2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 25-29, 2020, Las Vegas, NV, USA (Virtual)

An interactive drink serving social robot: Initial System Implementation

Lakshadeep Naik[1], Oskar Palinko[1], Avgi Kollakidou[1], Leon Bodenhagen[1] and Norbert Krüger[1]

***Abstract*— This paper describes the system architecture for an autonomous interactive drink serving robot developed for use in public spaces and at social events. The proposed system design focuses on finding the balance between technological readiness and social readiness levels thus enabling a technology that can be deployed in real-world environments along with social acceptance. We describe different components required for designing such a system and discuss both their technical feasibility and social acceptance aspects. We also present a behavior tree based software architecture for integrating these components, which results in modular design and promotes their re-usability.**

I. INTRODUCTION

Dehydration can have adverse effects on our mental, cognitive and emotional health[1], [2]. Thus, it is important to drink water frequently to prevent dehydration. It is advised to drink more than 2 litres of water every day even when we are not thirsty[3]. However, studies[4] have shown that just being conscious about these facts is not sufficient to motivate people to drink water as there are no immediate rewards. An effective way of solving this is by constantly motivating people and reducing friction[4], [5] i.e. make it as easy as possible for people to drink water. Drinking water regularly is even more important for elderly individuals[6]. In elderly care homes, health care workers have to spend their time encouraging people to drink more water.

II. OBJECTIVES

We present a mobile robot which motivates and encourages people to drink more water while offering them a cup. The robots primary task is to detect people, approach them in a socially acceptable manner, interact with them and offer them water.

The targeted application poses various challenges in the robotics context:

- Reliable drink serving
- Safe and socially aware navigation around humans
- Approaching humans in a socially acceptable manner
- Understanding human intentions and interacting with them intuitively
- A modular and reactive behavior control architecture that simplifies safety certification process and enables long term deployments

Our work focuses on finding solutions to these problems that are socially acceptable as well as ready to deploy in real-world environments. In the following sections, we present our approach to solving these challenges and discussions about future work.

III. SYSTEM OVERVIEW

A. Hardware

We use the SMOOTH robot for the given task[7]. It is a penguin-shaped[8] mobile social robot as shown in Fig. 1. Even a simple task such as pouring water, grasping or releasing a glass of water is still a challenging task for robots[9], [10]. Thus, they cannot be deployed reliably in un-controlled environments. Hence, instead of using a manipulator arm, we have developed a drink serving tray which can be installed on the robot as an add-on module as shown in Fig. 1 . It has infrared sensors for each cup on the bottom of the tray which provides feedback to the robot on how many glasses were taken from the tray. Intuitive interaction methods[11] are then used to motivate and encourage people to take a glass of water from this tray.

Fig. 1. SMOOTH robot with an add on tray module for serving drinks

B. Behavior control architecture

In complex software systems such as robotics, it is desirable to design systems from reusable and maintainable components. This becomes even more important during long term deployment and for the safety certification process. In recent years, behavior trees have emerged as a reliable solution for reactive and modular architectures in robotics[12]. They enable to create behaviors that are configurable, composable, reusable and reactive[13]. Our behavior control architecture has been designed using a behavior tree. Behavior tree nodes implement simple behaviors such as go to goal, turn by specific degrees, detect people, group people, plan, get attention, interact with people etc. In some cases, these nodes communicate with external components to provide the desired functionality. The system uses four main external components: perception, social grouping, socially aware

*This work was supported by the SMOOTH project (6158-00009B) of the Innovation Fund Denmark.

[1]SDU Robotics, Maersk Mc-Kinney Moller Institute, University of Southern Denmark, Odense, Denmark
lana,ospa,avko,lebo,norbert@mmmi.sdu.dk

978-1-7281-6211-9/20/$31.00 ©2020 IEEE

navigation and human-robot interaction. All these simple behaviors are combined using standard nodes in the behavior tree to execute drink serving behavior.

