Edge-Coordinated On-Road Perception for Connected Autonomous Vehicles Using Point Cloud

Jiawei Hou*, Peng Yang*, Tian Qin*, and Wen Wu†
*School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China
†Frontier Research Center, Peng Cheng Laboratory, Shenzhen, China
Email: *{jerry_hou, yangpeng, qin_tian}@hust.edu.cn, †wuw02@pcl.ac.cn

***Abstract*—In this paper, we tackle the problem of selecting connected autonomous vehicles (CAVs) with the most valuable point cloud data for edge-coordinated on-road perception. Through extensive experiments, we find that adding a CAV for collaborative perception yields diminishing gain in understanding the on-road environments, while the generated point cloud data size grows linearly with the number of employed CAVs. Meanwhile, both vehicular mobility and diversified road topology lead to the dynamics of data size of the captured point clouds. Based on those findings, we then formulate an optimization problem that maximizes the utility of collaborative perception at edge nodes. Considering the submodularity of collaborative perception utility and heterogeneity of point cloud data size from individual vehicles, a CAV candidate selection algorithm is proposed. The marginal gain of processing the point cloud data of each candidate is firstly evaluated and ranked, based on which a subset of CAVs are selected subject to bandwidth capacity. Finally, experimental results on an open dataset are presented to demonstrate the superiority of the proposed algorithm under dynamic traffic conditions and bandwidth fluctuations.**

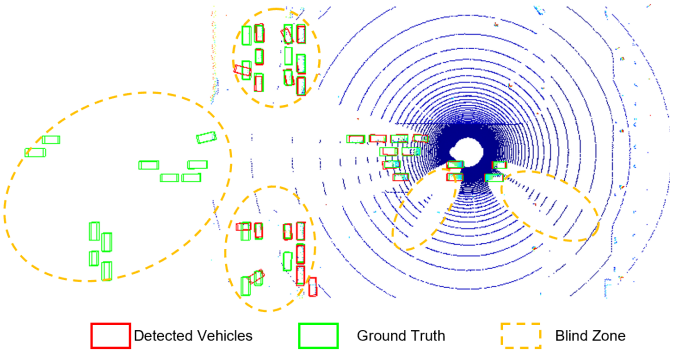

Fig. 1. A two-dimensional illustration of LiDAR generated point cloud. The ego vehicle fails to detect objects in the blind zone and that are out of the detection range.

I. Introduction

The development of artificial intelligence and on-board sensors has fostered a new era of autonomous driving, with unprecedented level of automation [1]–[3]. Mounted with advanced wireless interface and various sensors, e.g., mmWave Radar and LiDAR, connected autonomous vehicles (CAVs) are enabled to perceive on-road environment via multi-modal visionary sensors, leading to a complementary and comprehensive perception from visible light, point cloud, infrared images, etc. Among them, point cloud data (PCD) contains additional depth information, compared with visible light images. Nonetheless, while on-board sensors provide essential perceptual inputs, their limited physical range and sensitivity to occlusion can compromise their performance. As shown in Fig. 1, even equipped with the most advanced LiDAR, a single ego vehicle may fail to detect other vehicles in the distance and manifest perceptual blind zones.

To address this limitation and improve driving safety, collaborative perception has emerged as a promising solution, drawing significant attention from both academia and industry. By leveraging complementary perceptual information from neighboring CAVs [4], collaborative perception promotes more holistic and high level awareness of environment perception. According to the form of shared information, conventional wisdom on collaborative perception can be broadly categorized into raw-level [4], [5], feature-level [6], [7], and semantic-level [8]. In specific, feature-level collaborative perception shares intermediate features extracted by deep neural networks, which is implicit and neglects the heterogeneous on-board deep neural networks across CAVs. Semantic-level collaborative perception only merges processed results from CAVs, which is efficient but can result in poor perception performance. While a considerable amount of research works concentrate on the two aforementioned stages, collaborative perception in raw-level stands out due to its intrinsic superiorities: on the one hand, it preserves all captured information and thus reserves the possibility of achieving the highest perception performance, on the other hand, thanks to the granularity of raw data, it enables semantic perception tasks, such as 3D object detection and 3D semantic segmentation.

