Collaborative Perception in Multi-Robot Systems: Case Studies in Household Cleaning and Warehouse Operations

Bharath Rajiv Nair
Dept. of Mechanical Engineering
Columbia University
New York, USA
bharathnair2637@gmail.com

Abstract—This paper explores the paradigm of Collaborative Perception (CP), where multiple robots and sensors in the environment share and integrate sensor data to construct a comprehensive representation of the surroundings. By aggregating data from various sensors and utilizing advanced algorithms, the collaborative perception framework improves task efficiency, coverage, and safety. Two case studies are presented to showcase the benefits of collaborative perception in multi-robot systems. The first case study illustrates the benefits and advantages of using CP for the task of household cleaning with a team of cleaning robots. The second case study performs a comparative analysis of the performance of CP versus Standalone Perception (SP) for Autonomous Mobile Robots operating in a warehouse environment. The case studies validate the effectiveness of CP in enhancing multi-robot coordination, task completion, and overall system performance and its potential to impact operations in other applications as well. Future investigations will focus on optimizing the framework and validating its performance through empirical testing.

Index Terms—Collaborative perception, multi-robot systems, deep learning, robot perception, autonomous mobile robots

I. INTRODUCTION

In recent years, the field of robotics has seen significant progress, particularly in the context of multi-robot systems. Collaborative perception, where multiple robots and sensors in the environment share and integrate sensory data to create a unified understanding of their environment, has emerged as a crucial area of research. This collaborative approach to perception enhances the efficiency, accuracy, and safety of robotic operations, making it highly applicable across a variety of domains such as industrial automation, autonomous driving, household cleaning and environmental monitoring [1]–[3].

The paradigm of collaborative perception enhances the capabilities of a robot to perceive it's environment by compensating for the limits of its individual sensing capabilities such as limited range and field of view. This in turn enables the robots to coordinate their actions more effectively and perform individual tasks with greater efficiency. This paper presents collaborative perception frameworks designed to facilitate multi-robot collaboration, focusing on two specific case studies: household cleaning using a team of cleaning robots and autonomous mobile robots (AMRs) operating in warehouse environments.

In the first case study, we explore a scenario where a team of cleaning robots, including a robotic vacuum cleaner, a staircase cleaning robot, and a trash picking robot, work together to clean a home environment. Overhead cameras are strategically placed to provide a global view of the environment. Using this collaborative perception setup, the robots can coordinate their actions, identify areas that require cleaning, and delegate tasks dynamically based on the changes in the environment and the capabilities of each robot.

The second case study examines AMRs in a warehouse setting, where individual robots may not be aware of all obstacles in its planned path due to occlusions and limitations in its sensing range. By utilizing overhead sensors, the framework allows these robots to receive real-time updates about obstacles, enabling them to plan and replan their paths efficiently. A comparative analysis between collaborative and standalone perception approaches demonstrates the advantages of this framework in optimizing operational metrics such as task completion time and energy efficiency which in turn increases throughput and reduces costs.

This paper aims to describe the collaborative perception framework, detail the specific features implemented in each case study, and present experimental results demonstrating the effectiveness of the framework. Through these case studies, we highlight the potential of collaborative perception in enhancing multi-robot systems and its implications for future research and practical applications.

