DoorBot, a Platform for Interacting with Humans and Monitoring Crowd

Saeid Amiri
and Azin Shamshirgaran

Abstract With the achievements in robotics and artificial intelligence, robots may be able to replace humans in many tasks and jobs in future. There has been extensive research on mobile autonomous robots in these areas. Using a mobile robot and taking advantage of human-detection algorithms, we propose a robot agent that can successfully detect and track humans, and detect their intention of interaction based on their movement trajectory. The long-term goal for this project is to use the mobile robot assisting people in public places such as museums, hotels, airports, etc. To do so, we trained a classifier which is based on recurrent neural network using a publicly available dataset as an input. To obtain human trajectories, we have used human detection and tracking algorithms using RGB-D information. To evaluate the accuracy of our system, we conducted an experiment on human intention prediction using RGB-D vision at Cleveland State University main classroom building and the results are promising for further improvement.

1 Introduction

As robots become more prevalent in the everyday life, it is estimated that many jobs will be taken by the robots by the year 2030. Studies show that humans who work in shifts are more prone to heart coronary diseases [8] and excess of low HDL-cholesterol [4]. Given the mentioned facts, we believe leverage of robots in 3rd shift jobs can be beneficial to workers. In this project, we have introduce *DoorBot*, a robot that can be replaced for doormen at places such as museums, hotels, hospitals, etc in the 3rd shift. This idea might not be economic with current prices of robots, but in a decade or two, we anticipate cheap prices for such robots. Doormens' main responsibilities include finding interested people or people who need help, greeting with them, monitoring the entrance and departure of people.

Keywords: Human Detection, Human-Robot Interaction

Figure 1: Segway RMP110 in CSU Engineering building.

However, it is quite important for our agent to be able to initially recognize people's intentions. If they are not walking towards the robot, most likely they do not want to be distracted by it. On the other hand, if they are seeking help or walking towards the robot, they may need assistance and robot can initiate the interaction with them by stepping forward or starting a conversation. By using a public dataset, we have trained a classifier to detect human's intention of interaction using their trajectory as the classifier input. To obtain human's trajectories, we have used human detection and tracking algorithms using RGBD information.

In this report, we are going to implement this idea. First, the human detection with RGB-D sensor has been described. Then, we give details how we get the trajectories of detected people. And, we will train the classifier with these trajectories as an input. Finally, experiments and results are provided at the end to evaluate our implementation to show how the classifier detect interested people based on the trajectory.

1.1 Hardware Segway RMP110 produced by Stanley Innovation was our main research platform. The robot is equipped with Sick Tim laser sensor with a detection range of 25 meters[6]. Figure 1 shows this platform at the engineering building of Cleveland State university. We used Kinect camera to receive RGBD information needed for human detection.

1.2 Software Robot Operating System (ROS) [9] is the open-source framework in the robotics and its modular design can facilitate working with multiple packages and adding extra features to our project. Keras[1], which is an open-source Python library was used for training classifiers that will be discussed with more details in the next section.

