Towards Collaborative Perception in Automated Driving: Combining Vehicle and Infrastructure Perspectives

Joao-Vitor Zacchi
Safety Reliability and Availability
Fraunhofer IKS
Munich, Germany
joao-vitor.zacchi@iks.fraunhofer.de

Mario Trapp
Institute Director
Fraunhofer IKS
Munich, Germany
mario.trapp@iks.fraunhofer.de

***Abstract*—Environment perception constitutes a foundational block for autonomous systems such as automated driving systems. Enhancing such features is imperative to breach the barrier of complex environments such as urban scenarios. Occlusions, appearances, and disappearances are a few of the difficulties traditional tracking algorithms may face in an urban context that hinders their performance. Moreover, approaches that deal with the data association problem are still physically limited by the point-of-view of the ego vehicle. In order to address these issues, we propose in this position paper a framework to merge different perspectives enabling collaborative perception and thus to enhance the dependability of the environment perception of automated vehicles in complex scenarios. To this end, each participant, i.e., automated vehicles and infrastructure, sends their perception results to the framework. A perception result includes Bayesian Occupancy Filter providing probabilistic information about object positions. Moreover, the results might include an additional classification of the objects, enabling us to optimize predicting future trajectories of the objects, which is particularly important for non-automated participants such as human-driven cars or pedestrians. The framework facilitates a more complete and clarified view of the context to enhance decision-making of the individual vehicles.**

***Index Terms*—Collaborative perception, Automated driving system, Bayesian occupancy filter**

I. Introduction

To enable an automated vehicle to move safely, it must have a complete navigation system. Navigation is responsible for carrying out motion planning that guides the car from an origin to a destination and ensures safety by detecting and avoiding obstacles. During this whole process, the car must also know its location in the world and control its movement. To that end, a central component of the vehicle's navigation systems is its perception chain.

The potential of faults in perception chains is of great interest since it can directly endanger passengers and the vehicle's surrounding. Even though there have been improvements in the reliability of perception for ADS, these systems are still flawed and subject to faults, mainly when submitted to complex scenarios such as a crossing in urban areas. Therefore, relying on single ego-perspective sensors might prove insufficient to assure the proper behaviour of the system.

Through communication, future ADS might be able to share their single perception of the environment and greatly enhance their perception range and accuracy. More precisely, the ego-vehicle can exchange its state and sensor information and receive those from other vehicles and the traffic infrastructure, thus enhancing its local world model. Therefore, the collaborative perception can be view as a high-order sensor fusion where each participant works as a virtual sensor providing data to one another.

In this paper, we propose a probabilistic framework enabling collaborative perception. To this end, the framework uses the results of different vehicles and the infrastructure's perception chain for compiling a more complete and more reliable integrated world model as the diverse sources use different hardware and software and see the scene from different perspectives. The fusion of the single sources is based upon Bayesian Occupancy Filters. [1]. As an extension to occupancy grids, which do not contain classification information on detected obstacles, the framework additionally uses object lists from single sources facilitating the classification and tracking of objects. Using these two types of information, the framework can associate behavior models for predicting a probabilistic future trajectory-space, e.g., to predict which way a pedestrian will go in the near future. Being integrated as a service of the roadside infrastructure, the framework can additionally be tailored to the concrete traffic routing, such as to a concrete crossing, so that additional semantic and statistical information for the local area can be incorporated, leading to more accurate results compared to comparably generic assumptions and information used by the ego-vehicles.

This paper is structured as follows: In Section II, the related work in collaborative perception is summarized, and research problems are formulated. Section III presents the vision of the framework and explains its components. Finally, section IV discusses the current state of the proposed framework, addresses concluding remarks, and gives an outlook on future work.