Although collaborative perception at raw-level offers distinctive advantages, performing analytic tasks on PCD can be intensive in both computation and storage [9], [10]. In this context, edge computing technology becomes promising by leveraging additional computing resource on edge nodes (ENs), e.g., the roadside units [11]–[13]. Deployed in close proximity to the road, ENs are able to employ CAVs within its coverage for data collection, analyze the integrated data and broadcast the resulting

978-1-6654-9380-2/23/$31.00 © 2023 IEEE

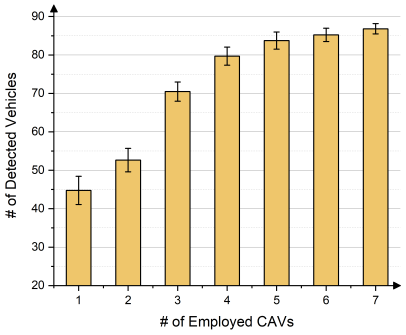

Fig. 2. The perceptual performance versus # of employed CAVs.

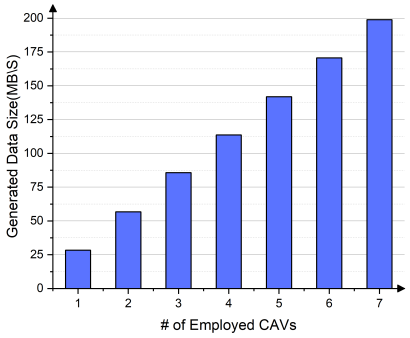

Fig. 3. The generated PCD size versus # of employed CAVs.

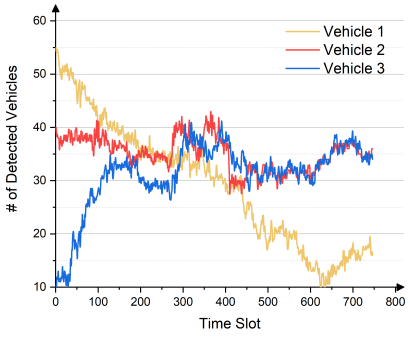

Fig. 4. The perceptual capability of individual CAVs across one period.

collaborative perception messages (CPMs) in a sequential manner. While edge-coordinated vehicle-to-infrastructure (V2I) communication can provide more bandwidth resources than current DSRC-based vehicle-to-vehicle (V2V) communications [14], the substantial transmission of PCD from CAVs to EN still poses challenge to existing V2I communication capability, according to Intel [15], each CAV can generate approximately 4000 GB of data per day, equivalent to that generated by 3000 mobile users. Therefore, a bandwidth-efficient collaborative perception framework with the aid of edge computing is desirable.

In this paper, we first conduct extensive experiments and reveal that, while ENs can perceive the environment from a broader perspective by integrating collected information from CAVs within their coverage range, the mobility of CAVs makes their individual PCD value to the edge-coordinated perception dynamic and unpredictable. Furthermore, the information redundancy generated by adjacent CAVs can result in submodular integrated contributions, i.e., the utility of integrating information from different CAVs at the ENs subjects to diminishing returns, a CAV's perceived PCD is more valuable to the EN when fewer CAVs are employed. Motivated by such insights, we propose a framework that enables EN to adaptively employ CAVs in order to maximize perceptual performance in edge-coordinated collaborative perception. The contribution of this paper can be summarized in threefold:

- We explore the correlation between collaborative perceptual performance and the number of employed CAVs. We also analyze the influence of vehicular mobility on the PCD value of individual CAVs. The experimental results reveal the submodularity of collaborative perceptual performance, as well as the heterogeneity and dynamic of individual CAVs' PCD value.
- We formulate an optimization problem to maximize the total utility function of the EN with a latency constraint. Accordingly, a CAV employment algorithm is proposed, which evaluates the marginal gain of each CAV candidate and selects the optimal subset.
- Extensive experiments on an open benchmark dataset are carried out for performance evaluation. The proposed algorithm shows superiority over other baselines under dynamic traffic conditions and bandwidth fluctuations.

The remainder of this manuscript is organized as follows. Section II presents the motivating experiments and corresponding observations. Section III introduces the system model and algorithm design, followed by performance evaluations in Section IV. Finally, the paper is concluded with future work in Section V.

