Collaborative Grid Mapping for Moving Object Tracking Evaluation

Rémy Huet, Antoine Lima, Philippe Xu, Véronique Cherfaoui, Philippe Bonnifait

Collaborative Grid Mapping for Moving Object Tracking Evaluation

Rémy Huet[1], Antoine Lima[1], Philippe Xu[1], Véronique Cherfaoui[1], Philippe Bonnifait[1]

[1]Heudiasyc UMR CNRS 7253, Université de technologie de Compiègne

2023-07-24

Perception of other road users is a crucial task for intelligent vehicles. Perception systems can use on-board sensors only or be in cooperation with other vehicles or with roadside units. In any case, the performance of perception systems has to be evaluated against ground-truth data, which is a particularly tedious task and requires numerous manual operations. In this article, we propose a novel semi-automatic method for pseudo ground-truth estimation. The principle consists in carrying out experiments with several vehicles equipped with LiDAR sensors and with fixed perception systems located at the roadside in order to collaboratively build reference dynamic data. The method is based on grid mapping and in particular on the elaboration of a background map that holds relevant information that remains valid during a whole dataset sequence. Data from all agents is converted in time-stamped *observations* grids. A data fusion method that manages uncertainties combines the background map with observations to produce dynamic reference information at each instant. Several datasets have been acquired with three experimental vehicles and a roadside unit. An evaluation of this method is finally provided in comparison to a handmade ground truth.

1 Introduction

In order to navigate safely, intelligent vehicles must perceive their environment. Perception systems have become more and more complex and evaluating them under real conditions is a difficult task that generally requires manual annotation and processing. In addition to being a time-consuming task, ground-truth annotation of perception systems is error-prone, as occlusions can limit perceivable objects. Therefore, this task can be especially

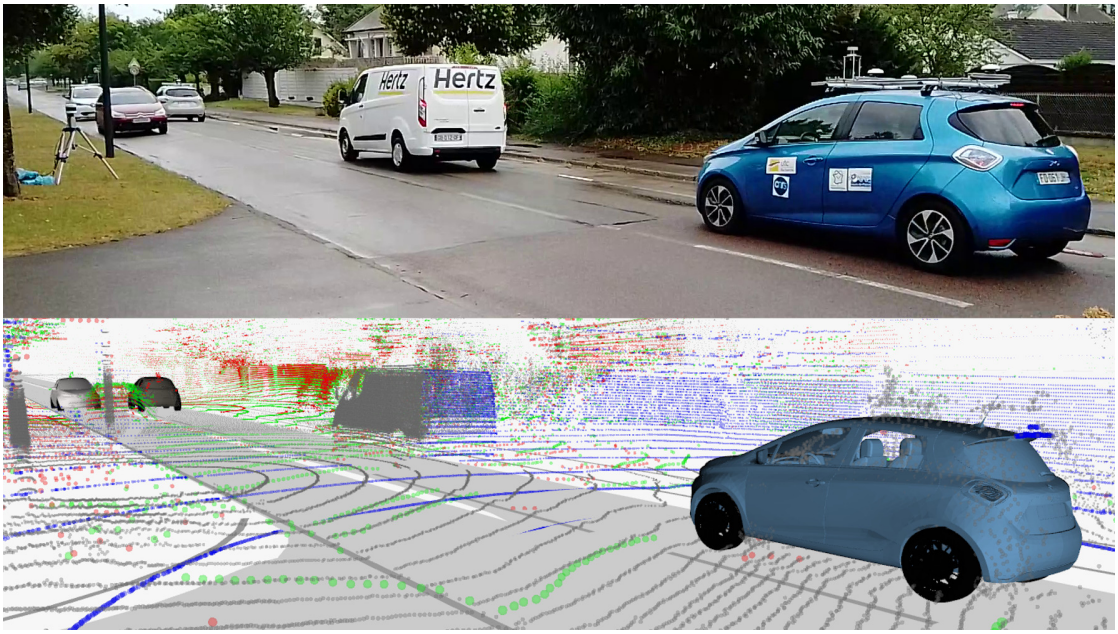


Figure 1: LiDAR and GNSS data gathered by several moving vehicles and a roadside sensor before processing.

challenging in dynamic and cluttered environments. One way to address this issue is to use and record complementary points of view from remote sources of information that can be other road users or roadside units. The objective is then to merge in a post-processing stage all the available data through space and time into a coherent reference.

In this paper, we propose a method that builds occupancy grid maps on recorded driving sequences for the evaluation of perception and tracking systems using several sensors that are either embedded on vehicles or fixed in the environment. Each agent participating in the mapping registers the time-stamped data of its sensors, the synchronization of the clocks being provided by Global Navigation Satellite System (GNSS) receivers.

To represent and fuse data from various perception systems in a common frame, bird eye view grids can be used. As such, our method provides both a grid of the background containing all static elements and time-stamped perception grids in which moving road users are detected. Such grids are a first step towards the generation of a reliable ground truth for object tracking evaluation built semi-automatically.

Moreover, a dataset is released with real data coming from three vehicles and a road-side sensor, including three short scenarios of interest involving a dozen road-users. To our knowledge, it is the first dataset with real data and multiple moving points of view for collaborative perception. A video illustrating the dataset recording and the method is given[1].

[1] https://hal.science/hal-04172876

The article is organized as follows. Section 2 gives an overview of related work on post-processing for object tracking and on occupancy grids. Section 3 introduces the notion of *background map* and its representation as an evidential occupancy grid. In section Section 4, we provide examples about how to build this map with high-accuracy GNSS receivers, on-car and roadside LiDAR sensors. In Section 5, we explain how we construct the final grids and how the background map we created helps to take into account past and future observations. Finally section Section 6 presents experimental results and evaluation of our method on a multi-vehicle dataset recorded for this purpose.

2 Related Work

Perceiving other road users is a complex but mandatory task for autonomous vehicles. It allows the navigation system to find a safe trajectory in the complex environment of the roads. Road users can be represented and tracked object-wise (Vo et al. 2015) but a dense mapping approach is particularly adapted to planning and reactive control in opened and uncontrolled environments (Ziegler and Stiller 2010). Introduced in (Elfes 1989), 2D probabilistic occupancy grids, that densely represent where the space is free or occupied, have widely gained traction in the perception community. One of their main drawback is the difficulty to represent movement of the environment, a limitation alleviated in (Coué et al. 2006) by using a 4D grid to represent the position and velocity of obstacles with probabilities. Another common extension is to use the theory of Dempster-Shafer (or theory of evidence) to represent multiple possible states of a cell (Moras, Cherfaoui, and Bonnifait 2011). Combining both tracking and multi-state is referred to as Grid-Based Tracking and Mapping (GTAM) (Steyer, Tanzmeister, and Wollherr 2018; Tanzmeister and Wollherr 2017). These approaches often use a particle filter to estimate the motion of obstacles in the grid.

