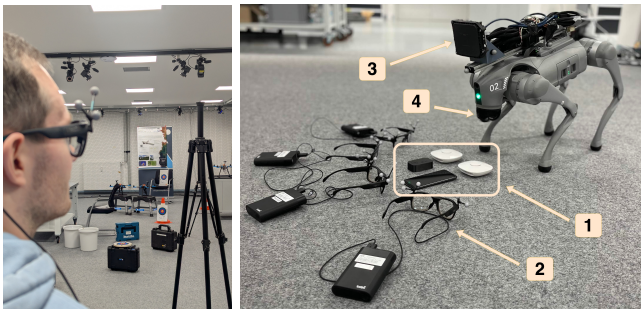
Collecting Human Motion Data in Large and Occlusion-Prone Environments using Ultra-Wideband Localization

Janik Kaden[1], Maximilian Hilger[2], Tim Schreiter[2,3], Marius Schaab[2],
Thomas Graichen[4], Andrey Rudenko[5], Ulrich Heinkel[1], and Achim J. Lilienthal[2,3]

***Abstract*— With robots increasingly integrating into human environments, understanding and predicting human motion is essential for safe and efficient interactions. Modern human motion and activity prediction approaches require high quality and quantity of data for training and evaluation, usually collected from motion capture systems, onboard or stationary sensors. Setting up these systems is challenging due to the intricate setup of hardware components, extensive calibration procedures, occlusions, and substantial costs. These constraints make deploying such systems in new and large environments difficult and limit their usability for in-the-wild measurements. In this paper we investigate the possibility to apply the novel Ultra-Wideband (UWB) localization technology as a scalable alternative for human motion capture in crowded and occlusion-prone environments. We include additional sensing modalities such as eye-tracking, onboard robot LiDAR and radar sensors, and record motion capture data as ground truth for evaluation and comparison. The environment imitates a museum setup, with up to four active participants navigating toward random goals in a natural way, and offers more than 130 minutes of multi-modal data. Our investigation provides a step toward scalable and accurate motion data collection beyond vision-based systems, laying a foundation for evaluating sensing modalities like UWB in larger and complex environments like warehouses, airports, or convention centers.**

I. INTRODUCTION

Understanding human motion is a cornerstone for intelligent robots to interact seamlessly with humans in shared spaces. Recent advances in human motion prediction do not only use geometric and velocity information but also leverage semantic and contextual cues for more accurate performance [1]. These approaches often depend on substantial amounts of high-quality data for training and evaluation, e.g., generated using a motion capture system [2] or high-resolution LiDAR sensors [3]. Acquiring this data is costly and limited by the volume covered by the motion capture system or by occlusions in the LiDAR data. To generate larger-scale datasets, it is necessary to establish new ways of collecting human motion data. Ultra-Wideband (UWB) technology is an increasingly adopted solution for precise indoor localization in the consumer market, as a growing number of smartphone manufacturers integrate UWB hardware into their devices [4], [5]. In 2023, native UWB positioning on a smartphone was demonstrated for the first time [6].



(a) Environment (b) Modalities

Fig. 1: **Overview of the datasets modalities and recording environment:** The UWB system (1), eye-tracking glasses (2) and the robot with radar (3) and LiDAR (4) sensors.

With the untracked navigation use case standardized by the FiRa consortium[1] in their 2.0 specification, GPS-comparable solutions for smartphones with indoor accuracy in the decimeter range are now possible. This standardization enables UWB device interoperability and supports a broad adoption and integration across consumer and industrial applications. The underlying Downlink Time-Difference-of-Arrival method allows scalable and private positioning directly on the user's device. Accuracy can be improved by fusing with additional device sensors, such as an Inertial Measurement Unit (IMU). Due to standardization and growing demand for precise indoor navigation, the availability of UWB-enabled smartphones will continue to increase. This will provide new digital experiences in venues like museums, which until now have relied on specially tailored solutions based on less accurate technologies like Bluetooth [7].

