Dataset Collection for Long-term Forecasting of Human Presence, Motion and Activity

Tomáš Vintr[1], George Broughton[1], Tomáš Rouček[1], Zhi Yan[2], Tomáš Krajník[1]

***Abstract*— Robots are expected to be able to navigate in crowded or human-populated environments. The dynamics of such environments are, however, the major obstacle to autonomous robot deployments. Development of models that can predict and forecast changes in human-populated areas is therefore crucial for autonomous robots intended to help people in their environment. In this work, we propose and discuss a new dataset collection that shall be executed in the following years. We aim to gather a new, long-term dataset composed of multiple sensory outputs from diverse types of environments. The data will also be post-processed to extract the semantic data and prepared for use by the robotic scientific community.**

I. Introduction

During the last decade we have seen much development in the world of robotics, but fully autonomous robotic deployments still remain elusive. One of the primary reasons for this is that while their navigation algorithms work well, their robustness outside of lab conditions can be less-than ideal, and therefore they still require some level of human oversight.

Traditional approaches to robot navigation in uncontrolled environments use static maps combined with reactive planning for unexpected events. However, many unexpected events are caused by human actions, especially movement through the robot's operational environment. Reactive re-planning using sense-plan-act frameworks, such as those used in the Robot Operating System (ROS) [1] *move_base* package, can lead to a perception of clumsiness [2], slow response, and frequent re-planning, resulting in negative emotions towards the robot [3]. Failing to navigate in human-populated or crowded environments is a major obstacle to robotic deployments and their acceptance in the long-term. Therefore, incorporating expected human movement into the robot's navigation and planning is crucial for autonomous robots intended to help people in their environment.

Addressing this, several large projects attempted to work on autonomous robots operating in human-populated environments [4], [5], [6]. These environments tend to be dynamic, with natural daily and seasonal changes [7], but more importantly, their dynamics are shaped by human actions [8]. A promising way to achieve long-term human-aware navigation is to include spatio-temporal maps into the navigation systems to support decision-making in advance of the navigational task. Robots which forecast the dynamics of human customs and adjust their decisions accordingly perform better at human-centric tasks [5], [9], [10], [11], [12], [13]. Furthermore, it has been shown that autonomous robots that take human habits into account are more likely to be accepted by society [3], [4], [14].

This work was supported by a joint project funded by the Czech Ministry of Education No. 8J23FR023 and PHC Barrande (No. 49275QM) project '3L4AR', the Bourgogne-Franche-Comté regional research project 'LOST-CoRoNa' and OP VVV MEYS RCI project CZ.02.1.01/0.0/0.0/16_019/0000765. [1]Faculty of Electrical Engineering, Czech Technical University, Czechia
[2]UTBM, CIAD, F-90010 Belfort, France

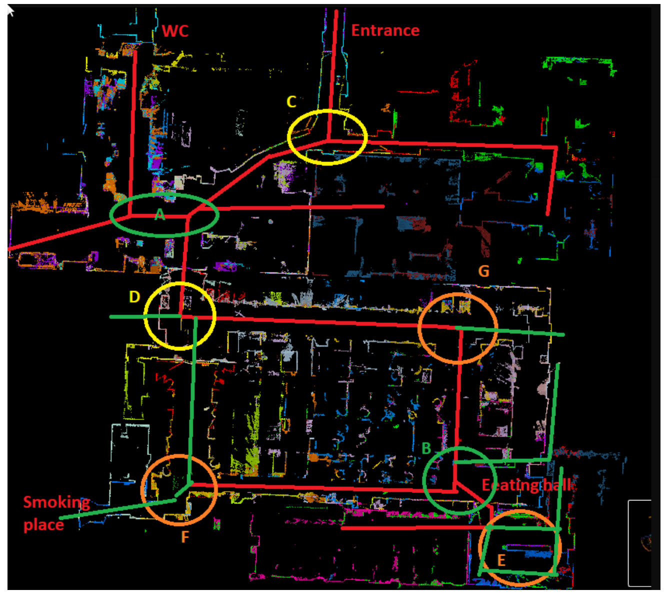

Fig. 1. Tesla factory lidar scan. Red lines highlight the usual path of people, while green ones are used less frequently. Green circles highlight the most interesting crossroads - A is the nearest crossroad to the bathroom, and B includes a resting area. Yellow circles highlight less frequent crossing with expectedly strong temporal patterns, and orange circles highlight places where people usually stay for some time.