II. RELATED WORK

Collaborative perception in multi-robot systems has been explored extensively, focusing on different data fusion techniques [4]. Early fusion [5] involves combining raw sensor data from multiple sources at the initial stages of data processing, providing rich environmental representations but requiring high-bandwidth communication. Intermediate fusion [6], [7] combines partially processed data, balancing data richness and
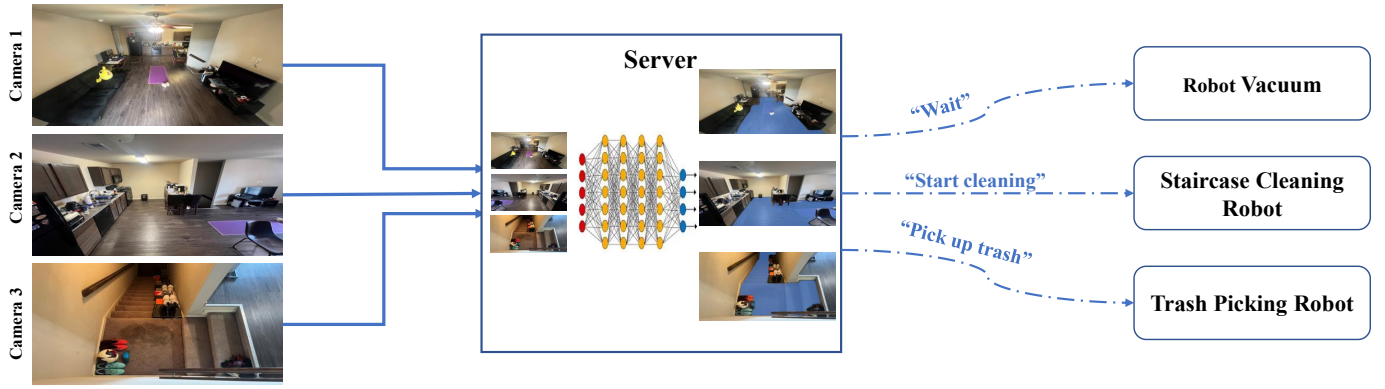
Fig. 1. Pipeline for multi-robot decision making based on collaborative perception.

communication efficiency. Late fusion [8], [9] shares high-level processed information, reducing communication load but potentially losing some detail in environmental representation.

Sharing perception data from multiple sources to a central server provides comprehensive information about the surroundings, facilitating better decision-making for the robots. However, this introduces security concerns, as the data being transmitted and stored centrally can be vulnerable to cyber-attacks. Robust data encryption and secure communication channels are critical to safeguarding the integrity and privacy of sensory data for such safety-critical applications [10], [11].

One of the most widely researched application of collaborative perception is in autonomous driving. In this context, vehicles and roadside infrastructure share sensory data to improve situational awareness and decision-making. Vehicle-to-Vehicle (V2V) [12], Vehicle-to-Infrastructure (V2I) [13] and Vehicle-to-Everything (V2X) [14] communication systems enable real-time data sharing, enhancing the ability of autonomous vehicles to detect and respond to obstacles, pedestrians, and other vehicles. This collaborative approach significantly enhances safety and efficiency in autonomous driving scenarios.

Other potential applications of collaborative perception in multi-robot systems include automated warehouses, smart home environments, human-robot collaboration, environmental monitoring, and search-and-rescue operations. While most collaborative perception is vision-based, utilizing cameras and LIDARs to gather visual data, it can also be tactile-based. Distributed tactile perception [15]–[17], through sensors embedded in robotic skins, provides a more comprehensive understanding of the environment and physical interactions. This enhanced sensory capability improves the robots' performance in tasks such as manufacturing, assembly, and quality inspection, allowing them to handle delicate and complex operations with greater accuracy and efficiency.

III. COLLABORATIVE PERCEPTION FRAMEWORK

The collaborative perception framework enables effective multi-robot collaboration through several integrated components. Each robot is equipped with perception modules that consist of sensors like cameras and Lidars. Additionally, the environment the robots operate in could be fitted with sensors as well. All or some of this data could be transmitted to a centralized server via low-latency communication network.

The centralized server can aggregate data from multiple sources and create a unified and comprehensive understanding of the environment. The server can also utilize the data for optimized task allocation for multi-robot systems. This setup also helps in reducing edge computation on the robots as some non-real-time, compute-intensive tasks could be performed on the server.

In this paper, we specifically analyze how sharing data from sensors embedded in the robot's working environment enhances robot productivity. Figure 1 illustrates a typical pipeline for multi-robot decision making based on collaborative perception.

