

# THÖR: Human-Robot Navigation Data Collection and Accurate Motion Trajectories Dataset

Andrey Rudenko<sup>1,2</sup>, Tomasz P. Kucner<sup>2</sup>, Chittaranjan S. Swaminathan<sup>2</sup>, Ravi T. Chadalavada<sup>2</sup>, Kai O. Arras<sup>1</sup> and Achim J. Lilienthal<sup>2</sup>

**Abstract**—Understanding human behavior is key for robots and intelligent systems that share a space with people. Accordingly, research that enables such systems to perceive, track, learn and predict human behavior as well as to plan and interact with humans has received increasing attention over the last years. The availability of large human motion datasets that contain relevant levels of difficulty is fundamental to this research. Existing datasets are often limited in terms of information content, annotation quality or variability of human behavior. In this paper, we present THÖR, a new dataset with human motion trajectory and eye gaze data collected in an indoor environment with accurate ground truth for position, head orientation, gaze direction, social grouping, obstacles map and goal coordinates. THÖR also contains sensor data collected by a 3D lidar and involves a mobile robot navigating the space. We propose a set of metrics to quantitatively analyze motion trajectory datasets such as the average tracking duration, ground truth noise, curvature and speed variation of the trajectories. In comparison to prior art, our dataset has a larger variety in human motion behavior, is less noisy, and contains annotations at higher frequencies.

**Index Terms**—Social Human-Robot Interaction, Motion and Path Planning, Human Detection and Tracking

## I. INTRODUCTION

UNDERSTANDING human behavior has been the subject of research for autonomous intelligent systems across many domains, from automated driving and mobile robotics to intelligent video surveillance systems and motion simulation. Human motion trajectories are a valuable learning and validation resource for a variety of tasks in these domains. For instance, they can be used for learning safe and efficient human-aware navigation, predicting motion of people for improved interaction and service, inferring motion regularities and detecting anomalies in the environment. Particular attention towards trajectories, intentions and mobility patterns of people has considerably increased in the last decade [1].

Datasets of ground level human trajectories, typically used for learning and benchmarking, include the ETH [2], Edinburgh [3] and the Stanford Drone [4] datasets, recorded

Manuscript received: September 10, 2019; Accepted November 28, 2019.

This paper was recommended for publication by Editor D. Lee upon evaluation of the Associate Editor and Reviewers' comments. This work has been partly funded from the European Union's Horizon 2020 research and innovation programme under grant agreement No 732737 (ILIAD) and by the Swedish Knowledge Foundation under contract number 20140220 (AIR).

<sup>1</sup>A. Rudenko and K.O. Arras are with Bosch Corporate Research, Stuttgart, Germany {andrey.rudenko, kaioliver.arras}@de.bosch.com

<sup>2</sup>A. Rudenko, T. Kucner, C. Swaminathan, R. Chadalavada and A. Lilienthal are with the Mobile Robotics and Olfaction Lab, Örebro University, Sweden {Tomasz.Kucner, Chittaranjan.Swaminathan, Ravi.Chadalavada, Achim.Lilienthal}@oru.se

Digital Object Identifier (DOI): see top of this page.

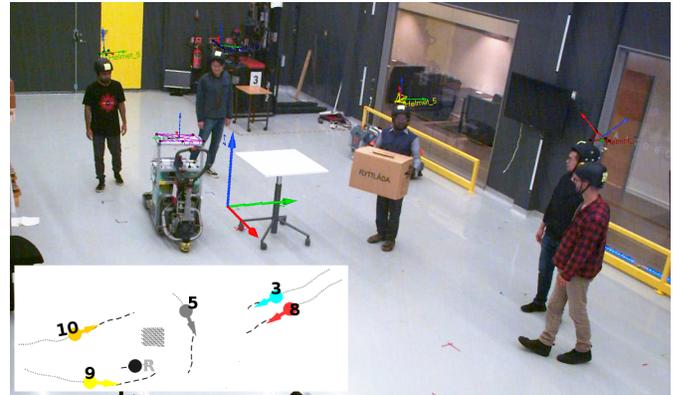


Fig. 1. Environment configuration. Participants, wearing tracking helmets, and the robot are moving towards their goals in a shared space, tracked by the Qualisys motion capture system (recorded motion in the bottom left corner).

outdoors, or the indoor ATC [5], L-CAS [6] or Central Station [7] datasets (see Table I). While providing the basic input of motion trajectories, these datasets often lack relevant contextual information and the desired properties of data, e.g. the map of static obstacles, coordinates of goal locations, social information such as the grouping of agents, high variety in the recorded behaviors or long continuous tracking of each observed agent. Furthermore, most of the recordings are made outdoors, a robot is rarely present in the environment and the ground truth pose annotation, either automated or manual, is prone to artifacts and human errors.

