Anticipating Human Behavior for Safe and Efficient Collaborative Mobile Manipulation

Simon Bultmann[1], Raphael Memmesheimer[2], Jan Nogga[2], Julian Hau[2], and Sven Behnke[2]

I. Introduction

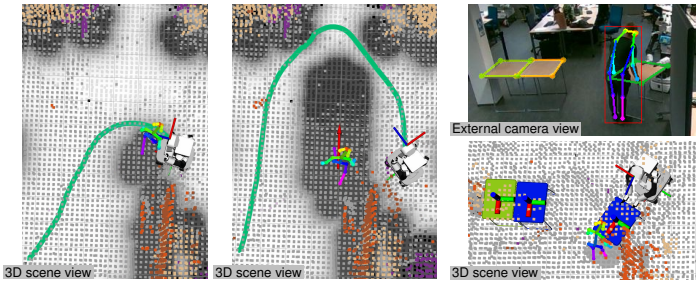
The ability to anticipate human behavior is essential for robots to interact safely and efficiently with humans. In this work, we integrate anticipatory behavior into the control of a mobile manipulation robot using a smart edge sensor network. The external sensors provide global observations, future predictions, and goal information, enhancing the robot's ability to navigate safely and collaborate effectively.

We present two key approaches to human behavior anticipation: (1) safe navigation using projected human motion trajectories from the smart edge sensor network into the robot's planning map, and (2) collaborative furniture handling, where the robot anticipates human intentions to achieve a predefined room layout. Fig. 1 illustrates these two scenarios in which we anticipate human behavior. By incorporating human trajectories observed and predicted by the smart edge sensor network into the robot's planning framework, we enable it to benefit from global context information and thus navigate more safely in dynamic human-centered environments. In the collaborative furniture handling scenario, anticipation combines compliant control with goal inference, enabling efficient human-robot interaction.

Our experiments demonstrate that integrating anticipatory behavior improves navigation safety and enhances collaboration efficiency. We showcase a system that utilizes human behavior anticipation to safely navigate while collaboratively achieving a target room layout, including the placement of tables and chairs. This work builds upon prior research on mobile robots as nodes in a smart edge sensor network [1], now focusing on anticipatory human behavior for real-time robotic action. These advancements contribute to making human-robot interactions more intuitive, safe, and effective.

II. Method

We use a PAL Robotics TIAGo++ [2] omnidirectional dual-arm mobile manipulator, shown in Fig. 2 (a), equipped with an Orbbec Gemini 335 RGB-D camera and a Zotac ZBOX QTG7A4500 computer mounted on the back of the robot. The robot is informed by a smart edge sensor network consisting of 25 sensor nodes, shown in Fig. 2 (b), with an Intel RealSense D455 RGB-D camera and an Nvidia Jetson Orin Nano compute board with an embedded GPU



(a) navigation without (l) and with (r) anticipation: path is adapted before robot can see the person

(b) collaborative manipulation: where to grasp and place table

Fig. 1. Two scenarios in which our robot anticipates human behavior.

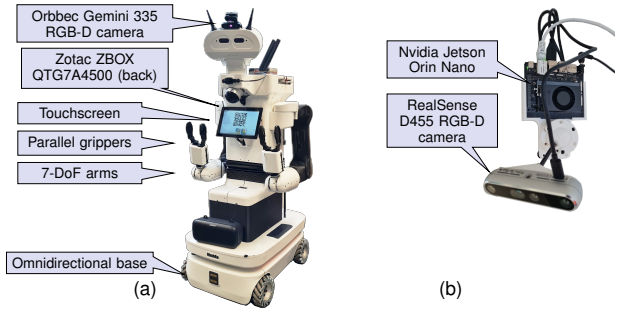


Fig. 2. Robot and sensor setup: (a) PAL Robotics TIAGo++ robot; (b) exemplary smart edge sensor.

for onboard semantic perception using lightweight CNNs [3], [4]. The smart edge sensors are mounted at a height of $\sim 2.5\,\mathrm{m}$, distributed over a lab space of $\sim 240\,\mathrm{m}^2$ size.