II. Motivation

Leveraging the technology of edge computing, the deployed ENs can provide stable and reliable computation and wireless communication resources to CAVs. However, as collaboration perception at raw-level is inherently bandwidth-intensive and the bandwidth resources on ENs are limited. It is prohibitive for the ENs to employ all CAVs for point cloud transmission. Motivated by the envision to utilize bandwidth efficiently, we investigate the relationship between the perceptual capability, bandwidth consumption, and the number of employed cooperative CAVs. Additionally, the dynamic and uncertainty of perceptual capability on individual CAV is analyzed. In particular, we conduct experiments on OPV2V [16], a large-scale simulated dataset which generates PCD at frequency of 10 Hz. 3D object detection is designed to be the analytics tasks in those experiments, considering it is the fundamental task in state-of-the-art autonomous driving system. Experimental results are presented as follows.

Firstly, we apply PointPillars [17] as 3D detection models and gradually increase the quantity of cooperative CAVs to 7. Fig. 2 illustrates the average number of detected vehicles in two urban scenes, each with around 700 frames. As adjacent CAVs share overlapping perceptual area, there is information redundancy in their generated PCD. Therefore, the marginal gain brought by each CAV decreases as the quantity of CAV increases, indicating that perceptual performance exhibits the characteristic of submodularity. However, as depicted in Fig. 3, the generated

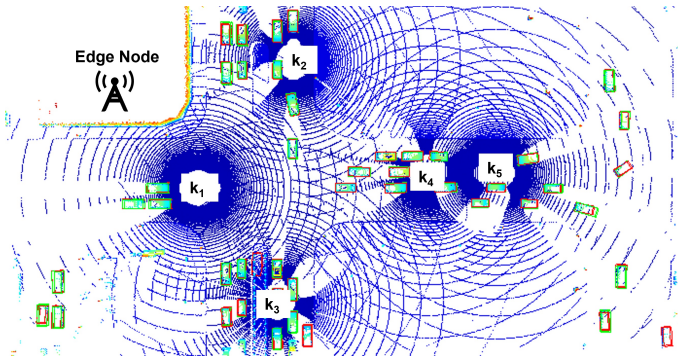

Fig. 5. Illustration for edge-coordinate collaboration perception.

PCD size increases linearly with the number of CAVs, implying that employing additional CAV for data upload when more CAVs have been employed incurs higher bandwidth and computation costs, yet, low perception gain.

Secondly, we explore the impact of time and heterogeneity on a single CAV's perceptual capability. Specifically, we record the number of vehicles detected by different CAVs without collaboration within a single period, as shown in Fig. 4. It can be observed that the perceptual capability of a CAV is dynamic and differs from other CAVs. This indicates that, when selecting a CAV with the most performance enhancement, the optimal candidate set is time-varying. Consequently, it is necessary to develop a framework which can maximize the perceptual capability of the system while jointly take the submodularity and dynamic characteristic into consideration.

III. Problem Formulation and Algorithm Design

In this paper, we consider a typical urban scenario in which a set of CAVs are driving across an intersection. This is because more than 50% of injury or fatal crashes are reported to occur at or near an intersection[1]. To support edge-assisted collaborative perception, we assume a set of CAV candidates $\mathcal{C} = \{1, ..., C\}$ are willing to contribute their PCD. Without loss of generality, we assume all CAVs can be accurately positioned and capture point cloud frames synchronously at the very beginning of each time slot. An EN with network management functionality is deployed to provide computation and communication resources. During each discrete time slot t, a set $\mathcal{K}_t \subseteq \mathcal{C}$ of CAVs are employed for collaborative perception. Specifically, they transmit the captured PCD to the EN, where data fusion and analytic tasks are performed sequentially. After processing, the EN broadcasts the generating CPMs to all CAVs before they make next driving decisions.

An illustration of the considered scenario is presented in Fig. 5. In this case, five CAVs $\mathcal{K}_t = \{k_1, k_2, k_3, k_4, k_5\}$ are employed by the EN to transmit their PCD. It can be observed that the blind zones and miss detection in Fig. 1 are solved through edge-coordinate collaboration perception.

[1]https://highways.dot.gov/research/research-programs/safety/intersection-safety

A. Computation Model

Analytic tasks on PCD are usually computation-intensive [17], [18], as a result, when designing collaborative perception system, the computational latency should be considered. Denote the computation capability of EN by f, the EN computation latency L_t^q is expressed as:

$$L_t^q(\mathcal{K}_t) = \frac{D(K_t)}{f}, \tag{1}$$

where $D(\cdot)$ and $K_t = |\mathcal{K}_t|$ is the required computation resource function for analytic tasks and number of employed CAVs, respectively. It is worth noting that $D(K_t)$ increases linearly since the computing complexity of deep neural networks is proportional to the range of input PCD.

