Social Robots for Road Safety: Pedestrian Crossing Assistance Use-Case

Andy Luis Flores Comeca
Karlsruhe Institute of Technology
Karlsruhe, Germany
andy.comeca@kit.edu

Nina Masarykova
Slovak University of Technology
Bratislava, Slovakia
nina.masarykova@stuba.sk

Matej Halinkovic
Slovak University of Technology
Bratislava, Slovakia
matej.halinkovic@stuba.sk

Marek Galiński
Slovak University of Technology
Bratislava, Slovakia
marek.galinski@stuba.sk

Pavel Laskov
University of Liechtenstein
Vaduz, Liechtenstein
pavel.laskov@uni.li

Alexey Vinel
Karlsruhe Institute of Technology
Karlsruhe, Germany
alexey.vinel@kit.edu

***Abstract*—Today, robots are being commonly used in industry, healthcare, household, and many other fields, but not in intelligent transportation systems. We are aiming to fill this gap by adapting the humanoid-like social robot ARI by enhancing its perception, decision-making, and interaction capabilities to manage complex traffic scenarios effectively. Key advancements include upgraded sensors, YOLO-based perception algorithms, and improved mechanisms to interact with pedestrians and vehicles. Through proof-of-concept demonstrations, the robot establishes its ability to detect pedestrian intent, evaluate vehicle behavior, and facilitate safe crossings. This contribution presents a step toward the integration of robotic systems into traffic management, contributing to improved safety and efficiency in challenging traffic environments.**

***Index Terms*—Social robots, intelligent transportation systems, vulnerable road users, pedestrian crossing, advanced driver assistance systems, road safety.**

I. Introduction

Robotic technology is increasingly being utilized across various sectors, including industry, healthcare, and household applications. However, its integration into intelligent transportation systems remains relatively unexplored. The implementation of robotics in road traffic management presents significant potential for enhancing safety and improving the overall efficiency of urban transportation networks. Such robots may be able to obtain a comprehensive situational awareness thanks to a diverse set of onboard cameras and sensors, develop a safe and an efficient coordination strategy in real-time by means of artificial intelligence, and act as a reliable and thoughtful helper thanks to a variety of interaction modalities.

The mobility of a robot can further amplify these features, which can eventually bring about 'coordination skills' that surpass the capabilities of the most advanced traffic lights and the most experienced police officers. The robot's integration with existing traffic management systems allows for an effective operation as a part of a larger traffic coordination network. Finally, the humanoid outlook and the social interaction capabilities make the robot an attractive solution for improved safety of especially vulnerable user groups, for instance, children.

Although the deployment of a robot-policeman for intelligent intersection management or steering autonomous driverless vehicles may appear rather futuristic, our paper aims to make the first step and to explore the feasibility of deployment of social robots for a use-case that may have an immediate practical impact: pedestrian crossing assistance. We identify significant potential for social robots in this role, as they can simultaneously communicate with both vehicle drivers and pedestrians while presenting traffic signals in ways that are intuitive to humans, such as replicating gestures used by traffic officers or emulating light patterns from traffic signals.

The importance of such a mediator becomes even more pronounced during disruptions or failures in standard traffic regulation infrastructure, where the robot could assist or temporarily replace human officers. Moreover, the robot could dynamically manage traffic by adapting regulations in real time, such as temporarily closing a pedestrian crossing to enable an emergency vehicle to pass swiftly and safely. Furthermore, in complex traffic scenarios involving a mixture of automated and human participants, social robots could play a critical role in maintaining traffic safety and optimizing traffic flow.