In this work, we investigate extension of the THÖR protocol for human motion data collection [8] to include UWB tracking of moving people. THÖR features a scripted indoor environment to generate goal-driven and natural human motion in crowded social spaces containing fixed obstacles and a moving robot. The participants draw random cards at the goal points, indicating the next target point. THÖR-MAGNI [2] further extends this protocol by enriching the

[1]Chemnitz University of Technology, Chair for Circuit and System Design, Chemnitz, Germany, firstname.name@etit.tu-chemnitz.de
[2]Munich Institute of Robotics and Machine Intelligence (MIRMI), Technical University of Munich, Germany, firstname.name@tum.de
[3]Center for Applied Autonomous Sensor Systems (AASS), Örebro, Sweden, firstname.name@oru.se
[4]Pinpoint GmbH, Chemnitz, Germany, thomas.graichen@pinpoint.de
[5]Bosch Corporate Research, Germany, andrey.rudenko@de.bosch.com
This work was supported by the EU Horizon 2020 No. 101017274 (DARKO) and by the Federal Ministry of Education and Research (BMBF) within the project ELFE as part of the WIR! program.

[1]https://www.firaconsortium.org/

environment with semantic contexts, such as areas of caution or one-way passages, and adds diverse tasks and activities for people, including several modes of interaction with the robot. It contains over 3.5 hours of motion data for 40 participants. Both THÖR and THÖR-MAGNI are open source[2] and are used to improve human motion prediction algorithms [1], e.g., based on deep learning [9], causal discovery [10] or physics-based methods [11]. In both datasets, 3D LiDAR data is also available.

In this paper, we extend the THÖR data collection setup from an industrial to a public museum environment. Museum layouts are designed to attract the visitor's visual attention [12], [13]. Hence, the obstacle setup in our dataset intends to steer the participants' visual attention similarly, which we can quantify by recording eye-tracking data. The dataset features multiple goal points (museum exhibits) and diverse static obstacles in the room, encouraging natural human motion behavior. In addition, a robot equipped with both LiDAR and radar sensors is utilized. The availability of velocity measurements and the penetration abilities makes radar an interesting sensing modality in dynamic and occluded environments. Uniquely for this recording, 2D UWB trajectories are available for up to three participants at the same time. Finally, accurate ground truth positions are recorded by a motion capture system.

The rest of the paper is organized as follows: Sec. II describes the recording modalities and the interaction scenarios. We show preliminary results in Sec. III. After completing post-processing and data curation, the data of the UWB-localization, the robot's sensors, the motion capture system, and the raw eye-tracking data will be made available[3].

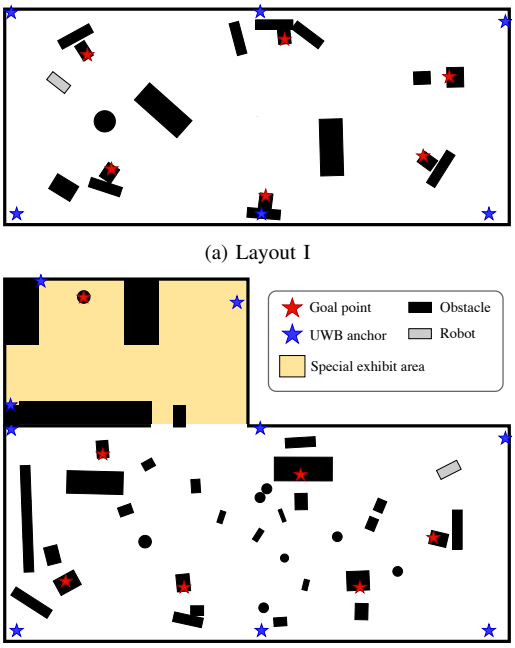
II. DATA COLLECTION

A. Room Setup

The data collection occurred in a robot lab at the German Aerospace Center (DLR). The main test area measures $15.64\,\mathrm{m} \times 6.68\,\mathrm{m}$ and is covered by the motion capture system and UWB. Six goal positions are marked within the test area. In some experiments, this room is complemented by a second room not covered by the motion capture system. This introduces an area without optical tracking to demonstrate the use of UWB as a standalone localization method. The room simulates a temporary exhibition space, similar to those commonly seen in museums where installations are frequently changed. We implement this setup in two layouts.