IV. CASE STUDY 1: COLLABORATIVE HOME CLEANING ROBOTS

In modern smart homes, the integration of multiple robots to handle various cleaning tasks can significantly enhance the efficiency and effectiveness of maintaining a clean living space. Collaborative perception, where robots and sensors in the environment share and integrate sensory data, allows such robots to operate more intelligently and cohesively. This case study explores the implementation of a collaborative perception system involving multiple cleaning robots in a home setting, guided by overhead cameras providing real-time RGB images to optimize the cleaning process. This framework provides the basis for working with these robots, although no physical robots were used in the experiments.

Fig. 2. View from overhead cameras

A. Scenario Description

A home cleaning scenario is set up with three overhead cameras installed in a section of a living space. Each camera monitors a different section of the space as shown in Fig. 2: one covers one side of the room, another covers the opposite side, and the third monitors the staircase going down to the door. In this scenario, we consider three different cleaning robots: Robotic Vacuum Cleaner (RVC) [18], Staircase Cleaning Robot (SCR) [19] and Trash Picking Robot (TPR) [20]. The overhead cameras send data to the server that processes the information to make suitable decisions based on conditions. The room is assumed to be well-lit, as variations in lighting (excessive or insufficient) can significantly impact the performance of deep learning models used for perception [21].

B. Collaborative Perception in Home Cleaning

The collaborative perception system leverages the overhead cameras to instruct specific robots based on the detected needs. The cameras send data in real-time to the server that processes the information. The server interprets the data coming in from the cameras and sends commands to individual robots based on the specific situation. The outline of the pipeline for multi-robot decision making based on collaborative perception is shown in Fig. 1.

The server runs a YOLOv8 [22] object detection model to detect the presence of dynamic obstacles like humans and pets. The server also runs a YOLOv8 [22] semantic segmentation model to segment cleanable floor area from the camera feed. Figure 3 shows an example of an output of the Machine Learning (ML) pipeline run on the server. The blue mask in Fig. 3 represents the cleanable area, while the red square indicates an obstacle that was detected. Figure 4 shows the output of the semantic segmentation model accurately segmenting cleanable area for both flat surfaces as well as staircases by avoiding the entities in those areas. The pretrained YOLOv8 models were fine-tuned by training on a dataset of 1000 images. These images were quickly annotated using the Segment Anything Model (SAM) [23] by Meta.
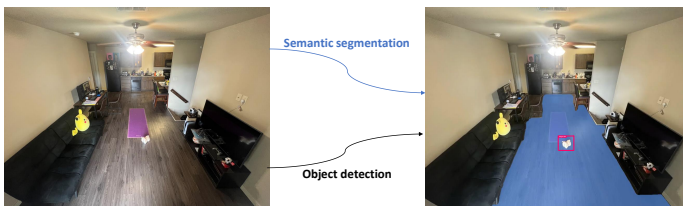

Fig. 3. Example output of ML pipeline run on the server

C. Conceptual Framework

Although physical experiments have not yet been conducted, the conceptual framework involves a monitored environment where the room sections are defined and observed by overhead cameras. To ensure safety and cleaning effectiveness, the robots could be programmed to halt operations when humans

(a) Cleanable area on the floor   **(b) Cleanable area on the staircase**

Fig. 4. Cleanable area segmentation results

or pets are detected in the scene. Task prioritization could be implemented so that if trash is detected, the TPR is dispatched first to clear the area. Once the trash is removed, the RVC and SCR proceeds to clean their designated areas, optimizing the cleaning process through coordination.

Potential performance metrics for future experiments include cleaning coverage, cleaning performance, time efficiency, and the effectiveness of the robots' coordination. This framework highlights the potential benefits of collaborative perception in enhancing the efficiency and safety of home cleaning operations through the integrated efforts of multiple robots.

V. CASE STUDY 2: AMRS IN WAREHOUSES

In warehouse environments, Autonomous Mobile Robots (AMRs) play a crucial role in optimizing logistics and inventory management. This case study explores the implementation of collaborative perception in a warehouse setting to enhance the efficiency and safety of robotic operations.

