6G-EWOC: Crowdsourced SLAM data fusion for Safe and Efficient ADAS Driving

José A. Lázaro,
Josep R. Casas,
Javier Ruiz-Hidalgo,
Martí Cortada Garcia,
Gerard Martín Pey,
Judit Salavedra Pujol,
Mingrui Wang
Signal Theory and Comm. Dept.
UPC - Univ. Politècnica de Catalunya
Barcelona, Spain
ORCID: 0000-0002-5592-8434
ORCID: 0000-0003-4639-6904
ORCID: 0000-0001-6774-685X
ORCID: 0009-0005-9258-6428
ORCID: 0009-0009-0958-1538
ORCID: 0009-0004-5927-7132
ORCID: 0000-0003-1055-2311
jose.antonio.lazaro@upc.edu

Eleni Theodoropoulou,
George Lyberopoulos
R&D Fixed and Mobile
Hellenic Telecommunications
Organization SA (OTE SA)
Athens, Greece
ORCID: 0000-0002-4789-6022
ORCID: 0000-0002-3645-902X

Pablo García,
Santiago Royo,
Jordi Riu
Beamagine SL
Castellbisbal, Barcelona, Spain
ORCID: 0000-0001-9265-7242
ORCID: 0000-0003-0136-8301

Carina Marcus,
Olof Eriksson
Research & Innovation
Magna Electronics
Vårgårda, Sweden
ORCID: 0000-0003-4147-1805
ORCID: 0009-0009-4182-9222

Bernhard Schrenk
Center for Digital Safety & Security
AIT Austrian Institute of Technology
Vienna, Austria
bernhard.schrenk@ait.ac.at
ORCID: 0000-0001-5374-0078

Josep M. Fàbrega
Packet Optical Net. & Serv. (PONS)
CTTC - Centre Tecn. Telecom. Cat.
Castelldefels, Barcelona, Spain
ORCID: 0000-0001-9285-9768

***Abstract*— The development of transport infrastructures for Advanced Driver Assistance Systems (ADAS) and autonomous vehicles operating efficiently and safely in congestion-free traffic flows is a major challenge for telecommunications technologies. Simultaneous Localization and Mapping (SLAM) plays a crucial role in ensuring uninterrupted journeys for emergency vehicles and increasing the safety of vulnerable road users in complex traffic scenarios. Accurate SLAM mapping for ADAS systems requires data from different sensor technologies –such as high-resolution cameras or Radio/Light Detection and Ranging (RaDAR/LiDAR)– to be effectively combined or fused. Sensor fusion results in high data throughput and low latency requirements. However, optimal mapping outcomes occur when processing systems fuse data from sensors positioned at diverse locations within the traffic scene. By crowdsourcing diverse sensors, we can multiply the view angles, mitigate occlusions and improve the overall scene coverage. Yet, this approach introduces additional challenges for communication systems within both the vehicles and the infrastructure. Addressing these challenges is essential for seamless development of safe and efficient ADAS driving techniques.**

Keywords—Autonomous Driving, Crowdsourced SLAM, Data Fusion, 3D Scene Completion, 3D Object Detection, Free-space optical communication, Optical communication terminals, 6G

I. INTRODUCTION

Road safety and efficient traffic management are significant concerns, with over 1 million fatalities annually [1] and economic impacts due to disruptions in transport infrastructure [2]. Autonomous Vehicles (AVs) present an opportunity to improve road safety, optimize traffic flow, reduce CO2 emissions, and enhance accessibility [3]. AVs are set to transform transportation systems, improving daily life by boosting efficiency and well-being. They will also provide mobility for those unable to drive and stimulate growth in sectors like software engineering, cybersecurity, vehicle maintenance, and fleet management. Additionally, self-driving, ride-sharing and delivery services may lower transportation and logistics costs, benefiting both consumers and businesses [4].

This work has received funding from the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe research and innovation Programme under Grant Agreement No. 101139182, and Spanish MICIU founded, TRAINER-B (PID2020-118011GB-C22).

However, to fully realize the potential of AVs, particularly at autonomy levels 4 and 5 [5], key challenges related to technology and regulations must be addressed. One essential task is ensuring AVs can perceive their environment accurately. Simultaneous Localization and Mapping (SLAM) is a crucial method for achieving this, as it creates optimized maps using real-time observations.

Occluded areas in busy driving scenes can not be addressed by simply adding more sensors to Advanced Driver Assistance Systems (ADAS), which is only feasible for high-end vehicles. Most sensors, such as LiDARs, cameras, and RADARs, rely on line of sight to gather information. Challenges like detecting a skater overtaking from behind or a vehicle hidden by a truck often cause sensor limitations. Instead of adding more sensors to the host vehicle (HV), it makes more sense to reduce their number and rely on crowdsourcing to fill in the gaps.

