Gesture-Based Optimal Planning for Autonomous Human-Following Shopping Carts

Victoria Hong
Merrill Edmonds, Tarik Yigit, Faiza Sikandar, and Jingang Yi
Department of Mechanical Engineering, Rutgers University, New Brunswick, N.J.

Introduction

In recent years, human-following autonomous systems have been a growing interest among the robotics community. These robots are designed to assist and work directly with users to increase human productivity and lower physical barriers for certain human-centric tasks [1]. With the ongoing pandemic, we believe exchanging traditional shopping carts for a human-following autonomous system will lower the transmission of bacteria and viruses that exist on high-touch surfaces and significantly reduce shopping effort for consumers. While there have been previous attempts to commercialize smart carts and research to produce autonomous shopping carts, their work fails due to relying on traditional methods to generate human-following trajectories and computer vision. In this work, we present an autonomous shopping cart.

Related Works

- **RGB and RGB-D cameras**, used for their precise and adaptive behavior, are common for human tracking and activity recognition [2]. Other methods, like color features and hue-saturation-value histograms, leg-based person detectors, Kinect sensors, and binocular cameras, fail to operate in varying settings with inconsistent lighting and the every-changing number of users and multi-robots in use.
- **Map-based techniques** have proven its success with human-robot and human-human interactions but demonstrated limitations when more complex interactions caused delays [3]. It is crucial for our system to be able to follow a target in crowded environments.
- **Adaptive Monte Carlo Localization (AMCL) self-localization system** was explored for pose estimates and path planning but found to have the possibility of failing due to frequent sudden changes in human trajectories and more complex dynamics [4]. The many limitations in distance and connection range confirmed the lack of robustness of this technology.

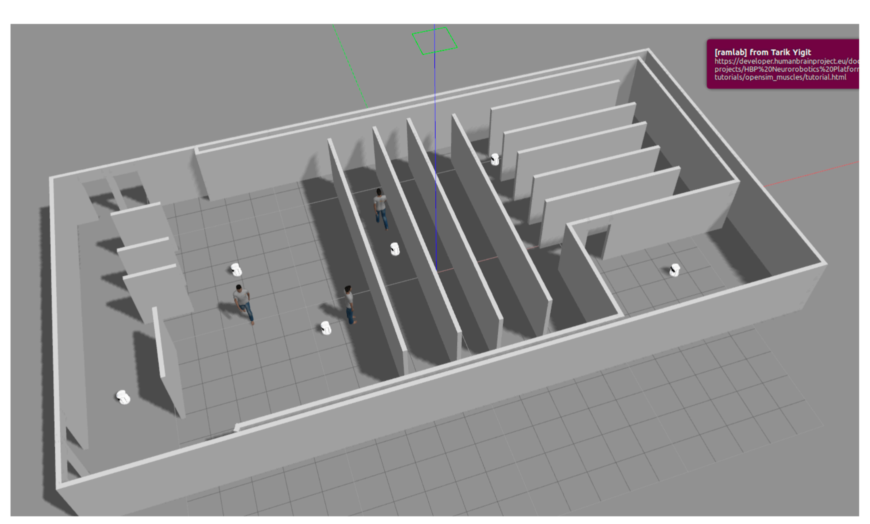


Fig. 1 One of the Gazebo worlds used in the supermarket simulations. A large group of robots can be simulated simultaneously on one machine with accurate physics and on-board controller architectures. Sensors are simulated individually for each robot using Gazebo/ROS plugins.

Problem Statement

The human-following shopping cart problem with gesture-based contactless positioning suggestions consists of three sub-problems: The estimation of user skeletal poses from sensor data for use in gesture-based positioning suggestions, the selection of target poses for each shopping cart, and the generation of optimal local and global trajectories that take the carts to their target individual poses. An equivalent constrained optimization problem is presented in (1), where $\boldsymbol{x}_i^{(g)}$ is the target pose.

