Cooperative Perception for Connected Autonomous Vehicles under Constrained V2V Networking

Ahmad Sarlak*, Hazim Alzorgan*, Sayed Pedram Haeri Boroujeni*, Rahul Amin†, and Abolfazl Razi*
*School of Computing, Clemson University, Clemson, SC, USA
†Tactical Networks Group, MIT Lincoln Laboratory, Lexington, MA, USA
Emails: {asarlak, halzorg, shaerib, arazi}@clemson.edu, rahul.amin@ll.mit.edu

***Abstract*—Cooperative Perception (CP) is a newly emerged technique that can be used to enhance the safety of Autonomous Vehicles (AVs) when relying merely on the AV's own camera may not yield accurate results. Examples of such conditions are delayed or low object detection accuracy under foggy weather, winding roads, and blocked camera vision by the front vehicle. A few CP methods have been recently proposed to enhance the quality of AV perception through cooperative perception. Despite their success, these methods have a few limitations, such as adopting unrealistic assumptions about perfect communication and unconstrained resources as well as having access to large training datasets. In this paper, we use the LTE Release 14 Mode 4 side-link communication to implement CP by selective V2V communication for situational awareness. Our method is based on a demand-response mechanism and solving an optimization problem to employ helper vehicles, considering networking performance metrics (such as packet drop rate), available resources (radio blocks in LTE-V), extended visual range (the total road segment covered), and more importantly the ultimate perception quality of the selected vehicles. Our preliminary results demonstrate a significant gain for the proposed method compared to a conventional single-camera vision.**

***Index Terms*—Cooperative perception, autonomous vehicles, vehicular networks, V2V communication.**

I. Introduction

Connected Autonomous Vehicles (CAVs) are a type of vehicle that utilize two complementary technologies: Artificial Intelligence (AI) for autonomous driving (e.g. SAE Levels 4 and 5) and wireless connectivity for exchanging messages between vehicles and roadside infrastructure. This technology enables a new feature, called Cooperative Perception (CP) with the main idea of fusing camera outputs from multiple AVs with potentially different imaging conditions or viewpoints to enhance AVs' perception or extended their visual range for safer and more informed autonomous driving [1].

Recent studies have explored the benefits of leveraging multiple visual feeds in enhancing perception quality and situational awareness through Vehicle to Vehicle (V2V) collaboration as well as Vehicle to Infrastructure (V2I) networking when Road Side Units (RSUs) join the cooperation. This approach involves sharing visual information (such as raw sensory data and input image), intermediate deep learning features, and detection outputs among nearby AVs [2]–[4].

Although these methods improve the accuracy of object detection for AVs by fusing visual information from the same scene, they may not suffice for adverse weather conditions. Recently, a domain adaptation technique is proposed in [5] to enhance the quality of CP-based object detection for AVs by addressing discrepancies in image style and object appearance. Their method achieves a much higher object detection accuracy in foggy weather conditions. However, this method requires a large training set of images taken from the same scenes under different conditions (e.g., clear, rainy, foggy). One general drawback of these methods is their unrealistic assumption of perfect communication and unconstrained availability of helper video feeds. Further, most of these methods do not cover situations like sharp curves in winding roads that require a more scrutinized camera selection based on their position and viewpoints.

This paper offers a cooperative perception approach for CAVs to enhance their situational awareness and safety for cases where the CAV's own perception is not reliable due to weather conditions (fog, low vision), road curvature, or limited vision range (masked by front vehicle). Our approach is communication protocol-agnostic and applicable to V2V networking in any protocol, including DSRC [6], LTE C-V2X (Cellular Vehicle-to-Everything) [7], and 5G-NR [8]. However, we consider C-V2X (Release 14, Mode 4: Sidelink communication for infrastructure-less V2V) for its practicality and proven performance [9], [10].