The optimal mapping approach involves integrating sensor data from all vehicles in the traffic scene, known as cooperative perception [6]. This method leverages roadside infrastructure and facilitates sensor data sharing between Connected Autonomous Vehicles (CAVs) [7][8]. By facilitating communication and information sharing with technologies such as those shown in Fig. 1, cooperative perception [9] can improve efficiency, reliability, and safety, benefiting both CAVs and vulnerable road users.

A CRS-SLAM approach (SLAM from CRowdSourced observations) gathers data from sensors on nearby vehicles and infrastructure, outperforming conventional Collaborative SLAM systems (C-SLAM) [10] by creating better maps of driving scenes. This paper proposes to improve ADAS systems but requires enhanced V2V (vehicle-to-vehicle) and V2I (vehicle-to-infrastructure) communications, with high throughput and low latency (raw LiDAR data is transmitted at Gbit/s speed) and low latency below 10 ms). Optical Wireless Communications (OWC) can complement Cellular-V2X (C-V2X) systems [10][11], adding capacity, fault tolerance, and mitigating interference, especially in high-density areas or intersections where 5G signals may face congestion.

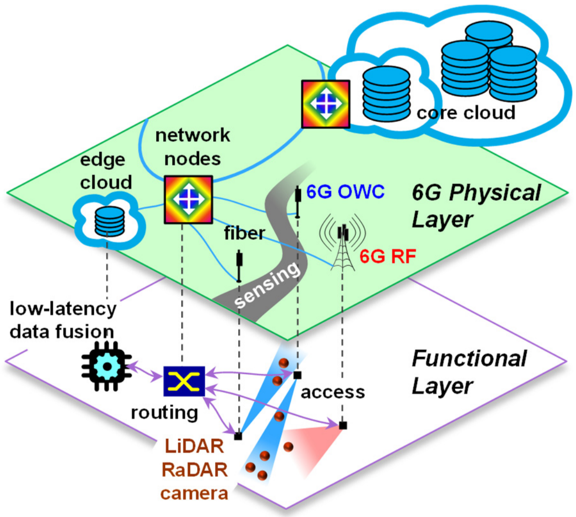

Fig. 1 6G addressing the challenge of autonomous driving requires the extensive deployment of communication and computation resources to facilitate the "collective perception" paradigm.

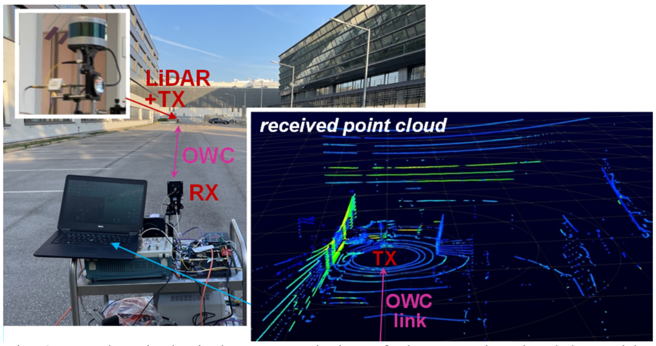

Fig. 2 Local optical wireless transmission of LiDAR point cloud data with fully analogue link layout based on light emitting diode and avalanche photoreceiver, ensuring low latency interconnection between multiple roadside users.

OWC does not introduce much complexity. Fig. 2 shows a LiDAR point cloud transmitted over a visible-light OWC link using a directly modulated high-power light emitting diode (LED) and an avalanche photodiode (APD) receiver. This analogue link layout avoids latency from signal processing and supports 100 m free-space transmission, ideal for local data exchange between vehicles. For longer-range connections with infrastructure, optical beamformers [12] can enable seamless links to 6G networks, providing beam steering without the need for mechanical components.

CRS-SLAM must flexibly negotiate data requests with the network orchestrator. Depending on communications capabilities, it can perform *early fusion* (using sensor data), *mid-level fusion* (combining features), or *late fusion* (merging object detections), in order to improve ADAS systems. In limited bandwidth scenarios, the system can adapt by using mid-level or late fusion, depending on the available data.

A proficient access network is essential for CRS-SLAM, ensuring seamless connectivity between ego-perception sensors and managing data requests from various users. The upcoming 6G technology will combine radio-frequency (RF) communications with efficient photonic tech, offering high bandwidth and transmission speeds [13]. In areas without RF signals, OWC [14] and wireless V2V and V2I systems will be critical for CRS-SLAM and ADAS reliability [15]. Efficient network operation and resource management, powered by AI, will also play a key role.