$$
\begin{aligned}
\underset{i=1,\dots,N}{\boldsymbol{x}_i(t)^*} &= \underset{\boldsymbol{x}_i(t)}{\operatorname{argmin}} \sum_{i=1}^{N} \left[\int_t^{t+T} \psi(\boldsymbol{x}_i(\tau), \boldsymbol{u}_i(\tau))\, \mathrm{d}\tau \right] && \text{(1a)} \\
\text{s.t.}\ \ & \psi(\boldsymbol{x}_i(t), \boldsymbol{u}_i(t)) = p_i \|\boldsymbol{x}_i(t) - \boldsymbol{x}_i^{(g)}\|, && \text{(1b)} \\
& \dot{\boldsymbol{x}}_i(t) = \boldsymbol{f}(\boldsymbol{x}_i(t), \boldsymbol{u}_i(t)), && \text{(1c)} \\
& \boldsymbol{x}_i(t) \in \mathcal{X}_i,\ \boldsymbol{x}_i(t+T) \in \mathcal{G}_i,\ \boldsymbol{u}_i(t) \in \mathcal{U}_i, && \text{(1d)}
\end{aligned}
$$

In this work, we focus on the first two subproblems of gesture-based positioning suggestions and optimal target pose selections. A suggested pose $\boldsymbol{x}_i^{(s)}$ must be determined based on user pose $\boldsymbol{\eta}_i$, which we assume is provided by a skeletal pose estimator. The optimal target pose $\boldsymbol{x}_i^{(g)}$ must then be selected to maximize utility provided to the user, while minimizing interference with other shoppers. The resulting optimal pose is then passed to the optimizer as part of constraint (1b).

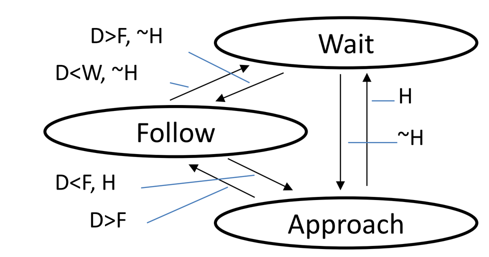
Proposed Solution

The proposed solution [5] divides the cart behavior into three distinct states based on user actions and robot proximity. A finite state machine with transition conditions shown in Fig. 2 then selects $\boldsymbol{x}_i^{(g)}$ from three potential poses, based on the current state: $\boldsymbol{x}_i^{(f)}$ if the robot is following the user, $\boldsymbol{x}_i^{(w)}$ if the robot is waiting for the user, and $\boldsymbol{x}_i^{(a)}$ if the robot is approaching the user so they can drop off the item they are holding. $\boldsymbol{x}_i^{(f)}$ is selected as the pose trailing the user by δ_f along their previous trajectory, so the robot follows the user at a comfortable distance. $\boldsymbol{x}_i^{(w)}$ is selected as a pose that is close to the user but out of the way of other shoppers and carts. A per-robot 2D costmap is constructed from (2)–(4) corresponding to a distance cost $||\boldsymbol{e}_s||$, which models proximity to the user, a pose cost $\boldsymbol{f}_c$, which models expected foot traffic, and a blocking cost $\boldsymbol{f}_b$, which models collisions or delays caused by waiting in another shopper's path.

$$
\boldsymbol{x}_i^{(g)\,*} = \underset{\boldsymbol{x}_i^{(g)}}{\operatorname{argmin}} \sum_i \|\boldsymbol{e}_s\| + \alpha_2 f_c(\boldsymbol{x}_i^{(g)}) + \alpha_3 f_b(\boldsymbol{x}_i^{(g)}), \quad (2)
$$

$$
f_c(\boldsymbol{x}_i^{(g)}) = \frac{1}{|\mathcal{H}|} \sum_{\boldsymbol{h}(t) \in \mathcal{H}} \frac{1}{T_h} \int_{t_0}^{t_f} K_{\mathcal{H}}(\boldsymbol{x}_i^{(g)} - \boldsymbol{h}(t))\, \mathrm{d}t, \quad (3)
$$

$$
f_b(\boldsymbol{x}_i^{(g)}) = \sum_{\boldsymbol{h}_j \in \mathcal{N}(\boldsymbol{x}_i(t))} \phi_\theta(\boldsymbol{x}_i^{(g)}, \boldsymbol{h}_j) K_b \quad (4)
$$



H = user is holding something
D = distance-to-user
F = following distance
W = waiting distance

Fig. 2 Transition conditions for the finite state machine that governs pose selection.