The use of robotics in traffic management and pedestrian safety has been explored in various contexts, providing a foundation for this study. Gong et al. [1] introduced a life-size traffic police officer robot *IWI*. The robot is equipped with cameras, capable of streaming real-time video to multiple end devices and communicating with traffic participants via hand gestures. The use of gestures is remotely controlled by a human. A similar concept was later adapted by Zhao et al. [2]

This work has been funded as part of the KIT Future Fields project "V2X4Robot" and Hilti Foundation. This paper is part of the CulturalRoad project, funded by the European Union under grant agreement No. 101147397. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

and Ghaffar [3]. Robotic police officers are intended to interact mostly with drivers; however, multiple solutions with robots that interact with pedestrians were proposed. Kumaran et al. [4] introduced a small robotic object mounted on a vehicle dashboard that communicates the intentions of autonomous vehicles near pedestrian crossings. Two simple gestures are used to give pedestrians the signal to cross or not cross the road. A study by Kulhandjian [5] proposed a simple robot based on the Pioneer 3-AT robot with a traffic light-inspired LED matrix mounted on a pole. The pole is also equipped with a camera and LiDAR sensor to assist pedestrians crossing the road. The YOLO [6] machine learning algorithm is used for efficient object detection and classification of pedestrians, cyclists, and vehicles. Red and green LED lights are used to signal both vehicles and pedestrians to stop or go. The robot actively moves to the center of the road when pedestrians are detected, stopping the cars.

Deployment of autonomous robots in outdoor environments is particularly challenging due to the high levels of environmental variability and unpredictability. Factors such as changing lighting conditions, adverse weather, and temporary obstructions can significantly affect their performance. Robotic platforms such as PAL Robotics' ARI, Boston Dynamics' Spot, KUKA's industrial robots, and Clearpath Robotics' OTTO Motors excel in specific applications such as inspection, industrial automation, and indoor logistics. However, they lack the situational awareness, adaptability, and system integration capabilities necessary for dynamic outdoor traffic environments. These limitations present a unique opportunity for the development of robotic solutions that are specifically tailored for traffic management.

To address this gap, we use the ARI humanoid robot as a foundational platform to develop a system capable of meeting traffic management requirements. The primary limitations of ARI in this context are:

1) Situational Awareness: ARI, like many other robots, is designed for controlled indoor environments and lacks the sensor fusion and data processing capabilities required for real-time situational awareness in outdoor traffic scenarios. Improving these capabilities is critical.
2) Adaptability: ARI's current adaptive learning algorithms are insufficient for handling rapidly changing traffic conditions. Real-time learning and scenario-based adaptability need to be integrated.
3) Integration: ARI's limited ability to interact with existing traffic management systems restricts its functionality in broader traffic orchestration networks. Enhancing communication protocols and coordination mechanisms is essential for optimizing traffic flow and safety.

In this study, we focus on the first item: enhancing ARI's situational awareness to enable its effective deployment in dynamic traffic environments. Our work involves upgrading ARI's sensors and perception algorithms to enable it to assist pedestrians with road crossing and improve his capabilities for traffic management. These enhancements aim to ensure reliable robot performance in outdoor scenarios, laying the groundwork for further advancements in robotic traffic management.

Recently, we have already introduced ARI in a virtual reality environment to gather feedback from participants, and evaluated the feasibility of deploying it for managing traffic at pedestrian crossings [7]. In the current work we not only bring our experiments from the virtual reality to the real-world, but also consider more elaborated pedestrian crossing assistance logic.

The paper is organized as follows. Section II introduces the developed system architecture, the proof-of-concept demonstrator is explained in Section III. Finally, concluding discussions and outlook comprise Section IV.

II. System Architecture

Social robot interaction relies on a consistent architecture to execute its actions, whether in indoor or outdoor environments. In a work by Stuede et al. [8], a social robot is deployed in public spaces using a layered architecture composed of hardware, human-robot interaction, localization, navigation, and behavioral trees. For our application, we designed an architecture with similar components, as shown in Fig. 2, but excluded the localization and navigation blocks, as they are not currently required.

The proposed system is divided into three main components. The first component is the perception process, which is responsible for identifying the intentions of pedestrians and detecting vehicles moving around a crossing island. The second component is the decision-making process, which employs a finite-state machine (FSM) to determine the robot's tasks based on the data provided by the perception process. Finally, the interaction process executes the tasks determined by the FSM, facilitating interaction with pedestrians and vehicles through motions and voice messages.

An important feature of this architecture is its ability to use voice messages for communication. This capability enhances the inclusivity of the system, improving usability for visually impaired individuals and ensuring clear and accessible communication in various traffic scenarios.