The objective of the 6G-EWOC project is to explore 6G technology capabilities for autonomous driving, aiming to enhance transportation efficiency and safety. The research focuses on OWC and sensing technologies to ensure a seamless connection between vehicles and the fiber-based, compute-enhanced 6G network. The project seeks to create an innovative access network that combines wireless/RF, fixed/wireless optical, low-energy photonic switching, programmable transceivers, and AI-enabled Software-Defined Networking (SDN). This infrastructure will manage communications efficiently and support advanced applications for Autonomous Vehicles. By establishing end-to-end connectivity between AI-based edge computation units, 6G-EWOC facilitates crowdsourced SLAM data fusion, enabling informed decision-making for safe and efficient ADAS systems.

II. The Role of 6G Technologies and Crowdsourced SLAM in Autonomous Driving

In the context of road safety and traffic optimization, real-time data collection from various sensor technologies –such as high-resolution cameras, radio-based detection (RaDAR) and light-based detection (LiDAR)– is crucial. However, to achieve efficiency gains, this information needs to be fused on a larger scale. Relying solely on data from individual vehicle sensors would limit the potential of the fusion approach, as a single viewpoint cannot cover all the necessary safety and efficiency information. In addition, fusing sensor data within the computing infrastructure of a single vehicle is challenging, as high-throughput data and computing resources are required. Therefore, sensor data should be collected from the on-board sensors of both vehicles and infrastructure wherever possible. Crowdsourcing of driving scene data is easier to achieve in busy urban environments where demand is naturally higher. This concept is commonly known as 'collective sensing' or 'collective perception' [16].

Instant access to a processing infrastructure plays a crucial role in the concept of "collective perception". It enables traffic optimization at the district level, aiming for congestion-free traffic flow while ensuring a swift and uninterrupted journey for emergency vehicles. It also protects vulnerable road users in difficult traffic scenarios, such as collisions during overtaking maneuvers. In addition, the expansion of the digital horizon for each individual vehicle user creates the conditions for collision-free, fully autonomous driving. Access to data (or processed data) from other vehicles and infrastructure becomes possible. This paradigm shift is initiated by using driving skills that could be superior to those of a human driver. Similar to the Internet, where individuals access shared information to make more informed decisions, crowdsourcing of sensor data from individual vehicles will be rewarded. This opens up new opportunities for the use of this information, including ADAS for safety, efficiency and automated driving.

In this context, advanced sensor technologies have demonstrated high sensitivity and resolution. Among these promising technologies, chip-scale LiDAR can capture a clear digital image of the environment with high update rates [17]. The combination of resolution, detection range and specialized signal processing engines enables a clear understanding of roadside events near the host vehicle. This capability improves the individual road user ability to recognize and understand objects, vehicles and pedestrians. It even extends to handling challenging scenarios, such as predicting pedestrian paths or recognizing child pedestrians who may not always follow logical behavior near roads.

However, far-reaching optimization of transport infrastructure requires a much larger situational picture –one that requires the distribution of road-side data collected in a much more distributed manner involving many road users. Here, advanced 6G technologies are expected to provide the necessary communication lines capable of transporting multimodal sensor data at high data rates of several GB/s/user, while also providing the necessary edge infrastructure that can process and fuse the collected data before feeding it back to road users with extremely low latency (Fig. 1).

In addition, practical applications require technologies such as Simultaneous Localization And Mapping (SLAM) technology to capture and map elements at different levels -be it the vehicle, the intersection, the road or a larger area. A crowdsourced SLAM system improves the mapping of objects in a driving environment by capitalizing on observations from hidden or less visible areas. This is made possible by V2V and V2I communication. Flexible SLAM can dynamically adapt to a variety of input observations from different input sources, including raw sensor data, mid-level features from neural models or straightforward detection and classification results from nearby vehicles. These flexible SLAM systems can operate either onboard the vehicle (V-Compute) or within the edge servers of the infrastructure (I(edge)-Compute).

III. Data Fusion for SLAM Generation

The goal of a CRS-SLAM (SLAM from CRowdSourced observations) approach is to create a comprehensive map of the objects within the driving scene. This map is generated from data (observations) coming 1) from the sensors of the host vehicle (HV), 2) from the sensors of nearby vehicles or even 3) from the sensors of the infrastructure. By "complete map of objects" we mean the accurate detection, localization and classification (including object type and status –dynamic or static) of all objects present in the scene.

When creating a CRS-SLAM map, each contributing agent in the scene must be precisely located in terms of position and direction (L). In addition, the data or detections from its sensors must be accurately mapped with respect to its location (M). The two tasks of Localization –L– and Mapping –M– are the fundamental components of any SLAM system.