A. Scenario Description

A warehouse with multiple racks is simulated. There are AMRs in the warehouse that receive missions to carry pallets or goods autonomously from arbitrary start points to arbitrary goal points depending on specific requirements. For standalone perception, the sensors are only attached to the AMRs. For collaborative perception, there are overhead Lidars, with one installed in every aisle in addition to the sensor suite on the AMRs. This setup allows for real-time sharing of perception data to improve navigation and task execution.

B. Collaborative Perception in Warehouses

The robots have a map of the warehouse facility that they use for path planning. The costmap_2d package [24] in Robot Operating System (ROS) is utilized for this purpose. Static obstacles, such as racks and poles, and dynamic obstacles, such as humans, other AMRs, and various entities, are added into the costmap. The picture of the warehouse setting and the corresponding costmap can be seen in Fig. 5. In the costmap, black represents free space, dark blue indicates navigable areas close to obstacles, cyan marks the outer boundaries of inflation zones, and bright magenta signifies obstacles. The visualization is done in RViz.

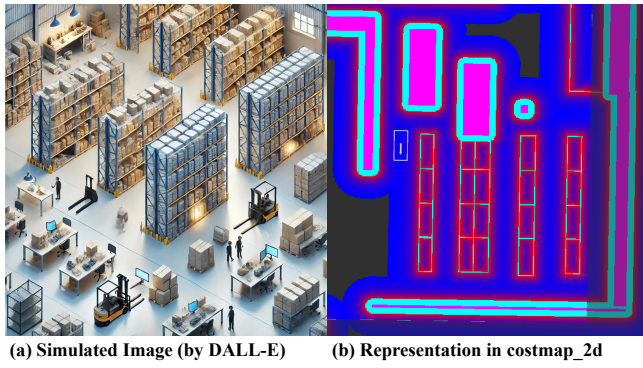(a) Simulated Image (by DALL-E)   (b) Representation in costmap_2d

Fig. 5. Warehouse environment used for analysis

In the standalone perception mode, the costmap is only updated when the individual robot perceives obstacles with its sensors. In this method, only the obstacles that are nearby, within the field of view and that are not occluded are added to the costmap.

In contrast, for the collaborative perception mode, the individual robots as well as the overhead sensors can add obstacles to the costmap in real-time. The perception data from the overhead sensors are shared to a central server that updates the costmap with any obstacles. The AMRs receive this costmap that contains all the static obstacles and dynamic obstacles seen by the overhead sensors. The AMRs then add the dynamic obstacles that it perceives as well to the local costmap.

Once the AMRs receive a mission to go to a goal point, the AMRs find an optimal path to reach its goal using the ARA* [25] algorithm in Search-Based Planning Library [26] of ROS.

C. Experimental Setup and Results

50 trials were set up in simulation for a single AMR to navigate from randomly selected feasible start points to randomly selected feasible goal points in the environment shown in Fig. 5. Two obstacles were randomly placed in different parts of the map for each trial. Three different metrics were considered for the comparative analysis between standalone perception and collaborative perception framework: Total distance traveled by the robot to reach its goal, Total time taken by the robot to reach its goal and Number of times replanning was necessary.

The results of one of the trials are discussed below and illustrated in Fig. 6. The AMR receives a mission to travel to the goal point indicated by the white circle. In the case of Standalone Perception (SP), the AMR plans a simple path shown in Fig. 6(a). As the AMR approaches the aisle, it encounters an obstacle blocking the path, stops, and replans a new path through the next aisle as shown in Fig. 6(b). The AMR now executes the new plan and encounters another obstacle as it turned into the next aisle, this caused the AMR to stop again and replan a third path to the goal as seen in Fig. 6(c). The AMR ultimately reaches the goal following the latest plan. In contrast, Fig. 6(d) illustrates the Collaborative

(a) Initial Path (SP)   (b) First Replan Path (SP)