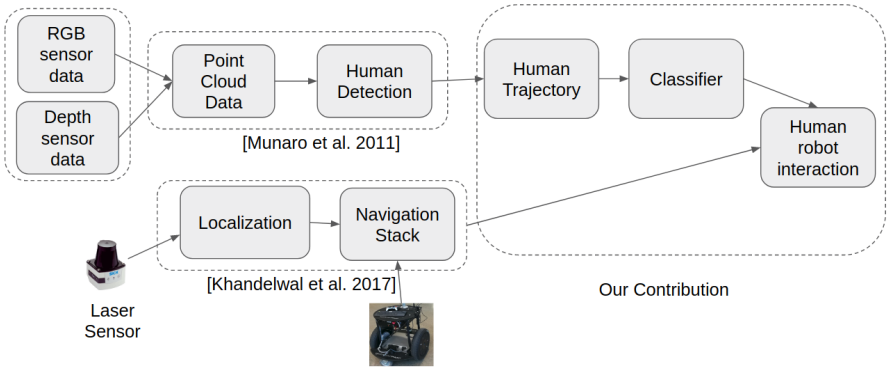

Figure 2: The software infrastructure of DoorBot

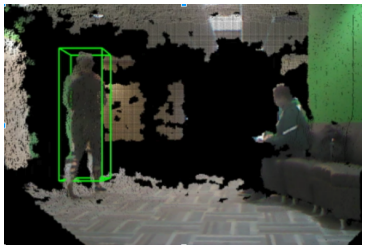

Figure 3: Example of human detection with RGB-D sensor.

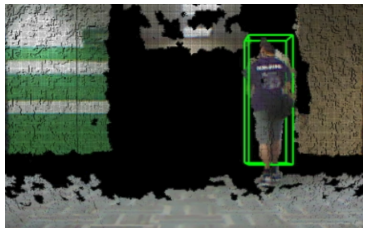

Figure 4: Example of human detection with RGB-D sensor.

2 Implementation

2.1 Overview Figure 2 shows the software infrastructure and the hierarchy of different packages. Below we will discuss how each module works:

2.1.1 Human-Detection: In order to be able to detect humans, we have used the algorithm developed by [7] and [10]. Paper [7] proposed a multi-people tracking algorithm with RGB-D data for static or mobile robots. Their algorithm is able to detect humans in ground plane in real time with only CPU computations. In their work, they leveraged an on-line learned classifier extracting features from color histogram to detect human features. Then it is improved by [10].

The main contributions of their works are a 3D sub-clustering method that allows to efficiently detect people very close to the background. Figure 3 and Figure 4 show examples of clustering of a detected human based on this paper approach.
Paper [2] used this method for Segway robot. This paper considered the problem of recognizing spontaneous human activities from a robot's perspective. They presented a novel dataset, where data is collected by an autonomous mobile robot moving around in a building and recording the activities of people in the surroundings.

The RGB-D data are processed by a detection module that filters the point cloud data, removes the ground and performs a 3D clustering of the remaining points.

Furthermore, they apply a HOG-based people detection algorithm to the RGB image of the resulting clusters in order to keep only those that are more likely to belong to the class of people. The resulting output is a set of detections that are then passed to the tracking module.

Since they make the assumption that people walk on a ground plane, the algorithm estimates and removes this plane from the point cloud provided by the voxel grid filter. The plane coefficients have been computed with a RANSAC-based least square method and remove all the inliers within a threshold distance. The ground plane equation is updated at every frame by considering as initial condition the estimation at the previous frame, thus allowing real time adaptation to small changes in the floor slope or camera oscillations typically caused by robot movements.

Once this operation has been performed, the different clusters are no longer connected through the floor, so they could be calculated by labeling neighboring 3D points on the basis of their Euclidean distances.
However, this procedure can lead to two typical problems: (i) the points of a person could be subdivided into more clusters because of occlusions or some missing depth data; (ii) more persons could be merged into the same cluster because they are too close or they touch themselves or, for the same reason, a person could be clustered together with the background, such as a wall or a table.

To solve the problem (i), after performing the Euclidean clustering, we merge clusters that are very near in ground plane coordinates, so that every person is likely to belong to only one cluster. For concerns problem (ii), when more people are merged into one cluster, the more reliable way to detect individuals is to detect the heads, because there is a one to one person-head correspondence and heads are the body parts least likely to be occluded.

From these considerations this paper implemented the following algorithm, that detects the heads from a cluster of 3D points and segment it into sub-clusters according to the head positions:

- for every cluster a height map is created along the direction corresponding to the image x axis.
- local maxima are searched for within the height map.
- only maxima farther than a threshold distance in ground plane coordinates are kept because people heads are not often nearer than the intimate distance, usually equal to 0.3m.
- a sub-cluster is created for every remaining maximum and points nearer than the intimate distance in ground plane coordinates are associated to it.
- sub-clusters with too few points or not enough high are discarded.