In this paper we present a human-robot interaction procedure, designed to collect motion trajectories of people in a generic indoor social setting with extensive interaction between groups of people and a robot in a spacious environment with several obstacles. The locations of the obstacles and goal positions are set up to make navigation non-trivial and produce a rich variety of behaviors. The participants are tracked with a motion capture system; furthermore, several participants are wearing eye-tracking glasses. “Tracking Human motion in the Örebro university” (THÖR) dataset<sup>1</sup>, which is released public and free for non-commercial purposes, contains over 60 minutes of human motion in 395k frames, recorded at 100 Hz, 2531k people detections and over 600 individual and group trajectories between multiple resting points. In addition to the video stream from one of the eye tracking headsets, the data includes 3D Lidar scans and a video recording from

<sup>1</sup>Available at <http://thor.oru.se>

Dataset	Location	Map	Goal positions	Groups	Head orientation	Eye gaze	Robot in the scene	Sensors for pose estimation	Frequency	Annotation
ETH [2]	Outdoors	✓	✓	✓				Camera	2.5 Hz	Manual
UCY [8]	Outdoors					✓		Camera	Continuous	Manual
VIRAT [9]	Outdoors	✓*						Camera	2.5,10 Hz	Manual
KITTI [10]	Outdoors	✓					✓	Velodyne and several cameras	10 Hz	Manual
Edinburgh [3]	Outdoors							Camera	6-10 Hz (variable)	Automated
Stanford Drone [4]	Outdoors	✓*						Camera	30 Hz	Manual
Town Center [11]	Outdoors	✓*						Camera	15 Hz	Manual
ATC [5]	Indoors				✓			Several 3D range sensors	10-30 Hz (variable)	Automated
Central station [7]	Indoors							Camera	25 Hz	Automated
L-CAS [6]	Indoors			✓			✓	3D LiDAR	10 Hz	Manual
KTH [12]	Indoors						✓	RGB-D, 2D laser scanner	10-17 Hz (variable)	Automated
THÖR	Indoors	✓	✓	✓	✓	✓	✓	Motion capture	100 Hz	Ground truth

\* Unsegmented camera image.

TABLE I  
DATASETS OF HUMAN MOTION TRAJECTORIES

stationary sensors. We quantitatively analyze the dataset using several metrics, such as tracking duration, perception noise, curvature and speed variation of the trajectories, and compare it to popular state-of-the-art datasets of human trajectories. Our analysis shows that THÖR has more variety in recorded behavior, less noise, and high duration of continuous tracking.

The paper is organized as follows: in Sec. II we review the related work and in Sec. III detail the data collection procedure. In Sec. IV we describe the recorded data and analyze it quantitatively and qualitatively. Sec. V concludes the paper.

## II. RELATED WORK

Recordings of human trajectory motion and eye gaze are useful for a number of research areas and tasks both for machine learning and benchmarking. Examples include person and group tracking [2], [13], [14], human-aware motion planning [15], [16], [17], [18], motion behavior learning [19], human motion prediction [20], [21], human-robot interaction [22], video surveillance [23] or collision risk assessment [24]. In addition to basic trajectory data, state-of-the-art methods for tracking or motion prediction, for instance, can also incorporate information about the environment, social grouping, head orientation or personal traits. For instance, Lau et al. [13] estimate social grouping formations during tracking and Rudenko et al. [21] use group affiliation as a contextual cue to predict future motion. Unhelkar et al. [25] use head orientation to disambiguate and recognize typical motion patterns that people are following. Bera et al. [26] and Ma et al. [27] learn personal traits to determine interaction parameters between several people. To enable such research in terms of training data and benchmarking requirements, a state-of-the-art dataset should include this information.

Human trajectory data is also used for learning long-term mobility patterns [28], such as the CLiFF maps [29], to enable compliant flow-aware global motion planning and reasoning about long-term path hypotheses towards goals in distant map areas for which no observations are immediately available. Finally, eye-gaze is a critical source of non-verbal information

about human task and motion intent in human-robot collaboration, traffic maneuver prediction, spatial cognition or sign placement [30], [31], [32], [33], [34].

Existing datasets of human trajectories, commonly used in the literature [1], are summarized in Table I. With the exception of [5], [6], [7], [12], all datasets have been collected outdoors. Intuitively, patterns of human motion in indoor and outdoor environments are substantially different due to scope of the environment and typical intentions of people therein. Indoors people navigate in loosely constrained but cluttered spaces with multiple goal points and many ways (e.g. from different homotopy classes) to reach a goal. This is different from their behavior outdoors in either large obstacle-free pedestrian areas or relatively narrow sidewalks, surrounded by various kinds of walkable and non-walkable surfaces. Among the indoor recordings, only [6], [12] introduce a robot, navigating in the environment alongside humans. However, recording only from on-board sensors limits visibility and consequently restricts the perception radius. Furthermore, ground truth positions of the recorded agents in all prior datasets were estimated from RGB(-D) or laser data. On the contrary, we directly record the position of each person using a motion capture system, thus achieving higher accuracy of the ground truth data and complete coverage of the working environment at all times. Moreover, our dataset contains many additional contextual cues, such as social roles and groups of people, head orientations and gaze directions.

## III. DATA COLLECTION PROCEDURE

In order to collect motion data relevant for a broad spectrum of research areas, we have designed a controlled scenario that encourages social interactions between individuals, groups of people and with the robot. The interactive setup assigns social roles and tasks so as to imitate typical activities found in populated spaces such as offices, train stations, shopping malls or airports. Its goal is to motivate participants to engage into natural and purposeful motion behaviors as well as to create a rich variety of unscripted interactions. In this section we detail the system setup and the data collection procedure.

