The Person Following Problem

Athulya Saravanakumar
FRI I Auton. Robots
University of Texas at Austin
athulya.saravanakumar1@gmail.com

Preetanmol Singh
FRI I Auton. Robots
University of Texas at Austin
preetanmol@utexas.edu

Nethra Middela
FRI I Auton. Robots
University of Texas at Austin
nmiddela@utexas.edu

Shreya Agrawal
FRI I Auton. Robots
University of Texas at Austin
shreyaagrawal@utexas.edu

***Abstract*—The Building Wide Intelligence (BWI Project) at the University of Texas at Austin enables robots to serve as tour guides of mapped buildings. Although these robots have robust navigation and conversational capabilities, they lack awareness of people in space around them, resulting in instances of people getting lost or left behind. This underscores the necessity for the robot to effectively track individuals walking with it. Using Azure Kinect Body Tracking SDK and the ROS Navigation stack, we've now enabled the bots to recognize people in space, follow them if triggered, and make consistent attempts at keeping the person in the frame. Our implementation utilizes a gestured based triggering mechanism to determine who is to be followed. The robot travels to the individual using ROS's built-in path planning algorithm, until it's sufficiently close enough to enter just frame centering mode, whereby it keeps the person centered in frame until they're far enough away again to translate to once again. Later, over a series of trials, we evaluate the performance of our program based on 3 main metrics: path accuracy, body accuracy tracking, and orientation accuracy. In natural walking conditions, the robot performs exceptionally well at consistently tracking and following the person.**

I. Introduction

In today's world, robots play diverse roles, ranging from delivering packages to guiding tours. As robots navigate through spaces, they must consider various factors, including the movements of people around them. Specifically, in the role of a tour guide, robots are tasked with leading visitors through different environments, be it a building or an outdoor area. However, many robots struggle to track the whereabouts of the people they're guiding. In unfamiliar surroundings, there's a high chance that individuals may become lost, confused, or simply need to pause along the way.

To enhance the effectiveness and adaptability of tour guide robots, it's essential to equip them with spatial awareness of people. Our proposed solution involves integrating computer vision and ROS (Robot Operating System) to track people's positions in real-time and follow them accurately. This means that when someone requests to be followed, the robot will utilize spatial body tracking capabilities and the ROS navigation stack, ensuring it stays aligned with the person's trajectory. By incorporating this, the robots can expand their skill set and navigate dynamic environments more proficiently.

The importance of implementing this solution lies in its ability to mirror human behavior more closely. Just as a human tour guide would pause if their group stopped, our robot, armed with these capabilities, can offer a more intuitive touring experience. Moreover, person-following algorithms have widespread applications. Many specialized robots need to be right next to their operator until they're needed. Constantly teleoperating such machines can prove exhaustive and distract one from the task at hand. Thus, autonomous following capabilities can be extremely useful. This may best be seen in a military environment, where terrain can be complex and unmapped and the optimal way is to just follow a person's lead. Evaluation criteria for our robot will include assessing the accuracy of its path tracking, body tracking, and frame orientation accuracy.

II. Background

Others have approached similar problems to recognize human movement and navigation, such as researchers at UT Austin. For example, one study approached the problem by using gaze and head orientation to predict where a person is expected to move. This study is based on the need to navigate through spaces with moving people, which led to research on the aspects of the way people move. In the study, evidence was collected that aimed to highlight the statistical connection between a person's gaze and their future movement [1]. In the work done by Sharma et al. [2], person-following is implemented using an ultrasonic and infrared camera, which is a newer approach to person-following as opposed to traditional methods. Our approach was slightly different in that we used joint body tracking with our camera's computer vision abilities. Additionally, Turing argues on whether machines can think and cites Ada Lovelace's thoughts on how "The Analytical Engine has no pretensions to originate anything. It can do whatever we know how to order it to perform" [3]. This means that the robot system will not know that it has to do something specific unless the action is emphasized. This philosophy lends itself well to our approach, as our key issue is that we want the robot to follow based on an external stimulus by a person, which is then converted to an order that the robot can understand and perform. Our approach takes these studies and insights into consideration and focuses on a similar problem of following a person based on their location in real time.

In relation to the gaze study done at UT, our approach is focused on the navigation aspect of real-time following rather than the predictive one. However, predictive analysis could be used to enhance our person-following capabilities in later stages of development. The ideas from the second study could help us improve our accuracy by using different camera types. As for the ideas from Turing's paper, our program is based on the idea of telling a robot how to perform actions through orders it can understand. These orders incorporate recognizing if a person's hand is raised through computer vision and instructing the robot to follow that person and also includes giving the robot translation and rotation matrices based on the person's position so it has the necessary format of information needed for navigation.