The detection, localization and classification tasks result from real-time processes that can combine the different sensors of each agent, such as the HV, nearby vehicles or infrastructure, at multiple levels. At the lower level, raw sensor data can be transmitted directly to the HV. At the middle level, features derived by individual agents from raw sensor data are used. At higher level, partial results from direct detections, localizations and classifications made by other processes with their sensors are transmitted and combined.

The object map is created from detection and classification tasks involving sensor fusion with deep learning techniques. Since the ultimate goal is to create a joint map, Bird Eye View (BEV) strategies are used that convert all sensor data (2D cameras, 3D lidar or 4D radar) into a flat environment with 2D features. This transformation enables AI models to create a semantic map (or vectorized map) of the driving scene.

In the following subsections, we present the first innovations and results of the 6G-EWOC project that push the boundaries of current technologies in 6G research while contributing to the state of the art for automotive perception.

A. *Multimodal data fusion*

Before looking at the benefits of crowdsourcing data from different sources, it is important to optimize the information extracted from all sensors in each HV. For this, effective fusion of multimodal data is essential. A common strategy is to integrate data from 3D sensors (such as LiDAR) and conventional RGB cameras. As a first step, we have evaluated the effectiveness of fusing RGB camera data with 3D LiDAR information for object detection.

We used nuScenes dataset [18], which contains data from six RGB cameras and one LiDAR, providing full 360-degree coverage around the vehicle. Using this dataset, we replicated the BEVFusion framework [19], a state-of-the-art multi-sensor fusion model developed for 3D object detection across ten different classes. BEVFusion significantly improves the accuracy and reliability of detecting and identifying objects by integrating data from multiple sensors around the vehicle.

BEVFusion integrates the multimodal information in an intermediate stage. The camera and LiDAR data are processed individually to extract modality-specific features, which are then combined into a Bird's Eye View (BEV) representation. This approach exploits the complementary strengths of both modalities by combining sparse depth features from LiDAR data with the high-resolution, semantically rich features from the cameras. After feature fusion, a two-stage object detector is applied to achieve 3D object detections as explained in [19]. Considering that the main focus is to aggregate data from multiple agents and fill in missing information through crowdsourcing, we decided to investigate the performance of various sensor configurations. Our goal is to find out which sensors have the greatest impact on the detection results.

The front configuration of the nuScenes vehicle (see Fig. 3) includes 180 degrees of overlapping coverage by the front LiDAR range and three forward-facing cameras. In contrast, the rear configuration employs three cameras with wider angles exceeding 180 degrees, along with the rear LiDAR range. The data in Table 1 compares the results of BEVFusion using all sensors (baseline), against separate front and rear detection. Table 1 shows that a configuration with fully overlapping sensor coverage (as in the front area) slightly outperforms the configuration with wider camera angles (rear area) at the cost of losing LiDAR points. Although the differences are small, they can be considered statistically significant given the large size of the nuScenes dataset.

B. *Low-level crowdsource fusion*

An effective approach to combining data from various sources is to merge them at the lowest possible level. In this context, sensor data (e.g., LiDAR and RGB data) can be transferred between sources to ensure that all relevant information is available to each agent involved. However, it should be noted that this has significant demands on bandwidth and latency, requirements that are difficult to meet with current standards.

One possible solution to overcome this challenge is to reduce the amount of information shared by downsampling the original sensor data. In our study, we investigated neural network models designed to reconstruct downsampled LiDAR data within the host vehicle. Specifically, we trained a neural network to generate a complete scene from downsampled data so that it can closely approximate the original raw data from nearby agents.

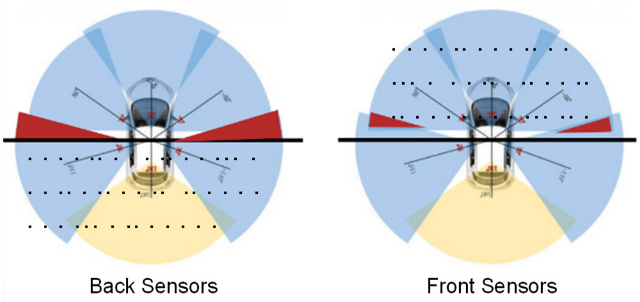

Fig. 3 Coverage of nuScenes Back Sensors and Front Sensors.

Table 1 Comparison of the performance of the model when using all sensors, only front sensors or only back sensors. mAP refers to mean Average Precision and NDS is the nuScenes Detection Score, which combines mAP with some true positive metrics (see section 3.1 in [18]).