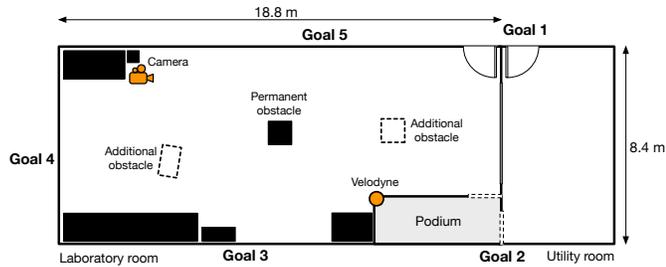


Fig. 2. Overview of the environment. The Qualisys motion tracking system is installed in a laboratory room, which is mostly empty except for some shelves and equipment along the walls. A permanent obstacle in the middle of the room is present in all recordings, while additional obstacles are only placed in the “Three obstacles” scenario (see Sec. III-B for details). The position of the camera is shown in the top left corner, and the position of the Velodyne in the bottom right.

### A. System Setup

Data collection was performed in a spacious laboratory room of  $8.4 \times 18.8$  m and the adjacent utility room, separated by a glass wall (see the overview in Fig. 2). The laboratory room, where the motion capture system is installed, is mostly empty to allow for maneuvering of large groups, but also includes several constrained areas where obstacle avoidance and the choice of homotopy class is necessary. Goal positions are placed to force navigation along the room and generate frequent interactions in its center, while the placement of obstacles prevents walking between goals on a straight line.

To track the motion of the agents we used the Qualisys Oqus 7+ motion capture system<sup>2</sup> with 10 infrared cameras, mounted on the perimeter of the room. The motion capture system covers the entire room volume apart from the most right part close to the podium entrance – a negligible loss due to the focus on the central part of the room. The system tracks small reflective markers at 100 Hz with spatial discretization of 1 mm. The coordinate frame origin is on the ground level in the middle of the room. For people tracking, the markers have been arranged in distinctive 3D patterns on the bicycle helmets, shown in Fig. 3. The motion capture system was calibrated beforehand with an average residual tracking error of 2 mm, and each helmet, as well as the robot, was defined in the system as a unique rigid body of markers, yielding its 6D head position and orientation. Each participant was assigned an individual helmet for all recording sessions, labeled 2 to 10. Helmet 1 was not used in this data collection.

For acquiring eye gaze data we used four mobile eye-tracking headsets worn by four participants (helmet numbers 3, 6, 7, and 9 respectively). However, in this dataset we only include data from one headset (Tobii Pro Glasses), worn by the participant with helmet 9. The gaze sampling frequency of Tobii Pro Glasses is 50 Hz. It also has a scene camera which records the video at 25 fps. A gaze overlaid version of this video is included in this dataset. We synchronized the clocks of each machine (the Qualisys system, the stationary Velodyne sensor and the eye-tracking glasses) with the same NTP time

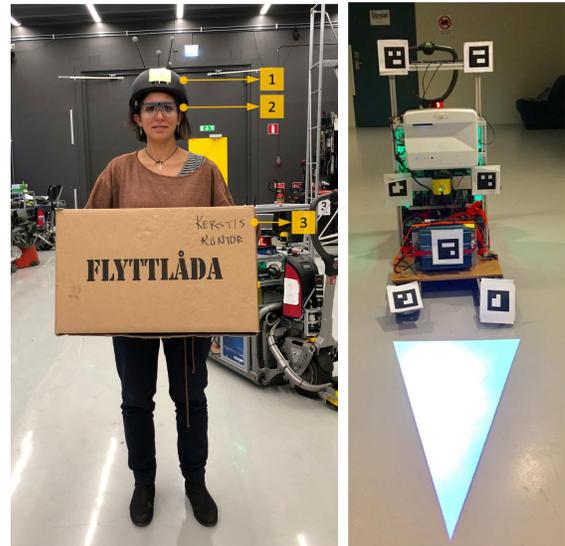


Fig. 3. Equipment used in our data collection: **Left:** (1) bicycle helmet with mocap tracking markers, (2) Tobii Pro Glasses, (3) boxes which were carried by the participants as a part of the tasks. **Right:** Linde CitiTruck robot projecting its current motion intent on the floor.

server. Finally, we recorded a video of the environment from a stationary camera, mounted in a corner of the room.

The robot, used in our data collection, is a small forklift Linde CitiTruck robot with a footprint of  $1.56 \times 0.55$  m and 1.17 m high, shown in Fig. 3. It was programmed to move in a socially unaware manner, following a pre-defined path around the room and adjusting neither its speed nor trajectory to account for surrounding people. For safety reasons, the robot was navigating with a maximal speed of  $0.34 \text{ m s}^{-1}$  and projecting its current motion intent on the floor in front of it using a mounted beamer [34]. A dedicated operator was constantly monitoring the environment from a remote workstation to stop the robot in case of an emergency. The participants were made aware of the emergency stop button on the robot should they be required to use it.

### B. Scenario Description and Participants’ Priming

During the data collection the participants performed simple tasks, which required walking between several goal positions. To increase the variety of motion, interactions and behavioral patterns, we introduced several roles for the participants and created individual tasks for each role, summarized in Fig. 4.

The first role is a *visitor*, navigating alone and in groups of up to 5 people between four goal positions in the room. At each goal they take a random card, indicating the next target. As each group was instructed to travel together, they only take one card at a time. We asked the visitors to talk and interact with the members of their group during the data collection, and changed the structure of groups every 4-5 minutes. There are 6 visitors in our recording. The second role is a *worker*, whose task is to receive and carry large boxes between the laboratory and the utility room. The workers wear a yellow reflective vest. There are 2 workers in our recording, one carrying the boxes from the laboratory to the utility room, and the other vice versa.

