Human Motion Prediction Based on Object Interactions

Lilli Bruckschen Nils Dengler Maren Bennewitz

***Abstract*— In this paper, we consider the problem of predicting the navigation goal of a moving human in an indoor environment. Knowledge about this goal can greatly increase the efficiency of robots acting in the same environment as interferences can be avoided and assistance quickly provided if necessary. Often the navigation goal depends on the previous action of the human and the object the human has interacted with before. Thus, the information about previous object interactions can be used to infer possible objects the human will interact with next, which in term can be used to predict the current navigation goal. We propose to learn a probability distribution of subsequent object interactions and present a framework that utilizes the learned transition model as well as observations of the human's location and pose for the prediction of their movement goal. As we show in various experiments, the information about transition probabilities of object interactions leads to reliable predictions of the navigation goal and improves the accuracy compared to prediction approaches that rely only on spatial information and do not consider object interactions. Furthermore, we demonstrate how the prediction can be used to realize foresighted robot navigation.**

I. Introduction

The prediction of human movements is of great importance for any robot that operates in the same environment. A service robot can utilize this prediction for anticipating when and where the human may need assistance with certain tasks as well as for avoiding interferences by inferring the trajectory of the human.

Previous approaches to movement prediction mainly focus on learning typical human trajectories in the environment [1], [2] or with respect to other humans and obstacles in the local surrounding [3] and neglect information about subsequent objects with which the human interacts. However, humans naturally interact with objects while moving through the environment, e.g., with a plate, table, cup, coffee machine, fridge, TV, laptop, etc.

In this work, we consider the problem of predicting the next destination of a moving human inside an indoor environment utilizing information about their interactions with objects in combination with observations about their movement. We propose to use the information about the previous object the human has interacted with to infer possible objects the human will interact with next, which in term can be used to predict the human's future motions and the navigation goal. Fig. 1 shows a motivating example of our approach in which our system predicts the human's navigation goal to be the table since an interaction with a cup is detected.

All authors are with the Humanoid Robots Lab, University of Bonn, Germany. This work has been supported by the DFG Research Unit FOR 2535 Anticipating Human Behavior.
978-1-7281-3605-9/19/$31.00 ©2019 European Union

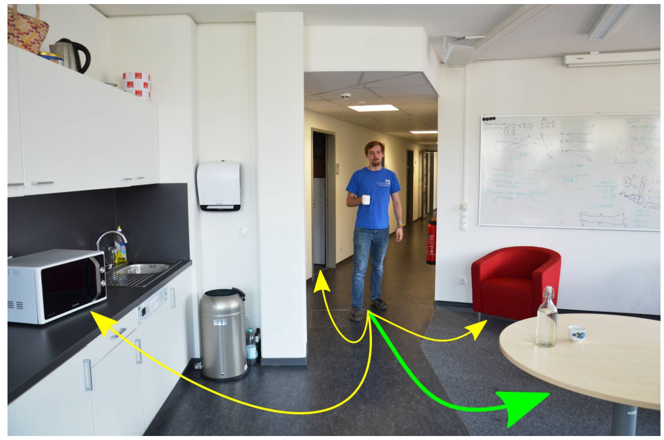


Fig. 1: The aim of our framework is to infer the navigation goal of a moving human. In this example, our system detects an interaction with a cup and predicts the table to be the most likely next object the human will interact with. The position of the table is therefore the most likely navigation goal of the human.

Our approach relies on learned transition probabilities of object interactions and uses this information in combination with observations about the current human's location and pose to predict the navigation goal using a recursive Bayes' filter. As we show in extensive experiments in different environments, the information about transition probabilities of object interactions in combination with observations of the human's location and pose yields reliable estimates about the navigation goal. Furthermore, we can improve the prediction accuracy compared to approaches that rely only on spatial information and which do not consider object interactions. To the best of our knowledge, our approach to human movement prediction is the first one that explicitly considers object transition probabilities to infer navigation goals. As we demonstrate in a real-world experiment, a mobile robot can use our prediction framework for foresighted navigation and positioning.

