Hierarchical Edge AI for Cooperative Driving

Kiminobu Makino*, Hiroshi Miyata†, Toru Furusawa*, Atsushi Hayami†, Hisayoshi Mizuno‡

*InfoTech, Toyota Motor Corporation, Tokyo, Japan
†Tokyo Office, Techno-Accel Networks Corporation, Tokyo, Japan
‡Development Div., Kousokuya Corporation, Kanagawa, Japan

***Abstract*—This paper presents vehicle-to-vehicle (V2V) cooperation with hierarchical edge AI and we demonstrate it with real-vehicle experiments. We propose a Wi-Fi and MQTT communication framework, implement it, and evaluate it through platooning experiments on public roads. We confirm that V2V cooperative processing meets real-time requirements while reducing communication to tens of kbps with sub-10 ms latency. For the demo, we show it with real-vehicle experiments, highlighting the novelty of our approach, its real-time capability, and its potential to engage the audience.**

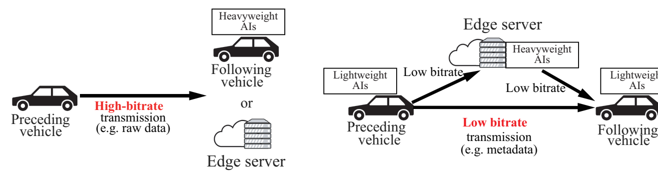
***Index Terms*—Hierarchical edge AI, Cooperative driving, V2V communications**

I. Introduction

This paper proposes vehicle-to-vehicle (V2V) cooperative systems leveraging hierarchical edge AI. Our contribution lies in designing and implementing a complete system that integrates the communication function, and in evaluating its communication and cooperative performance through real-vehicle experiments, based on hierarchical edge AI architecture as proposed in [1].

Recently, various driving assistance technologies—such as autonomous driving and remote driving—have been actively researched, developed, and deployed. However, many of these technologies rely heavily on the perception and decision-making capabilities of individual vehicles. To achieve a society with zero traffic accidents [2], it is essential to share information through V2V, vehicle-to-infrastructure, and vehicle-to-pedestrian. For such cooperative driving, information sharing among vehicles is indispensable [3]; however, available bandwidth for vehicle-to-something communications is limited. Achieving both scalability and real-time performance is challenging when transmitting large volumes of raw data, such as video. In hierarchical edge AI, it has been suggested that both communication and computation loads can be reduced by converting in-vehicle processing results into lightweight data, transmitting them, and completing the remaining processing on the receiving side [1]. However, little attention has been paid to comprehensive system design—including wireless communication and validation through real-vehicle experiments.

This paper proposes a communication framework based on Wi-Fi and the message queuing telemetry transport (MQTT) protocol. We implement the entire system, conduct real-vehicle experiments, and evaluate communication latency and feasibility of cooperative driving. Our results show that metadata transmission requires only tens of kbps and achieves sub-10 ms latency. In addition, the demonstration presents videos of the visualization UI from real-vehicle experiments, which illustrate the hierarchical edge AI.



(a) Cooperative drive w/o hierarchical edge AI (b) Cooperative drive with hierarchical edge AI

Fig. 1. System designs for V2V cooperation.

II. System Design

This section describes the design of the V2V cooperative driving system. For cooperative systems, it is essential to design the in-vehicle processing and communication functions.

Fig. 1 (a) shows the system design for V2V cooperative driving without hierarchical edge AI. Transmitting high-bitrate video directly causes bandwidth constraints and heavy processing loads.

By contrast, the hierarchical edge AI–based method, as shown in Fig. 1 (b), performs AI processing within the preceding vehicle and transmits the results as lightweight metadata. The following vehicle integrates the received metadata with its own information and performs the remaining AI processing. Edge servers integrate multi-vehicle data, reducing communication and offloading processing.

A. *Design for Preceding Vehicle Action Prediction*

This paper focuses on driving assistance for the following vehicle by predicting the actions of the preceding vehicle. Rather than transmitting the raw video captured by its in-vehicle camera to the following vehicle, the preceding vehicle first performs object detection using lightweight in-vehicle AI, and the detection results are then transmitted to the following vehicle. The following vehicle integrates the received data with its own object detection results and sequentially performs vehicle tracking and prediction, generating risk heatmap, and generating driving action. This allows the following vehicle to estimate the behavior of the preceding vehicle and use it to support driving assistance, such as route selection. Without hierarchical edge AI, the preceding vehicle must transmit the raw video to the following vehicle, which then performs all processing, including object detection, leading to high communication and processing loads. By contrast, using hierarchical edge AI enables real-time cooperation while reducing the communication load.

III. Proposed Method and Implementation

This section describes the proposed method and system implementation. Fig. 2 shows the overall system configuration.

A. *Proposed Method*

For our method, the object information detected by the Layer-1 AI in the preceding vehicle is converted into CSV metadata. Metadata contains fields such as frame number, obstacle class, confidence score, position, size, and ego-vehicle attributes. The metadata is transmitted to the following vehicle over V2V communication using the MQTT protocol [4].

B. *System Implementation*

This section describes the system implementation based on the proposed method. Both vehicles have a PC for hierarchical edge AI; the preceding vehicle also has a relay PC and Wi-Fi access point (AP, Wi-Fi 5, 2.4 GHz band).