III. Methodology

Our principal software utilizes Azure Kinect and Robot Operating System (ROS) on the BWI bots. The Azure Kinect camera and its Body Tracking SDK drive the initial parts of our system, including the hand-raise detection component and body tracking in the whole process. To trigger the following, the person must raise their right or left hand for 4 seconds or more. We used Azure Kinect's Body Tracking SDK with spatial joint information to recognize if a person in the frame has raised their hand, and accordingly trigger the program to select that person as the one the robot should currently follow.

After the person has been selected, our code evaluates whether they are within a certain distance threshold of the robot, which we specified to be 1.75 meters. If outside this threshold, the robot navigates until it's 1.25 meters away from the person. This is done through an abstraction provided by ROS's navigation stack, which enables robust path planning and obstacle avoidance, as requests to the move base node. Hereby, it enters the frame centering mode to pivot at its spot and center the person within the camera frame. In case they move away, all frames are consistently processed and the process repeats with appropriate updates being sent to the navigation stack, which regards only the latest goal. While centering the frame, the rotational angles are scaled based on the person's distance from the robot to improve centering. Empirically, an inverse square relation between distance and turning angle was found to be optimal for natural walking speeds. This methodology reduces overshooting or undershooting while also minimizing infinite self-feedback pivots. It's important to note that all transformations are done in relative to the robot. While localizing in-door, mapped environments is not required, it is recommended for better performance. Additionally, throughout the following process, if another person raises their hand, they will be the new person to be followed.

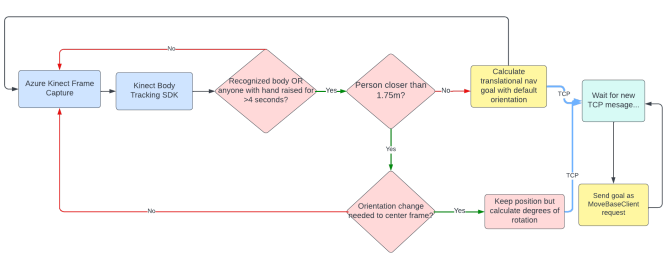

Fig. 1. Approach of Person Following Robot

IV. Experimental Setup

The three metrics we are testing our person-following algorithm on are path accuracy, body accuracy tracking, and frame orientation accuracy. For all of the tests, we ran 5 trials and saw how many times the robot succeeded at keeping track of the person and not losing them, along with getting to the right destination.

Path accuracy is a measure of whether the robot can follow the person down a given path successfully, which includes turning corners. We tested this by allotting different types of specific paths for a person to walk and saw if the robot strayed in direction/orientation or kept up successfully with the leading person. These paths include straight-line navigation, curve navigation, L-shape navigation, zigzag navigation, and obstacle navigation. For straight-line navigation, we tested two types; The first being when the person was standing 5 meters in front of the robot, and the other being when the person was moving in a straight line ahead of the robot.

The second metric we tested is body accuracy testing, which is a measure of whether the robot will accurately recognize when it should change the person it selected to follow given that a new person in the frame raises their hand. We tested this by including multiple people in the camera frame, having the robot select the person who raised their hand, making the robot follow them for some time, and then having a new person from the frame raise their hand.

Finally, our third metric, frame orientation accuracy, is a measure of how often the robot can accurately orient itself left or right based on the person's position. We tested this by counting the number of times the robot failed and succeeded in turning left and right respectively to where the person is located.

V. Results

In our evaluation of the robot's path accuracy, the robot successfully executed all test cases (5 out of 5) for approaching a person at a distance of 5 meters, following a person in a straight line, and tracking a person along a zigzag path. These results affirm the precision of linear transformations in our system.

Conversely, the robot's performance in more complex navigation scenarios was less consistent. In tests involving following a person on a circular path, the robot achieved a success rate of 4 out of 5. When navigating around obstacles while tracking a person, the success rate dropped to 1 out of 5. These outcomes suggest that rotational transformations are less robust than linear transformations. This degradation can also be attributed to the navigation stack's treatment of obstacles, where rerouting attempts are not sufficient, resulting in the abandonment of goals and eventual occlusion of the person from the frame. Given improved navigation planners, for example in ROS2, and higher positioning of Kinect cameras, we expect significant improvement in these tests. Additionally, the low success rate in obstacle-rich environments indicates a limitation in our algorithm's ability to maintain a memory of the person's identity—essential for continuous tracking after navigating past an obstacle.

The body tracking accuracy tests, involving the identification and tracking of two different individuals, yielded perfect results (5 out of 5). This demonstrates the effectiveness of our body tracking algorithm in distinguishing and seamlessly switching focus between multiple individuals, even when both are within close proximity.

The last test attempted to evaluate the robot's performance in keeping the person in frame when they're close and moving all around it. Across five trials, at various speeds and distances, it kept the person in the frame in 4 of the trials.