2.1.2 Training classifier: Once a human is detected, the robot stores the information of the human and at the end plots them on the 2D room map. Figure 5 shows some of the detected

Now that we could detect and track the humans, we get the relative position of the detected person from the robot's point of view. By having the human's trajectory,

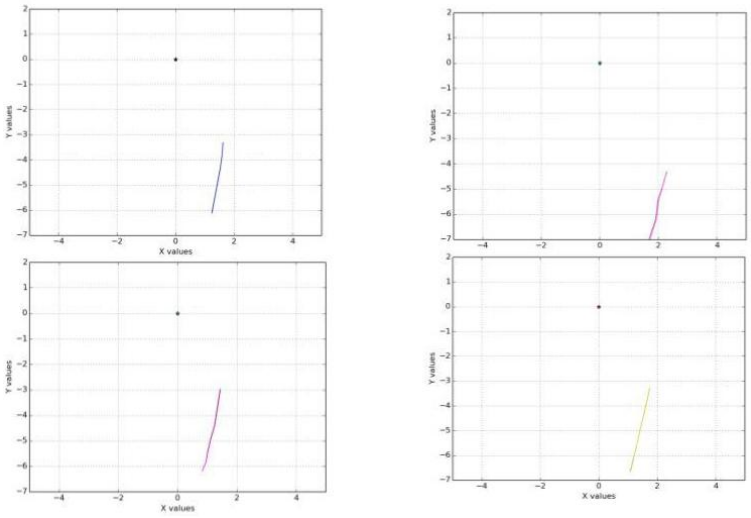

Figure 5: Trajectories of different people walking in the main classroom hallway.

the robot needs a classifier to distinguish what people are interested to interact with the robot and what people are not.
The public dataset of [5], contains the trajectory information of people walking in a shopping center in proximity of a robot. Each person's information includes their global position, their body and head orientation and the label that indicates whether they were intended to interact with the robot in the shopping center or not. These information are presented in time series. Therefore, we have a label classification problem.
Recurrent neural networks are a class of artificial neural network architecture that use iterative function loops to store information [3] and one of their applications is supervised sequence labeling. We used RNN-LSTM (long-short-term-memory) networks to avoid problems such as vanishing/exploding gradients that are common in such networks.
General classifier training requires tuning of different parameters. The number of memory units in the network was one of the main parameters that needed to be tuned. Eventually, an LSTM network with 50 memory units resulted in best precision and recall values on the test set. Also the input data varied in length. We resized all the input trajectory vectors to 30 timesteps by either interpolating or downsampling the data.

The classifier training implementation was done in Keras, an open-source python library used for fast and easy classifier training[1]. The data was split into training and testing sets. The feature vector contained only human relative position to the robot as our hardware is not able to distinguish human body or face orientation.

Table 1 Confusion matrix showing the distribution of correct and incorrect predictions of the robot.

	Ground truth: Human intended	Ground truth: Human not intended
Robot's prediction: Human intended	4 (30 %)	9 (70 %)
Robot's prediction: Human not intended	1 (2 %)	51 (98 %)

3 Experiments

In order to evaluate the performance of the the trained classifier and the human detection algorithm, we tested our system in the Main Classroom building at Cleveland State university where a great number of students walk in the hallways most of the time. The human detection algorithm was able to successfully detect people within the range of the depth sensor and was able to store and plot the trajectories. Overall, during experiments robot detected 65 people and in table 1 is the confusion matrix indicating the correctness of the robot's prediction.