Experiment	mAP (%)	NDS (%)
Baseline (all sensors)	0.68	0.71
Front cameras + Frontal 180 degree LiDAR	0.66	0.69
Back cameras + Back 180 degree LiDAR	0.63	0.68

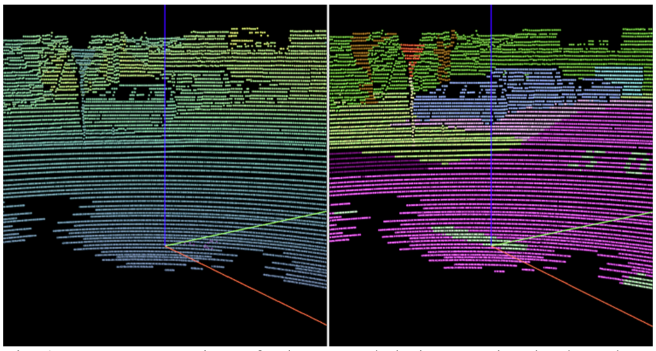

Fig. 4 Two representations of a downsampled LiDAR point cloud: on the left, a depth map (color coded); on the right, point cloud object labels.

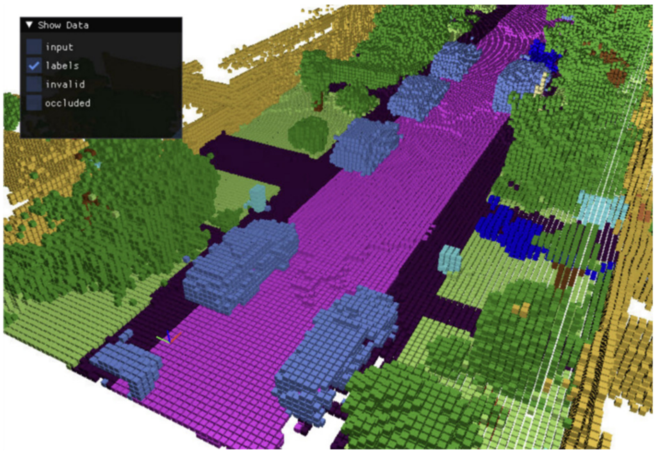

Fig. 5 LiDAR point cloud of Fig. 4 after applying scene completion

By using these advanced reconstruction techniques, we aim to preserve the integrity and utility of sensor data while reducing bandwidth and latency requirements. This approach facilitates efficient data sharing between different sources and improves the performance of the object detection system by exploiting the reconstructed high fidelity sensor data.

To achieve this, we have implemented a scene completion system following the SCPNet architecture [20]. In this solution, a teacher-student framework is proposed to distill dense semantic knowledge from full LiDAR point clouds (teacher) to downsampled – using a simple random sampling - LiDAR point clouds (student).

Experiments were conducted using the Semantic KITTI dataset [21] where downsampled point clouds in Fig. 4 are successfully completed and semantically segmented as shown in Fig. 5. Our proposal is to exploit the completed point cloud in the ego vehicle when only the downsampled versions are transmitted from neighbour agents, therefore reducing bandwidth and latency needs. The experiments demonstrate the competitive robustness of this approach for static objects.

However, the model faces challenges with smaller dynamic classes, such as bicyclists and motorcyclists. Despite these difficulties, qualitative visualizations of inferred scenes remain clear and comprehensible, allowing for straight-forward interpretation and understanding of the traffic scene.

C. *Intermediate-level crowdsource fusion*

Intermediate fusion [6] represents a compromise approach that uses processed intermediate feature representations of the individual agents. Instead of transmitting raw sensor data or final detection results, mid-level features are shared. These representations are created independently for each agent and contain data from LiDAR and RGB cameras. By sharing feature representations, the amount of information transferred is reduced, while retaining valuable content of the original sensor data. In addition, neural networks used to extract these feature representations can be trained to maximize the utility of the extracted data for neighboring agents. This approach optimizes the efficiency of data transfer and the preservation of information needed for collaborative tasks such as object detection and scene understanding.

A cooperative perception architecture based on feature fusion can achieve feature selection and information aggregation to improve perception accuracy. We extract valuable spatial and channel features with an adaptive feature fusion network [22] by exploiting the trainable feature selection module. These features are transformed into the same data format for fusion through projection or feature alignment.

An intermediate fusion approach as shown in Fig. 6 optimizes, integrates and processes information from various locations, achieving lightweight feature fusion with lower bandwidth requirements for precise real-time object detection.