Developing spatio-temporal predictive maps requires long-term datasets of natural human environments. The most known dataset of this kind is the ATC dataset of a shopping mall [15]. Although it consists of detections taken throughout one year, the data is collected only during opening hours on Wednesdays and Sundays. In addition, researchers from the Lincoln Centre for Autonomous Systems (L-CAS) gathered their own UoL dataset [16] that consists of several consecutive weeks of lidar point clouds and human detections [17]. However, local security rules did not allow researchers to gather the data out of working hours. Using a very similar system, researchers created the UTBM dataset [18] that was supposed to be long-term, similar to the ATC one, and continuous. The effort resulted in a dataset several months long covering entire days. However, due to political problems in the country that led to the universities closing, a large

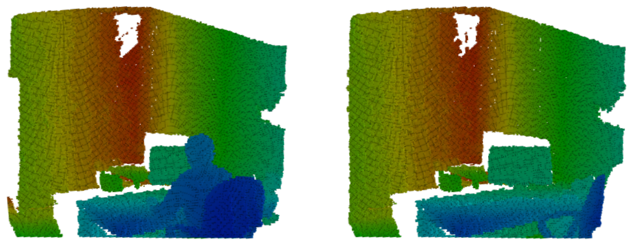

Fig. 2. MHT lecturer office dataset: snapshots of the person present and absent.

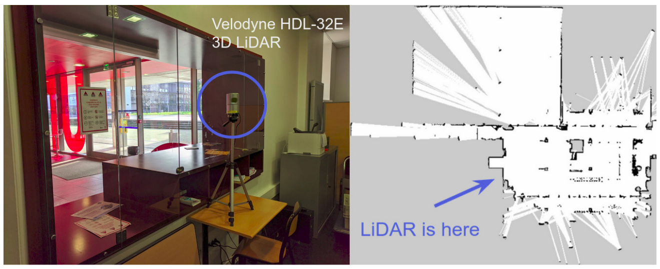

Fig. 3. UTBM dataset: Velodyne HDL-32E 3D lidar position near the UTBM main entrance hall (left) and a birds-eye view of the lidar's environment. Courtesy of [18].

part of the dataset consists of empty halls. Changes in security rules and the following COVID-19 epidemic led to an inability of researchers to continue collecting the data. The meaningful data now includes one month and one week of continuous data.

Despite the discontinuity in the ATC dataset, it is still commonly used [19], [20], [21]. From our experience, a more complex dataset is needed for developing an autonomous system that includes service robots with various tasks. The service robots do not only work during opening hours; they need to cooperate with technical and service staff who work in public places outside of working hours. For example, security robots need to understand differences in behaviour between customers and utility workers. Therefore, we are starting to gather continuous long-term datasets from a factory (see Fig. 1). As the data gathering and dataset preparation is planned for the next few years, we are looking for advice and possible cooperation with scientists from the long-term human motion prediction community.

II. Datasets

A. Our Current Datasets

For year-long forecasts, we usually use open data 'MHT building lecturer office' for testing long-term models. The data was gathered during the STRANDS project [5] that can be found in the project's web pages [22]. The data consists of preprocessed video frames with a frame rate of $0.2Hz$. Every frame consists of a set of depth values captured with a 320×240 RGB-D camera (see Fig. 2). The length of the video is more than 2 years with several few-days-long gaps. The video shows one of the lecturer's offices at the University of Lincoln.

The second dataset we usually use to test our models is the UTBM dataset [18]. The data is not directly accessible to the public, but only by request. It contains two long-term segments gathered during one year. The first segment was collected in one of the UTBM building halls of approximately $500m^2$. As shown in Fig. 3, a Velodyne 32-layer lidar was placed in the reception near the building door to ensure safe 24h operation. The spatial placement of the lidar was carefully determined to ensure maximum field-of-view (i.e., approximately $200m^2$) of the hall beyond the glass windows. The raw data of the lidar was recorded to ROS rosbags 24 hours per day for several months. However, due to political unrest, only one month (March 2019) includes normal student behaviour. The second segment, one week long, was collected in December 2019 in a similar way. There is also the third segment, still in the post-processing stage, collected in the same hall but with a 128-layer lidar outside but close to the reception. The data collection was performed for three consecutive days during March 2023.