Another important aspect of perception systems is their evaluation. There exist many single Point of View (PoV) datasets such as KITTI (Geiger, Lenz, and Urtasun 2012), SemanticKITTI (Behley et al. 2019) or NuScenes (Caesar et al. 2020) for the evaluation of automotive perception systems. They provide manually labeled datasets with delimited 3D bounding boxes. To extend evaluations to navigation, Ettinger et al. (Ettinger et al. 2021) provide a perception dataset with a focus on motion prediction. However, multi-PoV datasets are not yet widely developed with most being based either on simulated data (Xu et al. 2022; Yiming Li et al. 2022; Mao et al. 2022) or static sensors (Busch et al. 2022). Recently, the DAIR-V2X dataset (H. Yu et al. 2022) has been released with several points of view from real sensors. However, it is focused on vehicle-to-infrastructure communication and only contains one vehicle. A possible reason for this is the complexity of labeling a multi-PoV dataset, which is why several methods have been proposed to label them semi-automatically (Han et al. 2023).

Approaches for ground-truth generation are done offline on recorded datasets. They can take advantage from heavy-computational tasks such as multiple point clouds registra-

tion (Ye, Spiegel, and Althoff 2020) or the use of non-causal (i.e. depending on future information) algorithms (Erik Stellet, Walkling, and Marius Zöllner 2016; B. Yu and Ye 2020). In (B. Yu and Ye 2020), the tracks corresponding to other road users are created at the most certain point in the dataset before being propagated forward and backward by filtering and smoothing algorithms. Other approaches (Ye et al. 2021) are based on GTAM techniques and use offline post-processing to make a backward smoothing pass on the particle filter to help its convergence.

Our method builds on previously cited GTAM techniques (Steyer, Tanzmeister, and Wollherr 2018; Tanzmeister and Wollherr 2017; Ye et al. 2021). A finer frame of discernment allows us to better represent the environment to take advantage of segmentation algorithms to gather more information. We use the possibilities offered by offline post-processing to gather all the information of a recording in a "background map", that is then fused with observation grids.

3 Definition of the Background Map

3.1 Semantics of the Map

In general, the mapping of an environment holds information that is true over very long periods of time. Here, the information stored by the mapping is defined as the information that is reliable during the whole dataset. It represents statically occupied areas and passable areas, which can be either free or contain a moving object.

Occupancy of the space can be divided into two main classes: *immovable occupancy* that corresponds to non-movable features (e.g. buildings, trees or road signals), and *objects* that corresponds to movable features (i.e. vehicles or vulnerable road users). The distinction between those two classes can be made in a single measurement, while objects have to be refined into *static* or *dynamic* using multiple measurements. As the goal of the map is to determine some information that is valid for the whole dataset, static objects are defined as objects that do not move during the whole sequence. Dynamic objects are thus objects that moved at least once during the sequence.

3.2 Evidential Grid Representation

The map has to represent several classes and combinations, a problem well suited for the Dempster-Shafer Theory (DST) framework (Dempster 1968; Shafer 1976). In the DST, a *mass function* m assigns a non-negative value to each element of a power set 2^Θ where Θ is the *frame of discernment*, the exhaustive and atomic set of hypotheses about a variable, such that $\sum_{A \in 2^\Theta} m(A) = 1$.

This framework allows a fine representation of uncertainty by assigning some mass to a set of classes, including the whole frame of discernment to represent full uncertainty. This

is especially suited to the case of partial sensor information, which can be represented by a union of classes but not as an atomic hypothesis from the frame. Mass functions can also be combined using for example the conjunctive rule:

$$m_{1,2}(A) = \sum_{B \cap C = A} m_1(B) \cdot m_2(C), \ \forall A \in 2^{\Theta} \tag{1}$$

To represent this information, we project the 3D data into a 2D discrete Grid Map at the ground level. 2D grids provide an efficient framework for data representation for terrestrial vehicles and help for data fusion in a common spatial frame.

3.3 Frame of Discernment

The frame of discernment to characterize an environment is defined as follows:

$$\Theta = \{F, I, S, D\}, \tag{2}$$

where F stands for free space, I for immovable, S for static objects and D for dynamic objects, as defined in Section 3.1. For convenience, we add the following definitions:

- $M = \{S, D\}$ for movable objects that can be static or dynamic;
- $O = \{I, S, D\}$ for generic unclassified occupancy;
- $P = \{F, D\}$ which stands for *Passable* as defined in (Steyer, Tanzmeister, and Wollherr 2018), corresponding to zones that are either free or occupied by a dynamic object.

Finally, cells of the grid without any information will have all the mass assigned to Θ.

As the map only contains information that is valid during the whole dataset, its frame of discernment does not include time-dependent classes such as dynamic objects and free space. It is a 2D spatial discrete structure, in which each cell contains a mass function defined on the following frame of discernment:

$$\Theta_{\mathcal{M}} = \{P, I, S\} \tag{3}$$

Indeed, information about passable areas, immovable occupancy and static objects is time independent during a whole sequence, contrary to dynamic objects and free space that should not appear in the background map.

Compared to the frame of discernment used by previous approaches (Steyer, Tanzmeister, and Wollherr 2018; Tanzmeister and Wollherr 2017), ours allows a finer representation thanks to the distinction between *immobile* and *mobile* features. This distinction allows us to integrate class information from segmentation algorithms, as explained in the next section.