The obstacles in Layout I are designed to encourage the visual search behavior of participants' gaze. Two large tables block parts of the room; some of the goals are situated behind roll-ups and do not provide a direct line of sight of parts of the room. Fig. 4a shows a picture of the obstacle setup, and Fig. 2a depicts a map of the room layout. In Layout II, in contrast, many small obstacles are placed in the middle of the room to encourage navigating motion patterns. This forces the participants to choose between multiple different paths. A picture of the second layout is given in Fig. 4b, and Fig. 2b

[2]http://thor.oru.se
[3]https://doi.org/10.5281/zenodo.15211243



(a) Layout I

(b) Layout II

Fig. 2: **Top-down view of the indoor museum room setup.** Layout II also shows the additional "Special Exhibit" area.

illustrates a map of the layout. The robot is placed in a corner of the room in all the layouts mentioned above, except for the "Moving Robot" Scenario 6 introduced in Sec. II-B.

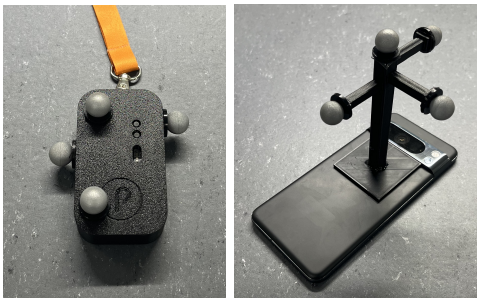
B. Scenario Description

In total, seven scenarios in two obstacle layouts are recorded to capture various movements and interactions. Each scenario includes multiple runs with varying numbers of participants, and each run is 3–5 minutes. The participants are assigned a random goal point with a deck of cards at the start of each run. Each participant draws one card indicating the next random goal target, returns it to the bottom of the card deck, and proceeds to the next goal point. The card decks are designed to favor longer and more complex paths between goal points.

Starting with an empty room, in Scenario 1, the participants are instructed to build up Layouts I or II, which were given by a floor map drawing. The participants thus not only have the Visitor role used in [2] and [8] but also have a similar role to the Carrier role where boxes had to be transported. After building up the layouts, regular test runs are conducted in Scenarios 3 and 4 with one to four simultaneous participants. Scenario 2 depicts the participants' regular (baseline) motion, where only the goal points are present in the volume. The "Special Exhibit" Scenario 5 includes the additional room depicted in Fig. II-A. The quadruped robot moves through the area in Scenario 6. After recording all runs for the layouts, the participants were instructed to deconstruct the layouts in Scenario 7.

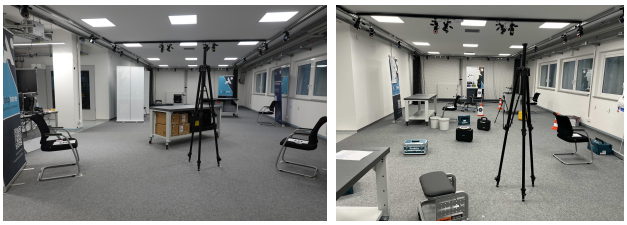
C. Sensing modalities

The dataset features a unique combination of sensing modalities. Our sensor setup is selected such that two claims



(a) TRACElet (b) Google Pixel 8 Pro

Fig. 3: **Two UWB receiver devices used in the runs.** Devices are equipped with IR-Markers for 6DoF MotionCapture



(a) Layout I (b) Layout II

Fig. 4: **The two obstacle layouts used in the recordings.** They feature a few large obstacles that need to be navigated around (a) and many small obstacles that can be stepped over (b). Retractable banners introduce occlusions in the environment, adding visual search behavior to the scenario.

can be verified prior to the larger-scale recordings: i) human motion data collection is possible with UWB, and ii) environment reconstruction in dynamic and occluded environments can be accomplished in a combination of radar and LiDAR. Motion capture data serves as accurate ground truth.