II. Related Work

To foster this approach, we base our work on a panel of previous work developed in various areas. Firstly, the approach of evaluating complex scenarios has been addressed for a long time in the field of autonomous vehicles. Once the vehicle is outside of a controlled environment - even with only a minor variance of participants and possible actions - the tracking of the environment becomes difficult. The problem of effective data association, i.e., the classification and tracking of detected objects and thus identifying appropriately targeted measures in complex scenarios such as crossings in the urban environment, remains a challenge. Moreover, it is notable that the interaction of non-connected participants with the automated and connected vehicles is an essential factor in building an effective system capable of navigating in scenarios where this is required.

From the cooperative perception community, the core of the work has been developed throughout the last decade. First approaches have a firm root in the communication domain with a first work [2] using VANETS to exchange raw image data in order to create a see-through system as an Advanced Driver-Assistance System (ADAS) in order to facilitate overtaking maneuvers. Still, on the level of communication, the German project Ko-PER investigated the communication requirements for collaborative perception (CP) such as transmission range and latency [3]. Extending the initial findings, the project also investigated techniques for temporal and spatial alignment of the messages exchanged between vehicles [4].

The author in [5] investigated the impact of CP on the control and motion planning of the vehicle. The results achieved with a see-through collision warning system showed that CP improved the safety and controllability of the vehicle. The work was further extended to address delay composition in the exchange between messages among vehicles [6].

From a computer vision perspective, several studies have been developed. The Ko-HAF project [7] investigated the safe identification of traffic signs and traffic disruption - the project built on sharing information among vehicles in order to optimize traffic flow and enhance safety.

At the sensor level, Chen et al. [8] proposed a low-level fusion architecture in order to unify different vehicles' point clouds for enhancing object detection accuracy. Extending its previous work, a feature-level fusion was proposed to compare the performance of the different levels. The work shows in both cases that the aforementioned data fusion enhances the number of objects detected by the vehicles. Recently, Arnold et al. [9] proposed a fusion architecture based on sensors installed in the infrastructure in order to provide enriched world models to vehicles in the infrastructure's range. This approach opposes the one in [8], which had the fusion executed at the vehicle level.

Both of the approaches mentioned above show improved results against individual perception. Nonetheless, by executing the merger inside the vehicles, [8] introduces a high burden and unnecessary redundant computations. As for [9], by not using the available perception in vehicles, it creates the necessity for permanent and reliable infrastructure sensors. Moreover, both solutions rely on artificial intelligence object detection algorithms, which are susceptible to unpredictable behaviour. This position paper, proposes an intermediate solution by profiting from both infrastructure and vehicle perception capabilities by sharing intermediate features through Bayesian Occupancy Filter (BOF). By these means, the proposed framework shall provide as a contribution:

- An extension of the classical Bayesian Occupancy Filter to merge different perspectives receiving inputs from different sources, thus enlarging the perception capability of the collective.
- When available, use classification inputs coming in parallel to the BOF to optimize path prediction of the observed objects and, as a result, provide a more precise assessment of the context.

III. Proposed Framework

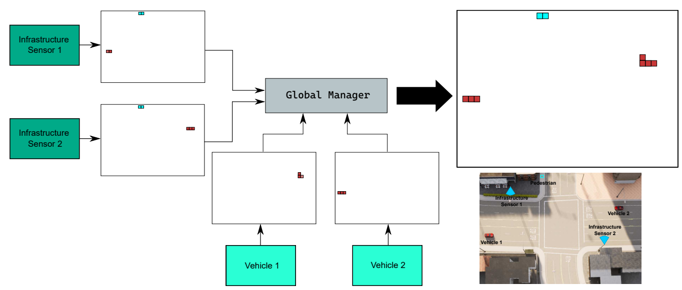

Figure 1. Framework overview

Sharing information among autonomous vehicles seems to be one way to enhance their perception capabilities and consequently minimize the risk associated with context misrepresentations. In this section, we present our vision of a framework that integrates perception of different ADS within a local area as well as infrastructure-based perception. We present its components and how they interact within the framework as a means to assure an enhanced world model. The main goal of this framework is to integrate simultaneously the dynamic evolution of the ADS, their perception chains, and those from infrastructure in order to guarantee that their interaction leads to an enhanced performance without causing any hazardous event. Figure 1 gives an overview of the collaborative perception cycle.

