ADVANCING AUTONOMOUS VEHICLE PERCEPTION: COOPERATIVE DETECTION VIA V2V COMMUNICATION

Ha Vu Thi Thuy
Posts and Telecommunications Institute of Technology, Hanoi, Vietnam

Abstract—An Autonomous Vehicle (AV) is a form of vehicle that can run by itself without calling for direct human intervention. For AVs to properly function in real-world settings, they need to be equipped with sophisticated perception and situational awareness. These abilities will enable them to effectively handle high-stress scenarios, make intelligent decisions, and ensure safety for users at all times. Having said that, the perception capabilities of AVs, which rely on sensors like cameras, LiDAR, and radar, have inherent limitations in range and detection accuracy. For instance, an AV may fail to detect objects obscured by other obstacles, either a moving or stationary one.

Cooperative perception is thus a technology that can revolutionise the development of AVs, by allowing connected and autonomous vehicles (CAVs) to share information about detected objects via vehicle-to-vehicle (V2V) communication. This approach not only can improve the accuracy and range of CAV detection, but also significantly enhances their awareness of the surrounding environment.

Our research introduces a cooperative perception mechanism to improve the accuracy in detecting objects around the environment of AV. The proposed simulation framework provides a comprehensive environment for evaluating traffic models, vehicle models, communication models, and object detection models. Simulations performed in real mobile scenarios show that collaborative perception can improve object detection accuracy by up to 35% compared to independent detection methods.

Keywords: Connected and Autonomous Vehicles, Cooperative Perception, vehicle-to-vehicle communication.

I. INTRODUCTION

In recent years, advancements in vehicle technology have led to the development of features such as automatic parking, cruise control, and autopilot systems, all designed to assist drivers and reduce the likelihood of accidents.

These innovations in autonomous vehicles have significantly contributed to road safety, addressing the fact that 94% of accidents are caused by human error [1]. Beyond enhancing safety, autonomous vehicles offer the potential to improve mobility for many people, including those with physical limitations that prevent them from driving conventional vehicles.

Despite these benefits, a critical challenge in autonomous driving remains the accurate detection and interpretation of objects encountered on the road, such as vehicles, traffic lights, signs, pedestrians, and railroad crossings. Achieving high precision in identifying and responding to these objects is essential for ensuring safe and efficient autonomous navigation. For human-driven and autonomous vehicles, capturing sensor data from blind spots is vital for preventing collisions and avoiding deadlocks. However, the environmental perception capabilities of local on-board sensors are limited in both coverage and detection accuracy [2]. Objects that are distant or obscured by other road elements may go undetected or be inaccurately classified, posing a significant challenge to the reliability and safety of autonomous systems.

Cooperative perception is an emerging technology aimed at enhancing road safety by allowing connected vehicles to share their raw or processed sensor data with nearby vehicles via Vehicle-to-Vehicle (V2V) communication. Connected and Automated Vehicles (CAVs) utilize V2V communications to augment the capabilities of their onboard sensors, thereby improving both safety and driving performance. Through the exchange of sensor information among CAVs, V2V communications help mitigate the limitations of individual sensor systems. By sharing this information, CAVs can extend their field of view beyond the range of their sensors and enhance detection accuracy. This process, also known as collective perception or cooperative sensing, has been standardized by the European Telecommunications Standards Institute (ETSI) [3].

To facilitate collaborative object detection, some studies have proposed and evaluated the exchange of raw sensor data between vehicles [4]. However, sharing raw data requires high bandwidth, compromising the system's scalability. Additionally, it increases system complexity and the computational load needed to process the large volumes of data received from multiple CAVs. Consequently, most research to date has focused on exchanging processed information about detected objects,

Contact author: Ha Vu Thi Thuy
Email: havt@ptit.edu.vn
Manuscript received: 8/2024, revised: 9/2024, accepted: 10/2024.

such as the locations, sizes, and classifications of critical objects like cars, pedestrians, and cyclists in 3D space, which enables more efficient and effective cooperative sensing.

In this paper, we present an approach that leverages cooperative perception to enhance 3D object detection and segmentation in urban environments. To facilitate the exchange of detected information, Connected and Automated Vehicles (CAVs) use Collective Perception Messages (CPMs), as defined by the European Telecommunications Standards Institute (ETSI) [3].