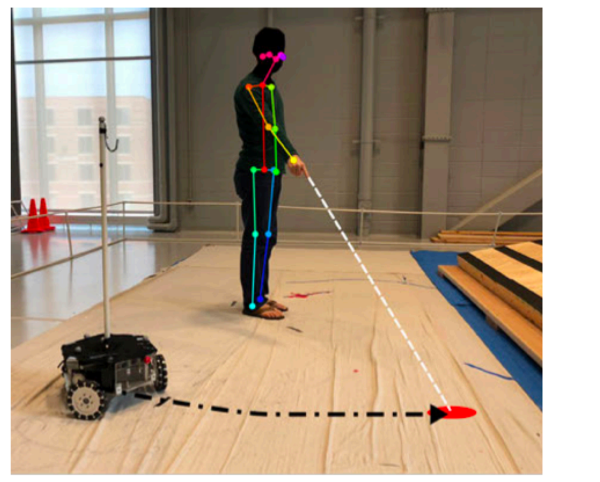


Fig. 3 Users provide pose suggestions by pointing at a position on the ground.

The overall waiting behavior for the cart can be tweaked by changing the cost weights in (2), where α_2 is the weight of the proximity layer relative to the distance layer, and α_3 is the weight of the blocking layer relative to the distance layer. Finally, $x_i^{(a)}$ is selected as a pose along a direct path between the robot and the user that is at least δ_a away from the user, such that the robot is within reaching distance. Once at $x_i^{(a)}$, the cart does not move until the user is either no longer holding an item or is moving away from the cart.

Performance Evaluation

The human-following shopping cart data was processed in two settings:

I. Multi-Cart Simulations

- A supermarket environment and 12 multi-robot team is simulated in ROS/Gazebo using modified Turtlebot3 platforms, with each robot equipped with a laser scanner, depth camera, and RGB camera plugins and access to the 2D supermarket floor plans.
- Robot poses are estimated using AMCL and then used to calculate optimal global trajectories. Obstacle avoidance and goal selection is done locally.

II. Human-Following Experiments

- A 3-wheeled omni-directional robot is constructed with forward-facing color and depth cameras for navigation and human pose and gesture recognition. Retro-reflective markers are also placed on the top surface of the robot and a set of Vicon Bonita cameras are used to collect ground truth robot odometry and human pose information.

We report results for the three robot behaviors listed below:

1. **State and Goal Switching (Fig.4)**: The robot first follows the user, then waits at a position the user points to, follows them again once they move further than $R_{\omega\to f}$, and finally, wait at a second position determined by the user.
2. **Non-Blocking Pose Selection (Fig.5)**: The three costmap layers described above are used to determine the minimal-cost waiting position across the entire supermarket. The costmap layers are combined into a single matrix, and the pose corresponding to the minimum matrix element is selected as the target pose.
3. **Gesture-based Suggestion Tracking (Fig.6)**: Pose suggestions are extracted from human finger pose estimates, transformed into the world frame, and sent to the conebot via wifi for online tracking. Finger estimates are taken over a couple of seconds and averaged using a 3rd order Savitzky-Golay filter with a 1.5s window.

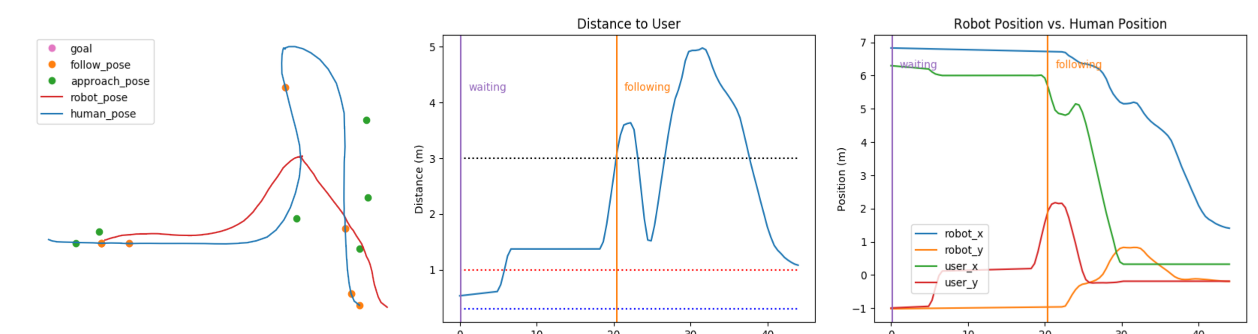


Fig. 4 State and goal switching with a simulated robot.