A. Robot Description

We chose the robot "ARI," a humanoid robot standing at 165 cm tall, designed to facilitate natural human-robot interactions across various applications. With its advanced suite of sensors and actuators, ARI is well-suited to perceive and respond to its environment in real time. This makes it a compelling choice for managing traffic scenarios, helping pedestrians cross streets, and interacting with autonomous vehicles.

ARI's sensory capabilities include an optional 2D YDLIDAR TG15, an 8MP RGB head camera, and two RGB-D cameras positioned on the front and rear of its torso. It also has a stereo-fisheye camera on the back torso and offers optional enhancements, such as thermal and touchscreen cameras. To handle complex processing tasks, ARI is equipped with an Intel i9 CPU, enabling it to run algorithms necessary for

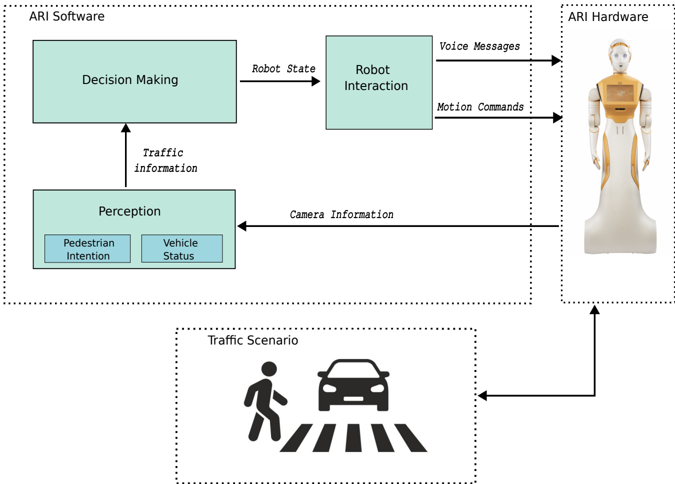


Fig. 1: The architecture of the implemented solution

dynamic traffic environments. ARI features a two-degree-of-freedom (DoF) head for movement, a differential mobile base with two DoF for navigation, and two robotic arms, each with five DoF, capable of executing a wide range of gestures for conveying intuitive signals and more engaging interactions with pedestrians and drivers. Its 10.1-inch touchscreen and integrated speakers and microphones further support clear communication.

For our traffic management application, we enhanced ARI's sensory capabilities by adding an extra RGB-D camera to the back of its head, improving its ability to monitor two-lane road. These sensors provide depth data, enabling ARI to anticipate collisions and execute arm motions and navigation plans with heightened safety.

B. Decision Making

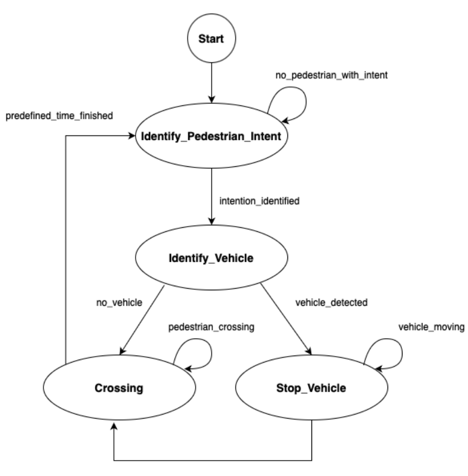


Fig. 2: The state machine describes the robot behavior

The state machine shown in Fig. 2 describes the robot's behavior for the proposed solution. The initial state, *Start*, is where the robot remains until it detects a pedestrian. Once a pedestrian is detected in front of the robot, the system transitions to the *Identify_Pedestrian_Intent* state. In this state, the robot determines whether the pedestrian intends to cross the street, ignoring those who do not.

If the pedestrian's intent to cross is identified by our custom perception algorithm described in Subsection II-D, the robot transitions to the *Identify_Vehicle* state. Here, the robot moves its head toward the road to detect any approaching vehicles. If no vehicles are detected, the robot proceeds to the *Crossing* state, where it communicates with the pedestrian using gestures and voice messages, signaling them to cross the street. The robot remains in this state for a predefined time period. Afterwards, it returns to the *Start* state.