B. Communication Model

Assume the OFDMA-based C-V2X technology based on cellular base stations is adopted for vehicular communications [19]. According to the access in C-V2X, the employed CAVs share wireless resources provided by the EN. Consequently, in each time slot t, the total V2I bandwidth W_t is divided into M orthogonal channels, each with W_t/M Hz.

Due to the mobility of CAVs, the wireless channel quality between CAV i and the EN exhibits variability. Given the coordinate of the EN and CAV i as (X_t, Y_t), $(X_{i,t}, Y_{i,t})$, respectively, the distance between CAV i and EN is:

$$d_{i,t} = \sqrt{(X_t - X_{i,t})^2 + (Y_t - Y_{i,t})^2}. \tag{2}$$

According to the Shannon formula, the V2I uplink data rate can be expressed as:

$$H_{i,t}(\mathcal{K}_t) = \begin{cases} \frac{W_t}{K_t}\log_2(1 + \frac{p_{i,t}g_{i,t}}{\frac{W_t}{K_t}n_0}), & K_t \leq M \\ \frac{W_t}{K_t}\log_2(1 + \frac{p_{i,t}g_{i,t}}{\frac{W_t}{M}n_0}), & K_t > M, \end{cases} \tag{3}$$

where the sub-channel path loss $g_{i,t} = \epsilon(d_{i,t})^{-\varrho}h_{i,t}$, ϵ, ϱ, h denotes path loss coefficient, path loss exponent, channel gain [19], respectively. n_0 is the noise density and $p_{i,t}$ is the transmitting power. Consequently, for CAV i, denoting its perceptual data volume by $S_{i,t}$, the PCD transmission latency $L_{i,t}^r$ is:

$$L_{i,t}^r(\mathcal{K}_t) = \frac{S_{i,t}}{H_{i,t}(\mathcal{K}_t)}. \tag{4}$$

Since the CPMs are on semantic-level and much smaller in size, compared with PCD, we consider the latency caused by broadcasting CPMs is negligible. At raw-level collaborative perception, only when all CAVs $i \in \mathcal{K}$ finishing transmitting,

the EN can start data fusion and analysis, therefore, the overall communication latency L_t^r can be represented as:

$$L_t^r(\mathcal{K}_t) = \max_{i \in \mathcal{K}_t} L_{i,t}^r(\mathcal{K}_t). \tag{5}$$

After receiving captured data from employed CAVs, the EN projects point clouds from the sender's view to its own view, which can be achieved by a coordinate transformation matrix, then the point clouds from multiple CAVs can be fused in EN.

C. Objective Function and Algorithm Design

As stated in Section II, when employing CAVs for collaborative perception, overlapping sensing areas among neighboring CAVs lead to information redundancy in their contributed PCD. In other words, the perception capability enhanced by collaborative CAVs is characterized in diminishing return: the value of a CAV candidate to EN is higher if fewer CAVs have been employed, which makes the reward of the EN submodular. In specific, we denote the utility of EN by U, for two CAV sets $\mathcal{A}$ and $\mathcal{B}$, suppose $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{C}$ and $C_k \in \mathcal{C}/\mathcal{B}$, define $U(\mathcal{A}|\mathcal{B}) \triangleq U(\mathcal{A} \cup \mathcal{B}) - U(\mathcal{B})$ as a utility set function which measures the marginal gain of employing set $\mathcal{B}$ on the basis of set $\mathcal{A}$. Then the submodularity can be described as:

$$U(C_k|\mathcal{A}) \geq U(C_k|\mathcal{B}). \tag{6}$$

Eq. (6) indicates that, given the EN have already employed a subset $\mathcal{A} \in \mathcal{C}$, the marginal benefit provided by employing a CAV does not increase as it employ more CAVs. In each time slot t, the EN needs to select a subset $\mathcal{K}_t \subseteq \mathcal{C}$ of CAVs to upload perceptual data. The objective function aims to maximize the utility of EN during a period of time $\mathcal{T} = 1, ..., T$, which is defined as

$$\max \quad \sum_{t=1}^{T} U(\mathcal{K}_t), \tag{7}$$

$$\textit{s.t.} \quad L_t^r(\mathcal{K}_t) + L_t^q(\mathcal{K}_t) \leq \tau, \ \forall t \in \mathcal{T}. \tag{8}$$

In Eq. (7), there is a trade-off between utility function and overall latency. Due to the monotonicity of submodular functions, the more CAVs employed by EN, the higher utility will be achieved. Nonetheless, the wireless channel will be congested, which leads to long transmission delay, correspondingly, the EN computation latency is also prolonged.