First, the preceding vehicle captures video from the in-vehicle camera and performs object detection using the Layer-1 AI, converting the results into metadata as described in Sect. III-A. YOLO v5s [5], a lightweight model, is employed as the Layer-1 AI. Metadata is sent to the relay PC's MQTT broker and received by the following vehicle via MQTT over Wi-Fi. The received metadata is integrated with the GNSS receiver output, after which Layer-2 AI and subsequent processing are performed. The Layer-2 AI uses the Kalman filter [6] to track and predict the preceding vehicle. The Layer-3 AI generates a risk heatmap highlighting areas with a high likelihood of obstacles, using Gaussian functions [7]. The Layer-4 AI calculates steering actions toward low-risk areas based on the risk heatmap [1].

IV. Experimental Evaluation

We conducted platooning experiments with two vehicles on public roads to evaluate communication performance and cooperative driving with a processing rate of 200 ms per frame (5 fps).

Fig. 3 shows the experimental results: (a) CSV metadata size distribution and (b) latency. The average size was 0.76 kB per frame (∼31 kbps), far smaller than video transmission, which requires several Mbps. The average latency was 9.5 ms, well below the 200 ms cycle, satisfying real-time requirements. Although outliers over 50 ms were observed, caused by Wi-Fi, the median latency of 8.2 ms confirms that real-time requirements were satisfied for the vast majority of frames.

Fig. 4 shows a UI of the following vehicle in the experimental results. Although details will be explained in the demonstration, the results confirm that the proposed hierarchical edge AI processing operates in real time.

V. Conclusion

We presented a V2V cooperative driving system with hierarchical edge AI, confirmed its effectiveness through experiments, and plan to extend it to larger-scale scenarios.

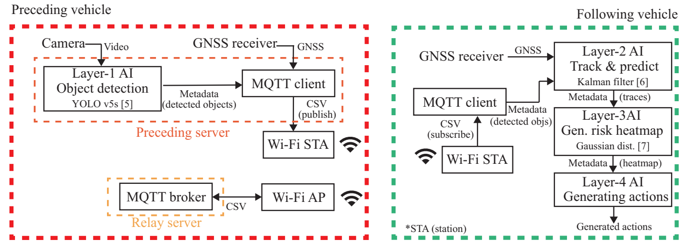
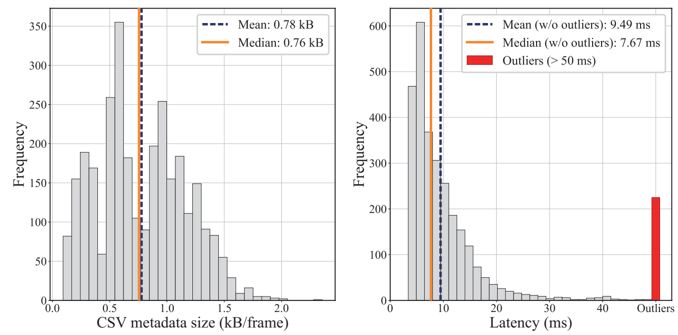


Fig. 2. Overall system configuration of the proposed method.



(a) Distribution of CSV metadata size (b) Distribution of latency

Fig. 3. Experimental results from real-vehicle field tests on public roads.

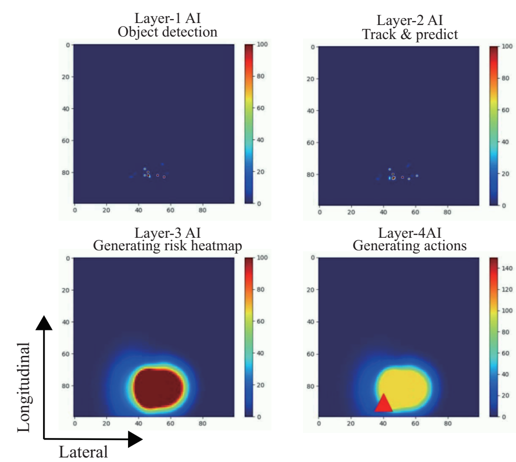


Fig. 4. Bird's-eye view of Layer-1-4 processing in the field experiment.

References

[1] H. Miyata, et. al, "Distributed Hierarchical AI Architecture for Meta-Described Local Dynamic Maps in Autonomous Driving," in Proc. *2024 IIAI-AAI*, pp. 595–600, Jul. 2024.

[2] Toyota Times, "NTT and Toyota Aim for Zero Traffic Accidents with AI-Powered Mobility Platform," Nov. 12, 2024. [Online]. Available: https://toyotatimes.jp/en/toyota_news/1066.html

[3] T. Huang et al., "Vehicle-to-Everything Cooperative Perception for Autonomous Driving," in Proc. *IEEE*, vol. 113, no. 5, pp. 443–477, May 2025.

[4] A. Banks and R. Gupta, "MQTT Version 3.1.1," OASIS Standard, Oct. 2014. [Online]. Available: https://docs.oasis-open.org/mqtt/mqtt/v3.1.1/os/mqtt-v3.1.1-os.html

[5] G. Jocher, et al., "YOLOv5: Real-Time Object Detection," Ultralytics, 2020. [Online]. Available: https://github.com/ultralytics/yolov5

[6] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, Mar. 1960.

[7] S. Thrun, et. al, *Probabilistic Robotics*. Cambridge, MA: MIT Press, 2005.