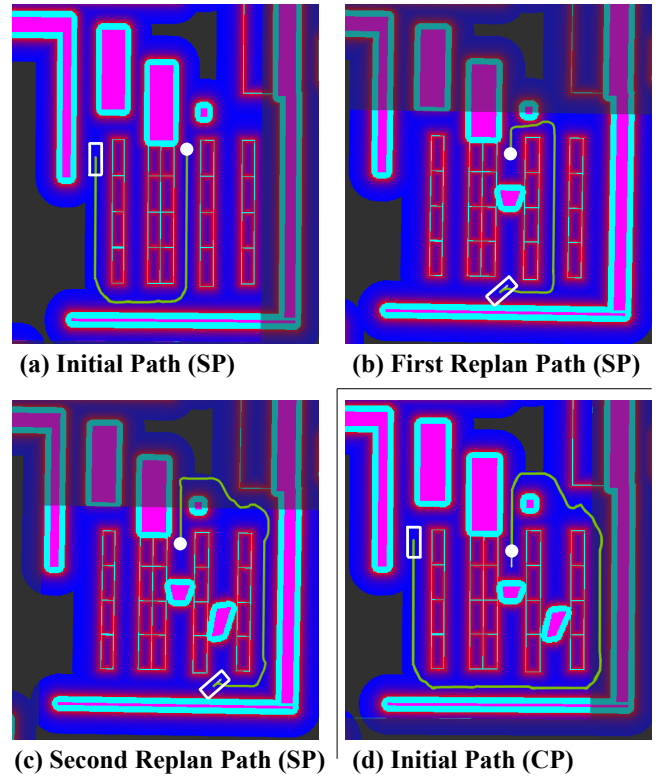(c) Second Replan Path (SP)   (d) Initial Path (CP)

Fig. 6. Multiple plans generated to reach goal by Standalone Perception (SP) vs Single plan generated to reach goal by Collaborative Perception (CP)

Perception (CP) scenario, where the overhead Lidars in each aisle updated the costmap with obstacles in real-time. This allows the AMR to plan an uninterrupted path to the goal without the need for replanning.
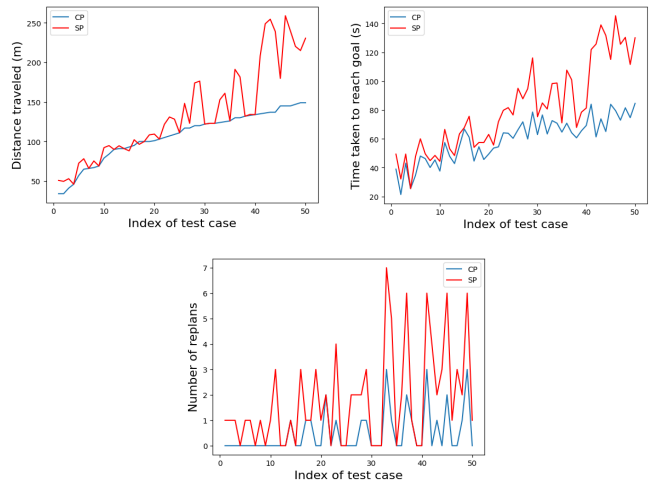

Fig. 7. Comparison of different metrics for Standalone Perception (SP) vs Collaborative Perception (CP)

Various metrics were tracked during the experiments to compare Collaborative Perception (CP) and Standalone Perception (SP) for AMRs. The distance traveled by AMRs to reach goal locations was recorded across multiple trials.

Results show that AMRs consistently travel shorter distances using CP, averaging 37.21 meters less than SP, except when obstacles are not on the shortest path to the goal. Similarly, the average time taken to reach the goal, which includes travel and path planning/replanning time, was reduced with CP by an average of 23.15 seconds. It was observed that the difference between the distance travelled and time taken to reach the goal using SP and CP increased the farther the goal location was from the start location. Furthermore, the frequency of replanning events was lower with CP, with an average reduction of 1.6 replans, demonstrating its efficiency in navigation. The results of the full set of trials is shown in Fig. 7. The results also indicate that performance improvements gained from using CP over SP become more pronounced when the AMRs operate in larger warehouse environments.