<sup>2</sup><https://www.qualisys.com/hardware/5-6-7/>

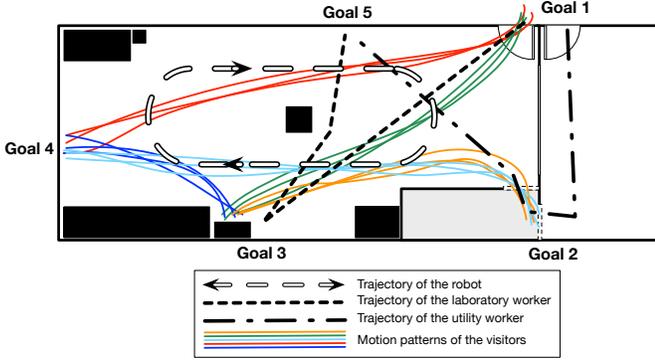


Fig. 4. Roles of the participants and their expected motion patterns. *Visitors*, walking alone and in groups, are instructed to navigate between goals 1,2,3 and 4. Their motion patterns are shown with colored solid lines. The *laboratory worker*, whose waiting position is at goal 3, picks up an incoming box at goal 1, registers its ID at goal 3 and then places it at goal 5. The *utility worker*, whose waiting position is at goal 2, picks up the box at goal 5, registers it at goal 2 with a new ID and places it at goal 1. The patterns of both workers are shown with dotted lines. The trajectory of the *robot*, circulating around the obstacle in the middle of the room, is shown with a thick hollow line.

The third role is the *inspector*. An inspector is navigating alone between many additional targets in the environment, indicated by a QR-code, in no particular order and stops at each target to scan the code. We have one inspector in our recording.

There are several points to motivate the introduction of the social roles. Firstly, with the motion of the visitors and the workers we introduce distinctive motion patterns in the environment, while the cards and the tasks make the motion focused, goal-oriented and prevent random wandering. However, the workers’ tasks allocation is such that at some points idle standing/wandering behavior is also observed, embedded in their cyclical activity patterns. Furthermore, we expect that the visitors navigating alone, in groups and the workers who carry heavy boxes exhibit distinctive behavior, therefore the grouping information and the social role cue (reflective vest) may improve the intention and trajectory prediction. Finally, motion of the inspector introduces irregular patterns in the environment, distinct from the majority of the visitors.

We prepared three scenarios for data collection with different numbers of obstacles and motion state of the robot. In the first scenario, the robot is placed by a wall and not moving, and the environment has only one obstacle (see the layout in Fig. 2). The second scenario introduces the moving robot, navigating around the obstacle (the trajectory of the robot is depicted in Fig. 4). The third scenario features an additional obstacle and a stationary robot in the environment (see Fig. 2 with additional obstacles). We denote these recording scenarios as *One obstacle*, *Moving robot* and *Three obstacles*, accordingly. In each scenario the group structure for the visitors was reassigned 4-5 times. Between the scenarios, the roles were also reassigned. A summary of the scenarios and durations is given in Table II.

Each round of data collection started with the participants, upon command, beginning to execute their tasks. The round lasted for approximately four minutes and ended with another call from the moderator. To avoid artificial and unnatural

Scenario, round	Visitors, groups		Workers		Inspector	Duration
	Helmet ID	2–10	Utility, lab			
One obstacle	1	6,7,5 + 8,2,4	3	9	10	368 sec
	2	2,5,6,7 + 8,4	3	9	10	257 sec
	3	6,7,8 + 4,5 + 2	3	9	10	275 sec
	4	2,4,5,7,8 + 6	3	9	10	315 sec
Moving robot	1	4,5,6 + 3,7,9	2	8	10	281 sec
	2	3,5,6,9 + 7,4	2	8	10	259 sec
	3	5,7,9 + 4,6 + 3	2	8	10	286 sec
	4	3,5,6,7,9 + 4	2	8	10	279 sec
	5	3,6 + 4,9 + 5,7	2	8	10	496 sec
Three obstacles	1	2,3,8 + 6,7,9	5	4	10	315 sec
	2	2,8,9 + 3,6,7	5	4	10	290 sec
	3	2,3,7 + 8,9 + 6	5	4	10	279 sec
	4	2,3,6,7,9 + 8	5	4	10	277 sec

TABLE II  
ROLE ASSIGNMENT AND RECORDING DURATION IN THE THREE SCENARIOS OF OUR DATA COLLECTION: (I) ONE OBSTACLE, (II) MOVING ROBOT, (III) THREE OBSTACLES

motion due to knowing the true purpose of the data collection, we told the participants that our goal is to validate the robot’s perceptive abilities, while the motion capture data will be used to compare the perceived and actual positions of humans. Participants were asked not to communicate with us during the recording. For safety and ethical reasons, we have instructed participants to act carefully near the robot, described as “autonomous industrial equipment” which does not stop if someone is in its way. An ethics approval was not required for our data collection as per institutional guidelines and the Swedish Ethical Review Act (SFS number: 2003:460). Written informed consent was obtained from all participants. Due to the relatively low weight of the robot and the safety precautions taken, there was no risk of harm to participants.