The MHT dataset is two years long but includes only one significant phenomenon connected to human customs. The dataset's length allows us to test a model's deterioration over months-long periods. However, testing the models on one phenomenon can lead to overfitting the chosen scenario. The UTBM dataset consists of human flows in the entrance hall, but it is limited only to a few weeks. Although the data consists of many different people, their usual behaviour is walking through or waiting in the hall, with rare exceptions of utility workers providing their services. Moreover, the usable data cover only one topologically trivial area - the hall.

B. Available Datasets from Other Teams

The ATC dataset [15] covers an area of 900 m^2 over one year. The tasks people are performing are connected with shopping. We can find there different types of behaviour like walking from one shop to another, reading some information, waiting, chatting and so on. However, as was said earlier, the datasets consist only of Thursdays and Sundays, and only of working hours. The models developed over such data need not comprise subtle differences between individual working days and Saturday and Sunday. They cannot be tested in an ability to provide apparent differences between days and nights. We had a similar experience working on the UoL dataset [16]. One can expect that out of the working hours, there is nothing to care much about. However, an autonomous system should be prepared to utilise the time when the halls are empty, and therefore it needs to know what is happening during the night to schedule its tasks accordingly. Moreover, during non-working hours, there is a higher possibility of a security breach [23] that must be distinguished from rare but expectable events or regular utility activities.

One can find other human-dynamics-centred datasets like L-CAS [17], THÖR [24], or Magni [25] datasets. Although interesting, these datasets do not meet our concept of long-term datasets. They do not focus on human customs nor model deterioration.

C. Dataset under Construction

We made an agreement with the management and owners of a factory that we have the freedom to place sensors usually used in mobile robotics. The factory consists of two large sheds and multiple rooms with an overall area of approximately $2500m^2$, more than twice the area of the ATC dataset. Figure 1 shows a cross-section of a lidar scan of the factory. Our team, in cooperation with the factory management, chose and highlighted the best areas for data collection. The lines highlight the frequent paths the employees usually use. The circles highlight important crossings and places where employees assemble.

We aim to gather a new, similarly long or longer dataset than the MHT building lecturer office composed of people working on multiple tasks in an environment with a more complicated structure than the UTBM dataset. We expect people to perform more types of tasks than in UTBM datasets. Besides walking, waiting, and cleaning the area, we expect activities like resting, eating, refilling their water containers, and, similarly to the MHT dataset, working in front of their tables. That can give us a more generalised perspective in analysing the deterioration of models over time and the rules of pedestrian flows. Contrary to the ATC dataset, we also plan to gather the data continually during all days of the week and outside the usual working hours.

The number of people present in the area will probably not exceed 50 people, which is less than in UTBM and ATC datasets. However, two companies are in the building fabricating different products while sharing changing rooms, bathrooms, and the official entrance. One company is present there for approximately three decades while the second one is starting up, hiring new employees, and preparing new operations. As the building is positioned between two cities, a non-public bus for employees connects with public transportation. Such transportation is felicitous because the entry time of a large section of employees is determined by the most prolonged delay in documented public transit. The delays are usually connected with weather changes that can be included in the models and allow researchers to interconnect the predictions of human customs with weather forecasting.

The management is in favour of the experiments. They see cooperation with researchers as an opportunity to raise the working morale and moderate discomfort due to experiments as an acceptable price. On the other hand, the workers positively perceive the experiments as a non-traditional event in an overall invariable work. As a result, it is possible to agree with the management to delay the bus or move a shift a little bit to test the ability of models to detect unexpected or anomalous events as a part of the experimental design. There is also a possibility to perform field experiments with piloted or supervised robots.

D. Challenges in Data Gathering

The factory is a harsh environment. There is heavy machinery creating strong magnetic fields. The data cables need to be carried through high-voltage pipes. Computers need to be placed far away from heavy machinery and reactive chemicals. Nothing can be placed near the gas tubes or hydrogen tanks. The environment is dusty, and the air temperature in some places exceeds 40 degrees Celsius. Multiple security fences around some machines block a WiFi communication between the computers. On the other hand, the ceilings are $4-8$ meters high, which can give sensors the required scope.

