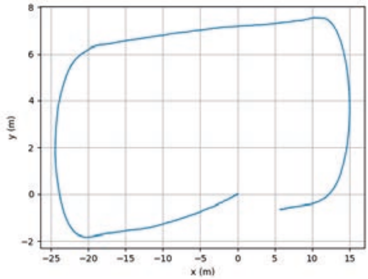
We examined various benchmarks for LiDAR-based localization and selected two solutions: HDL-Graph-SLAM [26] and LEGO_LOAM [27]

HDL-Graph-SLAM generates and optimizes a graph where nodes indicate the sensor positions, and the edges between these nodes represent the odometry constraints (the relative pose between nodes), generated in this case by a point cloud scan matching algorithm, such as GICP [28]. The graph is optimized so that the error function between constraints and positions is minimized. With the help of **floor detection** and **loop-closure**, the graph is optimized and a more precise map of the environment is produced. the LiDAR frames are afterwards reprocessed through a scan-matching algorithm, where these LiDAR frames are registered with the generated map to estimate the final trajectory.

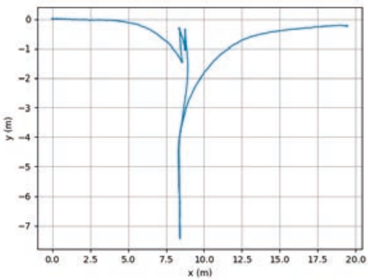
LEGO-LOAM is a LiDAR-based SLAM algorithm based on feature matching. This algorithm discretizes the search space according to the LiDAR point cloud structure, to extract 3D points that belong to edges and surfaces. It estimates the spatial transformation between two consecutive frames by finding correspondences between extracted edges and surface features. In addition, a LiDAR mapping module is used to refine the pose transformation. It matches the features extracted from the latest frame with all the features extracted from the oldest frames stored in a map, which is further optimized by a loop closure detection method.



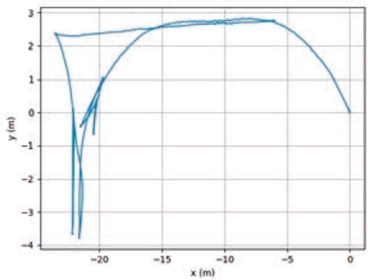
(a) Random trajectory (RT)



(b) Right circle trajectory (RCT)



(c) Forward parking (FP)



(d) Backward parking (BP)

Figure 5: Different tested parking scenarios trajectories

Table 4: LiDAR-based localization results on different trajectories

Method	Evaluation Metrics (m)	RT	RCT	FP	BP
LEGO-LOAM	RMSE	0.07	0.162	0.077	0.074
	MAX Error	0.1433	0.249	0.149	0.115
HDL-GS	RMSE	0.058	0.144	0.086	0.094
	MAX Error	0.169	0.252	0.159	0.191

Figure 5, represents an example of relevant trajectories that we used for testing our Reference localization system, and the table 4 represents the results for these different trajectories. We observe that while the root mean square error (RMSE) results for most of the trajectories are below 10 cm, the maximal error reaches 0.24 cm. This presents a challenge since a reference system should always maintain accuracy across all scenarios. Thus, further improvement of this system is necessary.

4.2.2 Data Annotation Methods

Annotating data is expensive and time-consuming, making evaluating and testing perception systems a challenge. We propose an active learning-based labeling tool and an indoor positioning system-based annotation approach to address this challenge.

Active Learning Based Semi-Auto Labeling Tool for Semantic Segmentation At Expleo, we created a semi-automatic labeling tool for semantic segmentation that employs active learning to reduce the manual effort of human annotation. Notably, we have incorporated the One-vs-All (OVA) method into active learning and found that it enhances diversity for active selection, leading to improved segmentation accuracy [29]. Using uncertainty, the Semi-Auto labeling tool suggests potential candidate points of an image and asks the human annotators for the true label. The human annotation is utilized to retrain the neural network. According to [29], this semi-automatic labeling tool achieves 96% performance in fully supervised learning with 100 pixels per image (0.06% of the entire dataset) on CityScapes [14]. We incorporate this semi-automatic labeling tool in our reference system pipeline to facilitate the prompt and efficient generation of image segmentation labels by human annotators.