If a vehicle is detected in the *Identify_Vehicle* state, the robot transitions to the *Stop_Vehicle* state, where it signals the vehicle to stop. Once the vehicle has come to a halt, the robot moves to the *Crossing* state, allowing the pedestrian to cross the street. If the vehicle departs while the robot is in the *Stop_Vehicle* state, the system transitions back to the *Identify_Vehicle* state to re-assess the situation.

C. Robot Interaction

The FSM in Fig. 2 defines ARI's states and transitions. It issues commands to ARI which depend on the state he is currently in, taking advantage of ARI's capabilities to physically interact with the environment through navigation, motion planning, and speech. For this experiment, ARI's mobile base is utilized to rotate its body, working together with arm motions to convey distinct signals. Additionally, voice messages are integrated into each state of the robot, enhancing the interaction and ensuring that both pedestrians and vehicles clearly understand its instructions as a street traffic manager.

In [3], a traffic robot tries to mimic police gestures with seven different arm motions. The robot ARI in [7] uses a more simple approach, by defining just two arm motions: stop and pass. In our solution, we combine the mobile base together with arm motion and voice message to interact with pedestrians and vehicles.

The following motions are defined for the various states: Initially, as illustrated in Figure 3a, ARI begins in the *Start State*, facing the vehicles. From this state ARI transitions into the *Identify_Pedestrian_Intent* state by rotating its head and looking towards the pedestrian. Once the *intention_identified* variable becomes active, ARI's head rotates to detect the vehicle and the robot transitions into the *Identify_Vehicle_State*.

To assist pedestrians in crossing the street, ARI extends one hand and rotates its base toward the crossing lane. Simultaneously, it emits a voice message informing pedestrians that they may proceed, as shown in Figure 3b. This motion is executed when either the *no_vehicle* or *vehicle_stopped* variable is active, transitioning ARI into the *Crossing State*.

Finally, to interact with vehicles, ARI faces the road and raises both hands, signaling drivers to stop, as shown in Figure 3c. The motion is executed when the *vehicle_in_front* variable is active, transitioning ARI to the *Stop_Vehicle* state.

(a) Start pose

(b) Crossing Pose

(c) Stop Pose

Fig. 3: ARI's actions for managing street traffic: (a) initial position, (b) assisting pedestrians, and (c) signaling vehicles to stop.

D. Perception

The state machine that controls ARI's behaviour is largely driven by custom perception algorithms designed to handle the complexities of close to real-time traffic crossing management. They are separated into two blocks, one to identify the intentions of pedestrians and another one to determine the safety of crossing. Both are based on the YOLOv11 [6], [9] object detection model. The medium size of the model was used, consisting of 20.1M parameters. The version of the model provided by Ultralytics is trained on COCO (Common Objects in Context) [10] dataset and achieves mean average precision (mAP) of 51.4, considering intersection over union (IoU) in range from 0.5 to 0.95 over all 80 classes. For person class the mAP reaches 0.84. In proposed solution, the bicycle, car, motorcycle, bus and track classes are detected, which reach mAP 0.65, 0.74, 0.8, 0.89 and 0.62 respectively. In order to guarantee that our perception algorithms can run with at least 10 FPS on CPU, we use the *medium* version of YOLOv11 compiled using the OpenVINO framework.

Considering our need to maximize the efficiency of perception, we designed an algorithm to identify a pedestrian's intention to cross with minimal computational overhead. When ARI is in the *Identify_Pedestrian_Intent* state, the field of view of both cameras is centered on the pedestrian crossings. This simplifies the problem of intention prediction, as we are processing data from a stationary ego-agent. We can thus infer several assumptions about the scene that let us increase the efficiency of our algorithm.

We avoid processing frame sequences by a deep learning model, relying on single-frame detections and post-processing instead. Detections are tracked over time using the Hungarian method [11]. Using the position history of each pedestrian, we determine their intention to cross by measuring the time spent on the border of the pedestrian crossing. Once a pedestrian has been stationary at the crossing for at least 3 seconds, a signal is sent to transition to the *Identify_Vehicle* state.