Intrinsically, Eq. (7) aims to maximizing a submodular function with cardinality constraint that is proven to be NP-hard [20]. We take advantage of greedy solution, which achieves promising performance with lower bound $(1 - 1/e)$ of optimal. More specifically, the EN keeps recording the historical performance from candidate set $\mathcal{C}$. In each time slot t, the EN sequentially employs CAVs according to their marginal gain (line 4) as long as Eq. (8) holds (line 5-9), ensuring itself with the highest utility enhancement. The details of the proposed Greedy CAV Employment Algorithm is shown in Algorithm 1.

Algorithm 1: Greedy CAV Employment Algorithm

Data: $\mathcal{C}$
Result: $\{\mathcal{K}_t\}, \forall t \in \mathcal{T}$
1 **for** $t \in \mathcal{T}$ **do**
2 Initialization: $\mathcal{K}_t, \mathcal{O}_t \leftarrow \emptyset$, $\mathcal{Z}_t \leftarrow \mathcal{C}$
3 **while** $L_t^r(\mathcal{K}_t) + L_t^q(\mathcal{K}_t) \leq \tau$ **do**
4 $\mathcal{O}_t \leftarrow \mathcal{K}_t \cup \arg\max_{i \in \mathcal{Z}} U_t(\mathcal{K}_t \cup \{i\})$
5 **if** $L_t^r(\mathcal{O}_t) + L_t^q(\mathcal{O}_t) \leq \tau$ **then**
6 $\mathcal{K}_t \leftarrow \mathcal{O}_t$
7 $\mathcal{Z} \leftarrow \mathcal{C} \setminus \mathcal{K}_t$
8 $L_t^r \leftarrow L_t^r(\mathcal{K}_t), L_t^q \leftarrow L_t^q(\mathcal{K}_t)$
9 **end**
10 **end**
11 Return: $\mathcal{K}_t$
12 **end**

IV. Performance Evaluation

This section provides a comparison of the proposed algorithm with several benchmarks under different network conditions. The performance of the algorithms is evaluated in terms of latency, followed by detailed description and analysis.

A. Parameter Setting

All experiments are conducted on three urban scenarios from OPV2V dataset, each with 750 frames and time duration $\tau = 100$ ms. The number of CAVs in each scene and corresponding vehicle speed fall in $[60, 100]$ and $[0, 20]$ km/h, respectively. For EN, there are always 7 CAV candidates to employ, which means $|\mathcal{C}| = 7$. Based on extensive experimental results, the relationship between computation resource consumption $D(\cdot)$ as well as data size $S(\cdot)$ and the employed set $\mathcal{K}_t$ are fitted by two linear functions: $S(\mathcal{K}_t) = 60 \times |\mathcal{K}_t|$ GFLOPS and $D(\mathcal{K}_t) = 0.028 \times |\mathcal{K}_t|$ MHz. Meanwhile, the EN is enabled with the computation capability of 3 TFLOPS.

Following works in [19], the channel gains of CAVs are exponentially distributed, i.e., $g_{i,t} \sim \text{Exp}(1)$. Accordingly, CAVs communicate with the EN with transmission power $p_{i,t} = 100$ mW and share the V2I bandwidth which varies from 8 MHz to 20 MHz. Our proposal is compared with the following four benchmarks.

- **Bandwidth-Thrift**: From the perspective of bandwidth-thrift, only one CAV with the most utility gain from set $\mathcal{C}$ will be employed. This methods is expected to utilize the least bandwidth.
- **Unlimited**: In each time slot, all of the CAVs in set $\mathcal{C}$ are employed to transmit their PCD. This benchmark does not take latency constraint into consideration and is expected to achieve the highest perception performance.