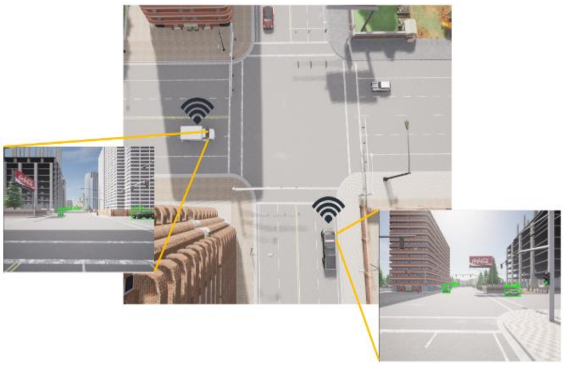

Figure 1. Detected data sharing in Cooperative Perception.

These message generation rules, which will be discussed in detail later, specify that a vehicle must send a CPM in three scenarios: at regular intervals, upon detecting a new object, or when an already detected object undergoes a significant update in terms of position or velocity. To design and develop a simulation environment that supports cooperative perception, we integrated multiple open-source software components, including a traffic model, vehicle model, communication model, and object classification model. This extended framework provides a foundation for further research into optimizing the system, particularly in addressing challenges related to communication efficiency, computational load, and perception accuracy.

The contributions of this paper are as follows:

1) We investigate various object detection methods in autonomous vehicles and how cooperative perception addresses their limitations.
2) We design and develop a simulation environment integrating multiple open software and tools to assess our scheme in a repeatable manner.
3) We evaluate scenarios in a simulated environment and demonstrate the effectiveness of cooperative perception in object detection.

II. PRELIMINARIES

A. Object detection in Autonomous Vehicles (AVs)

Object detection is a crucial task for autonomous vehicles, involving the identification and classification of objects in the vehicle's environment, such as pedestrians, vehicles, traffic signs, and obstacles. Recent advances in Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) have significantly enhanced these technologies. The object detection process typically includes several key steps:

- **Bounding Boxes**: The algorithm detects objects and encloses them within rectangular boxes on the image plane or 3D bounding boxes that define the object's dimensions and position.
- **Classification**: Each detected object is categorized into classes such as cars, pedestrians, or traffic signs.
- **Localization**: Specifies the position of detected objects using x, y, and z coordinates, providing information about the object's location, size, and orientation relative to the vehicle.

Applications of object detection in autonomous vehicles include traffic sign detection, lane line detection, obstacle avoidance, and spatial awareness.

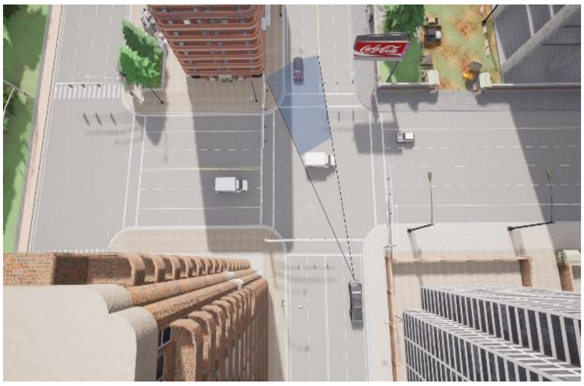

Figure 2. Occlusion issue.

There are two main categories of object detection algorithms: one-stage detectors and two-stage detectors. One-stage detectors perform object detection in a single step without requiring a preliminary stage to identify potential object regions. This makes them well-suited for real-time applications due to their shorter processing time. Notable one-stage detection algorithms include YOLO, SSD, RetinaNet, YOLOv3, YOLOv4, and YOLOR [5].

Two-stage detectors are more complex and consist of two parts: the first stage generates regions of interest (RoI), and the second stage performs regional classification and provides a precise description of the object's location based on the RoI. Prominent two-stage algorithms include Faster R-CNN, R-FCN, FPN, and Cascade R-CNN [6].

B. Cooperative Perception in Connected Autonomous Vehicles (CAVs)

A Connected Autonomous Vehicle (CAV) is a complex system comprising several components, including (i) a sensor-based perception system, (ii) a wireless communication interface, (iii) a map database, (iv) a navigation system, (v) an autonomous vehicle controller, and (vi) a localization component [7], [8]. This work focuses on the sensor-based perception system and the wireless communication interface.

Firstly, the sensor-based perception system integrates multiple sensors, such as cameras, LiDAR, radar, and GPS, to enable the vehicle to perceive its surrounding environment. It includes a sensor fusion component that combines data from these sensors to enhance accuracy and consistency. Additionally, advanced algorithms, AI, and machine learning are employed to interpret the observed data, ensuring safe navigation and operation on the roads.