Once ARI transitions to the new state, we store a reference frame from both cameras. These frames are crucial to guarantee the robustness of our vehicle detection system by filtering out parked cars. By computing the difference between the reference frame and the current frame we are able to filter out vehicle detections to only consider active traffic participants. The goal of our perception algorithm in this state is to determine whether it is safe to cross by making sure that there are no vehicles in the lanes being monitored or that the vehicles are yielding to pedestrians by stopping at the crossing. In this state, we also rely on single-frame object detections and tracking. We measure the spatial displacement of objects between frames to determine their velocity. These measurements are then used to determine whether the vehicle closest to the pedestrian crossing in each lane has come to a full stop. ARI only signals that it is safe to cross once we make sure that both lanes are empty or the leading vehicle has stopped.

III. Proof-of-Concept Demonstration

In our proposed setup, ARI is strategically positioned at the center of a crossing island, oriented in parallel with both lanes. This placement allows it to effectively manage interactions between pedestrians and vehicles on both sides of the crossing. An illustration of this setup can be seen in Figure 4. ARI is equipped with a rear-facing camera to monitor pedestrians on the opposite side of the zebra crossing, ensuring comprehensive situational awareness. Furthermore, ARI's ability to rotate its head enables it to observe vehicles in both lanes and pedestrians on both sides of the road, enhancing its capacity to navigate dynamic traffic conditions.

As described in Subsection II-B, ARI initiates vehicle detection only when it identifies a pedestrian's intention to cross the street. Once such a pedestrian is detected, ARI repositions its head to align its front and rear cameras towards the vehicles, ensuring targeted and effective monitoring. This approach allows ARI to accurately assess traffic situations, facilitating safe and efficient interactions between pedestrians and vehicles.

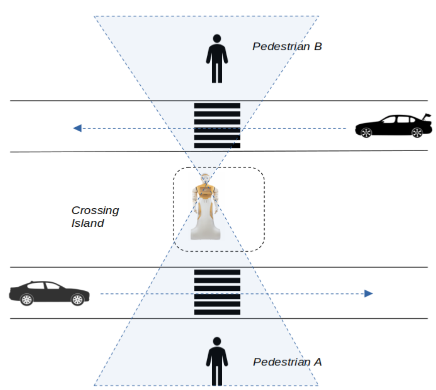


Fig. 4: ARI robot in the middle of the crossing island facing pedestrian and vehicle in both lanes

To illustrate ARI's functionality, we detail three scenarios encountered during its operation:

- **Scenario A:** ARI detects a pedestrian directly in front of it. The robot engages its forward-facing sensors to monitor both the pedestrian and the vehicles.
- **Scenario B:** ARI identifies a pedestrian on the opposite side of the zebra crossing using its rear-facing camera.
- **Scenario C:** ARI encounters multiple vehicles on the road simultaneously. The robot's perception algorithm processes this complex scenario, prioritizing pedestrian safety.

These scenarios highlight ARI's adaptability and advanced perception capabilities in managing diverse traffic conditions to ensure pedestrian safety.

A. Scenario A

In this scenario, ARI attempts to detect a pedestrian using its front-facing head camera, as illustrated in Fig. 5a. When a pedestrian is detected, the pedestrian detection algorithm determines their intention to cross the road. In response, ARI rotates its head to check for oncoming vehicles. When a vehicle is detected in the lane ahead, ARI signals the vehicle to stop, as depicted in Fig. 5b.

After confirming that the vehicle has come to a complete stop, ARI transitions to the Crossing state. In this state, ARI allows the pedestrian to safely cross the road, as shown in Fig. 5c.

B. Scenario B

In this scenario, ARI detects pedestrians using its rear-facing camera at the opposite zebra crossing, as illustrated in Fig. 5d. The process follows a similar procedure as described in Scenario A. ARI rotates its head to ensure the rear camera can monitor oncoming vehicles. However, to stop the vehicle effectively, ARI rotates its body toward the approaching car, as shown in Fig. 5e. Once the vehicle is stopped, ARI transitions to the Crossing state, allowing the pedestrian to safely cross the lane, as depicted in Fig. 5f.

C. Scenario C

In this scenario, the reaction of the robot is challenged by having several vehicles on the same road. ARI has to ensure safety by making sure either the lead cars in each lane are stopped or no vehicles are present. In Fig. 5g, ARI detects the green car in front. Even though the car leaves in Fig. 5h, it waits until there is no hazard to switch to the Crossing state, as shown in Fig. 5i.