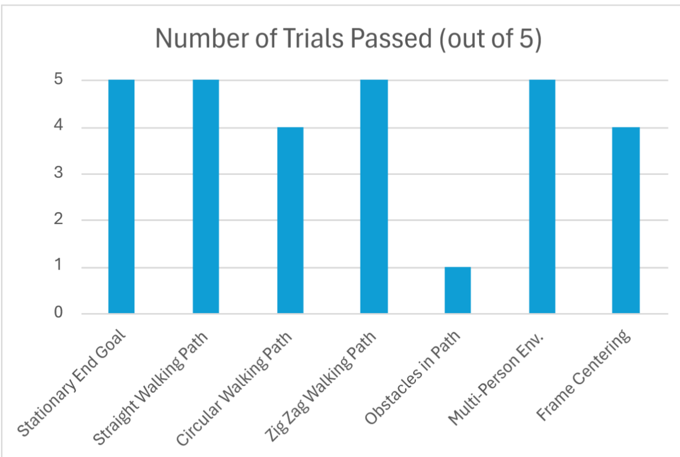

Fig. 2. Numerous Test Cases for Accuracy

VI. Discussion

Based on our evaluation, we have derived several key insights that dictate the optimal use conditions for our person-following algorithm. One significant observation is that the robot shows diminished performance when the person it is following lowers their hand. This highlights the necessity for the person to maintain their hand raised throughout the interaction, ensuring more robust tracking by the robot.

Furthermore, the robot struggled to maintain tracking when confronted with obstacles or navigating through narrow spaces. These scenarios suggest that our algorithm performs best in open, unobstructed environments where the robot can freely follow the subject without encountering physical barriers. This is particularly crucial as it underscores the importance of environment configuration in deploying our person-following technology effectively.

Additionally, our tests revealed that the robot follows a person more efficiently when they move at a steady to normal pace, rather than adjusting their walking speed or behavior excessively. This indicates that under normal walking conditions, the robot's tracking capabilities are sufficiently adaptive and robust, facilitating smooth operation without requiring the subject to alter their natural motion patterns.

In our body accuracy tracking tests, the robot demonstrated a high degree of adaptability by successfully switching its focus to a new individual when they signaled by raising their hand. This ability is pivotal, showing that the robot does not rely on facial recognition or continuous tracking of the same individual, which significantly enhances its applicability in dynamic environments where multiple potential subjects are present.

Overall, these findings affirm that, within specific operational parameters, our person-following algorithm is effective and can significantly aid in scenarios requiring autonomous robotic following. This understanding will guide further refinements and potential applications of our technology, ensuring that it meets practical needs while accommodating the inherent variability of human movement and environmental conditions.

VII. Conclusion

Our person-following algorithm was created to give robots spatial awareness that allows them to follow a specified person when they are prompted. We have accomplished this through the integration of Azure Kinect, TC/IP web sockets, and ROS navigation goals. We developed a feature using Azure Kinect such that raising a person's hand initiates the following algorithm. Additionally, we scaled the orientation angle amounts with respect to the distance of the person from the robot to allow for smoother movement.

In our testing, we evaluated our algorithm using the metrics of path accuracy, body accuracy tracking, and frame orientation accuracy. Within path accuracy, we looked at straight paths, circular paths, zig-zag paths, L-shaped paths, and paths with obstacles in them. Our findings have shown that our person-following algorithm was successful with straight and circular paths, however struggled in following a person when there were obstacles present. Also, we found that the body accuracy tracking was successful when two people switched. Our frame orientation accuracy was very accurate after implementing the angle scaling feature.

The significance of our project and our results is that we gave the robot more enhanced and efficient capabilities to recognize a person in space and follow them. This program can be implemented with tour guide and other service robots. In future work on this project, we can add varying velocities to help the robot follow a person more closely and accurately. We could also implement our algorithm into Dobby—an LLM-driven conversational robot that functions as a tour guide at UT Austin—to not lose track of the guests to whom Dobby is giving a tour.

VIII. Acknowledgment

The authors would like to thank Dr. Justin Hart and the peer mentors, without whom this project would not have been possible. They would also like to thank the University of Texas at Austin and the Freshman Research Initiative (FRI) program for allowing them this valuable opportunity to learn from the Autonomous Intelligent Robots Stream.

REFERENCES

[1] B. Holman, A. Anwar, A. Singh, M. Tec, J. Hart, and P. Stone, "Watch where you're going! gaze and head orientation as predictors for social robot navigation," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021.

[2] M. A. K. A. T. A. S. A. K. Sharma, A. Pandey and P. K. Yadav, "Human following robot," in *International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, 2021.

[3] A. M. Turing, "Computing machinery and intelligence," in *Mind 49*, 1950.