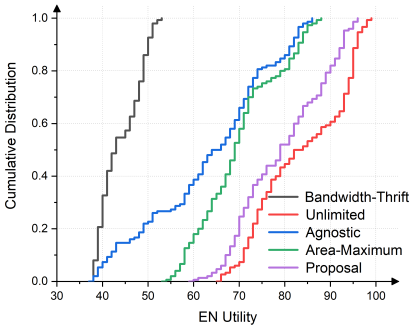

Fig. 6. Comparison of the cumulative distribution among different methods in normal channel condition.

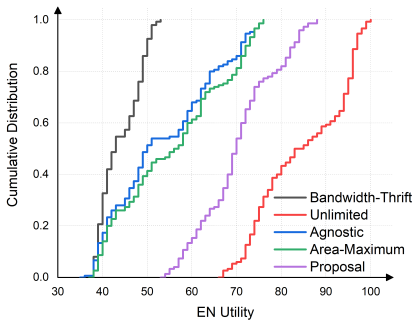

Fig. 7. Comparison of the cumulative distribution among different methods in poor channel condition.

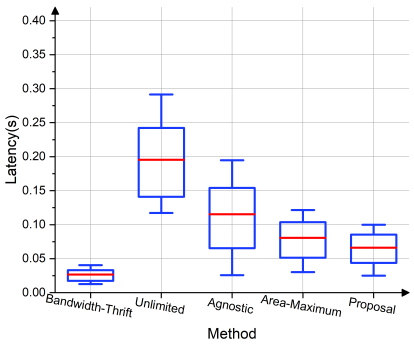

Fig. 8. Comparison among different methods on average latency

- **Agnostic**: In each time slot, the EN employs CAV candidates from set $\mathcal{C}$ in a round-robin manner, which shows randomness.
- **Area-Maximum**: In each time slot, the EN employs CAV candidates greedily based on maximization of sensing range instead of utility function. The LiDAR sensing range of each CAV is set as a rectangle with sides measuring [$140m$, $80m$].

B. Simulation Results

We first measure the utility of EN over time period $\mathcal{T}$ under normal V2I channel condition and analyze the cumulative distributions of all methods. Next, we consider a constrained channel condition in which the maximum bandwidth is limited from 20 MHz to 10 MHz.

As depicted in Fig. 6, the worst and best results are obtained by **Bandwidth-Thrift** and **Unlimited**, respectively, which aligns with our expectations. Moreover, our proposed method outperforms other benchmarks such as **Agnostic** and **Area-Maximum**, achieving a 50% probability of detecting more than 80 vehicles. The results are consistent with those observed in Fig. 7, where all methods except **Unlimited** experience degraded performance due to poor channel conditions. Nevertheless, our proposal still outperforms them. Although our proposal performs slightly inferior to **Unlimited** under both conditions, it's important to note that **Unlimited** does not account for latency, which will be illustrated later.

As vehicles are with high mobility, the effectiveness of the generated CPM is highly sensitive to latency. In Fig. 8, the average latency of different methods is presented. It is evident that both **Unlimited** and **Agnostic** methods are far from satisfactory due to their high mean value and unstable performance on latency. The Bandwidth-Thrift method has the least latency, but it under-utilizes bandwidth. Notably, in contrast to **Unlimited**, **Agnostic** and **Area-Maximum**, our proposed method not only achieves the minimal latency but also confines it within 0.1 seconds in all cases. This confirms the robustness of our proposal under dynamic traffic conditions and network fluctuations, which is a crucial aspect of collaborative perception.

V. Conclusion

In this paper, we have addressed the problem of employing the most valuable vehicles for efficient on-road perception in edge-coordinated autonomous vehicles using PCD. Our experimental results have revealed two significant insights, directly leading to the formulation of objective function which maximizes the utility of EN while meeting latency constraint for real-time autonomous driving. The performance advantages of the designed CAV candidates employing algorithm have been verified through extensive experiments on an open benchmark dataset. The outcome of this paper can provide effective solution to autonomous driving systems. For the future work, we will consider the scenario with cross-edge coordination for collaborative perception at scale.