Specifically, We use an on-demand request-response mechanism with three steps: (i) the AV with low vision (due to foggy weather, masked vision, or curvy roads) broadcasts its request to the front vehicles; (ii) nearby vehicles respond by notifying their intention to cooperate along with their estimated channel conditions, transmission budget, and sample images through a contention-based feedback mechanism, (ii) the AV solves an optimization problem to select optimal cameras considering both networking conditions and the added gain to the quality of cooperative perception. Our method is different than prior work from multiple perspectives. First, our method relies on fusing various scenes from different viewpoints with partial overlap. Secondly, our method considers imperfect communication and takes into account the quality of communication channels (e.g., packet drop rate, average delay, energy budget) and resource availability (the number of radio blocks in LTE-V2X, Mode 4) in the employed optimization. Thirdly, the added quality to the CP is included in the helper camera selection process, as opposed to prior CP methods that consider a pre-selected set of images. Finally, our method integrates the results of object detection of multiple cameras at later stages instead of fusing intermediate features, therefore does not require custom-trained DL networks. We use CARLA simulator to generate test cases. Our preliminary results show a significant gain in the perception quality using the proposed method.

II. Proposed Method

In our method, the ego vehicle broadcasts a request for selecting a subset of M out of N vehicles, denoted by $V_1, V_2, \ldots, V_N$. The vehicles notify their intention through Ack messages with profile info that includes their position (x_i, y_i), speed v_i, channel conditions), and sample images. Further information (e.g., the distance to the ego vehicle, approximate vision range, motion blur, field of view, the required transmission energy budget E, and the packet drop rate β) can

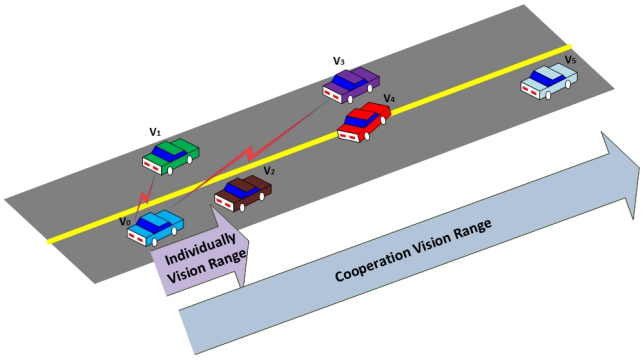

Fig. 1. Cooperative perception helps increase the vision range in masked vision. The ego (V_0) vehicle's vision is marked by the front car (V_2, and employing the green (V_1) vehicle's camera extends its visual range and situational awareness.

be estimated from the gathered information (position, channel conditions) to be further refined by inspecting the sample images. Another key factor is maximizing contextual information diversity since prior work and our own investigation show a higher gain when the information diversity (e.g., FoV, distance range) is maximized [11].

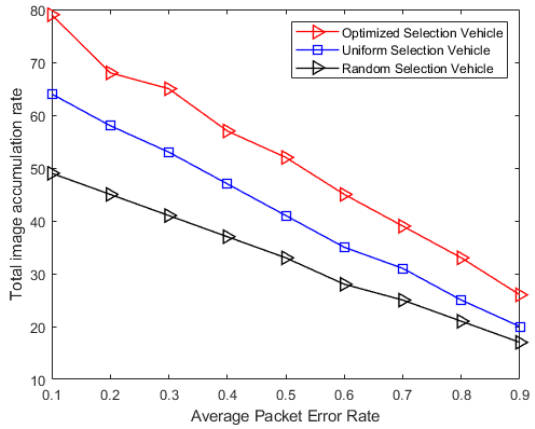

Fig. 2. Total image accumulation rate (with $N = 10$, $M = 5$) using different selection policies versus packet error rate.