Secondly, the wireless interface enables the system to transmit and receive information using mechanisms such as 4G/5G cellular interfaces, satellite communications, or embedded infrastructure. We assume that each connected vehicle is equipped with an On-Board Unit (OBU) that supports Dedicated Short-Range Communications (DSRC), Wireless Access in a Vehicular Environment (WAVE) protocol stack [9], and/or Cellular V2X (C-V2X).

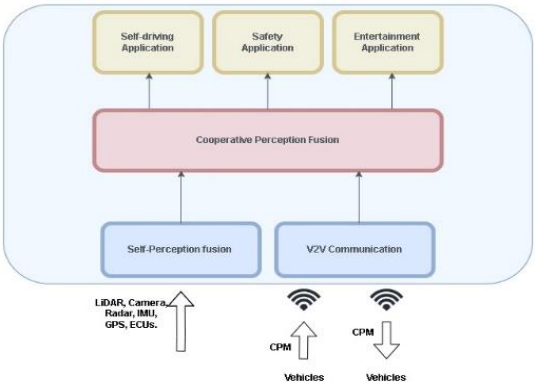

Figure 3. Model of sensor fusion.

Given that autonomous driving is both life-critical and safety-critical, the cooperative protocol must guarantee safety under practical conditions, ensuring that packet losses or delays do not compromise vehicle safety.

III. COOPERATIVE PERCEPTION PLATFORM FOR CAVS

A. Our System Model

Our Connected and Autonomous Vehicle (CAV) system is depicted in Fig. 3. In this model, each vehicle is equipped with three cameras for capturing Left, Front, and Right views. Vehicles locally integrate object detection information from multiple on-board sensors. Additionally, our system globally integrates cooperative perception data received through V2V communication networks from other vehicles, utilizing Cooperative Perception Messages (CPM). CPM messages include details such as the relative positions, orientation, type, and ID of detected objects.

The CPM generation rules in this study adhere to the ETSI ITS specifications [3]. Vehicles are required to evaluate each $T_{G}enCpm$ interval to determine if a new CPM should be generated. $T_{G}enCpm$ should be set between 100 ms and 1000 ms and can be adjusted by the Dynamic Channel Control (DCC) based on channel load. A new CPM is generated if a new object is detected, or if any of the following conditions are met:

1) The absolute difference (ΔP) between the current position of the object and its previous recorded position is greater than 4 meters.
2) The absolute difference (ΔS) between the current speed of the object and its previous recorded speed is greater than 0.5 meters per second, and the time difference (ΔT) between the current time and the last recorded time for the object is greater than 1 second.
3) A vehicle will include all newly detected objects and those that meet at least one of the aforementioned conditions (i.e., $\Delta P > 4$ meters, $\Delta S > 0.5$ meters per second, or $\Delta T > 1$ second) in a new CPM. Vehicles will generate a CPM every second regardless of whether any detected objects meet the specified conditions. Information about on-board sensors is included in the CPM only once per second.

B. *Cooperative perception platform for CAVs*

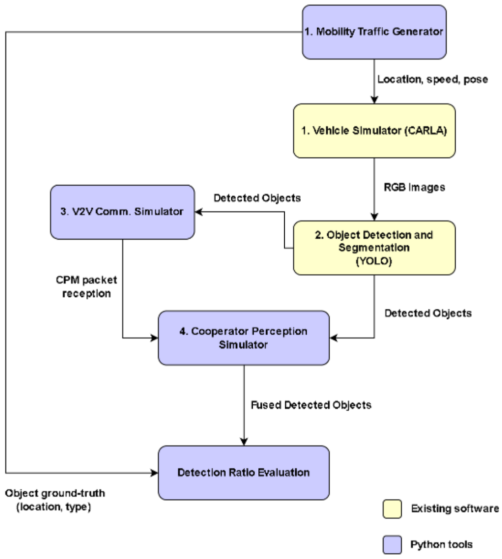

Figure 4. Software components in Cooperative perception platform for CAVs.

The software components in the collaborative cognitive platform for CAVs are detailed in Fig. 4.