C. Perception

Perception component is used for estimating 3D pose and gaze angle of the person. The 3D position of the person is estimated using a human pose estimation CNN on RGB-D data and orientation is estimated based on facing direction of people in the camera's field of view (FoV). A tracker is used to assign unique IDs to the detections as well as reducing noise on them, in order to provide temporal consistency[14]. Gaze angle of the person is estimated using OpenFace[15]. The 3D pose estimation of people is used for approaching the people, while the gaze estimation angle is used to determine if a person is interested in interacting with the robot.

D. Approaching humans

People maintain a certain distance while interacting with a person. In case of a close friend it is usually between 45 cm to 120 cm. While in the case of acquaintances, it is normally more than 120 cm[16]. Similarly, for static social groups there exists a concept of o-space which is an area reserved explicitly for interaction. It exists in the form of a circle as shown in Fig. 2. Anyone interested in joining the conversation is expected to place themselves just outside the o-space to express intention to join the interaction[16]. We identify F-formations among the detected people using a hierarchical clustering method which employs a customary distance function for the incorporation of the body orientation. Multiple approach points are then generated for approaching detected F-formations by incorporating various social rules as discussed above (see Fig. 2). Approach-points are prioritized based on social conventions. The robot selects the first way-point to which it can successfully plan the motion from the prioritized list of way-points and starts the interaction.

E. Socially-aware navigation

As robots are moving into social environments, it has resulted in the need for human-aware or socially-aware navigation[16], [17] which is the intersection between human-robot interaction and robot motion planning. While these ideas are still evolving, ROS navigation stack provides a stable solution for navigating in indoor environments[18]. However, since it's solely based on geometric details of the environment, it doesn't exhibit socially acceptable behaviors. We follow the approach described in [19] and added a social layer to enable human-aware navigation using a ROS navigation stack. The social layer adds costs to the cost-map based on the o-space of detected humans and their F-formations[20] as shown in Fig. 2. The robot then plans its motion using this cost-map.

F. Human-robot interaction

Human-robot interaction involves various steps such as getting the attention of the person, understanding a person's intention and then effectively communicating the robot's intentions to the person. We use speech to initiate the communication. The robot then tries to understand the person's intention for interaction by estimating their gaze angle and starts a conversation to offer a drink accordingly. In the case of F-formation where there are multiple people, the robot has to ensure that it is addressing every member of the group. We achieve this by using robot's eye gaze to select each person[11] as shown in Fig. 3. DialogOS - an open-source dialog manager[21] is used for managing the dialog.

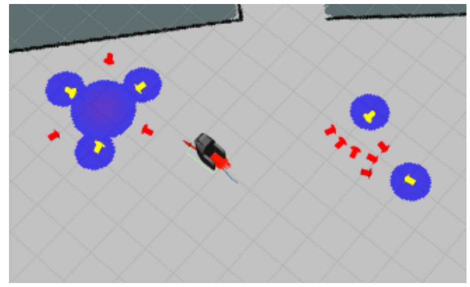

Fig. 2. Detected F-formations, social cost-map and approach points.(Yellow arrows indicates detected people, red arrows indicates planned approach points and blue circles indicate the o-space circles added to the costmap based on detected people and F-formations)

Fig. 3. Robot interacting with people using gaze and speech and offering them drinks

IV. CONCLUSIONS AND FUTURE WORK

We expect that the designed and implemented system for social interaction will be adequate for solving the drink serving scenario. In the near future, we will continue refining our interaction models contained in the behavior trees. We further plan to conduct human-robot interaction studies with invited participants and also "in-the-wild". We will continue upgrading our system based on information gathered during tests. We believe that SMOOTH will be successful at offering water to people and thus make sure that they keep hydrated.

REFERENCES

[1] P. Gopinathan, G. Pichan, and V. Sharma, "Role of dehydration in heat stress-induced variations in mental performance," *Archives of Environmental Health: An International Journal*, vol. 43, no. 1, pp. 15–17, 1988.

[2] M. G. Wilson and J. Morley, "Impaired cognitive function and mental performance in mild dehydration," *European Journal of Clinical Nutrition*, vol. 57, no. 2, pp. S24–S29, 2003.