4 Background Map Construction

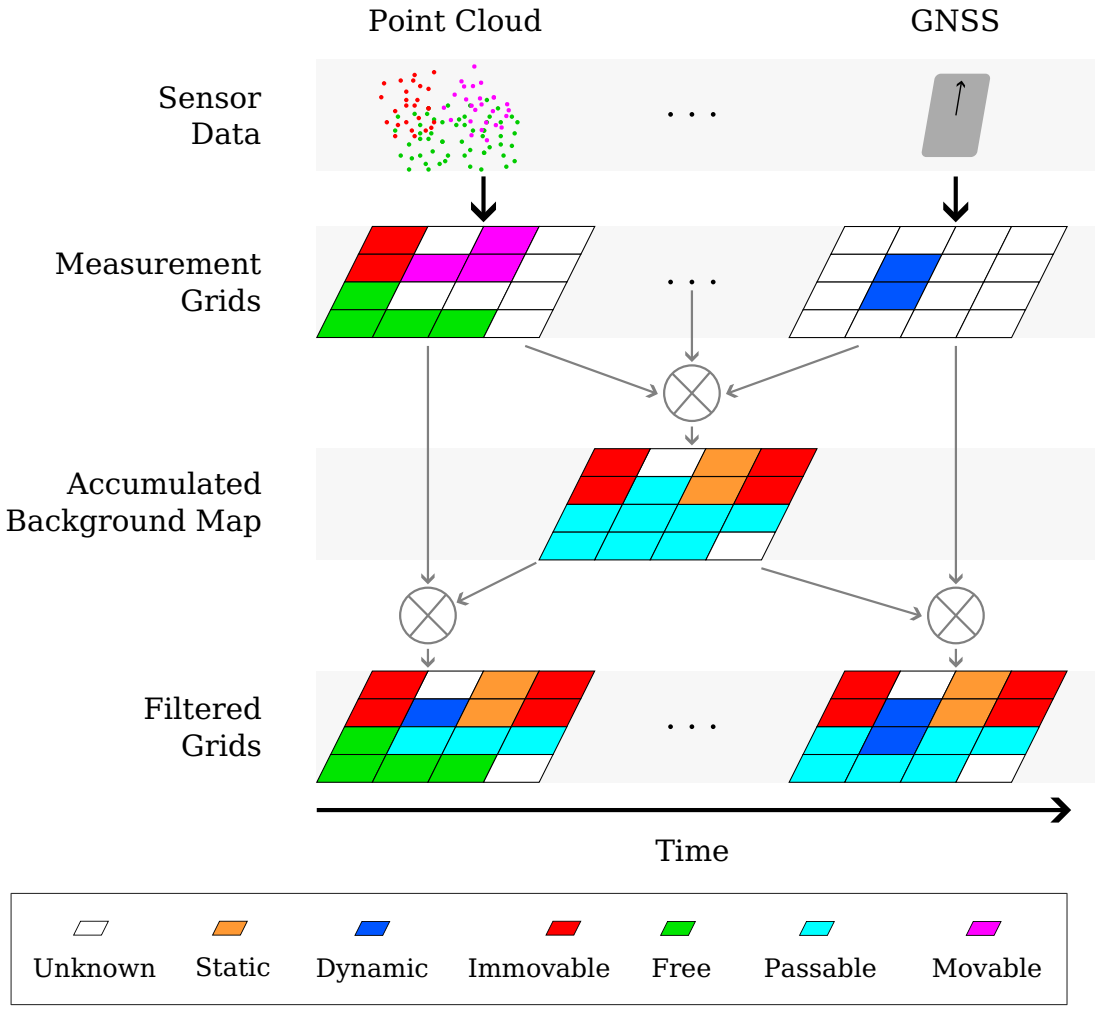


Figure 2: Schematic presentation of the method. Sensor data is turned into *Measurement Grids*. These grids are then fused into a *Background Map*. Finally, *Measurement Grids* are filtered through time and fused with the *Background Map* to generate *Filtered Grids* at each time step.

4.1 Collaborative Background Map Generation

The method to construct the background map takes different sources providing information about obstacles or free space in a common East-North-Up frame. GNSS and LiDAR sensors are used in this work, but other sensors could be adapted such as cameras or RADARs. The background map is generated in several steps, as illustrated in Figure 2.

4.1.1 Observation Grid Generation

As illustrated in Figure 2, observations are grids constructed from all sources separately at different instants by projecting sensor data in 2D. Depending on the sensor, different evidences can be derived such as free space F, occupancy O or dynamic objects D.

4.1.2 Map Size Deduction

In addition to being stored for further processing, observations are first used to determine the size of the background map as being wide enough to fuse all the observations.

4.1.3 Data accumulation

The map is then constructed in a second pass by accumulating the previously stored observations, as shown in Figure 2. Successive observations are accumulated on a per-cell basis. Each cell c stores the values of its consecutive observations m_c^t through time t in an observation buffer $\mathbf{m}_c = \{\dots, m_c^t, \dots\}$. The longest string of mostly free, mostly immovable and mostly static beliefs in $\mathbf{m}_c$ are then computed as w_f, w_i and w_s, respectively, to determine the overall state of cell c. For this, two thresholds t_f and t_o are defined as the minimal number of *consecutive* free and occupancy observations required to consider that a cell is free or occupied in a noise resilient manner:

$$s_c = \begin{cases} P & \text{if } |w_f| \geq t_f; \\ I & \text{else if } |w_i + w_s| \geq t_o \text{ and } |w_i| \geq |w_s|; \\ S & \text{else if } |w_i + w_s| \geq t_o \text{ and } |w_s| > |w_i|; \\ \Theta & \text{otherwise} \end{cases} \tag{4}$$

The thresholds are a trade-off between the completeness of the map and its coherence. Augmenting them will result in more unknown areas, while decreasing them could lead to unwanted passable or occupied evidence. To tune the thresholds, we ensure that the static and immovable obstacles are well mapped, and that there is no ghost (cells classified as static object because an object passed on them for too long) in the map. Figure 5 shows the effect of tuning the thresholds in our experiments. We keep them as low as possible while being consistent to avoid the creation of unknown zones in the map.

4.2 Segmentation of LiDAR Point Clouds

In the current implementation of the method, we use point clouds from rotating LiDARs on-board vehicles or installed on fixed supports at the roadside. This section describes

the processing applied to point clouds to extract information used in the generation of the map.

Raw point cloud data in itself is insufficient to derive the pieces of evidence required for the map generation process. However, it is processed to extract such evidence. For example, points hitting the ground support the free space hypothesis while points hitting a building support the infrastructure hypothesis. Such processing is described hereafter for each source and time step.

4.2.1 Ground Segmentation

A common task on point clouds is to segment non-ground points from ground ones. Such segmentation allows to make the difference between points on the ground supporting the free space evidence F from points above the ground, supporting the existence of an obstacle and thus some unclassified occupancy O.