1) ***CARLA Vehicle and Traffic Simulator***: CARLA (Car Learning to Act) [10] is an open-source simulator designed for autonomous driving research. It provides a high-fidelity environment for testing and developing autonomous driving systems. CARLA supports a wide range of vehicle and traffic scenarios, featuring various urban and rural environments with detailed buildings, roads, and vegetation to mimic real-world conditions. The simulator supports a range of sensors, including cameras, LiDAR, radar, and GPS, which can be attached to vehicles to collect data and evaluate sensor fusion algorithms and perception systems. CARLA simulates a variety of traffic participants, including vehicles,

pedestrians, and cyclists, which will be the targets of object detection in this study.

2) ***Object detection and segmentation module:*** For object detection, our platform uses YOLOv8 [11], an advanced version of the YOLO (You Only Look Once) series, which is a popular family of models designed for object detection tasks. YOLO models are distinguished by their ability to perform object detection in a single forward pass through the network, providing high-speed and accurate results. Since we are targeting autonomous driving tasks, objects other than vehicles, pedestrians, and cyclists will be filtered out.

3) ***V2V Communication Simulator:***
To simulate vehicular communications, we have developed a Python-based tool named V2V Communication Simulator. The V2V Communication Simulator provides parameters for communication range and interference distance in vehicular communications. In its simplified version, we exclude the impact of packet loss caused by network congestion.

4) **Cooperative Perception Simulator**:
First, CPM messages are generated by each CAV as described in our system model and broadcast to nearby neighbors. Second, we have developed an object fusion module that integrates detection information from onboard sensors with CPM data based on object IDs and their association with objects from previous frames. In the future, we plan to explore more advanced fusion methods to optimize the system in terms of communication and computational efficiency.

Figure 5. Town05 Map in Carla [10].

Figure 6. Images from the right, front, and left cameras.

IV. EXPERIMENTAL SETUP AND RESULTS

A. *Simulation Scenario*

To simulate vehicle mobility, we use the Carla Town05 map (Fig. 5), which features a squared-grid town layout with cross junctions and a bridge. This map includes multiple lanes in each direction and is characterized by numerous dual-lane urban roads intersecting at large junctions. These junctions provide access to a raised highway that forms a ring road around the town.

For evaluation, each vehicle is equipped with three RGB cameras—right, front, and left—calibrated to cover a 150-degree field of view. Example images from the three cameras are shown in Fig. 6. All connected vehicles have uniform dimensions: 1.76 meters in width, 4.54 meters in length, and 1.47 meters in height. The simulation scenario includes various vehicle types and objects of different sizes, including buses, trucks, cyclists, and pedestrians. Only vehicles are equipped with network interfaces, while bicycles are not connected. Object occlusion may occur due to the blocking views of buildings, cars, buses, and trees.

We evaluate the scenario 100 times with varying numbers of Connected and Autonomous Vehicles (CAVs). The evaluation involves comparing the number of correct detections with and without cooperative perception.

For V2V communication, we use the following equation for the Probability of Successful Packet Reception in wireless communication:

$$PSR = exp\left(-\lambda.\frac{S}{B}.d\right) \quad (1)$$

where:

- PSR is the Probability of Successful Reception, indicating the chance that a packet is received correctly.
- λ is the number of CPM (Cooperative Perception Messages) messages sent within a time window and the interference radius. This value is dynamic and depends on vehicle intensity and the CPM generation rule.
- S represents the Packet Size, measured in bits. Larger packets are more prone to errors, affecting the probability of successful reception.

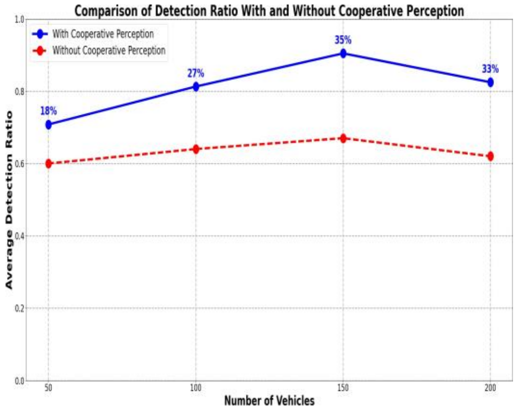

Figure 7. Average object detection ratio.

- *B* denotes the Bandwidth, measured in bits per second (bps). It defines the channel capacity and the amount of data that can be transmitted per unit time. Higher bandwidth generally improves the reception probability.
- *d* is the distance between the transmitter and receiver, measured in meters. Increased distance leads to greater signal attenuation, which can lower the probability of successful packet reception.