II. Related Work

Several approaches to human motion prediction exist. Vasquez *et al.* [4] propose to create a joint probability distribution to predict the movement of a human based on observed position changes, a pre-trained cost-based prediction model, and a gradient-based goal prediction function. In contrast to our approach, this system works only for short-term predictions.

A method for human motion prediction in indoor spaces using a maximum entropy approach is proposed by Ziebart *et al.* [2]. The authors argue that humans plan their movement according to cost functions that assign costs to environmental features, like surfaces or available spaces. Ziebart *et al.* aim to learn these functions based an observed data and then use the learned models to predict future movements. In this method, objects are only implicitly considered, i.e., as environmental features. Note that by predicting the destination of a human as in our approach, we can also infer the path the human will probably take using the assumption that humans operate based on a cost function as in [2].

Other existing motion prediction methods include velocity-based modeling of future human movements [5], [6] or learn social models to predict the behavior of humans in lively places [3], [7], [8] However, those approaches have been developed for short-term prediction of human motions and trajectory adaption of a mobile robot and not for foresighted navigation as in our application.

While object-based prediction, to our knowledge, has not been applied in existing motion prediction systems it is often used in the context of higher-level action prediction [9], [10]. In these frameworks, predicted actions are typically associated with objects, e.g., if a person holds a plate the next action will likely be setting the plate on some kind of surface or table. Those approaches have been used successfully in a local context but the authors did not consider general human-object interactions also including moving to other places.

In previous work, we learned a prediction model for moving humans in indoor environments consisting of several rooms. The predicted motions were used to learn a foresighted navigation policy for a service robot via Q-learning [1] and to find a user quickly if they cannot be located in the proximity of the robot [11]. In contrast to our current work, objects have not been explicitly used during the prediction, however, they inherently have an effect on many typical trajectories of humans and can be used to improve the prediction as we will show.

III. Prediction of the Navigation Goal

We consider the problem of predicting the navigation goal of a moving human in an indoor environment. The prediction is based on observations of the human's location and pose and knowledge about their typical transitions between objects. This knowledge can be collected by observing humans in indoor environment and learning a probability distribution called *interaction model* [12].

We represent the environment as an inflated grid map M, on which possible paths of the human can be computed. We model objects in the environment as tuples $o = (\tau_o, x_o)$ with known type τ_o and position x_o on M between which the human moves. To predict the current navigation goal, we use a recursive Bayes' filter and integrate observations of the human as well as the pre-learned object interaction model. In the next subsection, we first introduce the human-object interaction model before we describe our new framework to predict the navigation goal.

Fig. 2: Example for the goal prediction based *solely* on the interaction model after a detected worktop interaction. Objects are color coded with respect to their likelihood to be the next movement goal of the human (H). The darker the green the higher the likelihood. Object names are abbreviated: dining table (T), microwave (M), refrigerator (R), and sofa (S).

A. Interaction Model

We define the interaction model $I(\tau_a|\tau_b)$ as the distribution of the probability that after the human interacted with an object of type τ_b they will next interact with an object of type τ_a. We collected data of users in indoor environments and recorded their object interaction sequences to learn the interaction model. We plan to publish an approach to gather this data automatically from video streams. Note that to generalize well between different environments, we only consider object-interaction sequences to learn the interaction model and do not consider the actual trajectories of the human. The interaction model can then be used to predict future movements of the human based on the known locations of objects (see Fig. 2).

B. Formulation as a Bayes' Filter

Let τ_{L_t} be the type of the last observed object the human has interacted with and $S_t = (x_{h_t}, \theta_{h_t})$ be the human state at time t with x_{h_t} being the observed position and θ_{h_t} the orientation at time t (e.g., obtained from a pose estimation system [13]). We define the belief about the navigation goal at time t based on all previous observations as $bel(o_t) = p(o_t|S_{1:t}, \tau_{L_{1:t}})$. The belief is recursively computed as:

$$
\begin{aligned}
& bel(o_t) = p(o_t|S_{1:t}, \tau_{L_{1:t}}) && (1)\\
& = \eta \cdot p(S_t|o_t, S_{1:t\text{-}1}, \tau_{L_{1:t}}) \cdot p(o_t|S_{1:t\text{-}1}, \tau_{L_{1:t}}) && (2)\\
& = \eta \cdot p(S_t|o_t) \cdot p(o_t|S_{1:t\text{-}1}, \tau_{L_{1:t}}) && (3)\\
& = \eta \cdot p(S_t|o_t) \cdot \sum_{o_{t\text{-}1}} p(o_t|o_{t\text{-}1}, S_{1:t\text{-}1}, \tau_{L_{1:t}}) \\
& \qquad\qquad \cdot p(o_{t\text{-}1}|S_{1:t\text{-}1}, \tau_{L_{1:t\text{-}1}}) && (4)\\
& = \eta \cdot p(S_t|o_t) \cdot \sum_{o_{t\text{-}1}} p(o_t|\tau_{L_t}) \cdot p(o_{t\text{-}1}|S_{1:t\text{-}1}, \tau_{L_{1:t\text{-}1}}) && (5)\\
& = \eta \cdot p(S_t|o_t) \cdot \sum_{o_{t\text{-}1}} p(o_t|\tau_{L_t}) \cdot bel(o_{t\text{-}1}) && (6)
\end{aligned}
$$

with $p(S_t|o_t)$ as the observation model and $p(o_t|\tau_{L_t})$ as the motion model. For the derivation, we use Bayes' rule (Eq. (2)) with the normalizer η, the Markov assumption (Eq. (3), Eq. (5)), the law of total probability (Eq. (4)), and conditional independence.

The belief is updated at fixed time intervals as long as the human is visible and moving. At each time step, the object with the highest probability is assumed to be the navigation goal.

C. Motion Model

The motion model is based on the pre-learned interaction model $I(\tau_a|\tau_b)$, which is defined as a probability distribution that provides the probability that after the human interacted with an object of type τ_b they will next interact with an object of type τ_a (Sec. III-A). Given an interaction model we can define the motion model as follows:

$$p(o_t|\tau_{L_t}) = I(o_t|\tau_{L_t}) \quad (7)$$

If the robot has no knowledge about the previous human-object interaction, $I(o_t|\tau_{L_t})$ is replaced with the normalized sum over the probability of all possible object types $\eta' \cdot \sum_\tau I(o_t|\tau)$ with η' as normalizing parameter.

D. Observation Model

The observation model $p(S_t|o_t)$ is used to update the belief about the navigation goal based on observations about the human's pose. The closer the human gets to a possible navigation goal the higher the likelihood. Accordingly, we use the inverse distance of the A* path from the human's position x_{h_t} to the possible navigation goal x_{o_t} on the map in the observation model. Additionally, we compute the orientation θ_{opt_t} the human would have on the A* path from their current position x_{h_t} to x_{o_t} in the 2D map. We then compute the inverse angular distance of the human's orientation θ_{h_t} and θ_{opt_t}, both wrt. to the coordinate system of the 2D map.

Let $dist(x_{h_t}, x_{o_t})$ be the A* distance between x_{h_t} and x_{o_t} on M and let $distA(\theta_{h_t}, \theta_{opt_t})$ be the angular distance between θ_{h_t} and θ_{opt_t}. The observation model is then defined as:

$$\begin{aligned} & p(S_t|o_t) \\ & \quad = \quad p((x_{h_t}, \theta_{h_t})|o_t) \\ & \quad = \quad p(x_{h_t}|o_t) \cdot p(\theta_{h_t}|o_t) \\ & \quad = \quad dist(x_{h_t}, x_{o_t})^{-1} \cdot distA(\theta_{h_t}, \theta_{opt_t})^{-1} \end{aligned}$$

Now, we have defined all components of our recursive Bayes' filter that is used to predict the navigation goal of the human.

IV. Experimental Evaluation

To demonstrate the capabilities of our new prediction framework, we conducted a set of experiments. The first experiments were designed to show that based on the learned object interaction model we can reliably infer the navigation goal of the human using the presented Bayes' filter. We furthermore demonstrate the improvement compared to a system that does not use information about subsequent object interactions but only knowledge about the frequency of moving between common destinations on specific trajectories [14]. Finally, we provide experiments with a real robot observing a user and predicting its motions to realize foresighted navigation.