For this purpose, we use the method from (Jiménez et al. 2021) which gives very good results with an F1-score above 95% for obstacle detection on nuScenes dataset (Caesar et al. 2020). This method is based on gradient and height difference between consecutive points in a vertical scan of the LiDAR. Ground height is then estimated in a cylindrical voxel using loopy belief propagation to refine the segmentation previously done.

4.2.2 Class Segmentation

In parallel, points are classified to refine the unclassified occupancy O. This task (called class segmentation) has seen major improvements in the previous years, in particular based on neural networks (Zamanakos et al. 2021; Ying Li et al. 2020). We use the method from (Zhu et al. 2021) which achieved an *Intersection over Union* (IoU) score above 93% for car segmentation in both nuScenes and SemanticKITTI (Behley et al. 2019) datasets. This method is based on asymmetrical convolution networks on cylindrical voxels, that match to the polar nature of a rotating LiDAR. A per-point refinement is applied to segment the point cloud. The overall *mean IoU* (mIoU) is around 68% on SemanticKITTI and 76% on nuScenes. Other approaches such as (Chen et al. 2021; Dewan et al. 2016) propose instead to segment which point is moving across time or not, a task called *Moving-Object Segmentation*. Such a technique could be used to derive dynamic evidence, but would not provide a class for obstacles. Although class information is not directly used for ground-truth generation, it is a precious semantic information that we would like to keep.

Class segmentation thus provides evidence based on class, with for example the ground class supporting free space F, non-movable classes (i.e. buildings, fences or vegetation) supporting Infrastructure I and movable classes (i.e. cars, trucks or pedestrians) supporting movable M. As, the performance of the chosen deep learning method is not high

enough with our LiDARs, the segmented class has only been used to refine points labeled as obstacles by the ground segmentation of (Jiménez et al. 2021).

5 Final Map for Tracking Evaluation

Using the previously generated background map and observation grids, we compute ground-truth maps that characterize free space and dynamic objects (see Figure 2). These maps are obtained at each time step by fusing the background map and the observations with the conjunctive rule. This fusion allows to take into account information from the whole dataset given by the background map at each timestamp. As is, occupancy O seen at a time t in a passable area P will lead to the creation of dynamic cells D, as $O \cap P = D$. The observation of occupancy in immovable or static zone will result in the same immovable or static evidence, and free space will be added on top of passable zones when observed. In the case of incoherent information between the background map and the observation, we use the information from the map as an output as it results from information from the whole dataset and is less subject to noise than instantaneous observations.

The final result provided by the method is then a 2D evidential grid containing all classes defined in Section 3.2. Passable zones resulting from this fusion correspond to non-observed zones at this timestamp, and are just some information provided by the map. These grids can be used for moving object tracking evaluation as they contain cells occupied by dynamic objects. Static objects are contained in the map and their detection can then be evaluated, or one may choose to ignore them in the validation process if it is irrelevant.

6 Experiments

6.1 Collaborative Dataset

As highlighted in the review of related work, to our knowledge, no dataset provides all at once moving, multi-vehicle, with real data. To develop and test our method, we thus recorded a multi-vehicle dataset which is available online[2]. This dataset contains three scenarios involving three vehicles equipped with sensors and a roadside sensor. These scenarios have been made as typical use cases for cooperative perception (Ambrosin et al. 2019) with obstructions from buildings or other vehicles on the open road with other road-users. An overview of the experimental setup is shown at Figure 1.

The three experimental vehicles are equipped with a Velodyne VLP-32C LiDAR mounted on their roof and a Novatel SPAN-CPT IMU with Post-Processed Kinematics (PPK)

[2] https://datasets.hds.utc.fr/share/KrXdEBzaMnMmWbV

corrections providing centimeter-level localization accuracy. In addition, a Velodyne VLS-128 LiDAR sensor statically placed on a curb was used as a roadside sensor. All sensors are synchronized with GNSS time with a known extrinsic calibration.

6.2 Map Generation Using Real Data

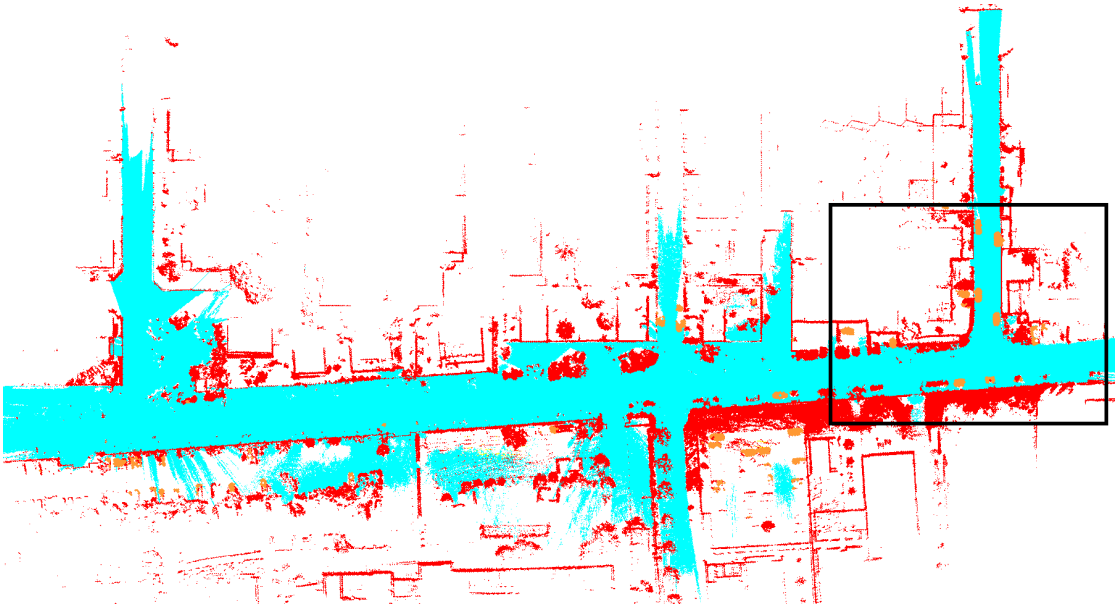


Figure 3: Generated map. Colors are the same as described in Figure 2. The black rectangle is the area zoomed-in for Figure 4.

Accurate poses and point clouds have been thus recorded and processed according to Section 4.1. In particular, the threshold t_o and t_f were manually tuned to 5 and 30, respectively (see Figure 5 for the effects of the threshold tuning). The thresholds appear to be dependant of the dataset, and can vary with the number of sensors or the number of passages in a given area.