B. *Simulation Results*

We present the average detection ratio to evaluate the system's effectiveness. If all vehicles on the road within a 100-meter range are correctly detected by a CAV, the ratio is 100%. Intuitively, objects may not be detected if they are outside the camera's field of view, occluded, or at a far distance. Detections shared from other CAVs may improve the ratio; however, if the V2V network becomes congested (with low PSR), many shared detections may fail.

Fig. 7 shows that the object detection success rate substantially improves with the use of cooperative perception. Specifically, with 50 CAVs, the success rate increases by 18%. This enhancement becomes more pronounced with higher vehicle numbers, with a 27% increase at 100 CAVs and a 35% increase at 150 CAVs. The improvement largely stems from cooperative perception's ability to address visibility issues caused by physical obstructions such as buildings at intersections and vehicles stopped at traffic lights. Cooperative perception enables vehicles to share detection information, effectively mitigating the blind spots created by these obstructions and providing a more comprehensive view of the environment. However, when the number of vehicles reaches 200, the success rate levels off and experiences a slight decline. This drop is due to network congestion, which hampers the efficiency of data sharing and negatively impacts overall detection performance.

In addition to object detection, we evaluate the impact of cooperative perception on instance segmentation, which involves delineating and classifying each object instance with pixel-level precision. Instance segmentation is more complex and computationally intensive than object detection, as it requires detailed pixel maps for each identified object, which increases the data transfer workload. Our results, shown in Fig.8, demonstrate that cooperative perception significantly enhances instance segmentation performance by enabling vehicles to share detailed segmentation data. This shared data helps create a more comprehensive and accurate environmental model, which is crucial for tasks such as precise object tracking and scene understanding in Autonomous Vehicle.

However, despite these benefits, the success rate for instance segmentation is impacted by network congestion more severely than object detection. We see that while cooperative perception initially improves instance segmentation performance, the advantage diminishes sooner—specifically when the number of vehicles reaches around 100.

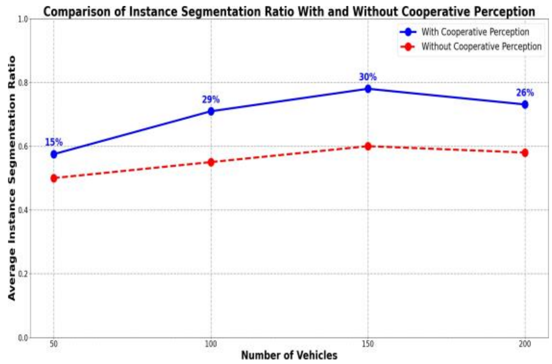

Figure 8. Average instance segmentation ratio

This is because the network becomes congested faster due to the high volume of data required for detailed pixel maps. The increased data transfer demands lead to reduced efficiency in network communication, resulting in a noticeable decline in instance segmentation success rates earlier than what is observed with object detection. Thus, while instance segmentation offers substantial benefits for enhancing the accuracy of Autonomous Vehicle systems, its effectiveness is limited by network capacity constraints, highlighting the need for optimized data management and network solutions in cooperative perception systems.

V. CONCLUSION

The benefits of cooperative perception in autonomous vehicle driving have been evaluated in this study, along with its effects on object detection tasks. Also, we built a platform that combines many open-source software tools for modeling perception, communication, and traffic in automobiles. Scalability is provided by this platform for further algorithmic improvements meant to improve the system as a whole.

Our analysis showed that shared perception data may successfully reduce occlusion problems, especially in towns and cities. Results also emphasized the necessity of efficient network congestion control and a number of other elements that can affect system performance, including effective data creation and selection. These results highlight the significance of ongoing cooperative perception research and development to enhance autonomous driving capabilities.

ACKNOWLEDGEMENTS

I would like to express my gratitude to Mr. Duc Nguyen Hong from the Self-driving car company ZMP Inc. Japan for his valuable advice and guidance, which helped me shape ideas and build the foundational model for the simulation in this work

REFERENCES

[1] Singh, Santokh. Critical reasons for crashes investigated in the National Motor Vehicle Crash Causation Survey. No. DOT HS 812 115. 2015.

[2] T. Higuchi, M. Giordani, A. Zanella, M. Zorzi, and O. Altintas, "Value- anticipating v2v communications for cooperative perception," in *2019 IEEE Intelligent Vehicles Symposium (IV)*, pp. 2087–2092, IEEE, 2019.

[3] "Intelligent transport systems (its); vehicular communications; basic set of applications; part 2: Specification of cooperative awareness basic service," Standard ETSI EN 302 637-2 V1.3.2, European Telecommunications Standards Institute (ETSI), Sophia Antipolis, France, 2014.