3.0.1 Implementation Challenges Quite similar to the imbalance of labels in our dataset, detected people were hardly interested to interact with the robot and come towards it. In a few cases, interested people approached the robot from behind where robot has no eyes to detect them. In some case, robot incorrectly detected people as intended to interact with the robot, however, it usually happened when humans were uncertain about their destination and robot felt that it can initiate the interaction.
Ideally, the robot was supposed to move towards the person to start interaction. However, the Kinect camera had to be always plugged in to an external power. We used extension cords to solve this issue but it was hazardous for the robot and people because robot needed to update its localization by turning around and the cord was being twisted around it. Also people might trip over the long cord if robot started moving in a crowded hallway. Therefore, the robot only played a greeting message "How can I help you?" to start interaction.

4 Conclusion

We were able to use a mobile robot to detect people and recognize their intention of interaction based on their trajectories. The experimental results showed that the robot can intelligently detect people's intentions. However, only a small number of people were interested to interact with the robot even in a crowded place. However, it is quite important to mention that more experiments are needed to better evaluate the performance of doorBot. The authors assume that doorBot can be more helpful in places such as museums where people do not walk too fast and are more in need of assistance.

5 Future Work

There is a huge potential to improve the current system in hardware and software. In hardware, by adding multiple RGB-D sensors, robot can have wider field of view to detect people even from behind. Also the evaluation can be tested with other types of RNN classifiers to get better results in terms of accuracy. The implementation can

be even applied to humanoid robots that can probably be more aesthetically pleasant to the people.

References

[1] Chollet, F. et al., "Keras," https://keras.io, 2015. 2, 5

[2] Gori, I., J. Sinapov, P. Khante, P. Stone, and J. Aggarwal, "Robot-centric activity recognition in the wild," in *International Conference on Social Robotics*, pp. 224–234. Springer, 2015. 3

[3] Graves, A., "Supervised sequence labelling," pp. 5–13 in *Supervised sequence labelling with recurrent neural networks*, Springer, 2012. 5

[4] Karlsson, B. H., A. K. Knutsson, B. O. Lindahl, and L. S. Alfredsson, "Metabolic disturbances in male workers with rotating three-shift work. results of the wolf study," *International archives of occupational and environmental health*, vol. 76 (2003), pp. 424–430. 1

[5] Kato, Y., T. Kanda, and H. Ishiguro, "May i help you?: Design of human-like polite approaching behavior," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pp. 35–42. ACM, 2015. 5

[6] Khandelwal, P., S. Zhang, J. Sinapov, M. Leonetti, J. Thomason, F. Yang, I. Gori, M. Svetlik, P. Khante, V. Lifschitz et al., "Bwibots: A platform for bridging the gap between ai and human–robot interaction research," *The International Journal of Robotics Research*, vol. 36 (2017), pp. 635–659. 2

[7] Munaro, M., F. Basso, and E. Menegatti, "Tracking people within groups with rgb-d data," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, pp. 2101–2107. IEEE, 2012. 3

[8] Nakamura, K., S. Shimai, S. Kikuchi, K. Tominaga, H. Takahashi, M. Tanaka, S. Nakano, Y. Motohashi, H. Nakadaira, and M. Yamamoto, "Shift work and risk factors for coronary heart disease in japanese blue-collar workers: serum lipids and anthropometric characteristics," *Occupational medicine*, vol. 47 (1997), pp. 142–146. 1

[9] Quigley, M., K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "Ros: an open-source robot operating system," in *ICRA workshop on open source software*, volume 3, page 5. Kobe, Japan, 2009. 2

[10] Zhang, S., D. Lu, X. Chen, and P. Stone, "Robot scavenger hunt: A standardized framework for evaluating intelligent mobile robots.," in *IJCAI*, pp. 4276–4277, 2016. 3

Amiri
Department of Computer Science
Cleveland State University
2121 Euclid Ave
Cleveland OH 44115
USA
s.amiri@vikes.csuohio.edu

Shamshirgaran
Department of Computer Science
Cleveland State University
2121 Euclid Ave
Cleveland OH 44115
USA
ashamshirgaran@vikes.csuohio.edu.edu