IV. Discussion and Outlook

The study demonstrates the potential for leveraging humanoid robots, such as ARI, to mediate traffic scenarios by addressing key challenges in perception, decision making, and human-robot interaction. The proof of concept demonstrations validate ARI's ability to manage interactions between pedestrians and vehicles, ensuring safe crossings and promoting smoother traffic flow. The system was able to process data from both sides of the crossing that provides a comprehensive understanding of the scene with limited computational resources.

Although the proposed system successfully improves the capabilities of ARI for outdoor use, several limitations and open questions were observed during the experiments. ARI's sensor suite is quite limited, which restricts the range at which the area surrounding the crossing can be monitored. The reliance on single-frame object detection may also limit robustness in more complex, higher density traffic conditions. When considering longer ranges and high density traffic situations, the visibility of humanoid robots might be insufficient depending on its size and colour. Further improvements in this area should also be made, for example by attaching high visibility elements. Such elements can also be used to improve the familiarity of the robot's gestures for road users.

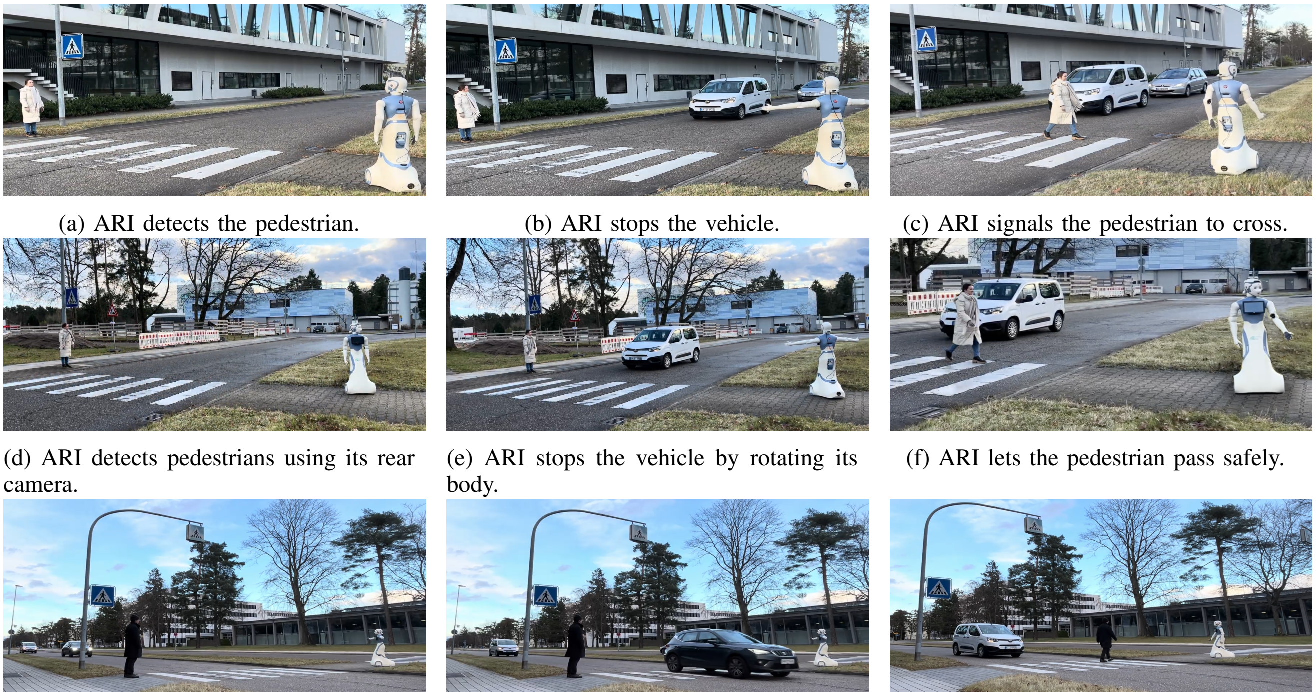
Future work should focus not only on improving the weaknesses and limitations mentioned above but also on utilizing all potential strengths of the platform. Humanoid robots often have more than one interface for interaction with road users. In case of ARI it is gestures, voice and a screen. These interfaces make it well suited for a variety of traffic situations. It is also modifiable in terms of hardware and can be easily integrated with a V2X unit enabling communication with other connected road participants. This can enhance its situational awareness thanks to cooperative perception, greatly increasing perception range. It would also allow the robot to notify vehicles in advance and also limit their speed in critical situations to further improve the safety of vulnerable road users. ARI can also be equipped with a Jetson AGX Orin which would improve its capabilities for processing more complex perception algorithms. These algorithms could then provide ARI with information such as the predicted trajectories of vehicles and pedestrians, allowing it to monitor if a pedestrian has finished crossing or act more proactively in critical situations. Predicting the intent of pedestrians could be accomplished by more complex end-to-end models or by implementing additional heuristics like monitoring the speed of pedestrians and their poses. Some control of the process could be handed of to the pedestrians themselves. Thanks to ARI's microphone array they could issues commands or additional information, for example they could verbally inform ARI of their intention to cross.