A. Perception inputs

Two inputs are considered for the merger at the global manager (GM): from vehicles and infrastructure. Each ADS is considered to be fully autonomous w.r.t. perception of the environment and controlling actions in to ensure the successful completion of its actions towards its environment. Each ADS is assumed to act safely without requiring the infrastructure to realize its safety concept. Additionally, we assume that each ADS is assured to provide safe interfaces to collaborate with the other systems safely. The safe integration of safe systems of systems is still challenging, but this particular aspect is not within the scope of our framework. Instead, we reuse safe integration concepts as, for example, suggested in [10] and [11].

In order to provide constant perception enhancement in critical scenarios such as high traffic crossings, the framework also relies on inputs provided by the surrounding infrastructure. The benefits of outside AV sensing has been shown in [9] and [12].

B. Bayesian Occupancy Filter

Bayesian Occupancy Filters (BOF) are the main format to be exchanged between sensors (vehicles and infrastructure) and the global manager. Many approaches use an object list, providing a list of classified objects, their position, and possibly their estimated expected trajectory. However, this requires that the data received from sensors can be associated with the objects, which is hardly possible in crowded scenes where many different objects are pass by each other so that it becomes challenging to identify and track objects. Meaning that given the object list of the detected objects, it must control if a sensor input matches one of the existing objects. Moreover, it must also, since the tracking is independent for each object, add and delete them to the object list. Extending these challenges to complex scenarios such as urban crossings makes the classical object-tracking intractable due to the high dynamicity of the targets. That is why [13] introduce BOF as a method to circumvent this so-called data association problem and to take into account the uncertainties associated with the dynamic evolution of the context perceived by the autonomous system. Initially implemented in Advanced Driver Assistance Systems, the BOF has been used in various other applications as described in [14].

Using BOF allows the framework to initially abstract from the notions of tracking and detection. The filter relies on an original use of occupation grids [15]. In this representation, the environment is cut into cells. The occupancy probability of a cell by any object is estimated from the observations of the sensors. In this case, the grids are defined in an ego-centric frame of reference (i.e., centered on the vehicle), representing the state space of the objects present in the environment (position and relative speed). The area defined by these grids constitutes an additional feature compared to the classical uses of occupancy grids (position only).

In the context of the framework, the output of the filter is the core data exchanged by the agents. Thus, regardless of the perception chain implemented or specific algorithms run, the filter becomes a standard interface to be exchanged with the global manager.

C. Data Fusion

Based on the BOF-data provided by the single participants, the core of the framework is realized within the Global Manager (GM). The Global Manager is the core of the framework. It is responsible for:

- Detecting new collaborating agents in the scenario. That is expected to happen over communication channels. Advances in the V2X community [16] lead to believe that resilient communication in automated driving scenarios is not a distant goal to be achieved. Although, challenges remain in the way of the technological deployment [17].
- Merge the individual outputs of the perception chains in order to evaluate the scenario. By doing so, the GM compiles a more complete picture of how the context may evolve and if an agent's predicted behaviour might impair other participant's safety.
- Creating a digital representation of the collaborating systems, ADS or infrastructure, referred here as Dynamic Agent (DA). The DA allows the GM to have precise information about the situation and intentions and thus spares the burden of predicting their behaviours.
- Feedback the enlarged world model to the participating ADS so they can improve their decision-making with a more accurate perception of the environment.