## IV. RESULTS AND ANALYSIS

### A. Data Description

The THÖR dataset includes over 60 minutes of motion in 13 rounds of the three scenarios. The recorded data in `.mat`, `.bag` and `.tsv` format contains over 395k frames at 100 Hz, 2531k human detections and 600+ individual and group trajectories between the goal positions. For each detected person the 6D position and orientation of the helmet in the global coordinate frame is provided. Furthermore, the dataset includes the map of the static obstacles, goal coordinates and grouping information. We also share the Matlab scripts for loading, plotting and animating the data. Additionally, the eye gaze data is available for one of the participants (Helmet 9), as well as Velodyne scans from a static sensor and the recording from the camera. We thoroughly inspected the motion capture data and manually cleaned it to remove occasional helmet ID switches and recover several lost tracks. Afterwards we applied an automated procedure to restore the lost positions of the helmets from incomplete set of recognized markers. In Fig. 5 we show the summary of the recorded trajectories.

### B. Baselines and Metrics

The THÖR dataset is recorded using a motion capture system, which yields more consistent tracking and precise

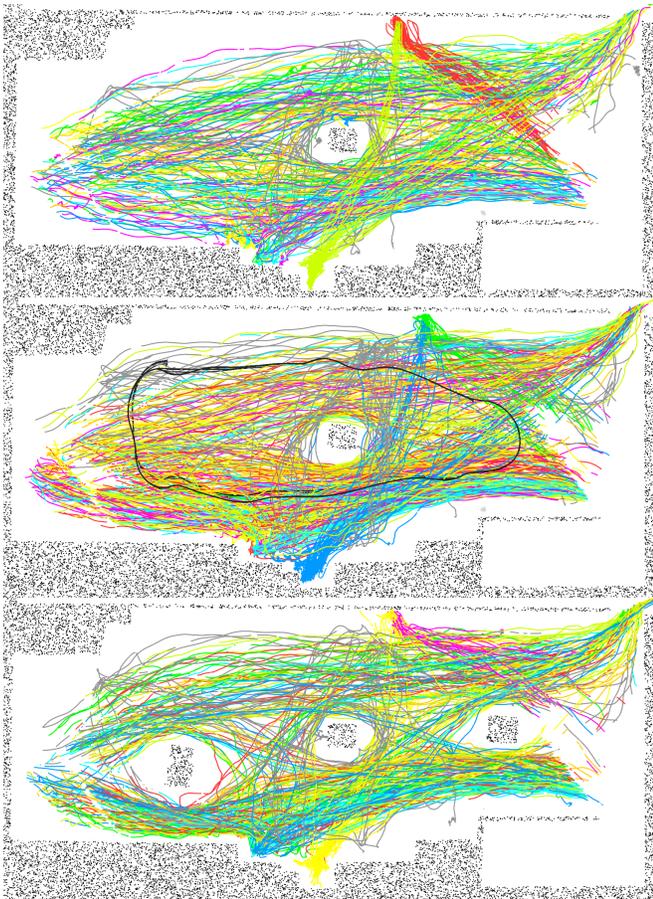


Fig. 5. Trajectories of the participants and the robot, recorded in the “One obstacle” scenario (top), “Moving robot” scenario (middle) and “Three obstacles” scenario (bottom). The robot’s path in the middle image is shown in black.

estimation of the ground truth positions and therefore higher quality of the trajectories, compared to the human detections from RGB-D or laser data, typically used in existing datasets. For the quantitative analysis of the dataset, we compare the recorded trajectories to the several datasets which are often used for training and evaluation of motion predictors for human environments [1]. The popular ETH dataset [2] is recorded outdoors in a pedestrian zone with a stationary camera facing downwards and manually annotated at 2.5 Hz. The Hotel sequence, used in our comparison, includes the coordinates of the 4 common goals in the environment and group information for walking pedestrians. The ATC dataset [5] is recorded in a large shopping mall using multiple 3D range sensors at  $\sim 26$  Hz over an area of  $900 \text{ m}^2$ . This allows for long tracking durations and potential to capture interesting interactions between people. In addition to positions it also includes facing angles. In this comparison we used the recordings from 24th and 28th of October and 14th of November. The Edinburgh dataset [3] is recorded in a university campus yard using a camera facing down with variable detection frequency, on average 9 Hz. For comparison we used the recordings from 27th of September, 16th of December, 14th of January and 22nd of June.

Metric	THÖR	ETH	ATC	Edinburgh
Tracking duration [s]	$16.7 \pm 14.9$	$9.4 \pm 5.4$	$39.7 \pm 64.7$	$10.1 \pm 12.7$
Trajectory curvature [ $\text{m}^{-1}$ ]	$1.9 \pm 8.8$	$0.18 \pm 1.48$	$0.84 \pm 1.43$	$1 \pm 3.9$
Perception noise [ $\text{m s}^{-2}$ ]	0.12	0.19	0.48	0.81
Motion speed [ $\text{m s}^{-1}$ ]	$0.81 \pm 0.49$	$1.38 \pm 0.46$	$1.04 \pm 0.46$	$1.0 \pm 0.64$
Min. dist. between people [m]	$1.54 \pm 1.60$	$1.33 \pm 1.39$	$0.61 \pm 0.16$	$3.97 \pm 3.5$

TABLE III  
COMPARISON OF THE DATASETS

For evaluating the quality of recorded trajectories we propose several metrics:

- 1) *Tracking duration* (s): average length of continuous observations of a person, higher is better.
- 2) *Trajectory curvature* ( $\text{m}^{-1}$ ): global curvature of the trajectory  $\mathcal{T}$ , caused by maneuvering of the agents in presence of static and dynamic obstacles, measured on 4s segments based on the first  $\mathcal{T}_t = (x_1, y_1)$ , middle  $\mathcal{T}_{t+2s} = (x_2, y_2)$  and last  $\mathcal{T}_{t+4s} = (x_3, y_3)$  points of the interval:  $K(\mathcal{T}_{t:t+4s}) = \frac{2(x_2 - x_1)(y_3 - y_1) - (x_3 - x_1)(y_2 - y_1)}{\|x_2 - x_1, y_2 - y_1\| \|x_3 - x_1, y_3 - y_1\| \|x_3 - x_2, y_3 - y_2\|}$ . The choice of 4s path segments is motivated by the standard motion prediction horizon in the related work [23]. Higher curvature values correspond to more challenging, non-linear paths.
- 3) *Perception noise* ( $\text{m s}^{-2}$ ): under the assumption that people move on smooth, not jerky paths, we evaluate local distortions of the recorded trajectory  $\{\mathcal{T}_t\}_{t=1 \dots M}$  of length  $M$ , caused by the perception noise of the mocap system as the average absolute acceleration:  $\frac{1}{M} \sum_{t=1}^M |\ddot{\mathcal{T}}_t|$ . Less noise is better.
- 4) *Motion speed* ( $\text{m s}^{-1}$ ): mean and standard deviation of velocities in the dataset, measured on 1s intervals. If the effect of perception noise on speed is negligible, higher standard deviation means more diversity in behavior of the observed agents, both in terms of individually preferred velocity and compliance with other dynamic agents.
- 5) *Minimal distance between people* (m): average minimal euclidean distance between two closest observed people. This metric indicates the density of the recorded scenarios, lower values correspond to more crowded environments.

### C. Results

The results of the evaluation are presented in Table III. Our dataset has sufficiently long trajectories (on average 16.7s tracking duration) with high curvature values ( $1.9 \pm 8.8 \text{ m}^{-1}$ ), indicating that it includes more human-human and human-environment interactions than the existing datasets. Furthermore, despite the much higher recording frequency, e.g.