VI. Discussion

Using the Collaborative Perception framework provides a more complete information about the robot's working environment. This richer information in turn leads to significant improvements in overall productivity of the team of robots and better task efficiency for individual robots.

The case study on Collaborative Home Cleaning Robots demonstrates that Collaborative Perception enables efficient task allocation, optimizing cleaning paths and minimizing redundant efforts. This process saves time and enhances safety by detecting humans and pets, ensuring secure operations. Furthermore, the automation of cleaning tasks reduces the need for manual intervention, allowing residents to maintain cleaner living spaces with minimal effort.

The case study on AMRs in warehouses quantitatively proved that leveraging CP by utilizing overhead sensors significantly reduces the distance traveled and time taken by AMRs to complete tasks. This increase in efficiency leads to higher throughput, which more than compensates for the additional cost of the sensors. Moreover, the enhanced operational efficiency boosts overall productivity, resulting in faster processing times and reduced labor costs.

This paper focuses on the sharing of environmental sensor data into a centralized server, which in turn provides actionable insights to individual robots. The framework can be extended to include data sharing between robots, further enhancing the overall system's situational awareness. However, expanding data sharing to this extent introduces challenges such as managing high bandwidth requirements, addressing latency issues, and ensuring data security. Additionally, implementing lossless data compression techniques will be essential to handle the increased data load optimally while maintaining data integrity. Future work will need to address these challenges to fully realize the potential of collaborative perception in multi-robot systems.

The collaborative perception framework has practical applications beyond home cleaning and warehouse automation. In healthcare, it can enhance patient monitoring and robotic assistance in hospitals. In agriculture, it can optimize farming operations through precise monitoring of crops and coordination of agricultural robots. In public safety, it can improve search and rescue missions by providing real-time data and coordination among rescue robots. In surveillance, it can enhance monitoring and security operations by providing comprehensive environmental awareness, enabling quicker response times, and improving the accuracy of threat detection. Thus, the implementation of collaborative perception can bring substantial benefits, transforming how industries such as healthcare, agriculture, and public safety operate.

VII. Conclusion

The collaborative perception framework presented in this paper demonstrates significant potential for enhancing multi-robot collaboration through sensor data integration and centralized processing. By aggregating data from various sensors and utilizing advanced algorithms, the framework improves task efficiency, coverage, and safety. Simulation experiments conducted for the AMR case study showed promising results in optimizing robot operations in a warehouse environment. The addition of overhead sensors in the warehouse proved to improve efficiency which in turn improves throughput and revenue. Although physical experiments were not conducted for the home cleaning scenario, the conceptual framework illustrates improvements in cleaning efficiency and potential performance metrics provide a solid foundation for future research. Future work will focus on refining the framework and conducting empirical studies to validate its effectiveness across different applications, addressing challenges such as high bandwidth management, latency issues, and data security. This research underscores the transformative potential of multi-robot systems in enhancing operational efficiency and safety across multiple domains.

References

[1] Y. Han, H. Zhang, H. Li, Y. Jin, C. Lang, and Y. Li, "Collaborative perception in autonomous driving: Methods, datasets, and challenges," *IEEE Intelligent Transportation Systems Magazine*, 2023.

[2] J. Cortés and M. Egerstedt, "Coordinated control of multi-robot systems: A survey," *SICE Journal of Control, Measurement, and System Integration*, vol. 10, no. 6, pp. 495–503, 2017.

[3] F. Vicentini, "Collaborative robotics: a survey," *Journal of Mechanical Design*, vol. 143, no. 4, p. 040802, 2021.

[4] S. Liu, C. Gao, Y. Chen, X. Peng, X. Kong, K. Wang, R. Xu, W. Jiang, H. Xiang, J. Ma *et al.*, "Towards vehicle-to-everything autonomous driving: A survey on collaborative perception," *arXiv preprint arXiv:2308.16714*, 2023.

[5] 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 *Proceedings of the 28th annual international conference on mobile computing and networking*, 2022, pp. 133–146.