The map resulting from this processing is illustrated in Figure 3, Figure 4, respectively showing the whole generated map and a zoom on a point of interest (top-right intersection).

A proper evaluation of the map has been realized and shows interesting results: passable areas fit well with the road, parked cars are well identified as static and tree/buildings are well identified as immovable.

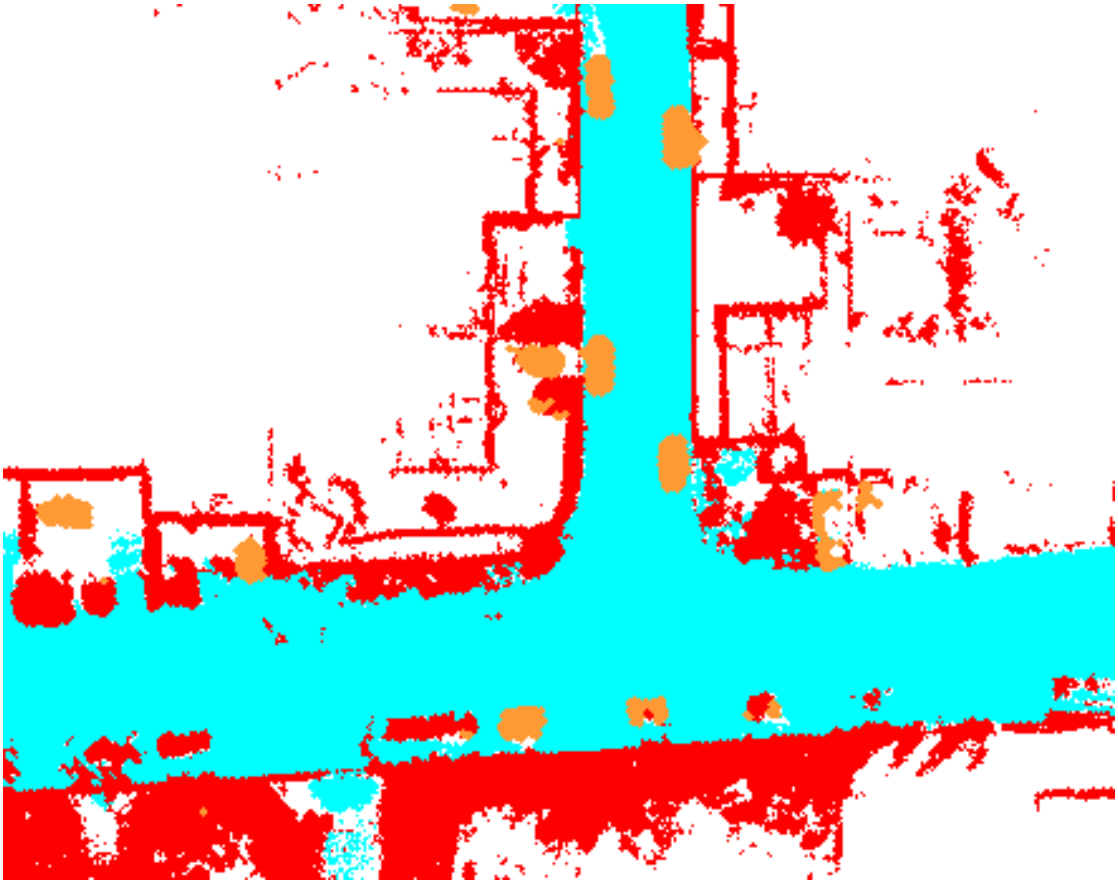
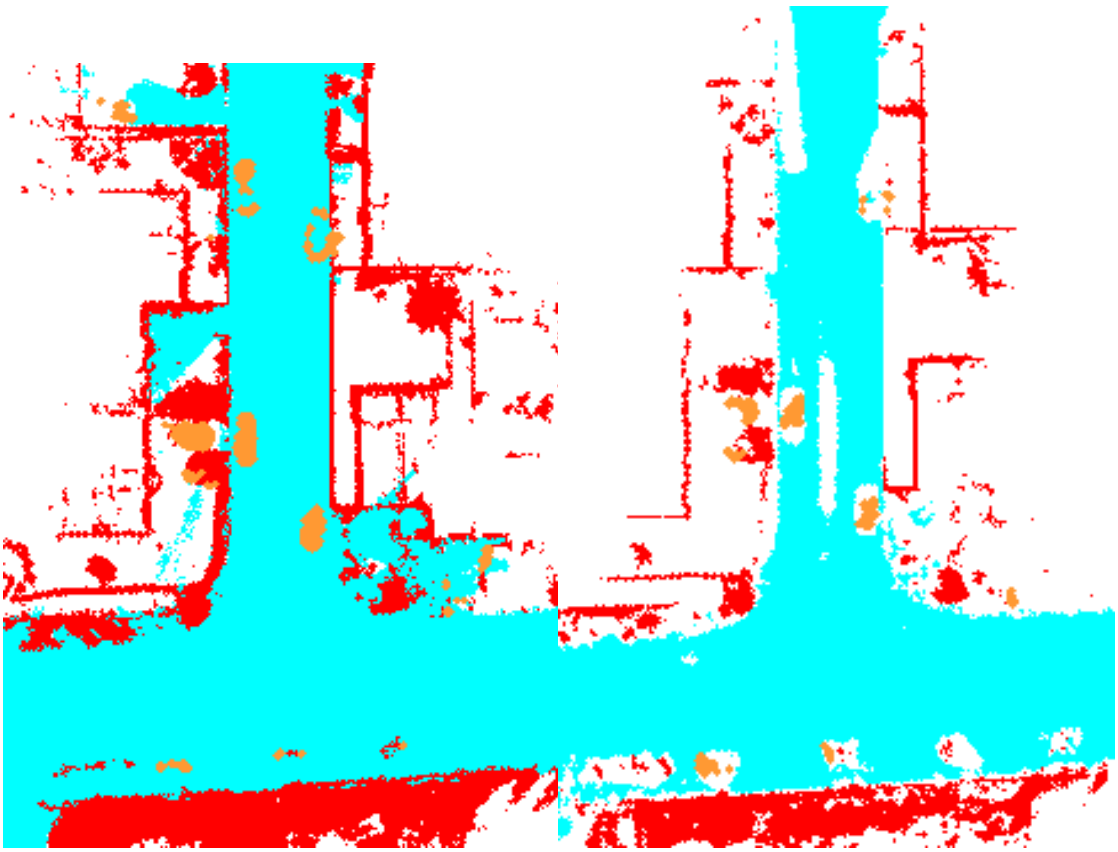


Figure 4: Zoom on the top-right intersection of the map. Uncertain zones (white), immovable occupancy (red), static objects (orange) and passable space (cyan) are visible.