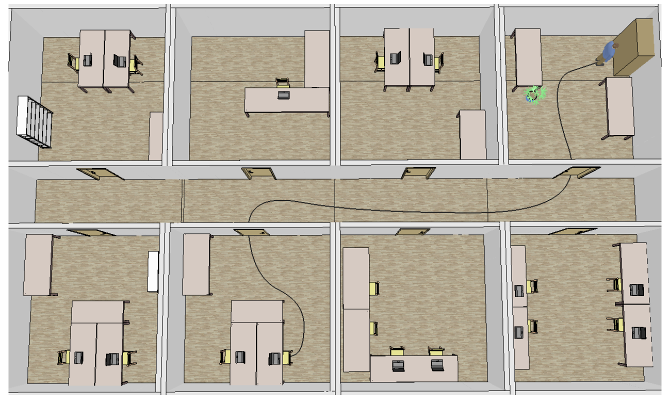


Fig. 3: Example trajectory and map from our dataset. The map corresponds to an office environment with the following object types: refrigerator, laptop, table, chair and cupboard. The human first interacts with the refrigerator and then moves to a table.

A. Data Collection

To test our prediction framework, we recorded a evaluation dataset of 64 human trajectories in 25 different simulated office or home environments with sizes between 100m^2 and 150m^2 that contained 24 objects on average. We created the maps using the V-REP editor [15]. Fig. 3 shows an example map. The trajectories were recorded by users who generated typical movements in office or domestic environments. Each trajectory started and ended at an object. Interaction objects included bottles, cups, microwaves, chairs, tables, beds, toilets, handbasins, bathtubs, washbasins, cupboards, wardrobes, refrigerators, sofas, and laptops. We recorded the simulated human's pose every time they moved one meter. To train the interaction model for our prediction framework we used a separate training dataset containing 161 human-object interaction sequences.

B. Quantitative Evaluation

For the quantitative evaluation of the goal prediction, we tested our approach on the recorded trajectories between two subsequent objects with and without knowledge about the last object interaction. For the first scenario, we assume that the robot observes the human movement and their initial interaction with an object. For the latter, the robot only observes the human during their movement and not the initial object interaction before the human started to move. In this scenario, we have to take into account all possible previous objects to compute a prediction (see Sec. III-C).

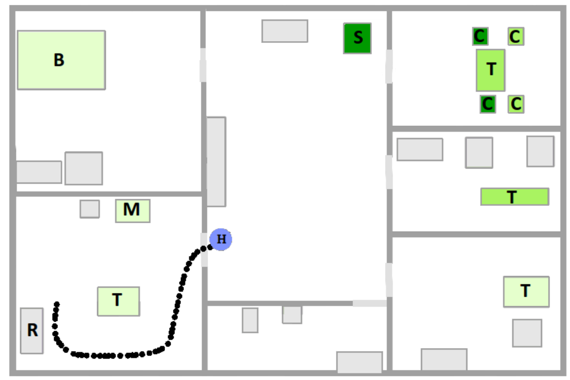
Evaluation results		
	Avg. Prediction Accuracy	Avg. Trajectory Length Until Correct Prediction
Last interaction observed	0.67	28%
Last interaction **not** observed	0.48	52%
Previous Approach as in [11]	0.36	69%

TABLE I: Results of the quantitative evaluation on 64 different human trajectories. As can be seen, the interaction model is essential to obtain a reliable prediction about the navigation goal. Furthermore, our new framework seriously outperforms our previous system.