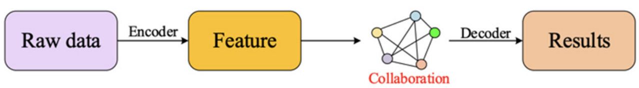

Fig. 6 Intermediate Fusion scheme

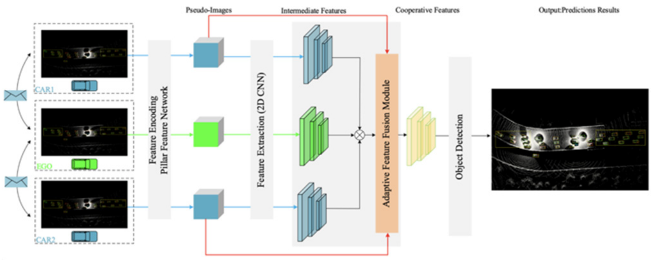

Fig. 7 Diagram of the 3D target detection cooperative perception network

By exploring feature fusion strategies, involving feature encoding, extraction and projection, we transform 3D point clouds into 2D pseudo-images rich in feature information (see section 3 in [22]). Finally, by adaptively fusing mid-level point cloud features obtained by I(edge)-Compute, the ego vehicle will receive and aggregate data from other CAVs, to achieve cooperative perception. Fig. 7 shows the schematic representation of the proposed 3D target detection network.

D. ***High-level crowdsource fusion***

High level (or late) fusion aims to directly combine detections performed by nearby vehicles and the infrastructure. This reduces bandwidth requirements and ensures communication efficiency, so that real-time data can be shared at high transmission speed and low latency.

High-level fusion transmits the results of algorithms that process raw sensor data and extracts the most meaningful and compacted information through late fusion to benefit from the perception of nearby agents. In our exploration of how high-level crowdsourcing of data and multi-ego pose detections can benefit connected mobility, we conduct simulations to address the CRS-SLAM problem in complex urban scenes from the nuScenes dataset [18], hoping that host vehicles (HVs) improve its awareness of the environment thanks to this late fusion.

The high-level crowdsourced fusion uses the detected 3D objects as SLAM observations. Therefore, we propose an online CRS-SLAM approach that estimates, at the current time step, the location of the agent together with the map *solely* from these observations (detection data). The approach is based on EKF (Extended Kalman Filter) and ICP (Iterative Closest Point). Each user builds his own map of detections and transmits it to the road participants –via V2V– and to the infrastructure –via V2I , or by C-V2X systems when available. This map is made up of distinct and individual points, each one referring to the center of each 3D Bounding Box associated to the detection of a 3D object.

Real-world datasets are crucial to evaluate the performance of the proposed approaches in various scenarios. While some synthetic datasets [7] address V2V or/ and V2I communication, there are few publicly available datasets with real data specifically designed for crowdsourced SLAM. To address this limitation, we adapt one of the largest multimodal open-source datasets, nuScenes [18], for the simulation of multi-vehicle scenarios. Our strategy involves downsampling by several factors the vehicle scene positions within the sequence according to the number of vehicles that must share the data with the HV. By assigning each sample of a series of consecutive frames to each road user, we create scenarios that resemble platooning.

Table 2 CRS-SLAM results for some scenes using the MDE metric. Scene-0061 is a complex scenario where the car moves through an intersection and turns left, with few detected cars and several roadside objects along the road. Scene-0238 is a simpler scenario where the vehicle crosses an intersection following a linear trajectory at low speed. When noise is added, it is normally distributed with standard deviation of 0.5 meters for positions and 0.2 radians for rotation angles, both for estimated ego-pose and detections.

Scene	**Detected Objects**	**Traffic Participants**	**MDE [m] (w/o noise)**	**MDE [m] (w/ noise)**
0061	Cars & Roadside Objects	-	2.49	3.19
		V2	0.62	0.63
		V1 & V2	0.99	0.64
0238	Cars	-	0.46	1.42
		V2	0.12	0.66
		V1 & V2	0.12	0.68

To investigate the extent to which CRS-SLAM from multiple ego-poses outperforms SLAM from a single ego-pose, we evaluate the performance of simulations that adopt the above strategy. We consider scenarios with up to three collaborating vehicles. Since localization and mapping are interdependent tasks, we measure the localization error using the Mean Displacement Error (MDE) metric. In order to isolate the effect of object detection and simulate the detection accuracy, we input as observations the nuScenes Ground Truth *annotations* without (w/o) and with (w/) zero mean Gaussian noise added. The simulated noise is added to estimated ego-poses, detection distances and rotation angles.

Table 2 shows the results obtained for a couple of scenes. The Traffic Participants column indicates whether the HV solves the SLAM problem using only its own detections (-), performs late fusion with the detections provided by the vehicle in front (V2), or with the detections of two vehicles that are behind and in front (V1 and V2).