Our optimization is based on a set of transmission Key Performance Indicators (KPIs) calculated for the set of selected vehicles. Specifically if the binary vector $\vec{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_N)$ represents the selected vehicles, we calculate the following metrics: (1) **Effective throughput:** $\zeta_i = R_{ch}/E[R_i] = R_{ch}(1-\beta_i)$, where R_{ch} is the rate of channel, $E[R_i] = 1/(1-\beta_i)$ is the average transmission per packet under drop rate β_i. The overall throughput which is simply the sum of the individual throughputs $f_1(\vec{\alpha}) = \sum_{i=1}^{\mathrm{N}} \alpha_i \zeta_i = \sum_{i=1}^{\mathrm{N}} \alpha_i R_{ch}(1-\beta_i)$. (2) **Transmission energy:** $f_2(\vec{\alpha}) = \sum_{i=1}^{\mathrm{N}} \alpha_i \frac{P_i d_i}{R_{ch}(1-\beta_i)}$, where P_i denotes the average transmission power in the process of transmitting, d_i is the distance between the ego and selected vehicle V_i. (3) **Motion Blur:** drop in the captured image quality due to the instantaneous relative speed v_i' between the vehicle V_i and the target object. We use $L_i = \frac{v_i' T[f\cos(\varphi) - QG\sin(\varphi)]}{v_i' TQ\sin(\varphi) + zp}$ following [12], with parameters defined therein. $f_3(\vec{\alpha})$ quantifies this loss for selected vehicles. (4) **Visual Range:** We also estimate the visual range of front cameras based on their locations (x_i, y_i) and velocities v_i Particularly, if $V_{i_1}, V_{i_2}, \cdots, V_{i_M}$ are the M selected vehicles in order, with individual visual range F_{i_j} and relative distances to front vehicle $\hat{d}_{i_j} = \sqrt{(x_{i_j} - x_{i_{j+1}})^2 + (x_{i_j} - x_{i_{j+1}})^2}$, then the effective visual range is $\hat{F}_{i_j} = \min(F_{i_j}, \hat{d}_{i_j})$, and the total visual range is $f_4(\vec{\alpha}) = \sum_{j=1}^{M} \hat{F}_{i_j}$. (5) **Perception Quality:** This is perhaps the most important factor that aims to quantify the added perception quality (e.g., object detection accuracy in IoU sense) by adding each camera to the pool. In this work, we use a heuristic method by inspecting the added CP quality using the sample images included in Ack Messages. The goal is to maximize $f_5(\vec{\alpha})$. As opposed to previous factors, this factor is not simply the summation of individual gains, and all potential $N!/M!(N-M)!$ sets should be investigated. To lower complexity, we use coalition game theory where the Shapely value quantifies the marginal gain of each helper camera when joining all potential coalitions.

Once the ego vehicle received help proposals through Ack messages, solve the following optimization problem

$$\arg\max_{\vec{\alpha}} f(t) = \sum_{i=1}^{5} w_i f_i(\vec{\alpha}, t) \tag{1}$$
$$\text{s.t.} \sum_{j=1}^{N} \alpha_j \leq M, \quad \text{for} \alpha_j \in \{0, 1\}$$

in an interval-by-interval manner, where $f_i(\vec{\alpha})$ are the the aforementioned quality terms under selection $\vec{\alpha}$ at interval t.

III. Simulation

Sample results for sample scenarios with $N = 10$ total vehicles with random relative positions and velocities, $M = 5$ selected vehicles are shown in Figures 2 and 3. These results reflect the accuracy of traffic light detection under foggy weather condition as an exemplary application using traffic light dataset [13] and scenarios produced by CARLA simulator.

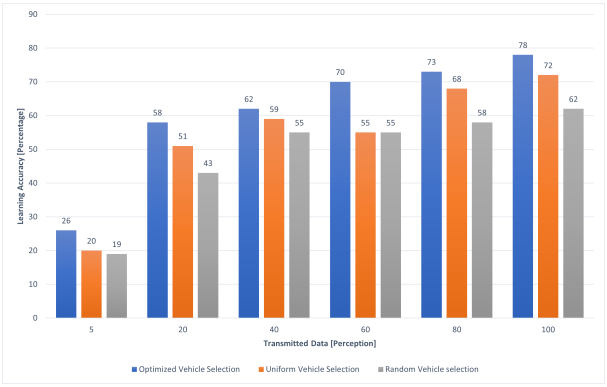

Fig. 3. Online learning accuracy of the system using different scheduling policies for different portions of transmitted samples.

