Active Tracking and Re-Identification for Mobile Robot Person Following

The University of Texas at Austin
College of Natural Sciences

The University of Texas at Austin
Computer Science

Authors: Parth Chonkar, Pranav Varanasi, Joseph Moyalan, Nevyn Duarte, Blake Holman, Nick Wille, Haresh Karnan
Dr Justin Hart.
Department of Computer Science, College of Natural Sciences, The University of Texas

Abstract

In this work, we are developing a system that enables a robot to follow a certain person in a crowded area. With many people in present, most algorithms have difficulty differentiating between the tracked person and other people in the scene. Although facial recognition is effective, the target's face is usually obstructed during person following. Additionally, other people can block the target completely. Trajectory-based person-following algorithms are useful until a need for explicit re-identification arises, which occur when targets leave the image frame captured by the robot's camera. We improve on existing person-following systems by combining a trajectory-based approach with a triplet loss metric for target re-identification. This approach improves the system's ability to re-identify people who have walked out of frame. While the target is in frame, the trajectory based tracker allows for accurate tracking.

Research Goal

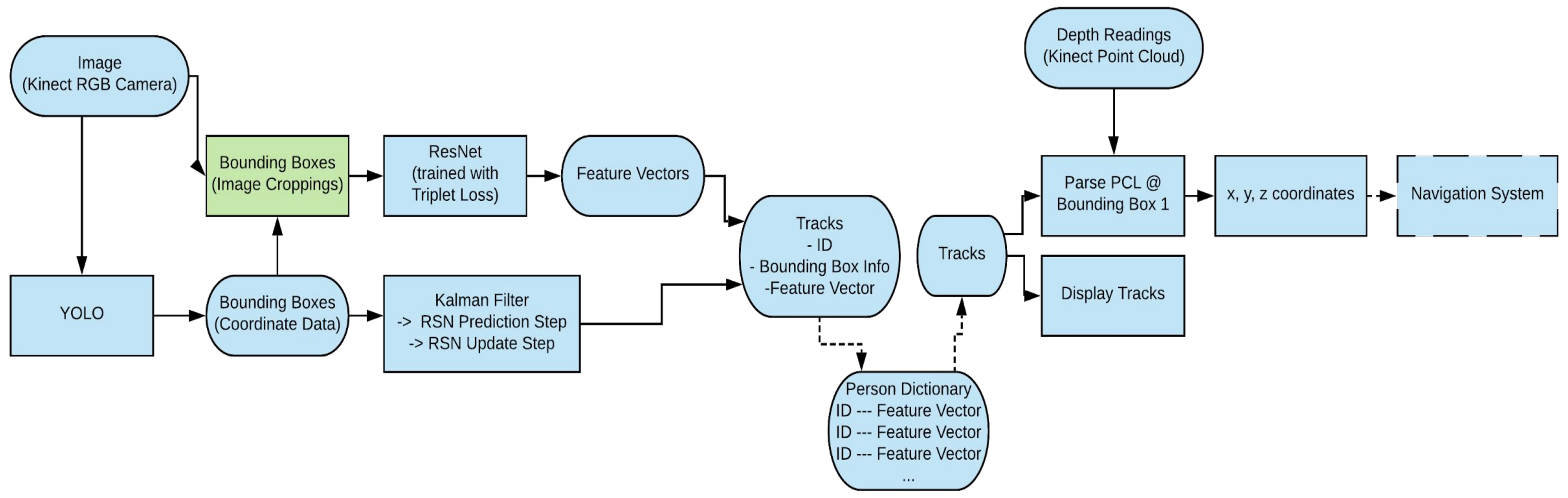
Our goal is to allow robots to follow a person through a crowd by maintaining a person's identity despite other people occasionally obstructing the view or the target going out of frame.