[4] R. Fukatsu and K. Sakaguchi, "Automated driving with cooperative perception using millimeter-wave v2v communications for safe overtaking," *Sensors*, vol. 21, no. 8, p. 2659, 2021.

[5] "Object detection in 2022: The definitive guide. available online: https://viso.ai/deep-learning/object-detection/.,"

[6] Z. R. . W. X. Du, L., "Overview of two-stage object detection algorithms," *Journal of Physics: Conference Series (Vol. 1544, No. 1, p. 012033).*, 2020, May.

[7] M. Liu, Q. Yang, J. Liu, X. Zhang, and Z. Wei, "A survey on connected and autonomous vehicles: Technologies, development, and challenges," *IEEE Access*, vol. 8, pp. 12345–12360, 2020.

[8] J. Zhang, H. Li, B. Yu, Q. Zhang, and S. Wu, "Connected and autonomous vehicles: A review and a roadmap," *IEEE Transactions on Intelligent Vehicles*, vol. 6, no. 2, pp. 292–307, 2021.

[9] F. B. Bastani, H. G. Wang, and S. Lee, "A comprehensive review of communication technologies for connected and autonomous vehicles," *IEEE Access*, vol. 9, pp. 56450–56475, 2021.

[10] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Proceedings of the 1st Annual Conference on Robot Learning (CoRL)*, pp. 1–16, 2017.

[11] A. G. e. a. Glenn Jocher, Ayush Chaurasia, "Yolo by ultralytics: Yolov8." https://github.com/ultralytics/ultralytics, 2023. GitHub repository.

NÂNG CAO NHẬN THỨC CỦA XE TỰ HÀNH: PHÁT HIỆN HỢP TÁC QUA GIAO TIẾP V2V

Tóm tắt: Xe tự hành (Autonomous Vehicle - AV) là loại phương tiện có khả năng di chuyển mà không cần sự can thiệp trực tiếp của con người. Để những chiếc xe này trở thành hiện thực, chúng phải được trang bị các khả năng nhận thức tiên tiến và nhận biết tình huống để có thể quản lý hiệu quả các kịch bản áp lực cao trong thế giới thực, đưa ra các quyết định thông minh và đảm bảo các hành động an toàn nhất có thể mọi lúc. Tuy nhiên, khả năng nhận thức của các phương tiện cá nhân, dựa vào các cảm biến như camera, LiDAR, và radar, vốn có giới hạn về phạm vi bao phủ và độ chính xác trong việc phát hiện. Chẳng hạn, một chiếc xe có thể không phát hiện được các đối tượng bị che khuất bởi các chướng ngại vật di chuyển hoặc đứng yên khác.

Do đó, nhận thức hợp tác là một công nghệ có thể cách mạng hóa sự phát triển của AV, bằng cách cho phép các phương tiện tự động và kết nối (CAV) chia sẻ thông tin về các vật thể được phát hiện thông qua giao tiếp giữa xe với xe (V2V). Cách tiếp cận này không chỉ có thể cải thiện độ chính xác và phạm vi phát hiện CAVs mà còn nâng cao đáng kể nhận thức của họ về môi trường xung quanh.

Nghiên cứu của chúng tôi giới thiệu cơ chế nhận thức hợp tác nhằm nâng cao độ chính xác trong việc phát hiện các vật thể xung quanh môi trường của AV. Khung mô phỏng được đề xuất cung cấp một môi trường toàn diện để đánh giá các mô hình giao thông, mô hình phương tiện, mô hình truyền thông và mô hình phát hiện đối tượng. Mô phỏng được thực hiện trong các tình huống di động thực tế cho thấy nhận thức cộng tác có thể cải thiện độ chính xác của việc phát hiện đối tượng lên tới 35% so với các phương pháp phát hiện độc lập.

Từ khóa: Phương tiện tự động và kết nối , nhận thức hợp tác, giao tiếp giữa các phương tiện.

Ha Vu Thi Thuy graduated from the Faculty of Mathematics-Informatics at Hanoi University in 1993 and obtained a Master's degree in Information Technology in 2001 at Vietnam National University, Hanoi. In 2017, she received a PhD in Telecommunications Engineering from the Posts and Telecommunications Institute of Technology. She is currently a lecturer in the Faculty of Telecommunications. Her areas of interest include network performance analysis and evaluation, peer-to-peer overlay networks, and multimedia communication data compression and processing.