1) UWB Indoor Positioning: The UWB system is a FiRa 2.0 compliant indoor localization system by the German company Pinpoint GmbH[4]. In total, nine anchors (called SATlets) are installed at known locations in the room to provide localization coverage. These are comparable to GNSS satellites that send out information that the receivers use to calculate their own position. For the main area seen in Fig. 4, six SATlets were placed on the room's walls at a height of about $2.5\,\mathrm{m}$. For dedicated "UWB-only" runs, which included a special exhibit outside the motion capture coverage, three SATlets were added to cover this area. The placement is shown in Fig. 2, depicted by the blue stars. The UWB anchors are battery-powered and synchronize wirelessly, eliminating the need to run cables throughout the room and significantly reducing setup time. After manually measuring the room with a laser distance meter having an accuracy of $\pm\,3\,\mathrm{mm}$, the Pinpoint app "EasyPlan" was used to configure the SATlets. This includes the position in the local coordinate system and some UWB-specific physical layer (PHY) parameters, such as the channel or preamble. Previous work has shown that the UWB PHY impacts the performance of UWB connections between devices [14]. Lower-frequency configurations generally support longer communication ranges, while higher-frequency settings offer better resilience against interference from other wireless technologies, such as WiFi 6E. Given the short distances in our exhibit area, we selected a FiRa-compatible configuration optimized for robustness against potential interference.

Three different UWB receiver devices were used to capture the trajectory of the participants: a TRACElet (battery-powered tag) and two smartphones (Samsung Galaxy S24+ and Google Pixel 8 Pro). The TRACElet was attached to a lanyard the participants wore during the runs, and the smartphones were handheld. The recorded 2D UWB positions represent the raw data obtained directly from the devices at $4\,\mathrm{Hz}$, without additional sensor fusion.

2) Eye tracking: Four wearable eye-tracking glasses (Tobii Pro Glasses 3) were used to capture the gaze information of the participants. Each pair of glasses consists of a head unit worn like a regular pair of glasses. If the participant needed a visual aid, corrective lenses were attached. We used the Tobii Motion Capture Marker sets and the Vicon integration to simultaneously track the participants and record the gaze information using the Vicon Tracker software[5]. This ensures the data is synchronized and the gaze information can be used in the motion capture coordinate system. The glasses with the recording units and the motion capture marker sets attached are shown in Fig. 1.

3) Robot with LiDAR and radar: The dataset features a UniTree Go2 quadruped robot, equipped with a UniTree L1 LiDAR with built-in IMU and a Bosch Off-Highway Premium radar. The low-cost LiDAR has a non-repetitive scan pattern. It produces 21.6k points per second, substantially lower than the 2.6m points per second available in similar datasets [2]. The sparsity of the data makes human motion tracking and Simultaneous Localization and Mapping (SLAM) in dynamic settings more challenging because less geometric context is available to detect humans. This challenge can be addressed using our proposed data collection setup. The radar sensor outputs a sparse point cloud comprising 3D geometrical information and radial velocity obtained through the Doppler effect. In particular, the velocity information can be leveraged for human motion tracking, for mapping of dynamics [15], and to detect moving objects in SLAM. Additionally, radar waves can penetrate through some materials, which benefits mapping in occluded environments. The maps estimated by SLAM, in turn, can be used to understand the recorded human trajectory data, especially if semantic attributes are deducted from the point clouds during [16] or after mapping [17]. The point clouds and IMU measurements were recorded in ROS2 bag files.

4) Motion capture: A Vicon motion capture system is used to obtain a highly accurate ground truth for the participants, the robot, and the UWB devices. It consists of 26 cameras installed in the ceiling of the room. These cameras were calibrated prior to the recordings and covered the entire volume of the recording room, including the participants' eye-tracking glasses and the robot on the floor. All tracked

[4]https://pinpoint.de/en

[5]https://help.vicon.com/space/Tracker310

TABLE I: Overview of all scenarios

Scenario	Description	Participants	Time recorded
1	Build-up	3	16 min + 6 min
2	Baseline	1 – 4	8 × 3 min
3	Layout I	1 – 4	8 × 3 min
4	Layout II	1 – 4	8 × 3 min
5	Special Exhibit	3	6 × 4 min
6	Moving Robot	3	2 × 5 min
7	Deconstruction	3	2 × 4 min
Total			136 min

objects are equipped with IR markers, as seen in Fig. 1 and Fig. 3, for 6 DoF-based tracking of rigid bodies with the motion capture software. This software supports the direct integration of the Tobii eye-tracking glasses. Hence, the motion capture system recordings include the poses of the rigid bodies and the position of the left and right eye, pupil diameter, left and right gaze, and gaze position for all participants. The data is stored in a CSV format.