Therefore, the global manager's core task is the data fusion, i.e., to merge the Bayesian Occupancy Filters' output received from the different vehicles and the infrastructure to detect potential obstacles. In order to merge perspectives from both vehicles and infrastructure sensors, a coordinate transformation must take place first so that all inputs are available in a common reference coordinate system. This can be achieved through the inverse extrinsic matrix of the perception chains. Given a set X_g of global coordinates (x_g, y_g, z_g), one can obtain the sensor i coordinates through the extrinsic matrix M^{ext}:

$$M_i^{ext} = [R_i | t_i]$$

Being R_i and t_i the rotation matrix and translation vector of sensor i. To obtain the global coordinates, the inverse process can be directly applied using homogeneous coordinates:

$$X_g = \begin{bmatrix} x_g \\ y_g \\ z_g \\ 1 \end{bmatrix} = M_i^{ext^{-1}} \begin{bmatrix} x_s \\ y_s \\ z_s \\ 1 \end{bmatrix}$$

It is important to notice that these matrices depend on the position and orientation of the sensor in the global environment. Thus, delays and noise in the computation of these parameters may disturb the transformation. Although currently not taking these factors into account, addressing them will be subject to future work.

Once the BOFs have been transformed into a joint coordinate system, the framework merges the different result into one global world model. To this end, we extended the BOF algorithms but still following the principles of Bayesian Filters. Instead of merging temporal sequences of sensor input, we merge spatial alternatives representing the same point in time.

Merging different BOF-based perception data from different sources and different perspectives allows us to yield more accurate detection results without using machine learning-based algorithms, which hence simplifies the according safety case. However, if the single vehicles prefer using their own perception system as the primary perception channel, the integrated results provided by the global manager can serve as additional estimated ground truth the vehicles can integrate into their monitoring architecture.

D. Model refinement with classification

However, BOFs do not only represent the current situation, but it also supports a prediction of the future using the probabilities of position and velocity, which are inherent information provided by the BOF grids. Therefore, in order to predict the future, the Global Manager propagates the probabilities provided in the integrated grid for inferring how the grid will evolve in the next step. However, using a standard BOF approach leads to a comparably low accuracy as the prediction assumes that the participants could, in principle, move in any theoretically possible direction. Though initially abstracting from tracking and detection with the BOF, we use classification information as optional, additional inputs for enhancing the prediction provided by the framework. By assigning a type to the object, the framework is then allowed to constrain the possibilities of predictions and generate a more accurate assessment of the situation. For example, it is clear that a car is likely to follow the street and will not drive onto a walkway. In the same way, typical paths of pedestrians can be considered by the framework. To this end, the framework additionally profits from the advantage to be tailored to the specific traffic situation, such as a concrete crossing, so that it can incorporate semantic and statistical knowledge about the specific crossing into the prediction models.

In this way, although still relying on the BOF's robustness to identify the surroundings, the framework can then fine-tune the model to provide a better representation.

IV. Conclusion and Future Work

As we shift towards a world with increased reliance on autonomous systems, their perception capabilities become paramount for safety and performance. Current approaches rely on perception chains associated with the individual vehicle and thus are limited to their physical constraints. Given the uncertainties and physical limitations of complex scenarios, their deployment is a challenge.

To address these limitations, we proposed a framework to merge different perspectives to augment the context knowledge of individual vehicles while maintaining their independence regarding their decision-making capabilities. We aim to achieve this by extending the Bayesian Occupancy Filter to a collaborative scale by sharing outputs of individual perception chains. Moreover, we aim to fine-tune the initial prediction models in the grids with the classification outputs to constrain path prediction.

Currently, we have implemented individual perception chains on individual vehicles in a simulated environment with CARLA [18]. As next step, we will introduce the BOF computation and merger to evaluate the approach against individual perception chains and other collaborative approaches.

Acknowledgment

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 812.788 (MSCA-ETN SAS). This publication reflects only the authors view, exempting the European Union from any liability. Project website: http://etn-sas.eu/.

References

[1] C. Coue, C. Pradalier, C. Laugier, T. Fraichard, and P. Bessière, "Bayesian Occupancy Filtering for Multitarget Tracking: An Automotive Application," vol. 25, no. 1, pp. 19–30.