[6] 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*. Springer, 2020, pp. 605–621.

[7] 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)*. IEEE, 2022, pp. 2583–2589.

[8] 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, 2012, pp. 270–275.

[9] D. Chen and P. Krähenbühl, "Learning from all vehicles," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17 222–17 231.

[10] Y. Li, Q. Fang, J. Bai, S. Chen, F. Juefei-Xu, and C. Feng, "Among us: Adversarially robust collaborative perception by consensus," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 186–195.

[11] J. Tu, T. Wang, J. Wang, S. Manivasagam, M. Ren, and R. Urtasun, "Adversarial attacks on multi-agent communication," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 7768–7777.

[12] H. Ngo, H. Fang, and H. Wang, "Cooperative perception with v2v communication for autonomous vehicles," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 9, pp. 11 122–11 131, 2023.

[13] Y. Mo, R. Vijay, R. Rufus, N. d. Boer, J. Kim, and M. Yu, "Enhanced perception for autonomous vehicles at obstructed intersections: An implementation of vehicle to infrastructure (v2i) collaboration," *Sensors*, vol. 24, no. 3, p. 936, 2024.

[14] T. Huang, J. Liu, X. Zhou, D. C. Nguyen, M. R. Azghadi, Y. Xia, Q.-L. Han, and S. Sun, "V2x cooperative perception for autonomous driving: Recent advances and challenges," *arXiv preprint arXiv:2310.03525*, 2023.

[15] S. A. Hassan and C. M. Oddo, "Tactile sensors for material recognition in social and collaborative robots: A brief review," in *2022 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. IEEE, 2022, pp. 1–5.

[16] S. Eom, P. Abbaraju, Y. Xu, B. R. Nair, and R. M. Voyles, "Embedded neuromorphic architecture for form+ function 4-d printing of robotic materials: Emulation of optimized neurons," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 12 185–12 191.

[17] Z. Ye, G. Pang, K. Xu, Z. Hou, H. Lv, Y. Shen, and G. Yang, "Soft robot skin with conformal adaptability for on-body tactile perception of collaborative robots," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 5127–5134, 2022.

[18] iRobot, "Roomba i5+ self-emptying robot vacuum," https://www.irobot.com/en_US/roomba-i5plus-self-emptying-robot-vacuum/I555020.html, accessed: July 28, 2024.

[19] A. V. Le, P. T. Kyaw, R. E. Mohan, S. H. M. Swe, A. Rajendran, K. Boopathi, and N. H. K. Nhan, "Autonomous floor and staircase cleaning framework by reconfigurable stetro robot with perception sensors," *Journal of Intelligent & Robotic Systems*, vol. 101, pp. 1–19, 2021.

[20] J. Bai, S. Lian, Z. Liu, K. Wang, and D. Liu, "Deep learning based robot for automatically picking up garbage on the grass," *IEEE Transactions on Consumer Electronics*, vol. 64, no. 3, pp. 382–389, 2018.

[21] P. Maruschak, I. Konovalenko, Y. Osadtsa, V. Medvid, O. Shovkun, D. Baran, H. Kozbur, and R. Mykhailyshyn, "Surface illumination as a factor influencing the efficacy of defect recognition on a rolled metal surface using a deep neural network," *Applied Sciences*, vol. 14, no. 6, p. 2591, 2024.

[22] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics yolov8," 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[23] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo *et al.*, "Segment anything," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 4015–4026.

[24] Open Source Robotics Foundation, "costmap_2d - ROS Wiki," https://wiki.ros.org/costmap_2d, 2024, [Accessed: July 28, 2024].

[25] M. Likhachev, G. J. Gordon, and S. Thrun, "Ara*: Anytime a* with provable bounds on sub-optimality," *Advances in neural information processing systems*, vol. 16, 2003.

[26] Open Source Robotics Foundation, "sbpl_lattice_planner - ROS Wiki," https://wiki.ros.org/sbpl_lattice_planner, 2024, [Accessed: July 28, 2024].