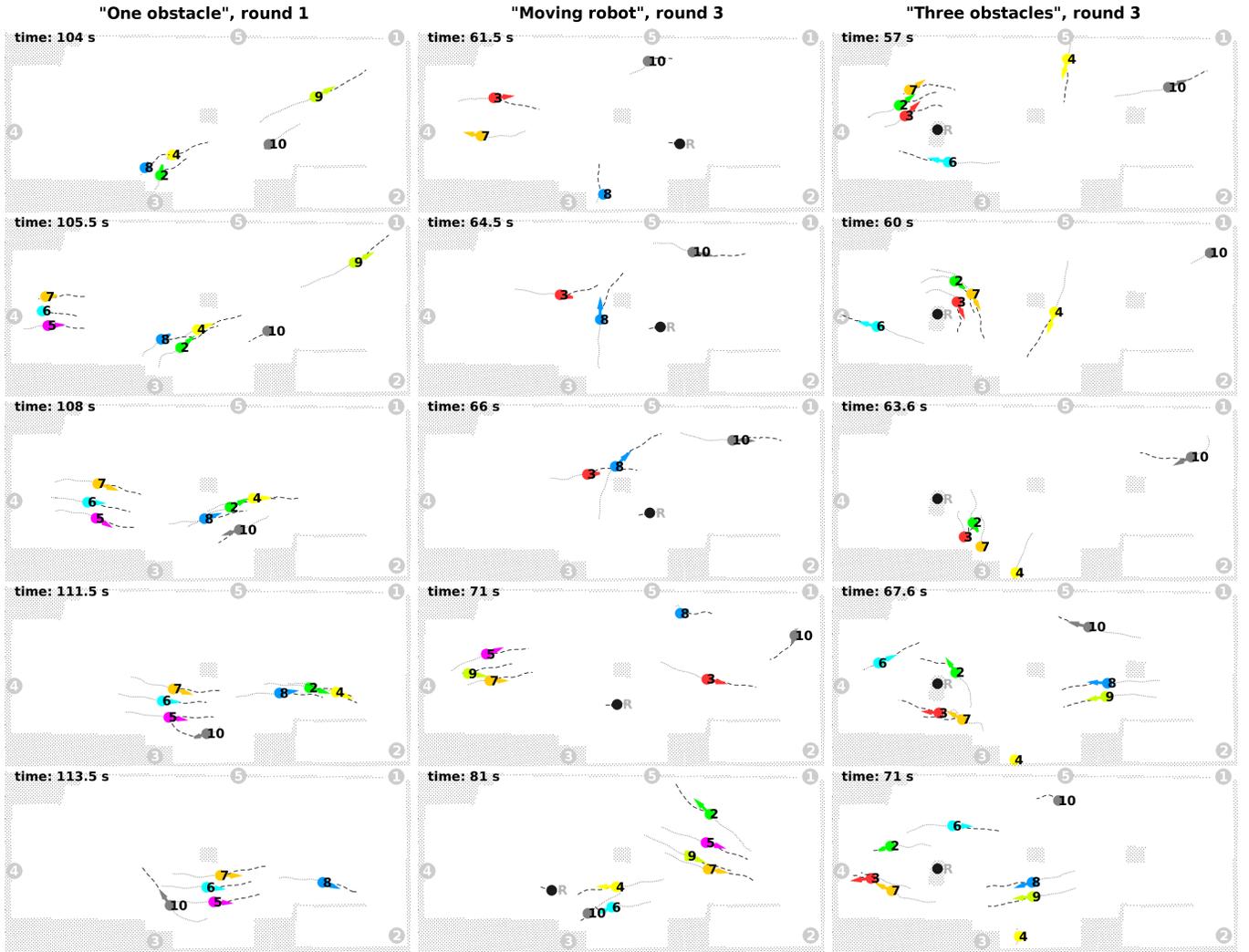


Fig. 6. Social interactions in the THÖR dataset with color-coded positions of the observed people. The current velocity is shown with an arrow of corresponding length and direction. The past and the future 2 s trajectories are shown with dotted and dashed lines respectively. Goal locations are marked with gray circles. **Left column:** at 104 sec the group (2,4,8) starts moving from the goal point, taking the *line formation* in the constrained space due to the presence of standing person 10. Later, at 111.5 sec, person 10 has to adjust the path and slow down while the group (5,6,7) proceeds in the *V formation* [35], engaged in communication. **Middle column:** person 8 is leaving the resting position at 61.5 sec and adapts the path to account for the motion of the robot, taking a detour from the optimal way to reach the goal 5. At 66 seconds person 8 crosses person 3, who has to slow down, as compared to the velocity at time 61.5 and 71. The same maneuver of taking a detour due to the presence of the robot is performed by the group (5,7,9) at time 71. **Right column:** Group (2,3,7), navigating in a constrained environment, at 57 sec has to make a detour around the obstacle while heading to goal 3. On the way back to goal 4 the group splits at 67.6 sec, and reunites later on.

100 Hz (THÖR) vs.  $\sim 26$  Hz (ATC), the amount of perception noise in the trajectories is lower than in all baselines. The speed distribution of  $\pm 0.49 \text{ m s}^{-1}$  shows that the range of observed velocities corresponds to the baselines, while the lower average velocity in combination with a high average curvature confirms higher complexity of the recorded behaviors, because comfortable navigation in straight paths with constant velocity is not possible in presence of static and dynamic obstacles. Finally, the high variance of the minimal distance between people ( $1.54 \pm 1.60 \text{ m}$  THÖR vs.  $0.61 \pm 0.16 \text{ m}$  ATC) shows that our dataset features both dense and sparse scenarios, similarly to ETH and Edinburgh.

An important advantage of THÖR in comparison to the prior art is the availability of rich interactions between the participants and groups in presence of static obstacles and the

moving robot. In this compact one hour recording we observe numerous interesting situations, such as accelerating to overtake another person; cutting in front of someone; halting to let a large group pass; queuing for the occupied goal position; group splitting and re-joining; choosing a sub-optimal motion trajectory from a different homotopy class due to a narrow passage being blocked; hindrance from walking towards each other in opposite directions. In Fig. 6 we illustrate several examples of such interactions.

## V. CONCLUSIONS

In this paper we present a novel human motion trajectories dataset, recorded in a controlled indoor environment. Aiming at applications in training and benchmarking human-aware intelligent systems, we designed the dataset to include a

rich variety of human motion behaviors, interactions between individuals, groups and a mobile robot in the environment with static obstacles and several motion targets. Our dataset includes accurate motion capture data at high frequency, head orientations, eye gaze directions, data from a stationary 3D lidar sensor and an RGB camera. Using a novel set of metrics for the dataset quality estimation, we show that it is less noisy and contains higher variety of behavior than the prior art datasets.

#### ACKNOWLEDGMENTS

The authors would like to thank Martin Magnusson for his invaluable support with the motion capture system and Luigi Palmieri for insightful discussions.