[2] C. Olaverri-Monreal, P. Gomes, R. Fernandes, F. Vieira, and M. Ferreira, "The See-Through System: A VANET-enabled assistant for overtaking maneuvers," in *2010 IEEE Intelligent Vehicles Symposium*. IEEE, pp. 123–128.

[3] A. Rauch, F. Klanner, and K. Dietmayer, "Analysis of V2X communication parameters for the development of a fusion architecture for cooperative perception systems," in *2011 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, pp. 685–690.

[4] A. Rauch, F. Klanner, R. Rasshofer, and K. Dietmayer, "Car2X-based perception in a high-level fusion architecture for cooperative perception systems," in *2012 IEEE Intelligent Vehicles Symposium*. IEEE, pp. 270–275.

[5] S.-W. Kim, W. Liu, M. H. Ang, E. Frazzoli, and D. Rus, "The Impact of Cooperative Perception on Decision Making and Planning of Autonomous Vehicles," vol. 7, no. 3, pp. 39–50.

[6] S.-W. Kim, B. Qin, Z. J. Chong, X. Shen, W. Liu, M. H. Ang, E. Frazzoli, and D. Rus, "Multivehicle Cooperative Driving Using Cooperative Perception: Design and Experimental Validation," vol. 16, no. 2, pp. 663–680.

[7] S. Krug, "Ko-HAF - Kooperatives Hochautomatisiertes Fahren," p. 270.

[8] Q. Chen, S. Tang, Q. Yang, and S. Fu, "Cooper: Cooperative Perception for Connected Autonomous Vehicles based on 3D Point Clouds."

[9] E. Arnold, M. Dianati, R. de Temple, and S. Fallah, "Cooperative Perception for 3D Object Detection in Driving Scenarios Using Infrastructure Sensors," pp. 1–13.

[10] E. Cioroaica, T. Kuhn, and B. Buhnova, "(Do Not) Trust in Ecosystems," in *2019 IEEE/ACM 41st International Conference on Software Engineering: New Ideas and Emerging Results (ICSE-NIER)*. IEEE, pp. 9–12.

[11] J. Rushby, "Trustworthy Self-Integrating Systems," in *Distributed Computing and Internet Technology*, ser. Lecture Notes in Computer Science, N. Bjørner, S. Prasad, and L. Parida, Eds. Springer International Publishing, vol. 9581, pp. 19–29.

[12] B. Lv, H. Xu, J. Wu, Y. Tian, Y. Zhang, Y. Zheng, C. Yuan, and S. Tian, "LiDAR-Enhanced Connected Infrastructures Sensing and Broadcasting High-Resolution Traffic Information Serving Smart Cities," vol. 7, pp. 79 895–79 907.

[13] C. Coue, T. Fraichard, P. Bessière, and E. Mazer, "Using Bayesian Programming for Multi-Sensor Mu1ti-Target Tracking in Automotive Applications," p. 6.

[14] M. Saval-Calvo, L. Medina-Valdés, J. Castillo-Secilla, S. Cuenca-Asensi, A. Martínez-Álvarez, and J. Villagrá, "A Review of the Bayesian Occupancy Filter," vol. 17, no. 2, p. 344.

[15] H. Moravec and A. Elfes, "High resolution maps from wide angle sonar," in *Proceedings. 1985 IEEE International Conference on Robotics and Automation*, vol. 2. Institute of Electrical and Electronics Engineers, pp. 116–121.

[16] D. Elliott, W. Keen, and L. Miao, "Recent advances in connected and automated vehicles," vol. 6, no. 2, pp. 109–131.

[17] I. Din, B.-S. Kim, S. Hassan, M. Guizani, M. Atiquzzaman, and J. Rodrigues, "Information-Centric Network-Based Vehicular Communications: Overview and Research Opportunities," vol. 18, no. 11, p. 3957.

[18] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun. CARLA: An Open Urban Driving Simulator.