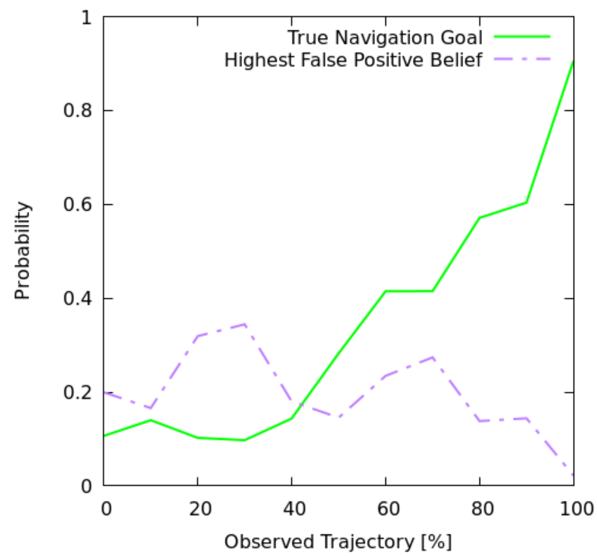
As evaluation metric, we use the prediction accuracy, i.e., after each update step we checked if the predicted goal was the true navigation goal of the human. If so we counted the prediction as correct. The accuracy was then computed by dividing the number of correct predictions by the total number of integrated observations for the trajectory. As second metric, we use the average length of the trajectory that needed to be observed until our system continuously returns the correct navigation goal. The belief about the navigation goal is updated once every second as long as the human is visible and moving.

As can be seen in Tab. I, our system is able to confidently infer the true navigation goal after observing only 28% of the trajectory on average if the last object interaction is known and 52% of the trajectory otherwise. Furthermore, we achieve an average prediction accuracy on the trajectory of 0.67 with the interaction model and of 0.48 without.

We also compared our approach to a prediction based on a hidden Markov model (HMM) as in our previous work [11]. This approach uses previously learned trajectories between fixed destinations on a map of the environment. In contrast to that, our approach is based on transition probabilities between object types, where the actual object positions need to be known but can vary. In this way, we achieve a more general solution that additionally needs less training data. As can be seen from Tab. I, our new framework is able to outperform our previous HMM based prediction, especially if the last object interaction is observed, but also if not.



(a) Example trajectory for which the evolution of the goal probabilities is shown below. Object names are abbreviated: dining table (T), microwave (M), bed (B), chair (C), and sofa (S).



(b) Evolution of the belief about the navigation goal with respect to the percentage of the observed length of a typical trajectory, which is depicted above.

Fig. 4: Example of a typical trajectory, in this case the human first interacts with a refrigerator and then moves to a chair. (a) Trajectory observed so far at the point where the prediction is correct for the first time. (b) Corresponding evolution of the belief about navigation goal.

C. Qualitative Evaluation

We further analyzed how the probability of the true navigation goal changes over time. Fig. 4 shows the evolution of the probability with respect to the percentage of the observed length of a typical test trajectory for which the last object interaction is known. As can be seen, the prediction becomes correct after around 40% of the trajectory has been observed and increases further with more observations of the human. Only in the first third of the trajectory another object is falsely assumed to be the navigation goal.

A concrete example of the evolution of the goal probabilities over time can be seen in Fig. 5. Here, we show the belief about the navigation goal for four different timesteps. The further the human moved on the trajectory, the more navigation goals could be excluded. In Fig. 5 (a), the prediction mostly depended on the interaction model since the human just started moving. In Fig. 5 (b), the information about the human's pose was used to update the belief and according to our observation model, the goal probabilities of the objects in the lower half of the map decreased. Fig. 5 (c) then shows a situation in which a false navigation goal was the most likely navigation goal. This goal already had a high likelihood from the interaction model, that was further increased by the movement of the human towards the door which would lead to the room in which the object was positioned. However, after integrating further observations and updating the belief, our system corrected the prediction