D. Recording Procedure

One experimenter operated the motion capture, UWB, and robot recording software to ensure the recording quality and check the status of all systems. We followed a precise workflow for each recording to have reproducible results. An eye tracker calibration routine was carried out for each participant by looking into a calibration card to achieve the best possible results. A successful calibration was indicated in the motion capture system. The recording of the motion capture, UWB, and robot systems was then started.

III. RECORDED DATA

As described in Sec. II-B, each recording lasts at least 3 minutes. The values for each scenario are given in Table I. This results in over 130 minutes of multi-modal data: eye-gaze data from up to four eye-tracking glasses, 2D UWB trajectories, poses of the motion capture system, and radar and LiDAR point clouds captured by the robot. Fig. 5 shows the trajectories in both layouts with four participants and a static robot. It is visible that the higher number of obstacles leads to more complex trajectories between the goals.

To evaluate the accuracy of the UWB-based localization, we compare the trajectories recorded by the motion capture system and the UWB system. We calculate the mean 2D displacement error using the root mean square error to quantify the system's precision. Fig. 6 shows 1 minute of movement in Scenario 3, recorded using the Pixel smartphone and having an average error of 41 cm.

To showcase the mapping capabilities of the robots' sensors, we used the trajectories recorded with the motion capture system in Scenario 4 to accumulate LiDAR and radar points in a voxel grid. We filter out dynamic points belonging to moving participants using the tracked eye-tracking glasses. For the radar point cloud, noise points outside the experimental volume are discarded. Fig. 7 depicts the recorded maps. Even with the sparse measurements, the LiDAR map shows accurate geometry. Compared to that, the radar map is more sparse and contains no ground reflections. Future work will investigate if the same maps can be estimated solely from the robot's onboard sensors.

(a) Scenario 3 (Layout I) (b) Scenario 4 (Layout II)

Fig. 5: **Trajectories of four participants, recorded with the motion capture system.** Results of a 3-minute run, colors corresponding to unique participants.

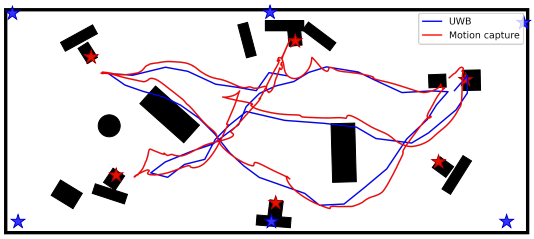
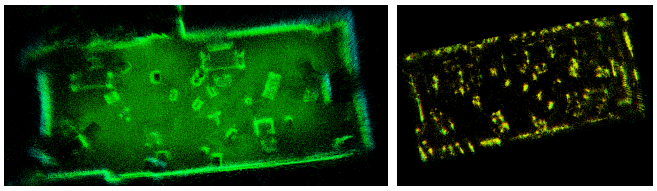


Fig. 6: **Top-down view of the trajectories recorded by the motion capture and the UWB system.** The figure illustrates 1 minute of movement from one participant holding the Google Pixel smartphone in hand during Scenario 3.



(a) LiDAR map (b) Radar map

Fig. 7: **Point cloud maps of Layout II generated based on ground truth trajectories.** The LiDAR map captures more geometric details compared to the radar map.

IV. CONCLUSION AND FUTURE WORK

We present a novel set of human motion recordings in a simulated contextually rich indoor museum-like environment. The dataset features a unique set of technological modalities combining motion capture, UWB indoor localization, eye-tracking glasses, radar, LiDAR, and a moving quadruped robot. With that, we pave the way toward large-scale dataset recordings in real-world settings, leveraging consumer hardware. Future work will thoroughly evaluate the performance of the UWB localization system compared to the ground truth motion capture. Furthermore, leveraging the potential of UWB people tracking and 3D environment reconstruction from on-board robot sensors, we aim to investigate the possibility to collect the gaze data in crowded environments without motion capture systems.