#### REFERENCES

- [1] A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and K. O. Arras, "Human motion trajectory prediction: A survey," *arXiv preprint arXiv:1905.06113*, 2019.
- [2] S. Pellegrini, A. Ess, K. Schindler, and L. van Gool, "You'll never walk alone: Modeling social behavior for multi-target tracking," in *Proc. of the IEEE Int. Conf. on Computer Vision (ICCV)*, 2009, pp. 261–268.
- [3] B. Majecka, "Statistical models of pedestrian behaviour in the forum," *Master's thesis, School of Informatics, University of Edinburgh*, 2009.
- [4] 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 Comp. Vision (ECCV)*. Springer, 2016, pp. 549–565.
- [5] D. Bršćić, T. Kanda, T. Ikeda, and T. Miyashita, "Person tracking in large public spaces using 3-d range sensors," *IEEE Trans. on Human-Machine Systems*, vol. 43, no. 6, pp. 522–534, 2013.
- [6] Z. Yan, T. Duckett, and N. Bellotto, "Online learning for human classification in 3D LiDAR-based tracking," in *Proc. of the IEEE Int. Conf. on Intell. Robots and Syst. (IROS)*, 2017, pp. 864–871.
- [7] B. Zhou, X. Wang, and X. Tang, "Understanding collective crowd behaviors: Learning a mixture model of dynamic pedestrian-agents," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 2871–2878.
- [8] A. Lerner, Y. Chrysanthou, and D. Lischinski, "Crowds by example," in *Computer Graphics Forum*, vol. 26, no. 3. Wiley Online Library, 2007, pp. 655–664.
- [9] S. Oh *et al.*, "A large-scale benchmark dataset for event recognition in surveillance video," in *Proc. of the IEEE Conf. on Comp. Vis. and Pat. Rec. (CVPR)*, 2011, pp. 3153–3160.
- [10] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.
- [11] B. Benfold and I. Reid, "Stable multi-target tracking in real-time surveillance video," in *Proc. of the IEEE Conf. on Comp. Vis. and Pat. Rec. (CVPR)*, 2011, pp. 3457–3464.
- [12] C. Dondrup, N. Bellotto, F. Jovan, and M. Hanheide, "Real-time multi-sensor people tracking for human-robot spatial interaction," in *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA), Works. on ML for Social Robo.* IEEE, 2015.
- [13] B. Lau, K. O. Arras, and W. Burgard, "Tracking groups of people with a multi-model hypothesis tracker," in *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2009.
- [14] T. Linder, S. Breuers, B. Leibe, and K. O. Arras, "On multi-modal people tracking from mobile platforms in very crowded and dynamic environments," in *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2016.
- [15] A. F. Foka and P. E. Trahanias, "Probabilistic autonomous robot navigation in dynamic environments with human motion prediction," *Int. Journal of Social Robotics*, vol. 2, no. 1, pp. 79–94, 2010.
- [16] H. Bai, S. Cai, N. Ye, D. Hsu, and W. S. Lee, "Intention-aware online pomdp planning for autonomous driving in a crowd," in *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*, May 2015, pp. 454–460.
- [17] L. Palmieri, T. P. Kucner, M. Magnusson, A. J. Lilienthal, and K. O. Arras, "Kinodynamic motion planning on gaussian mixture fields," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 6176–6181.
- [18] C. S. Swaminathan, T. P. Kucner, M. Magnusson, L. Palmieri, and A. J. Lilienthal, "Down the cliff: Flow-aware trajectory planning under motion pattern uncertainty," in *2018 IEEE/RSSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 7403–7409.
- [19] B. Okal and K. O. Arras, "Learning socially normative robot navigation behaviors with bayesian inverse reinforcement learning," in *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2016.
- [20] S.-Y. Chung and H.-P. Huang, "Incremental learning of human social behaviors with feature-based spatial effects," in *Proc. of the IEEE Int. Conf. on Intell. Robots and Syst. (IROS)*, 2012, pp. 2417–2422.
- [21] A. Rudenko, L. Palmieri, A. J. Lilienthal, and K. O. Arras, "Human motion prediction under social grouping constraints," in *Proc. of the IEEE Int. Conf. on Intell. Robots and Syst. (IROS)*, 2018.
- [22] P. A. Lasota, T. Fong, and J. A. Shah, "A survey of methods for safe human-robot interaction," *Foundations and Trends in Robotics*, vol. 5, no. 4, pp. 261–349, 2017.
- [23] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social LSTM: Human trajectory prediction in crowded spaces," in *Proc. of the IEEE Conf. on Comp. Vis. and Pat. Rec. (CVPR)*, 2016, pp. 961–971.
- [24] S.-Y. Lo, S. Alkoby, and P. Stone, "Robust motion planning and safety benchmarking in human workspaces," in *SafeAI@ AAAI*, 2019.
- [25] V. V. Unhelkar, C. Pérez-D'Arpino, L. Stirling, and J. A. Shah, "Human-robot co-navigation using anticipatory indicators of human walking motion," in *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2015, pp. 6183–6190.
- [26] A. Bera, T. Randhavane, and D. Manocha, "Aggressive, tense, or shy? Identifying personality traits from crowd videos," in *Proc. of the Int. Conf. on Artificial Intelligence (IJCAI)*, 2017, pp. 112–118.
- [27] W.-C. Ma, D.-A. Huang, N. Lee, and K. M. Kitani, "Forecasting interactive dynamics of pedestrians with fictitious play," in *Proc. of the IEEE Conf. on Comp. Vis. and Pat. Rec. (CVPR)*, 2017, pp. 4636–4644.
- [28] S. Molina, G. Cielniak, T. Krajník, and T. Duckett, "Modelling and predicting rhythmic flow patterns in dynamic environments," in *Annual Conf. Towards Autonom. Rob. Syst.* Springer, 2018, pp. 135–146.
- [29] T. P. Kucner, M. Magnusson, E. Schaffernicht, V. H. Bennetts, and A. J. Lilienthal, "Enabling flow awareness for mobile robots in partially observable environments," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 1093–1100, 2017.
- [30] A. Doshi and M. M. Trivedi, "On the roles of eye gaze and head dynamics in predicting driver's intent to change lanes," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 3, pp. 453–462, 2009.
- [31] O. Palinko, F. Rea, G. Sandini, and A. Sciutti, "Robot reading human gaze: Why eye tracking is better than head tracking for human-robot collaboration," in *Proc. of the IEEE Int. Conf. on Intell. Robots and Syst. (IROS)*, 2016.
- [32] H. Admoni and B. Scassellati, "Social eye gaze in human-robot interaction: a review," *Journal of Human-Robot Interaction*, vol. 6, no. 1, pp. 25–63, 2017.
- [33] P. Kiefer, I. Giannopoulos, M. Raubal, and A. Duchowski, "Eye tracking for spatial research: Cognition, computation, challenges," *Spatial Cognition & Computation*, vol. 17, no. 1-2, pp. 1–19, 2017.
- [34] R. T. Chadalavada, H. Andreasson, M. Schindler, R. Palm, and A. J. Lilienthal, "Bi-directional navigation intent communication using spatial augmented reality and eye-tracking glasses for improved safety in human-robot interaction," *Robotics and Computer-Integrated Manufacturing*, vol. 61, p. 101830, 2020.
- [35] M. Moussaïd, N. Perozo, S. Garnier, D. Helbing, and G. Theraulaz, "The walking behaviour of pedestrian social groups and its impact on crowd dynamics," *PloS one*, vol. 5, no. 4, p. e10047, 2010.