IPS Reference Based Annotation For the annotation process, we use the IPS that allows for accurate labeling regardless of weather conditions. Equipped with a receiver antenna, the positioning system can provide precise information, including coordinates, velocity, and heading angle of the object of interest. To effectively develop and test detection methods, reliable ground truth data is crucial, a necessity fulfilled by this positioning system. This system offers significant advantages in the domain of pedestrian and vehicle detection tasks that are based on vision-based sensors like camera, LiDAR, and radar sensors. A notable benefit is its ability to maintain accuracy even in inclement weather conditions, such as rain and fog, where other sensors may produce unreliable or incomplete data. Furthermore, it can generate both 3D and 2D bounding boxes for training object detection models using images and point clouds. See Figure 6 for an illustration of the data collection setup and annotation process.

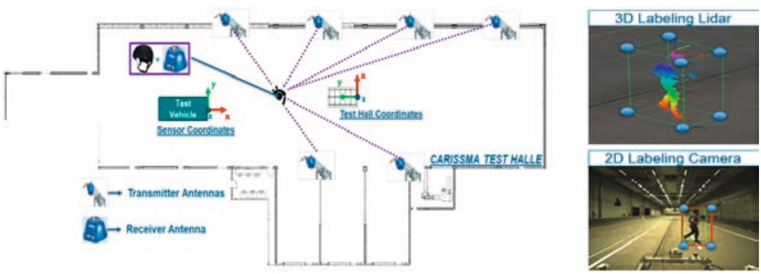


Figure 6: Overview of the annotation framework at ISAFE.

To conduct the annotation process, first, create a 3D bounding box that encloses the entire body of the pedestrian across all data frames in a specific experiment. This can be accomplished using IPS measurements for each coordinate (X, Y, and Z in meters). The reference point of the Z-axis remains at ground level. Using this data, the height (corresponding to the Z dimension of the 3D bounding box) is calculated. Subsequently, predetermined values for the X and Y dimensions are applied. In a second step, the corners of the 3D bounding box project onto the target image (2D space) employing the camera parameters and calibration values obtained from Chapter 3.3. The minimum and maximum values within the image boundaries establish the final 2D bounding box, as illustrated in figure 6. This method can be used regardless of weather conditions. However, as depicted in figure 6, the resulting bounding box (indicated in red), may be larger than the actual size of the object depending on the viewpoint and perspective. Furthermore, this technique can produce complete body bounding boxes even when the pedestrian is partially or entirely obstructed.

Complementarity of Annotation Methods Considering our context, the presented methods can complement each other to offer a more time-efficient approach to data generation. In real-world scenarios where no positioning system is available, the active learning-based method could prove useful for annotation generation. In contrast, in edge cases with harsh weather, the labeling method aided by the IPS presents a viable option. Hence, both methods offer a viable approach to gather supplementary scenarios, both real and in a controlled setting, to create a comprehensive dataset encompassing all possible test cases with variations in weather and scenarios to facilitate complete training and assessment of algorithms.

5 Conclusion

The evaluation of various pre-existing datasets for the development of a long-range automated vehicle revealed open issues in their application due to several factors, including localization in an indoor environment and the acquisition and annotation of adverse weather conditions. We propose a method of generating complementary data, taking advantage of the closed indoor environment found at ISAFE, and emphasize the benefits of an IPS reference-based annotation process. Table 5 shows our method and its possible features in regard to state-of-the-art datasets mentioned in chapter 2. The presented sensor setup is composed of a relatively diverse set of sensors in comparison with state-of-the-art datasets. For multi-sensor calibration, we implemented a solution adequate for our sensor setup. Through a test bench, we showed that the reached accuracy is still not enough in comparison with manual calibration and this can impact using sensor fusion for different use cases. For time synchronization, we showed that our framework provides more stability. Furthermore, we presented our reference systems. In future works, we will further investigate multi-LiDAR-based localization to enhance the accuracy of our reference localization system. This will give us more insight into the needed accuracy for sensor calibration. For the latter, we plan to investigate target-less sensor calibration methods.