An overview of the data processing architecture of the proposed approach is given in Fig. 3. The sensor network gathers semantic observations of the scene, i.e. detections and keypoints of persons, robots, and objects, as well as semantic point clouds [3]–[5]. The sensor views are fused into an allocentric 3D semantic scene model on a central backend comprising a volumetric semantic map of the static environment as well as dynamic human, robot, and object models. The sensors receive semantic feedback to incorporate global context, e.g. about occlusions, into their local perception [3], [5].

The robot augments its local perception, manipulation, and navigation capabilities with global context information received as semantic feedback from the backend of the smart edge sensor system. The robot's localization is initialized and tracked by the external smart edge sensors. Additionally, it receives feedback about persons who are in its vicinity but are out of sight of its internal sensors, e.g. due to occlusions or limited Field of View (FoV), and their predicted

This work was funded by grant BE 2556/16-2 (Research Unit FOR 2535 Anticipating Human Behavior) of the German Research Foundation (DFG)

[1]Department of Computer Science, University of Freiburg, Germany; bultmann@cs.uni-freiburg.de

[2]Autonomous Intelligent Systems Group, University of Bonn, Germany; memmesheimer@ais.uni-bonn.de

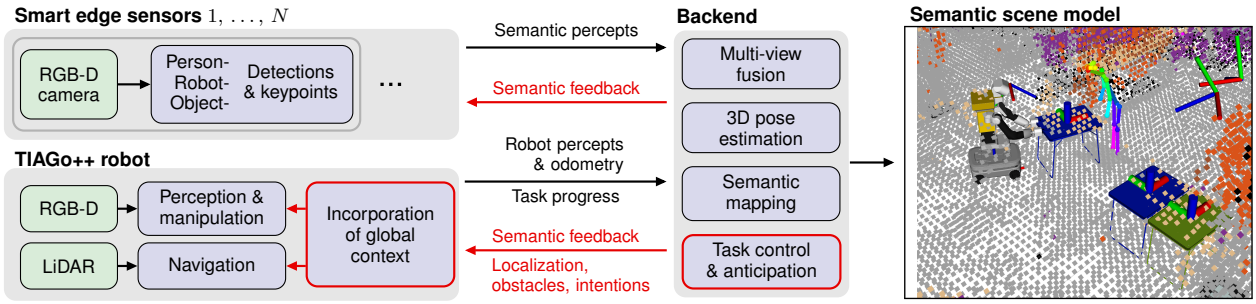


Fig. 3. Data processing architecture of the developed approach. A network of smart edge sensors supervises the work space. It detects persons, robots, and objects and estimates their pose. The backend fuses local percepts and controls the task. Both the robot and the sensors incorporate semantic feedback.

movement, as well as the intended target configuration of the manipulation task. This enables anticipatory human-aware robot navigation where the robot can preemptively adjust its navigation path, e.g. to persons appearing from behind occluders or to reach the intended target pose for picking up or placing an object.

The smart edge sensor network from our prior works [3]–[5] provides an allocentric 3D scene model, tracking humans, robots, and objects for safe and efficient interaction. We extend the semantic perception to separate closely positioned furniture instances using a fine-tuned YOLOv8 model [6], extracting keypoints for precise pose estimation. The robot employs onboard perception with MM-Grounding-Dino [7] and Nano-SAM [8] for local open-vocabulary object recognition. Grasp poses for efficient manipulation are calculated from the object masks and point cloud segments.

Anticipatory human-aware navigation integrates person tracking and velocity predictions into the robot's dynamic cost map for obstacle avoidance, enabling foresighted path adaptation. During collaborative furniture handling, the robot anticipates human grasp and placement intentions, dynamically adjusting its goal poses. A compliant control mode allows human-guided transport of tables, while chairs are autonomously manipulated by the robot alone.

Our system demonstrates improved safety, efficiency, and intuitiveness in human-robot collaboration through predictive and adaptive behavior.

III. Experiments

We evaluate our system through structured experiments on anticipatory human-aware navigation and collaborative furniture transport.