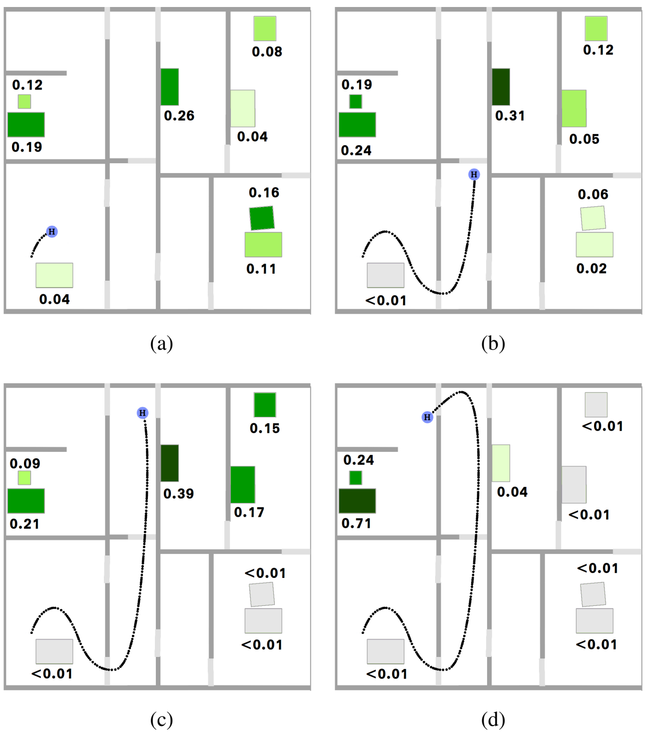


Fig. 5: Evolution of the belief over time for an example human trajectory. As can be seen, the initial belief based on the interaction model (a) is continuously updated with new observations (b), (c), (d). Possible goals are shown in green with their probabilities to be the navigation goal: the darker the green the higher the probability. Doors are colored in grey and walls in black. The human is depicted as blue dot with their trajectory as dashed line.

once the human moved away from the door and estimated the true navigation goal (Fig. 5 (d)).

D. Application

To demonstrate the capabilities of our approach for foresighted robot navigation we tested our framework on a Robotino mobile platform [16] in an office environment[1]. The robot uses a representation of the environment with discretization of 0.75 meters to decide where to place itself in order to be close to the human in case assistance is needed but not to be in their way. The robot updates every five seconds the belief about human's navigation goal based on new observations using the presented Bayes' filter.

Avoiding interferences with the human as well as foresighted robot navigation is important as humans react negatively to robots that are too close to them or follow them constantly [17], [18]. Accordingly, we designed a utility function that gives high utility values for robot positions that do not interfere with the human and their predicted path while still being close to the potential goal. Using x_r as the position of the robot, the utility function is defined as:

$$U(x) = \sum_{o} Bel(o) \cdot (dist(x_r, x) + dist(x, x_o))^{-1} \qquad (8)$$

The utility of a possible position thereby depends on its distance to the robot and possible navigation goals of the human weighted by their probabilities, i.e. the closer the distances the higher the utility. To ensure that the robot does not interfere with the human we set the utility of points with a distance up to 1.5m to the human or to their $A*$ path to the currently most likely navigation goal to zero.

Fig. 6 illustrates an example experiment. Here, the robot was in a corridor where it observed a human-object interaction with a cup. The robot then updated the belief about the navigation goal and computed a new position based on Eq. (8) (see Fig. 6 (a)). During their movement the robot regularly stopped and updated both the prediction of the navigation goal as well as the utility values of its own placement positions (see Fig. 6 (b)). In Fig. 6 (c), the human has entered the room containing the navigation goal and the robot correctly updated its belief. The robot would not enter the room itself since positions near the human have a utility of zero and no collision-free path can be found to the other side of the room. However, if the human called the robot to help them, the robot would be there immediately due to its foresighted positioning.

Thus, this real-world experiment shows how a mobile robot can use our prediction framework for foresighted navigation. As a result, the robot can avoid interferences but is still always close to the human and can be called to assist them if needed.

V. Conclusion

In this paper, we presented an approach to predict the navigation goal of a moving human. As novelty, we proposed to use knowledge about typical human-object interaction sequences. To learn the corresponding object interaction model, we observed humans in indoor environments and learned transition probabilities of human-object interactions. We then utilized this information in combination with observations about the movement and pose of the human to infer their navigation goal using a recursive Bayes' filter.

As we demonstrated in various experiments, our framework can reliably predict the navigation goal of a moving human even after observing only a fraction of its trajectory to the goal. Furthermore, we show that our system outperforms a traditional HMM based prediction approach that relies on previously learned trajectories between fixed destinations.

Finally, we performed an experiment in which a mobile robot uses the new prediction framework for foresighted navigation by computing favorable positions based on the predictions of the human's navigation goal. With this approach, the robot is able to position itself close enough to the human to provide assistance quickly if needed while avoiding interference with the predicted path of the human.

[1] A video of one of the experiments is available online at `https://www.hrl.uni-bonn.de/bruckschen19ecmr.mp4`.