(a) (b)

Figure 5: Effects of threshold tuning for the map generation. In (a), free space threshold is too low, leading noise to create passable areas inside vehicles (cyan regions inside orange borders). In (b), free space and occupancy thresholds are too high, leading to a complete lack of information and the creation of unknown zones (white).

6.3 Fusion of Map and Observations

The observations contain the perception (LiDAR) from our vehicles and their position and orientation as shown in Figure 6.

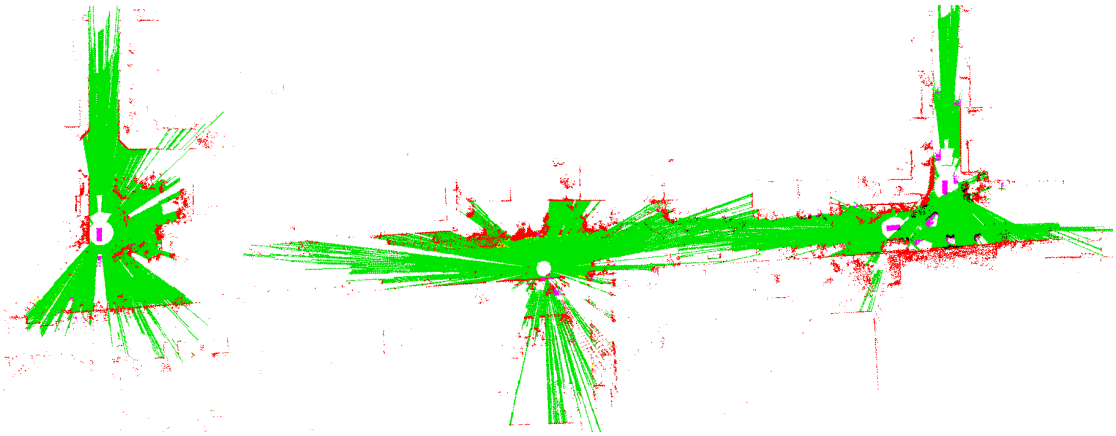


Figure 6: Fused observation grid. It contains all the LiDAR observation at a given timestamp and the position of the vehicles. Black cells represent conflicts between point of views, green cells represent free space.

The result of the fusion is illustrated in Figure 7 and Figure 8 at a given time step and is evaluated in Section 6.4. Qualitatively, one can see that the free space is observed on top of the passable zones of the map. Obstacles observed on top of passable zones are correctly classified as dynamic objects. On the opposite, occupancy observed over statically occupied zones correctly classifies them as static S and thus prevents the creation of dynamic cells.

6.4 Evaluation of the Method

Existing datasets are not relevant to evaluate this method, as they either lack moving vehicles or real data. A particular sequence of the dataset introduced in Section 6.1 has thus been manually labeled to this end. It consists of 5 to 10 objects over 50 seconds or 500 frames. The manual labeling methodology is as follows: based on accumulated point clouds, bird-eye-view bounding boxes have been delimited by a human operator for several key frames of the dataset. Bounding boxes have then been synchronized with the output of the above method by interpolation. The center $g_k^t = \langle x_k^t, y_k^t \rangle$ of each bounding box k for each time t has then been concatenated to form the ground truth $\mathcal{G} = \{g_k^t\}_{t=1,2,\dots}$. Finally, a Region of Interest R is also delimited by the human operator. Inside R, the human operator certifies that no objects were missed, but objects outside of it were ignored.

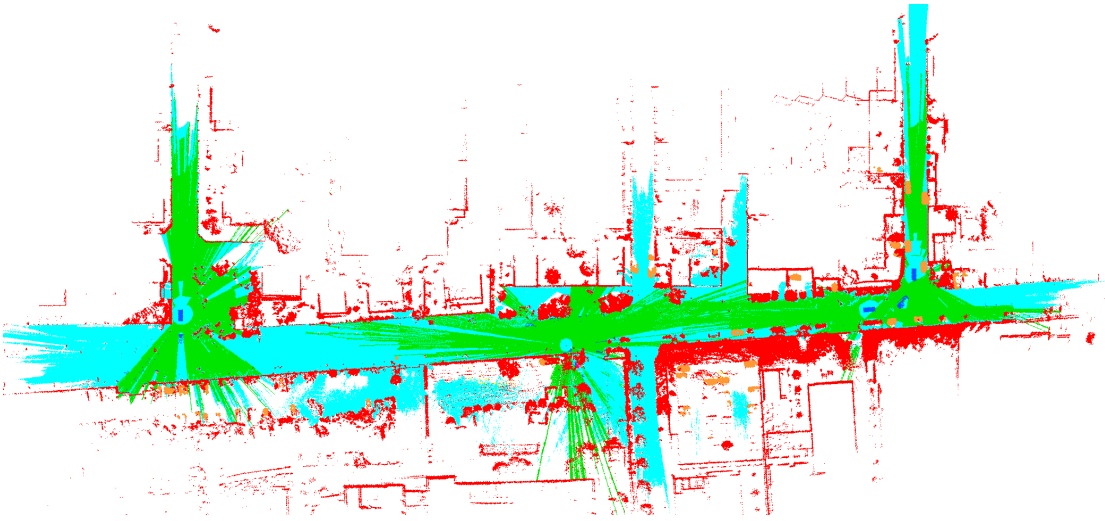


Figure 7: Fusion of map and observations. The colors are the same as explained in Figure 2.