ACKNOWLEDGMENT

The authors would like to thank the German Aerospace Center (DLR) for supporting this work by providing their facilities. In particular, we thank Thomas Wiedemann and Patrick Hinsen for their support during the dataset recording. We also thank Martin Magnusson for the fruitful 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," *The International Journal of Robotics Research*, vol. 39, no. 8, pp. 895–935, 2020.

[2] T. Schreiter, T. R. de Almeida, Y. Zhu, E. Gutierrez Maestro, L. Morillo-Mendez, A. Rudenko, L. Palmieri, T. P. Kucner, M. Magnusson, and A. J. Lilienthal, "THÖR-MAGNI: A large-scale indoor motion capture recording of human movement and robot interaction," *The International Journal of Robotics Research*, vol. 44, no. 4, pp. 568–591, 2024.

[3] M. Ehsanpour, F. Saleh, S. Savarese, I. Reid, and H. Rezatofighi, "Jrdb-act: A large-scale dataset for spatio-temporal action, social group and activity detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 20 983–20 992.

[4] Qorvo, "Qorvo Delivers Ultra-Wideband in Google Pixel 6 Pro," https//www.qorvo.com/newsroom/news/2021/qorvo-delivers-ultra-wideband-in-google-pixel-6-pro, Dec. 2021.

[5] Estimote. Why Apple's 2nd Gen UWB Chip is Exciting? [Online]. Available: https://blog.estimote.com/post/728317898359259136/whats-exciting-about-u2-the-second-gen-uwb-chip

[6] Qorvo, "Qorvo® Demonstrates UWB Indoor Navigation on Commercial Smartphones at Embedded World 2023," https://www.qorvo.com/newsroom/news/2023/qorvo-demonstrates-uwb-indoor-navigation-on-commercial-smartphones-embedded-world-2023, Mar. 2023.

[7] Kyuder Tsedenov. Indoor Navigation: A Comprehensive Guide. [Online]. Available: https://navigine.com/blog/indoor-navigation-a-comprehensive-guide/

[8] A. Rudenko, T. P. Kucner, C. S. Swaminathan, R. T. Chadalavada, K. O. Arras, and A. J. Lilienthal, "THÖR: Human-Robot Navigation Data Collection and Accurate Motion Trajectories Dataset," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 676–682, 2020.

[9] T. R. de Almeida, A. Rudenko, T. Schreiter, Y. Zhu, E. G. Maestro, L. Morillo-Mendez, T. P. Kucner, O. M. Mozos, M. Magnusson, L. Palmieri, K. O. Arras, and A. J. Lilienthal, "THOR-Magni: Comparative Analysis of Deep Learning Models for Role-Conditioned Human Motion Prediction," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 2200–2209.

[10] L. Castri, S. Mghames, M. Hanheide, and N. Bellotto, "Causal discovery of dynamic models for predicting human spatial interactions," in *International Conference on Social Robotics*. Springer, 2022, pp. 154–164.

[11] A. Rudenko, L. Palmieri, W. Huang, A. J. Lilienthal, and K. O. Arras, "The atlas benchmark: An automated evaluation framework for human motion prediction," in *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2022, pp. 636–643.

[12] J. Krukar, "Walk, Look, Remember: The Influence of the Gallery's Spatial Layout on Human Memory for an Art Exhibition," *Behavioral Sciences*, vol. 4, no. 3, pp. 181–201.

[13] M. L. Harvey, R. J. Loomis, P. A. Bell, and M. Marino, "The Influence of Museum Exhibit Design on Immersion and Psychological Flow," *Environment and Behavior*, vol. 30, no. 5, pp. 601–627.

[14] J. Kaden, E. Markert, and U. Heinkel, "Performance Investigation for IEEE 802.15.4z-compliant SiP-assisted Ranging," in *2024 IEEE 37th International System-on-Chip Conference (SOCC)*, 2024, pp. 1–6.

[15] 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.

[16] X. Chen, A. Milioto, E. Palazzolo, P. Giguère, J. Behley, and C. Stachniss, "Suma++: Efficient lidar-based semantic slam," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 4530–4537.

[17] L. Sun, Z. Yan, A. Zaganidis, C. Zhao, and T. Duckett, "Recurrent-octomap: Learning state-based map refinement for long-term semantic mapping with 3-d-lidar data," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3749–3756, 2018.