[3] I. of Medicine (US). Panel on Dietary Reference Intakes for Electrolytes and Water, *DRI, dietary reference intakes for water, potassium, sodium, chloride, and sulfate*. National Academy Press, 2005.

[4] W. Wood, *Good Habits, Bad Habits: The Science of Making Positive Changes That Stick*. Pan Macmillan, 2019.

[5] W. Wood and J. M. Quinn, "Habits and the structure of motivation in everyday life." 2005.

[6] M. Ferry, "Strategies for ensuring good hydration in the elderly," *Nutrition reviews*, vol. 63, no. suppl_1, pp. S22–S29, 2005.

[7] W. K. Juel, F. Haarslev, E. R. Ramírez, E. Marchetti, K. Fischer, D. Shaikh, P. Manoonpong, C. Hauch, L. Bodenhagen, and N. Krüger, "Smooth robot: Design for a novel modular welfare robot," *Journal of Intelligent & Robotic Systems*, vol. 98, no. 1, pp. 19–37, 2020.

[8] E. Marchetti, W. K. Juel, R. M. Langedijk, L. Bodenhagen, and N. Krüger, "The penguin–on the boundary between pet and machine. an ecological perspective on the design of assistive robots for elderly care," in *International Conference on Human-Computer Interaction*. Springer, 2019, pp. 425–443.

[9] F. F. Goldau, T. K. Shastha, M. Kyrarini, and A. Gräser, "Autonomous multi-sensory robotic assistant for a drinking task," in *2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR)*. IEEE, 2019, pp. 210–216.

[10] C. C. Kemp, A. Edsinger, and E. Torres-Jara, "Challenges for robot manipulation in human environments [grand challenges of robotics]," *IEEE Robotics & Automation Magazine*, vol. 14, no. 1, pp. 20–29, 2007.

[11] O. Palinko, K. Fischer, E. Ruiz Ramirez, L. Damsgaard Nissen, and R. M. Langedijk, "A drink-serving mobile social robot selects who to interact with using gaze," in *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 2020, pp. 384–385.

[12] M. Colledanchise and P. Ögren, *Behavior trees in robotics and AI: An introduction*. CRC Press, 2018.

[13] D. Faconti, "Mood2be: Models and tools to design robotic behaviors," 2019.

[14] W. Juel, F. Haarslev, N. Krüger, and L. Bodenhagen, "An integrated object detection and tracking framework for mobile robots," in *Proceedings of the 17th International Conference on Informatics in Control, Automation and Robotics - Volume 1: ICINCO*. SCITEPRESS Digital Library, 2020, pp. 513–520.

[15] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, "Openface 2.0: Facial behavior analysis toolkit," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 2018, pp. 59–66.

[16] T. Kruse, A. K. Pandey, R. Alami, and A. Kirsch, "Human-aware robot navigation: A survey," *Robotics and Autonomous Systems*, vol. 61, no. 12, pp. 1726–1743, 2013.

[17] X.-T. Truong and T. D. Ngo, "Toward socially aware robot navigation in dynamic and crowded environments: A proactive social motion model," *IEEE Transactions on Automation Science and Engineering*, vol. 14, no. 4, pp. 1743–1760, 2017.

[18] R. L. Guimarães, A. S. de Oliveira, J. A. Fabro, T. Becker, and V. A. Brenner, "Ros navigation: Concepts and tutorial," in *Robot Operating System (ROS)*. Springer, 2016, pp. 121–160.

[19] D. V. Lu, D. Hershberger, and W. D. Smart, "Layered costmaps for context-sensitive navigation," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2014, pp. 709–715.

[20] F. Setti, C. Russell, C. Bassetti, and M. Cristani, "F-formation detection: Individuating free-standing conversational groups in images," *PloS one*, vol. 10, no. 5, p. e0123783, 2015.

[21] A. Koller, T. Baumann, and A. Köhn, "Dialogos: Simple and extensible dialog modeling," 2018.