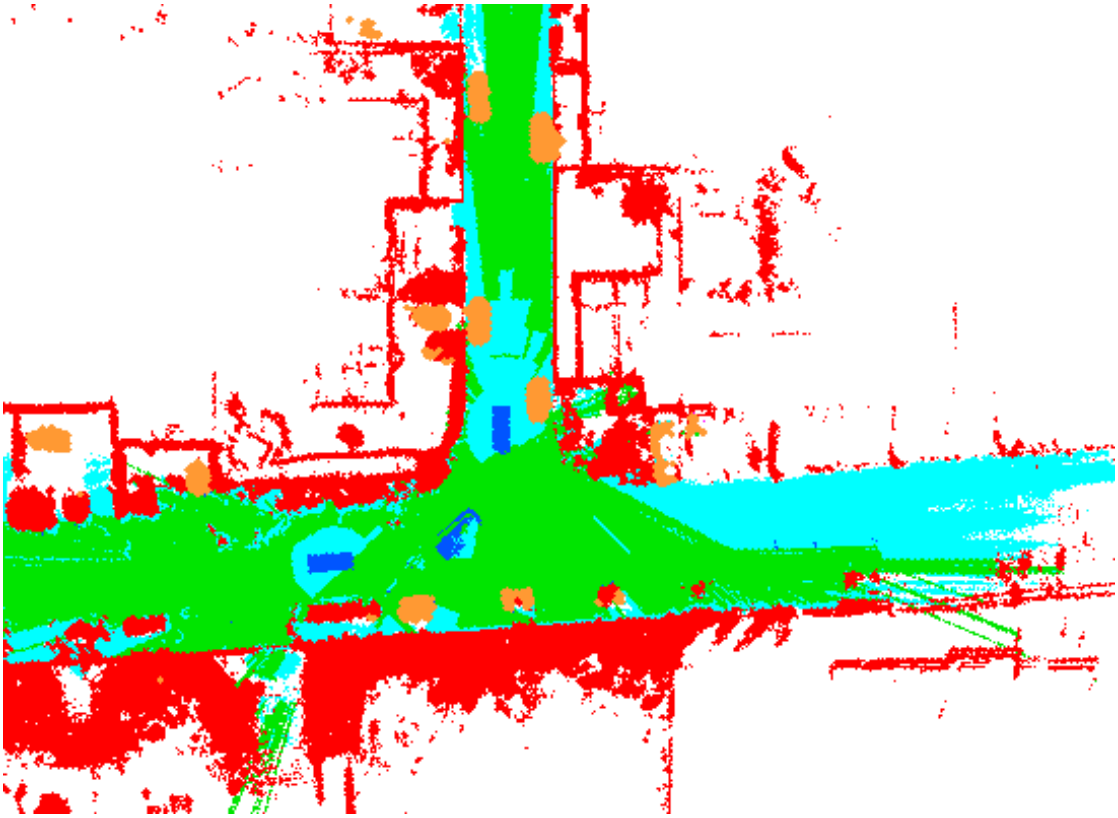


Figure 8: Zoom on Figure 7. Dynamic objects create blue cells unless static evidence (orange cells) is present in the background map.

As our interest is about object existence and not spatial accuracy, we evaluate our method object-wise. To do this, objects are first extracted from the background and final maps from of Section 4.1 and Section 5. The extraction is done by clustering static cells in the background map and dynamic cells in final grids using the DBSCAN algorithm (Ester et al. 1996). The center $o_j^t = \left\langle x_j^t, y_j^t \right\rangle$ of each cluster j for each time t are then concatenated to form the object list $\mathcal{O} = \left\{ o_j^t \right\}_{t=1,2,\dots}$. In a second time, objects of $\mathcal{O}$ that fall within R are compared to $\mathcal{G}$. Objects and ground truth are associated for each time step using the Hungarian algorithm and Euclidean distances. Metrics are then derived from these associations, with associated objects being True Positives (TP), un-associated objects being False Positives (FP) and un-associated ground-truth being False Negatives (FN). Classical metrics such as precision $\left(\frac{TP}{TP+FP}\right)$, recall $\left(\frac{TP}{TP+FN}\right)$ and F1-score $\left(\frac{2TP}{2TP+FP+FN}\right)$ are provided in Table 1. We believe that these scores are satisfactory as a first pass to generate ground-truths, but that human verification still remains mandatory.

Table 1: Evaluation of the proposed method

Precision	Recall	F1-score
0.95	0.97	0.96

7 Conclusion and Future Work

In this work, we proposed a novel approach to take advantage of post-processing for ground truth semi-automatic generation by introducing a complete frame of discernment for environment representation and a *background map* constructed using the whole dataset. The fusion of the observation grids with the map allows the instantaneous classification of objects as dynamic when they are on passable areas, in contrast to other GTAM techniques (Steyer, Tanzmeister, and Wollherr 2017; Tanzmeister and Wollherr 2017). Moreover, we provided a dataset with multiple real moving points of view, on which our method proved to have very good results.

Further improvements can be made such as tracking at object level above the clustered dynamic cells or the generation of parametric free space above the grid. We believe that tracking the dynamic objects using filtering and smoothing algorithms would lead to better results as it would remove some noise from the detection and would help to follow dynamic obstacles in occluded environments. Additionally, the current implementation was made using binary masses. A further task would be to implement finer masses using confidence provided by the sensors or processing methods. This would allow a finer representation of uncertainty during the fusion of different information sources.

Acknowledgment

This work has been carried out within the SIVALab laboratory between Renault and Heudiasyc and co-funded by the Région Hauts-de-France. The authors would like to thank C. Zinoune for lending the VLS-128 LiDAR and the colleagues of the lab who participated to the acquisition of the dataset, in particular S. Bonnet.

References

Ambrosin, Moreno, Ignacio J Alvarez, Cornelius Buerkle, Lily L Yang, Fabian Oboril, Manoj R Sastry, and Kathiravetpillai Sivanesan. 2019. "Object-Level Perception Sharing Among Connected Vehicles." In *IEEE Intelligent Transportation Systems Conference*.

Behley, J., M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall. 2019. "SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences." In *Proc. Of the IEEE/CVF International Conf. On Computer Vision (ICCV)*.

Busch, Steffen, Christian Koetsier, Jeldrik Axmann, and Claus Brenner. 2022. "LUMPI: The Leibniz University Multi-Perspective Intersection Dataset." In *2022 IEEE Intelligent Vehicles Symposium (IV)*, 1127–34. https://doi.org/10.1109/IV51971.2022.9827157.

Caesar, Holger, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. 2020. "nuScenes: A Multimodal Dataset for Autonomous Driving." In *CVPR*.

Chen, Xieyuanli, Shijie Li, Benedikt Mersch, Louis Wiesmann, Jurgen Gall, Jens Behley, and Cyrill Stachniss. 2021. "Moving Object Segmentation in 3D LiDAR Data: A Learning-Based Approach Exploiting Sequential Data." *IEEE Robotics and Automation Letters*.