Backround

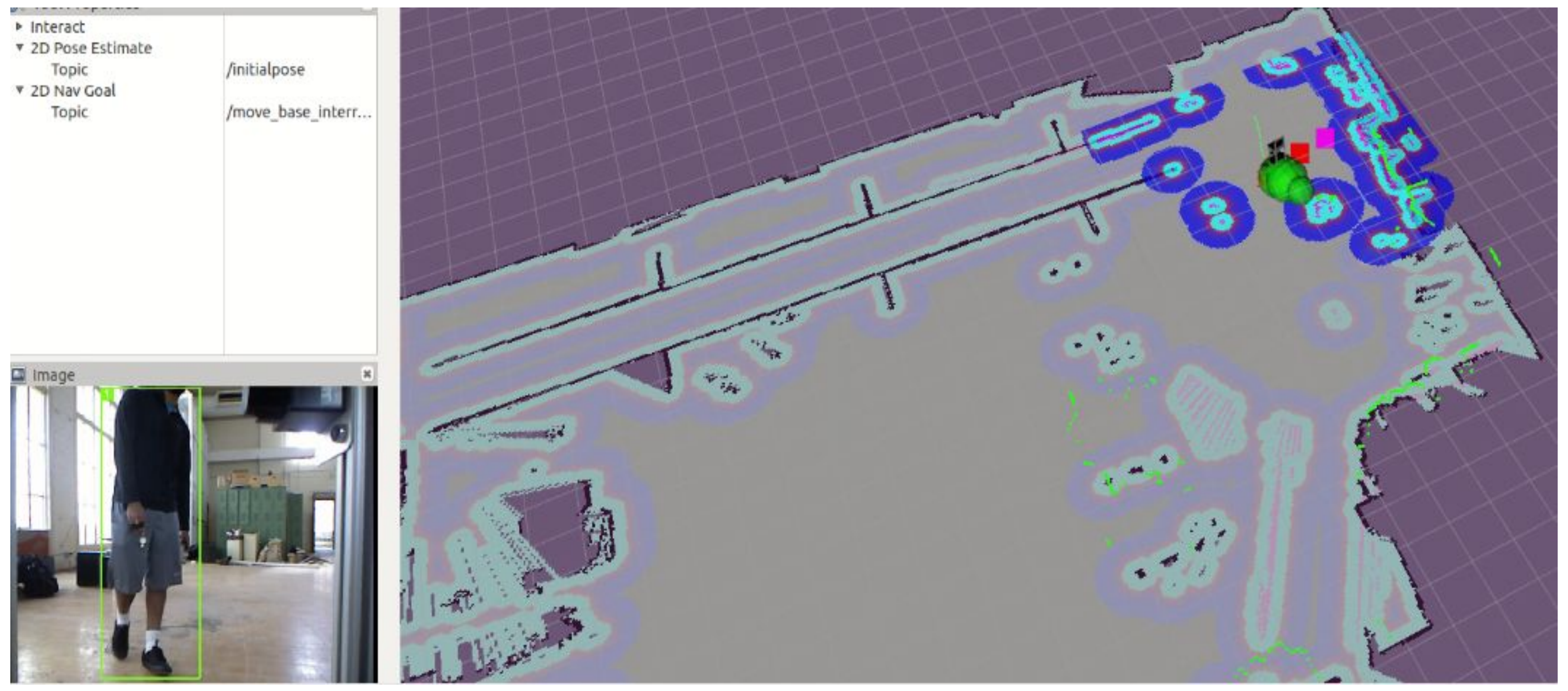
Recent approaches to this problem make use of detection based tracking and detection-free tracking etection-based tracking identifies in order to track a target based on its appearance. On the other hand, detection-free tracking tracks a target based on cues such as trajectory. Our method uses DeepSORT, a detection based tracker that relies on a kalman filter. DeepSORT uses a neural network to calculate association metrics for identity matching and uses this in conjunction with detection trajectories to track and uniquely identify persons in a video stream. Our novel contribution is to augment this technique with the Triplet Loss function, which anchors with a known positive match, and compares it against two negative matches to train neural networks and create feature embeddings from a set of images. The distance from the anchor to the positive embedding is minimized, and the distance from the anchor to the negative embedding is maximized, increasing the accuracy of the identifications.