In the navigation experiments, the robot integrates real-time semantic feedback from smart edge sensors to anticipate human movements beyond its onboard sensor range, enhancing safety in environments with frequent occlusions. Results reported in Table I demonstrate a significantly increased minimum safety distance—at least $50\,\mathrm{cm}$—compared to only $8\,\mathrm{cm}$ in the worst case without anticipation.

For collaborative furniture handling, the robot anticipates human intent for object pickup and placement, using compliant control to assist in table carrying. Compared to a baseline without anticipation, task execution is 26 seconds faster on average, with reduced pose error.

TABLE I

Average and worst-case person–robot safety distance.

	w/o anticipation		w/ anticipation	
	S1	S2	S1	S2
avg.	0.23 m	0.19 m	0.71 m	**0.79 m**
worst	0.08 m	0.12 m	0.50 m	**0.61 m**

Results from five experiment runs for two subjects (S1, S2).

Finally, in a continuous furniture rearrangement task, the system autonomously identifies objects for transport, anticipates target placements, and ensures safe navigation while interacting with a human partner. The results validate the effectiveness of our approach in improving safety, efficiency, and intuitiveness in human-robot collaboration. The experiments are illustrated in the attached video and a full video is available online[1].

IV. Conclusions

We present approaches to incorporate allocentric semantic context information from smart edge sensor network observations to anticipate human behavior on two levels: (1) in the context of human-aware navigation to improve safety, by projecting predicted human trajectories into the planning map of a mobile robot, and (2) in the context of collaborative mobile manipulation for improving efficiency, by anticipating intentions to work towards a desired goal.

Both approaches are evaluated in real-world experiments and compared against non-anticipatory baseline approaches utilizing only on-board sensors and a graphical user interface for human-robot interaction. Our approach demonstrates safer human-aware navigation and improved efficiency for human-robot collaboration with a mobile manipulation robot. We show that the robot anticipates persons emerging from behind occlusions and preemptively adjusts its navigation path to maintain a safe distance by incorporating semantic feedback of human pose observations from external sensors.

An integrated demonstration shows our approach's potential for collaborative human-robot interaction, achieving the complex task of setting a room layout with tables and chairs.

Directions for future work include implementing anticipatory human-aware navigation also on a higher planning level, instead of using the local obstacle cost map, taking long-term goals and intents of the persons into account.

[1] www.ais.uni-bonn.de/videos/GRC_2025_Bultmann

REFERENCES

[1] S. Bultmann, R. Memmesheimer, and S. Behnke, "External camera-based mobile robot pose estimation for collaborative perception with smart edge sensors," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 8194–8200.

[2] J. Pages, L. Marchionni, and F. Ferro, "TIAGo: The modular robot that adapts to different research needs," in *International Workshop on Robot Modularity, IROS*, vol. 290, 2016.

[3] S. Bultmann and S. Behnke, "3D semantic scene perception using distributed smart edge sensors," in *International Conference on Intelligent Autonomous Systems (IAS)*, 2022, pp. 313–329.

[4] J. Hau, S. Bultmann, and S. Behnke, "Object-level 3D semantic mapping using a network of smart edge sensors," in *6th IEEE International Conference on Robotic Computing (IRC)*, 2022, pp. 198–206.

[5] S. Bultmann and S. Behnke, "Real-time multi-view 3D human pose estimation using semantic feedback to smart edge sensors," in *Robotics: Science and Systems (RSS)*, 2021.

[6] G. Jocher, A. Chaurasia, and J. Qiu, *Ultralytics YOLOv8*, `https://github.com/ultralytics/ultralytics`, version 8.0.0, 2023.

[7] X. Zhao, Y. Chen, S. Xu, X. Li, X. Wang, Y. Li, and H. Huang, "An open and comprehensive pipeline for unified object grounding and detection," *preprint arXiv:2401.02361*, 2024.

[8] NVIDIA-AI-IOT, *NanoSAM*, `https://github.com/NVIDIA-AI-IOT/nanosam`, 2024.