From these results, we conclude that having access to the detections of another vehicle significantly improves the localization accuracy compared to performing SLAM without crowdsourcing, as the MDE is reduced by a factor of 4 times without noise, and by a factor of 5 to 2 with noise. However, adding a second vehicle (V1) that shares the data with the HV does not lead to a significant change in the MDE of the SLAM of the HV. Our results concur with earlier findings in [22], where increasing the number of cooperating participants does not necessarily lead to an improvement in performance.

The analysis of the behaviour in different scenes shows that crowdsourcing of detections in connected mobility is particularly helpful in complex scenarios. In particular in intersections, where the MDE is reduced by a significant factor both with and without noise. Furthermore, using data shared by a second vehicle increases the accuracy of their proposed vehicle detection algorithm, but adding more cooperating participants, as already found by other authors, just provides marginal benefits not leading to more significant improvements.

IV. Outlook and Future Directions

The sharing of information on the Internet has led to breakthroughs such as Wikipedia and numerous other technologies and applications that benefit from the vast amount of data available online. To ensure that a vehicle can accurately perceive its surroundings, simply adding more sensors to each vehicle's ADAS systems is insufficient. Much like information sharing on the Internet, integrating sensor data from all Connected and Autonomous Vehicles (CAVs) and traffic infrastructure can facilitate the transition to fully autonomous vehicles through cooperative and collective perception systems. This approach has several requirements:

- Facilitating communication and information sharing with technologies such as those proposed in the 6G-EWOC project, which aims to develop 6G technologies, network tools, and applications to support autonomous driving.
- Developing new approaches for environment perception based on data crowdsourcing, such as the proposed CRS-SLAM, which is flexible enough to adapt to different transmission conditions for low-, mid-, and high-level communications available in the scene.
- Changing societal attitudes towards what perception data sharing could mean to improve efficiency, reliability, and safety, benefiting both vehicles and other (particularly vulnerable) road users.

The 6G-EWOC project aims to develop these 6G technologies, network tools, and applications to support autonomous driving through V2V and V2I communication of observations of objects in hidden or invisible areas.

In this paper, we present preliminary results demonstrating the flexibility of the crowdsourced SLAM approach within the context of the improved communication technologies of the 6G-EWOC project. Future work will involve a technical analysis to adapt the specific characteristics of the flexible CRS-SLAM to the four Use Cases (UC) defined in 6G-EWOC [11]: Intersection drive-through; Improvement/ Optimization of Traffic Flow, Prioritization of Emergency Vehicles and Improved Safety for Vulnerable Road Users.

The detailed requirements, constraints, and AI resources available for each of these UCs will be analyzed and reported, leading to a comprehensive description of the 6G-EWOC architecture supporting flexible CRS-SLAM for each UC. This also requires the integration of advanced technologies, such as high-capacity optical fiber fronthaul connectivity, efficient edge data center technologies, photonic integrated circuits enabling Joint Communication and Sensing (JCAS), AI-based processing and optimization of network resources, and Optical Wireless Communication (OWC). Notably, this paper describes the first real-time, Gbit/s LiDAR point cloud transmission with a fully analogue, lowest latency, OWC link at 100 meters using an optical beamformer.

References

[1] WHO, *Global Status Report on Road Safety 2023*. Geneva: World Health Organization, 2023. [Online]. https://www.who.int/teams/social-determinants-of-health/safety-and-mobility/global-status-report-on-road-safety-2023

[2] TomTom, "Annual TomTom Traffic Index: Major Traffic Trends Shaping Cities Globally in 2023," Amsterdam, Press Rel., Jan. 2024. [Online]. https://www.tomtom.com/newsroom/press-releases/general/425369773/major-traffic-trends-shaping-cities-globally-in-2023

[3] F. Duarte and C. Ratti, "The Impact of Autonomous Vehicles on Cities: A Review," *Journal of Urban Technology*, vol. 25, no. 4, pp. 3–18, Oct. 2018, doi: 10.1080/10630732.2018.1493883

[4] T. Litman, "Autonomous Vehicle Implementation Predictions: Implications for Transport Planning," Victoria Transport Policy Institute, Victoria, Canada, Dec. 2023. [Online]. vtpi.org/avip.pdf

[5] SAE Intl, *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles - SAE International*, *J3016:*, USA., Apr. 2021. sae.org/standards/content/j3016_202104

[6] E. Arnold, M. Dianati, R. de Temple, and S. Fallah, "Cooperative Perception for 3D Object Detection in Driving Scenarios Using Infrastructure Sensors," *IEEE Trans. Intel. Transportation Sys.* vol. 23, no. 3, pp. 1852–1864, Mar. 2022, doi: 10.1109/TITS.2020.3028424