Table 5: State-of-the-art datasets. Abbreviations scenarios: (U)urban, (SU)suburban, (H)highway and (P)parking; conditions: (G)good, (N)night, (R)rain, (F)fog and (S)snow; sensors: (M)mono/(S)stereo (C)camera, (M)mechanical/(SS)solid-state (L)LiDAR and (R)radar; synchronization (Sync.): (H)hardware and (S)software; calibration: (M)manual, (TL)target-based and (TB)target-less; (-)no information available or not available in the dataset.

Dataset	Scenarios				Conditions					Sensors			GPS/IMU	Annotation			Classes	Sync.	Calib.
	U	SU	H	P	G	N	R	F	S	C	L	R		2D	3D	Sem.			
KITTI [9]	✓	✓	✓	-	✓	-	-	-	-	✓	✓	-	✓	-	✓	✓	28	H/S	TB
nuScene [2]	✓	-	-	-	✓	✓	✓	✓	-	✓	✓	✓	✓	-	✓	✓	23	H/S	TB/TL
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Our Method	scenario selectable in regard to possibilities of a testing area				✓	✓	✓	✓	-	2 SC	3 ML/ 3 SSL	6 R	IPS	IPS for moving objects 2D & 3D				S	TB

6 Acknowledgements

The research leading to these results is funded by the Bavarian Ministry of Economic Affairs, Energy and Technology and by Expleo Germany GmbH within the project: "SAFE2P - Increase Safety for Pedestrians and Parking" (DIK-350).

References

[1] Muhammad Khalid, Kezhi Wang, Nauman Aslam, Yue Cao, Naveed Ahmad, and Muhammad Khurram Khan. From smart parking towards autonomous valet parking: A survey, challenges and future works. *Journal of Network and Computer Applications*, 175:102935, 2021.

[2] Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. *arXiv preprint arXiv:1903.11027*, 2019.

[3] Mina Alibeigi, William Ljungbergh, Adam Tonderski, Georg Hess, Adam Lilja, Carl Lindstrom, Daria Motorniuk, Junsheng Fu, Jenny Widahl, and Christoffer Petersson. Zenseact open dataset: A large-scale and diverse multimodal dataset for autonomous driving, 2023.

[4] Marcel Sheeny, Emanuele De Pellegrin, Saptarshi Mukherjee, Alireza Ahrabian, Sen Wang, and Andrew Wallace. Radiate: A radar dataset for automotive perception. *arXiv preprint arXiv:2010.09076*, 2021.

[5] Xian Zhong, Shidong Tu, Xianzheng Ma, Kui Jiang, Wenxin Huang, and Zheng Wang. Rainy wcity: A real rainfall dataset with diverse conditions for semantic driving scene understanding. In Lud De Raedt, editor, *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 1743–1749. International Joint Conferences on Artificial Intelligence Organization, 7 2022. Main Track.

[6] Zhi Yan, Li Sun, Tomas Krajnik, and Yassine Ruichek. EU long-term dataset with multiple sensors for autonomous driving. In *Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020.

[7] De Jong Yeong, Gustavo Velasco-Hernandez, John Barry, and Joseph Walsh. Sensor and sensor fusion technology in autonomous vehicles: A review. *Sensors*, 21(6):2140, 2021.

[8] Jean-Luc DÃ©ziel, Pierre Merriaux, Francis Tremblay, Dave Lessard, Dominique Plourde, Julien Stanguennec, Pierre Goulet, and Pierre Olivier. Pixset : An opportunity for 3d computer vision to go beyond point clouds with a full-waveform lidar dataset, 2021.

[9] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The KITTI dataset. *International Journal of Robotics Research*, 32(11):1231 – 1237, September 2013.

[10] Andras Palffy, Ewoud Pool, Srimannarayana Baratam, Julian F. P. Kooij, and Dariu M. Gavrila. Multi-class road user detection with 3+1d radar in the view-of-delft dataset. *IEEE Robotics and Automation Letters*, 7(2):4961–4968, 2022.

[11] Tuomas Välimäki, Bharath Garigipati, and Reza Ghabcheloo. Motion-based extrinsic sensor-to-sensor calibration: Effect of reference frame selection for new and existing methods. *Sensors*, 23(7):3740, 2023.

[12] Chongjian Yuan, Xiyuan Liu, Xiaoping Hong, and Fu Zhang. Pixel-level extrinsic self calibration of high resolution lidar and camera in targetless environments. *IEEE Robotics and Automation Letters*, 6(4):7517–7524, 2021.