References

[1] L. Gong et al.,"Real-time human-in-the-loop remote control for a life-size traffic police robot with multiple augmented reality aided display terminals," 2017 2nd International Conference on Advanced Robotics and Mechatronics (ICARM), Hefei and Tai'an, China, 2017, pp. 420-425, doi: 10.1109/ICARM.2017.8273199.

[2] L. Zhao et al., "A Bionic Arm Mechanism Design and Kinematic Analysis of the Humanoid Traffic Police," 2019 IEEE 9th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER), Suzhou, China, 2019, pp. 1606-1611, doi: 10.1109/CYBER46603.2019.9066510.

[3] F. Ghaffar, "Controlling Traffic with Humanoid Social Robot," arXiv preprint arXiv:2204.04240, Systems Design Engineering, University of Waterloo, Waterloo, Canada, 2022. [Online]. Available: https://doi.org/10.48550/arXiv.2204.04240.



(a) ARI detects the pedestrian.
(b) ARI stops the vehicle.
(c) ARI signals the pedestrian to cross.
(d) ARI detects pedestrians using its rear camera.
(e) ARI stops the vehicle by rotating its body.
(f) ARI lets the pedestrian pass safely.
(g) ARI detects a car in front.
(h) ARI ensures no hazards are present.
(i) ARI allows the pedestrian to cross.

Fig. 5: Grid of ARI robot scenarios across different situations: Scenario A, Scenario B, and Scenario C.

[4] S.C. Kumaran, A. Oberlender, A. Grishko, B. Megidish, H. Erel, "To Cross or Not-to-Cross: A Robotic Object for Mediating Interactions Between Autonomous Vehicles and Pedestrians," 2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN).

[5] H. Kulhandjian, "Smart Robot Design and Implementation to Assist Pedestrian Road Crossing," 2024

[6] J. Redmon et al., "You only look once: Unified, real-time object detection, " 2016 Proceedings of the IEEE conference on computer vision and pattern recognition, doi: 10.1109/CVPR.2016.91

[7] M. Schrapel, M. Bied, B. Bruno, and A. Vinel, "Experiencing Social Robots for Traffic Guidance using Virtual Reality Videos," Mensch und Computer 2024 - Workshopband, Gesellschaft für Informatik e.V., 2024, doi: 10.18420/muc2024-mci-demo-318.

[8] M. Stuede, K. Westermann, M. Schappler and S. Spindeldreier, "Sobi: An Interactive Social Service Robot for Long-Term Autonomy in Open Environments," 2021 European Conference on Mobile Robots (ECMR), Bonn, Germany, 2021, pp. 1-8, doi: 10.1109/ECMR50962.2021.9568838.

[9] G. Jocher, J. Qiu. "Ultralytics YOLO11 (Version 11.0.0)," 2024 Retrieved from https://github.com/ultralytics/ultralytics

[10] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick (2014). Microsoft coco: Common objects in context. In Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13 (pp. 740-755). Springer International Publishing.

[11] H.W. Kuhn, "The Hungarian method for the assignment problem" 1955 Naval Research Logistics, 2: 83-97, doi: 10.1002/nav.3800020109