(a) (b) (c)

Fig. 6: Application example of our approach to foresighted navigation. Possible navigation goals of the human are shown in green and possible placement positions for the robot are shown in blue. The darker the color the higher the belief/utility. (a) The robot (R) observes a moving human (H) that interacts with a cup. Based on this information, it updates the belief about the navigation goal and computes a new position for itself taking into account the human's most likely navigation goals while avoiding interferences with the human and their predicted path. (b) The prediction as well as the robot's placement position are updated with new observations. (c) The human enters a room and the robot adapts its position to be close to the human in order to enable quick reaction when called for assistance. Note that the robot ignores positions in direct proximity to the human even if they lead to a higher utility to minimize interferences.

Acknowledgments

We would like to thank Sabrina Amft, Sandra Höltervennhoff, Sophie Jenke, Jenny Mack, Saskia Rabich, and Mosadeq Saljoki for their help during our experiments.

References

[1] A. Bayoumi, P. Karkowski, and M. Bennewitz, "Learning foresighted people following under occlusions," in *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2017.

[2] B. D. Ziebart, N. Ratliff, G. Gallagher, C. Mertz, K. Peterson, J. A. Bagnell, M. Hebert, A. K. Dey, and S. Srinivasa, "Planning-based prediction for pedestrians," in *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2009.

[3] H. Kretzschmar, M. Spies, C. Sprunk, and W. Burgard, "Socially compliant mobile robot navigation via inverse reinforcement learning," *Intl. Journal of Robotics Research (IJRR)*, vol. 35, no. 11, 2016.

[4] D. Vasquez, "Novel planning-based algorithms for human motion prediction," in *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2016.

[5] S. Kim, S. J. Guy, W. Liu, R. W. Lau, M. C. Lin, and D. Manocha, "Predicting pedestrian trajectories using velocity-space reasoning," in *Algorithmic Foundations of Robotics X*. Springer, 2013.

[6] S. Lefèvre, D. Vasquez, and C. Laugier, "A survey on motion prediction and risk assessment for intelligent vehicles," *Robomech Journal*, vol. 1, no. 1, 2014.

[7] M. Kuderer, "Socially compliant mobile robot navigation," Ph.D. dissertation, Universität Freiburg, 2015.

[8] A. Robicquet, A. Sadeghian, A. Alahi, and S. Savarese, "Learning social etiquette: Human trajectory understanding in crowded scenes," in *Proc. of the Europ. Conf. on Computer Vision (ECCV)*. Springer, 2016.

[9] T. Lan, T.-C. Chen, and S. Savarese, "A hierarchical representation for future action prediction," in *Proc. of the Europ. Conf. on Computer Vision (ECCV)*, 2014.

[10] H. S. Koppula, R. Gupta, and A. Saxena, "Learning human activities and object affordances from RGB-D videos," *Intl. Journal of Robotics Research (IJRR)*, vol. 32, no. 8, 2013.

[11] A. Bayoumi, P. Karkowski, and M. Bennewitz, "People finding under visibility constraints using graph-based motion prediction," in *Proc. of the Int. Conf. on Intelligent Autonomous Systems (IAS)*, 2018.

[12] L. Bruckschen, S. Amft, J. Tanke, J. Gall, and M. Bennewitz, "Detection of generic human-object interactions in video streams," in *Proc. of the International Conference on Social Robotics (ICSR)*, 2019, to appear.

[13] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime multi-person 2d pose estimation using part affinity fields," 2018.

[14] A. Bayoumi, P. Karkowski, and M. Bennewitz, "Speeding up person finding using hidden markov models," *Robotics and Autonomous Systems*, vol. 115, 2019.

[15] M. F. E. Rohmer, S. P. N. Singh, "V-rep: a versatile and scalable robot simulation framework," in *Proc. of The International Conference on Intelligent Robots and Systems (IROS)*, 2013.

[16] F. D. GmbH and C. KG, "Robotino manual," https://www.festo-didactic.com.

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

[18] K. Caine, S. Šabanovic, and M. Carter, "The effect of monitoring by cameras and robots on the privacy enhancing behaviors of older adults," in *Proc. of the ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI)*, 2012.