Unlike in the UTBM dataset, human flows are not the only dynamics in the environment. There are also chairs moving from one place to another. Some cupboards are used, and some working positions are occupied only on different days. Large machines are run occasionally, and some rooms are accessible only for one or a few specific people during their particular tasks. Those dynamics are different from human flows but tightly bound to people's activity. Therefore, the structure of the area is not purely static but ”reacts” and changes its shape according to the actual organisation of the work. A successful autonomous system should predict these small structural changes because they are connected with the differences in human flows.

E. Privacy Issues

The privacy and data protection rules limit personal or otherwise sensitive data manipulation only inside the building. Although the management does not plan to use the personal data from the measurements, they find the sensors' placement as motivational components for the workers. The workers are used to working with expensive products, and their only concern about the data gathering is not to be identified by the sensors when they smoke near the hydrogen tanks.

The internet connection to the factory is stable but slow for data transfer. The data must be transferred from the factory manually, but the computers and sensors can be handled using a secure connection from outside. Personal or sensitive data needs to be processed or anonymised inside the compound.

III. Data Collection

Our primary way to collect data of human motion tends to be with lidar sensors. These have the advantage of being small, light-weight, and not intrusive on the environment. They also yield more fruitful data than highly bespoke sensors such as door position detectors, and detecting people passing through an area with a single beam. Furthermore, compared to cameras, lidar data is much easier to process as properties such as depth can be directly estimated (especially at long range), they're insensitive to lighting conditions, whilst also manipulation of raw data does not open privacy issues. Moreover, they have a large field-of-view. However this raises the question of whether the community would agree with these assessments for the task of human detection, or perhaps whether it would be advantageous to include other sensors in the data collection process.

The current arrangement for data collection involves having a series of computers set up running ROS. Each of these is running the necessary drivers of the lidar sensor, in this case a 128-layer Ouster sensor giving 3D coverage of the environment at $10Hz$. The sensors have slightly overlapping fields-of-view, to ensure the consistent ability to track individuals. The sensor is left running continuously, with another ROS node collecting the data, compressing it, and saving it to disk.

The throughput of the raw lidar data amounts to just over 100MB per second per sensor, an unwieldy size for this kind of usage. Consequently the run-time compression of the data is essential. While lossless lz4 compression offered approximately a 50% saving compared to the raw data, this kind of compression is still not enough for long-term use. As a result, a bespoke lossy compression was added.

The compression algorithm works by taking a key-frame from the lidar, and then in subsequent frames only storing lidar points with significant changes in their spatial position. These are found by matching the corresponding lidar beams between the frames, and finding the euclidean change in position. Jitter inevitably is present in the data, so there is some thresholding, but the suppression of all static background lidar points is not a significant challenge. Furthermore, when storing only key-frames and the deltas, traditional compression algorithms can still be used on the data to reduce the file-size further. As a result, the volume of data being stored was able to be reduced to below 0.1% of the original file size.

The data that saved as a result of this process has been modified from the raw data. Visibly, the jitter present on the walls of rooms is completely absent in the data processed in this way, while people are still visible moving around the environment. While very effective in making the dataset a manageable size, it is not clear whether the community would appreciate the data being processed in such a way. While generally the raw data has the benefit of being closer to reality, some compromise needs to be found, so the question exists if this kind of compression is acceptable for the community. We believe that such an approach will not affect the quality or usability of the dataset, as the dynamic objects will be captured in full, real-time detail, while static background objects such as walls, less likely to be of interest, will still be captured, but less often.

The rest of the data collection pipeline is run offline. The compressed data can be downloaded from each of the machines periodically, and then the data can be worked on. Firstly, they are decompressed individually back into a usable form. As mentioned, the background lidar points will be completely static, but any dynamic areas of the pointcloud will be present at the full frame rate.

Then, these dynamic points are passed to human detection algorithms to extract the semantic data from the scene. The detection algorithms are those used in the aforementioned UTBM data collection [17]. Specifically, the Adaptive Clustering algorithm is implemented on each frame of the pointcloud to divide it into different segments, ideally with each segment corresponds to an object. The segmentation performance of the algorithm for pedestrians is considered to be state-of-the-art [26]. Then the segments are filtered using a human volumetric model. On the other hand, since the UTBM building is completely closed at night and ensures that no one is stranded inside, we use the evening data as a reference background and then match it with daytime data to remove any false positive samples efficiently. As a result, human samples can be easily extracted from the point cloud without using any machine learning methods. It is worth noting that, there is an assumption here that the moving objects in the UTBM building fitted in the human volume model are all human beings.