Methodology

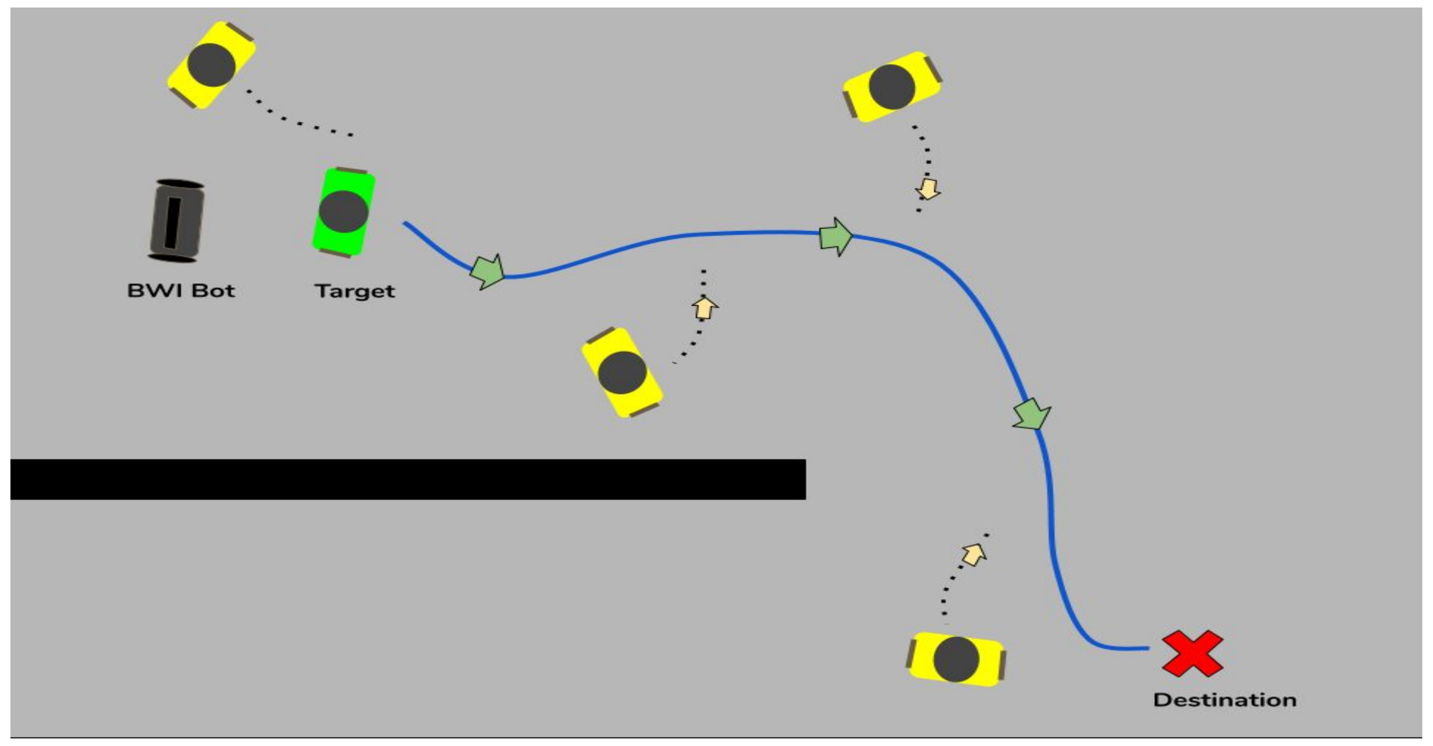
The tracking system uses information from an RGB-D camera, which provides color and depth information. The target and other people are tracked in the camera-frame using a modified DeepSORT system. While their trajectories remain within the frame, their corresponding bounding-box images are used to store their triplet embeddings, which are generated using a feature-embedding network trained on the MARS dataset using the triplet loss metric. The MARS dataset consists of image frames of over 1200 videos of walking pedestrians, where the images are cropped to identify each person. When trajectories are lost from exiting the frame or occlusion, the triplet metrics are used for re-identification. The target also has trajectory tracker in relation to the ground as well as the camera-frame.



To test this software, we are using the Building Wide Intelligence Project's BWI Bot platform [1]. We use a series of LIDAR sensors to track the positions of people and objects in the environment and the tcolor camera mounted to the front of the robot as our image feed.



Planned Experiments



We need to test two basic scenarios that our robot must be able to deal with when following a person:
Scenario A: The person goes around the corner of a hallway, which obstructs them from the robot's point of view.
Scenario B: The person walks through a crowd of people who walk between the robot and the person.
In both scenarios, a trial is considered successful if the robot is able to position itself within a half-meter of a person at the destination. We will compare these results to a successful baseline person following system, Monocular Person Tracking and Identification with On-line Deep Feature Selection for Person Following Robots [5].

References

1) BWIBots: A platform for bridging the gap between AI and human-robot interaction research: 2017 May, 19 2019 https://journals.sagepub.com/doi/full/10.1177/0278364916688949f

2) Zhai, et al. "In Defense of the Classification Loss for Person Re-Identification." ArXiv.org, 29 Nov. 2018, https://arxiv.org/abs/1809.05864.

3) MARS: A Video Benchmark for Large-Scale Person Re-Identification. Zheng, LiangBie, ZhiSun, YifanWang, JingdongSu, ChiWang, ShengjinTian, Qi.(2016). https://www.researchgate.net/publication/308277502

4) Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, et al. Deep Residual Learning for Image Recognition" 2015, http://arxiv.org/abs/1512.03385.

5) Luo, et al. "Multiple Object Tracking: A Literature Review." ArXiv.org, 22 May 2017, https://arxiv.org/abs/1409.7618