[13] Jakob Geyer, Yohannes Kassahun, Mentar Mahmudi, Xavier Ricou, Rupesh Durgesh, Andrew S. Chung, Lorenz Hauswald, Viet Hoang Pham, Maximilian Mühlegg, Sebastian Dorn, Tiffany Fernandez, Martin Jänicke, Sudesh Mirashi, Chiragkumar Savani, Martin Sturm, Oleksandr Vorobiov, Martin Oelker, Sebastian Garreis, and Peter Schuberth. A2D2: Audi Autonomous Driving Dataset. 2020.

[14] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016.

[15] Tamas Matuszka, Ivan Barton, Ádám Butykai, Péter Hajas, Dávid Kiss, Domonkos Kovács, Sándor Kunsági-Máté, Péter Lengyel, Gábor Németh, Levente Pető, Dezső Ribli, Dávid Szeghy, Szabolcs Vajna, and Balint Viktor Varga. aimotive dataset: A multimodal dataset for robust autonomous driving with long-range perception. In *International Conference on Learning Representations 2023 Workshop on Scene Representations for Autonomous Driving*, 2023.

[16] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, Vijay Vasudevan,

Wei Han, Jiquan Ngiam, Hang Zhao, Aleksei Timofeev, Scott Ettinger, Maxim Krivokon, Amy Gao, Aditya Joshi, Yu Zhang, Jonathon Shlens, Zhifeng Chen, and Dragomir Anguelov. Scalability in perception for autonomous driving: Waymo open dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[17] Joris Domhof, Julian FP Kooij, and Dariu M Gavrila. A joint extrinsic calibration tool for radar, camera and lidar. *IEEE Transactions on Intelligent Vehicles*, 6(3):571–582, 2021.

[18] Lipu Zhou, Zimo Li, and Michael Kaess. Automatic extrinsic calibration of a camera and a 3d lidar using line and plane correspondences. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5562–5569. IEEE, 2018.

[19] Andreas Geiger, Frank Moosmann, Ömer Car, and Bernhard Schuster. Automatic camera and range sensor calibration using a single shot. In *2012 IEEE international conference on robotics and automation*, pages 3936–3943. IEEE, 2012.

[20] Ying Weng and Yiming Zhang. A survey of secure time synchronization. *Applied Sciences*, 13(6):3923, 2023.

[21] Puck lidar sensor (vlp-16). velodyne lidar. https://velodynelidar.com/products/puck/dc Accessed: 2023-08-17.

[22] Robosense blackpearl: Hemispherical super wide fov, short-range blind spot lidar. https://www.robosense.ai/en/rslidar/RS-Bpearl. Accessed: 2023-08-17.

[23] Ultra puck: The high-density, long-range image generated by the ultra puck makes it an industry favorite for robotics, mapping, security, driver assistance, and autonomous navigation. https://velodynelidar.com/products/ultra-puck/downloads. Accessed: 2023-08-17.

[24] Marsel Faizullin, Anastasiia Kornilova, and Gonzalo Ferrer. Open-source lidar time synchronization system by mimicking gnss-clock. In *2022 IEEE International Symposium on Precision Clock Synchronization for Measurement, Control, and Communication (ISPCS)*, pages 1–5. IEEE, 2022.

[25] Specification of time synchronization over can autosar cp r21-11. https://www.autosar.org/fileadmin/standards/R21-11/CP/AUTOSAR$_SWS_TimeSyncOverCAN.pdf. Accessed : 2023-08-17$.

[26] Kenji Koide, Jun Miura, and Emanuele Menegatti. A portable three-dimensional lidar-based system for long-term and wide-area people behavior measurement. *International Journal of Advanced Robotic Systems*, 16, 02 2019.

[27] Ji Zhang and Sanjiv Singh. Low-drift and real-time lidar odometry and mapping. *Autonomous Robots*, 41(2):401 – 416, February 2017.

[28] Aleksandr Segal, Dirk Haehnel, and Sebastian Thrun. Generalized-icp. In *Robotics: science and systems*, volume 2, page 435. Seattle, WA, 2009.

[29] Jing Gu. One-vs-all semi-auto labeling tool for semantic segmentation in autonomous driving. July 2023.