VI. Acknowledgement

This work is supported in part by the Natural Science Foundation of China under Grant 62001180, in part by the Young Elite Scientists Sponsorship Program by CAST under Grant 2022QNRC001, in part by the Natural Science Foundation of Hubei Province of China under Grant 2021CFB338, and in part by the Fundamental Research Funds for the Central Universities, HUST, under Grant 2021XXJS014.

References

[1] N. Lu, N. Cheng, N. Zhang, X. Shen, and J. W. Mark, "Connected vehicles: Solutions and challenges," *IEEE Internet Things J.*, vol. 1, no. 4, pp. 289–299, 2014.

[2] P. Yang *et al.*, "Edge coordinated query configuration for low-latency and accurate video analytics," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4855–4864, 2019.

[3] W. Wu, N. Chen, C. Zhou, M. Li, X. Shen, W. Zhuang, and X. Li, "Dynamic RAN slicing for service-oriented vehicular networks via constrained learning," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 7, pp. 2076–2089, 2021.

[4] Q. Chen, S. Tang, Q. Yang, and S. Fu, "Cooper: Cooperative perception for connected autonomous vehicles based on 3D point clouds," in *Proc. IEEE ICDCS,* 2020, pp. 514–524.

[5] E. Arnold *et al.*, "Cooperative perception for 3D object detection in driving scenarios using infrastructure sensors," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 3, pp. 1852–1864, 2020.
[6] Y. Hu, S. Fang, Z. Lei, Y. Zhong, and S. Chen, "Where2comm: Communication-efficient collaborative perception via spatial confidence maps," *arXiv:2209.12836*, 2022.
[7] T.-H. Wang *et al.*, "V2VNET: Vehicle-to-vehicle communication for joint perception and prediction," in *Proc. ECCV,* 2020, pp. 605–621.
[8] H. Yu *et al.*, "DAIR-V2X: A large-scale dataset for vehicle-infrastructure cooperative 3D object detection," in *Proc. IEEE/CVF CVPR,* 2022, pp. 21361–21370.
[9] C. Wang, P. Yang, J. Lin, W. Wu, and N. Zhang, "Object-based resolution selection for efficient edge-assisted multi-task video analytics," in *Proc. IEEE GLOBECOM,* 2022, pp. 5081–5086.
[10] S. Shi, J. Cui, Z. Jiang, Z. Yan, G. Xing, J. Niu, and Z. Ouyang, "VIPS: real-time perception fusion for infrastructure-assisted autonomous driving," in *Proc. ACM Mobicom,* 2022, pp. 133–146.
[11] W. Wu *et al.*, "Accuracy-guaranteed collaborative DNN inference in industrial IoT via deep reinforcement learning," *IEEE Trans. Ind. Informat.*, vol. 17, no. 7, pp. 4988–4998, 2020.
[12] F. Lyu *et al.*, "Online UAV scheduling towards throughput QoS guarantee for dynamic IoVs," in *Proc. IEEE ICC,* 2019, pp. 1–6.
[13] J. Chen, H. Wu, F. Lyu, P. Yang, Q. Li, and X. Shen, "Adaptive resource allocation for diverse safety message transmissions in vehicular networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 13482–13497, 2021.
[14] X. Zhang *et al.*, "EMP: Edge-assisted multi-vehicle perception," in *Proc. ACM Mobicom,* 2021, pp. 545–558.
[15] J. Zhang and K. B. Letaief, "Mobile edge intelligence and computing for the internet of vehicles," *Proc. IEEE*, vol. 108, no. 2, pp. 246–261, 2020.
[16] 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 *Proc. IEEE ICRA,* 2022, pp. 2583–2589.
[17] A. H. Lang *et al.*, "Pointpillars: Fast encoders for object detection from point clouds," in *Proc. IEEE/CVF CVPR,* 2019, pp. 12697–12705.
[18] P. Yang, J. Hou, L. Yu, W. Chen, and Y. Wu, "Edge-coordinated energy-efficient video analytics for digital twin in 6G," *China Communications*, vol. 20, no. 2, pp. 14–25, 2023.
[19] H. Jiang, X. Dai, Z. Xiao, and A. K. Iyengar, "Joint task offloading and resource allocation for energy-constrained mobile edge computing," *IEEE Trans. Mobile Comput.*, 2022. doi:10.1109/TMC.2022.3150432.
[20] A. Krause and D. Golovin, "Submodular function maximization," *Tractability*, vol. 3, pp. 71–104, 2014.