Coué, Christophe, Cédric Pradalier, Christian Laugier, Thierry Fraichard, and Pierre Bessière. 2006. "Bayesian Occupancy Filtering for Multitarget Tracking: An Automotive Application." *The International Journal of Robotics Research*.

Dempster, A. P. 1968. "A Generalization of Bayesian Inference." *Journal of the Royal Statistical Society. Series B (Methodological)*.

Dewan, Ayush, Tim Caselitz, Gian Diego Tipaldi, and Wolfram Burgard. 2016. "Motion-Based Detection and Tracking in 3D LiDAR Scans." In *IEEE International Conference on Robotics and Automation*.

Elfes, A. 1989. "Using Occupancy Grids for Mobile Robot Perception and Navigation." *Computer*.

Erik Stellet, Jan, Leopold Walkling, and J. Marius Zöllner. 2016. "Post Processing of Laser Scanner Measurements for Testing Advanced Driver Assistance Systems." In *International Conference on Information Fusion*.

Ester, Martin, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. 1996. "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise." In *International Conference on Knowledge Discovery and Data Mining*.

Ettinger, Scott, Shuyang Cheng, Benjamin Caine, Chenxi Liu, Hang Zhao, Sabeek Pradhan, Yuning Chai, et al. 2021. "Large Scale Interactive Motion Forecasting for Autonomous Driving: The Waymo Open Motion Dataset." In *IEEE/CVF International Conference on Computer Vision*.

Geiger, A., P. Lenz, and R. Urtasun. 2012. "Are We Ready for Autonomous Driving? The KITTI Vision Benchmark Suite." In *IEEE Conference on Computer Vision and Pattern Recognition*.

Han, Yushan, Hui Zhang, Huifang Li, Yi Jin, Congyan Lang, and Yidong Li. 2023. "Collaborative Perception in Autonomous Driving: Methods, Datasets and Challenges." *ArXiv* abs/2301.06262.

Jiménez, Víctor, Jorge Godoy, Antonio Artuñedo, and Jorge Villagra. 2021. "Ground Segmentation Algorithm for Sloped Terrain and Sparse LiDAR Point Cloud." *IEEE Access*.

Li, Yiming, Dekun Ma, Ziyan An, Zixun Wang, Yiqi Zhong, Siheng Chen, and Chen Feng. 2022. "V2X-Sim: Multi-Agent Collaborative Perception Dataset and Benchmark for Autonomous Driving." *IEEE Robotics and Automation Letters* 7 (4): 10914–21. https://doi.org/10.1109/LRA.2022.3192802.

Li, Ying, Lingfei Ma, Zilong Zhong, Fei Liu, Dongpu Cao, Jonathan Li, and Michael A. Chapman. 2020. "Deep Learning for LiDAR Point Clouds in Autonomous Driving: A Review." *IEEE Transactions on Neural Networks and Learning Systems* 32: 3412–32.

Mao, Ruiqing, Jingyu Guo, Yukuan Jia, Yuxuan Sun, Sheng Zhou, and Zhisheng Niu. 2022. "DOLPHINS: Dataset for Collaborative Perception Enabled Harmonious and Interconnected Self-driving." In *Asian Conference on Computer Vision*.

Moras, Julien, Véronique Cherfaoui, and Philippe Bonnifait. 2011. "Credibilist Occupancy Grids for Vehicle Perception in Dynamic Environments." In *IEEE International Conference on Robotics and Automation*.

Shafer, Glenn. 1976. *A Mathematical Theory of Evidence*. Princeton University Press.

Steyer, Sascha, Georg Tanzmeister, and Dirk Wollherr. 2017. "Object Tracking Based on Evidential Dynamic Occupancy Grids in Urban Environments." In *IEEE Intelligent Vehicles Symposium (IV)*.

———. 2018. "Grid-Based Environment Estimation Using Evidential Mapping and Particle Tracking." *IEEE Transactions on Intelligent Vehicles*.

Tanzmeister, Georg, and Dirk Wollherr. 2017. "Evidential Grid-Based Tracking and Mapping." *IEEE Transactions on Intelligent Transportation Systems*.

Vo, Ba-ngu, Mahendra Mallick, Yaakov Bar-shalom, Stefano Coraluppi, Richard Osborne, Ronald Mahler, and Ba-tuong Vo. 2015. "Multitarget Tracking." In *Wiley Encyclopedia of Electrical and Electronics Engineering*. John Wiley & Sons, Inc.

Xu, Runsheng, Hao Xiang, Xin Xia, Xu Han, Jinlong Li, and Jiaqi Ma. 2022. "OPV2V: An Open Benchmark Dataset and Fusion Pipeline for Perception with Vehicle-to-Vehicle Communication." In *International Conference on Robotics and Automation*.

Ye, Egon, Philip Spiegel, and Matthias Althoff. 2020. "Cooperative Raw Sensor Data Fusion for Ground Truth Generation in Autonomous Driving." In *IEEE International Conference on Intelligent Transportation Systems*.

Ye, Egon, Gerald Wursching, Sascha Steyer, and Matthias Althoff. 2021. "Offline Dynamic Grid Generation for Automotive Environment Perception Using Temporal Inference Methods." *IEEE Robotics and Automation Letters*.

Yu, Boqian, and Egon Ye. 2020. "Track-Before-Detect Labeled Multi-Bernoulli Smoothing for Multiple Extended Objects." In *IEEE International Conference on Information Fusion*.

Yu, Haibao, Yizhen Luo, Mao Shu, Yiyi Huo, Zebang Yang, Yifeng Shi, Zhenglong Guo, et al. 2022. "DAIR-V2X: A Large-Scale Dataset for Vehicle-Infrastructure Cooperative 3D Object Detection." In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Zamanakos, Georgios, Lazaros Tsochatzidis, Angelos Amanatiadis, and Ioannis Pratikakis. 2021. "A Comprehensive Survey of LIDAR-based 3D Object Detection Methods with Deep Learning for Autonomous Driving." *Computers & Graphics*.

Zhu, Xinge, Hui Zhou, Tai Wang, Fangzhou Hong, Yuexin Ma, Wei Li, Hongsheng Li, and Dahua Lin. 2021. "Cylindrical and Asymmetrical 3D Convolution Networks for LiDAR Segmentation." In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Ziegler, Julius, and Christoph Stiller. 2010. "Fast Collision Checking for Intelligent Vehicle Motion Planning." In *IEEE Intelligent Vehicles Symposium*.