[7] T.-H. Wang, S. Manivasagam, M. Liang, B. Yang, W. Zeng, and R. Urtasun, "V2VNet: Vehicle-to-Vehicle Communication for Joint Perception and Prediction," in *Computer Vision – ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox, J.-M. Frahm, Eds., Cham: Springer Intl. Publishing, 2020, pp. 605–621. doi: 10.1007/978-3-030-58536-5_36

[8] B. Cao, C.-N. Ritter, K. Alomari, and D. Goehring, "Cooperative LiDAR Localization and Mapping for V2X Connected Autonomous Vehicles," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct. 2023, pp. 11019–11026. doi: 10.1109/IROS55552.2023.10341513

[9] P.-Y. Lajoie, B. Ramtoula, F. Wu, G. Beltrame, "Towards Collaborative Simultaneous Localization and Mapping: A Survey of the Current Research Landscape," *Field Robotics*, 2(1), 2022. doi: 10.55417/fr.2022032

[10] 5G Automotive Association (5GAA), "Connected Mobility: C-V2X explained," 5GAA. [Online]. https://5gaa.org/c-v2x-explained

[11] 6G-EWOC: D2.1 "Report on use cases, user and system requirements", June/2024. https://6g-ewoc.eu

[12] F. Honz, and B. Schrenk,"Fiber-Based Focal Plane Array Beamformer as Air Interface of an Alignment-Tolerant Optical Fi-Wi-Fi Bridge," in *Proc. Europ. Conf. Opt. Comm. ECOC*, Frankfurt, Sep. 2024, accepted

[13] M. Z. Chowdhury, Md. Shahjalal, S. Ahmed, Y. M. Jang, "6G Wireless Communication Systems: Applications, Requirements, Technologies, Challenges, and Research Directions," *IEEE Open J. Comm Society* vol. 1 pp. 957–975 2020 doi: 10.1109/OJCOMS.2020.3010270

[14] Z. Vujicic *et al.*, "Toward Virtualized Optical-Wireless Heterogeneous Networks," *IEEE Access*, vol. 12, pp. 87776–87806, 2024, doi: 10.1109/ACCESS.2024.3417358

[15] J. A. Altabas, O. Gallardo, G. S. Valdecasa, M. Squartecchia, T. K. Johansen, J. B. Jensen, "DSP-Free Real-Time 25 GBPS Quasicoherent Receiver With Electrical SSB Filtering for C-Band Links up to 40 km SSMF," *Journal of Lightwave Technology*, vol. 38, no. 7, pp. 1785–1788, Apr. 2020, doi: 10.1109/JLT.2019.2963787

[16] E. Yurtsever, J. Lambert, A. Carballo, K. Takeda, "A Survey of Autonomous Driving: Common Practices & Emerging Technologies," *IEEE Access* 8, 2020. doi: 10.1109/ACCESS.2020.2983149

[17] M. R. Watts, C. Poulton, M. Byrd, and G. Smolka, "LiDAR on a Chip Enters the Fast Lane: Sensors for Self-Driving Cars and Robots will be Tiny, Reliable, and Affordable," *IEEE Spectrum*, vol. 60, no. 9, pp. 38–43, Sep. 2023, doi: 10.1109/MSPEC.2023.10234174

[18] H. Caesar *et al.*, "nuScenes: A Multimodal Dataset for Autonomous Driving," in 2020 *IEEE/CVF Conf. Computer Vision and Pattern Rec. (CVPR)*, pp. 11618–11628. doi: 10.1109/CVPR42600.2020.01164

[19] Z. Liu *et al.*, "BEVFusion: Multi-Task Multi-Sensor Fusion with Unified Bird's-Eye View Representation," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, May 2023, pp. 2774–2781. doi: 10.1109/ICRA48891.2023.10160968

[20] Z. Xia *et al.*, "SCPNet: Semantic Scene Completion on Point Cloud," in *2023 IEEE/CVF Conf. Comp. Vision and Pat. Recognition (CVPR)*, Jun. 2023, pp. 17642–17651. doi: 10.1109/ CVPR52729.2023.01692

[21] J. Behley *et al.*, "SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2019, pp. 9296–9306. doi: 10.1109/ICCV.2019.00939

[22] D. Qiao and F. Zulkernine, "Adaptive Feature Fusion for Cooperative Perception using LiDAR Point Clouds," in *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Jan. 2023, pp. 1186–1195. doi: 10.1109/WACV56688.2023.00124