Fig. 2 shows that our proposed method using optimization (1) archives a higher image collection rate than both uniform and random selection with a significant gain between 15% to 30%. This resulted in a considerable gain about 10% to 15% in terms of overall traffic light interpretation accuracy for the proposed method, provided in Fig. 3.

IV. Conclusion

We exploit the side-link feature of LTE-V to implement Cooperative Perception (CP) for AVs to enhance their situational awareness and perception quality under harsh weather conditions (e.g., foggy, rainy, stormy), curved roads, and partially obscured vision by front vehicles. Our method considers communication performance metrics and available resources while selecting vehicles that collectively provide the longest vision range and higher cooperative perception quality. Our method does not require a large multi-view training dataset (that is non-existent) and is compatible with any DL-based image processing since we fuse the end results. Our preliminary results show a significant gain (15% to 30%) gain in traffic light interpretation as an exemplary application showing the promising potential of CP to enhance driving safety.

References

[1] H. Zhang, G. Luo, J. Li, and F.-Y. Wang, "C2fda: Coarse-to-fine domain adaptation for traffic object detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 12633–12647, 2021.

[2] R. Xu, H. Xiang, Z. Tu, X. Xia, M.-H. Yang, and J. Ma, "V2x-vit: Vehicle-to-everything cooperative perception with vision transformer," in *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXIX*, pp. 107–124, Springer, 2022.

[3] R. Xu, H. Xiang, X. Xia, X. Han, J. Li, and J. Ma, "Opv2v: An open benchmark dataset and fusion pipeline for perception with vehicle-to-vehicle communication," in *2022 International Conference on Robotics and Automation (ICRA)*, pp. 2583–2589, IEEE, 2022.

[4] 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: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*, pp. 605–621, Springer, 2020.

[5] J. Li, R. Xu, J. Ma, Q. Zou, J. Ma, and H. Yu, "Domain adaptive object detection for autonomous driving under foggy weather," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 612–622, 2023.

[6] Y. L. Morgan, "Notes on dsrc & wave standards suite: Its architecture, design, and characteristics," *IEEE Communications Surveys & Tutorials*, vol. 12, no. 4, pp. 504–518, 2010.

[7] A. Bazzi, A. O. Berthet, C. Campolo, B. M. Masini, A. Molinaro, and A. Zanella, "On the design of sidelink for cellular v2x: A literature review and outlook for future," *IEEE Access*, vol. 9, pp. 97953–97980, 2021.

[8] M. H. C. Garcia, A. Molina-Galan, M. Boban, J. Gozalvez, B. Coll-Perales, T. Şahin, and A. Kousaridas, "A tutorial on 5g nr v2x communications," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 3, pp. 1972–2026, 2021.

[9] S. Zeadally, M. A. Javed, and E. B. Hamida, "Vehicular communications for its: Standardization and challenges," *IEEE Communications Standards Magazine*, vol. 4, no. 1, pp. 11–17, 2020.

[10] A. Bazzi, G. Cecchini, M. Menarini, B. M. Masini, and A. Zanella, "Survey and perspectives of vehicular wi-fi versus sidelink cellular-v2x in the 5g era," *Future Internet*, vol. 11, no. 6, p. 122, 2019.

[11] X. Chen, H. Li, R. Amin, and A. Razi, "Rd-dpp: Rate-distortion theory meets determinantal point process to diversify learning data samples," *arXiv preprint arXiv:2304.04137*, 2023.

[12] H. Xiao, J. Zhao, Q. Pei, J. Feng, L. Liu, and W. Shi, "Vehicle selection and resource optimization for federated learning in vehicular edge computing," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 11073–11087, 2021.

[13] M. B. Jensen, M. P. Philipsen, A. Møgelmose, T. B. Moeslund, and M. M. Trivedi, "Vision for looking at traffic lights: Issues, survey, and perspectives," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 7, pp. 1800–1815, 2016.