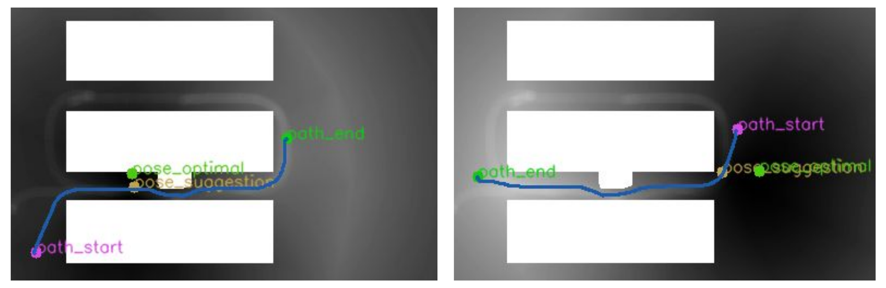


Fig. 5 Non-blocking pose selection demonstrated for a single robot.

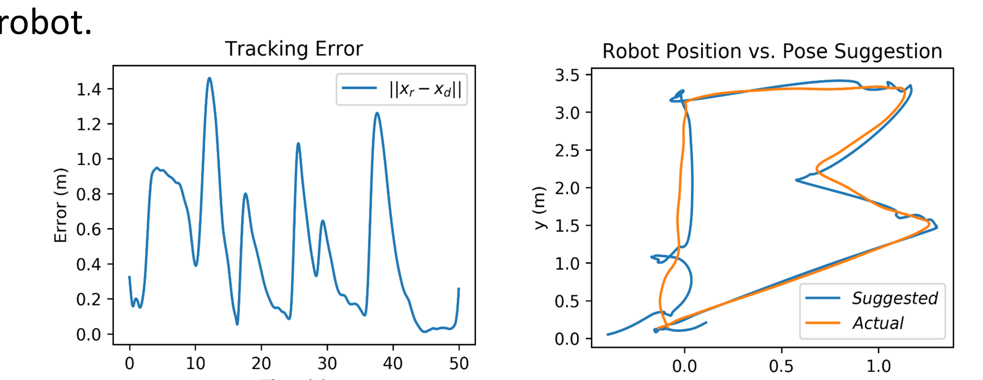


Fig. 6 Gesture-based suggestion tracking with a conebot.

Conclusion & Future Work

In this work, we presented a framework for autonomous human-following shopping carts in crowded environments with multiple users and robots. Our framework uses a finite state machine to switch between following, waiting, and approaching behavior. Users can provide pose suggestion using hand and arm gestures, which we extract from the user's skeletal estimate. We demonstrate our framework on both large-scale simulation in a supermarket setting, and with experiments using a 3-wheeled omni-directional robot.

Future work for this project includes:

- Fabrication of a prototype multi-robot shopping cart team based on the framework presented here
- On-site testing for validation

References

[1] W. He, Z. Li, and C. L. P. Chen, "A survey of human-centered intelligent robots: issues and challenges," IEEE/CAA Journal of Automatica Sinica, vol. 4, no. 4, pp. 602–609, 2017.
[2] P. Wang, W. Li, P. Ogunbona, J. Wan, and S. Escalera, "RGB-D-based human motion recognition with deep learning: A survey," Computer Vision and Image Understanding, vol. 171, pp. 118–139, 2018.
[3] C. Chen, Y. Liu, S. Kreiss, and A. Alahi, "Crowd-Robot Interaction: Crowd-Aware Robot Navigation With Attention-Based Deep Reinforcement Learning," in Proc. IEEE Int. Conf. Robot. Autom., Montreal, Canada, 2019, pp. 6015–6022.
[4] Z. Talebpour and A. Martinoli, "Risk-Based Human-Aware Multi-Robot Coordination in Dynamic Environments Shared with Humans," in Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst., Madrid, Spain, 2018, pp. 3365–3372.
[5] M. Edmonds, T. Yigit, V. Hong, F. Sikandar, and J. Yi, "Optimal Trajectories for Autonomous Human-Following Carts with Gesture-Based Contactless Positioning Suggestions," Submitted.

Acknowledgements

I would like to thank the Rutgers University Project SUPER STEM Summer Stipend Program for providing financial support for my research and want to acknowledge Dr. Jingang Yi for giving me the opportunity to work in his lab, Merrill Edmonds for teaching me skills that are necessary for my career in research, and Tarik Yigit and Faiza Sikandar for assisting me on this project.

RUTGERS