Once the raw human detections have been extracted from the dataset, a final stage of filtering is applied before the final data is ready. This filtering process removes any spurious detections of people from the scene, and fills in any missing detections where otherwise there was continuous detections. As the raw data was collected at 10Hz, data association for tracking is relatively straight-forward, however as the original data is present, it is possible to verify this manually, or apply different methods for human detection and tracking.

The data about human detections and tracking is therefore provided as a convenience for people wishing to work with the dataset, however there is still the option for people to try their own detection and tracking algorithms on the data should they wish to.

IV. Conclusions

In this paper, we have looked at why long-term datasets of people will be important for the long-term future of robotics. Currently, there is a lack of high-quality datasets available for this task. In an attempt to correct this, we have started collecting our own dataset based on 3D lidars, to be used for analysis of the typical presence and movement patterns of people. In order to make this dataset as beneficial to the community as possible, we would like to obtain feedback on what would ideally be present in such a dataset, ranging from technical details over additional sensors to the spatial and temporal scale of the measurements. The dataset details, along with the forms to provide potential feedback and suggestions, is available at `http://3l4ar.science`.

References

[1] M. Quigley, K. Conley, B. P. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "ROS: an open-source robot operating system," in *ICRA Workshop on Open Source Software*, 2009.

[2] D. Hebesberger, T. Koertner, C. Gisinger, and J. Pripfl, "A long-term autonomous robot at a care hospital: A mixed methods study on social acceptance and experiences of staff and older adults," *International Journal of Social Robotics*, vol. 9, no. 3, pp. 417–429, 2017.

[3] T. Vintr, Z. Yan, K. Eyisoy, F. Kubiš, J. Blaha, J. Ulrich, C. S. Swaminathan, S. Molina, T. P. Kucner, M. Magnusson *et al.*, "Natural criteria for comparison of pedestrian flow forecasting models," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 11 197–11 204.

[4] R. Triebel, K. Arras, R. Alami, L. Beyer, S. Breuers, R. Chatila, M. Chetouani, D. Cremers, V. Evers, M. Fiore *et al.*, "Spencer: A socially aware service robot for passenger guidance and help in busy airports," in *Field and Service Robotics: Results of the 10th International Conference*. Springer, 2016, pp. 607–622.

[5] N. Hawes, C. Burbridge, F. Jovan, L. Kunze, B. Lacerda, L. Mudrova, J. Young, J. Wyatt, D. Hebesberger, T. Kortner *et al.*, "The strands project: Long-term autonomy in everyday environments," *IEEE Robotics & Automation Magazine*, vol. 24, no. 3, pp. 146–156, 2017.

[6] S. Coşar, M. Fernandez-Carmona, R. Agrigoroaie, J. Pages, F. Ferland, F. Zhao, S. Yue, N. Bellotto, and A. Tapus, "Enrichme: Perception and interaction of an assistive robot for the elderly at home," *International Journal of Social Robotics*, vol. 12, no. 3, pp. 779–805, 2020.

[7] T. Krajnik, J. P. Fentanes, G. Cielniak, C. Dondrup, and T. Duckett, "Spectral analysis for long-term robotic mapping," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 3706–3711.

[8] G. D. Tipaldi, D. Meyer-Delius, and W. Burgard, "Lifelong localization in changing environments," *The International Journal of Robotics Research*, vol. 32, no. 14, pp. 1662–1678, 2013.

[9] J. M. Santos, T. Krajník, J. P. Fentanes, and T. Duckett, "Lifelong information-driven exploration to complete and refine 4-d spatio-temporal maps," *IEEE Robotics and Automation Letters*, vol. 1, no. 2, pp. 684–691, 2016.

[10] J. M. Santos, T. Krajník, and T. Duckett, "Spatio-temporal exploration strategies for long-term autonomy of mobile robots," *Robotics and Autonomous Systems*, vol. 88, pp. 116–126, 2017.

[11] L. Kunze, N. Hawes, T. Duckett, M. Hanheide, and T. Krajník, "Artificial intelligence for long-term robot autonomy: A survey," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 4023–4030, 2018.

[12] M. Hanheide, D. Hebesberger, and T. Krajník, "The when, where, and how: An adaptive robotic info-terminal for care home residents," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2017, pp. 341–349.

[13] L. Calderita, A. Vega, P. Bustos, and P. Núñez, "A new human-aware robot navigation framework based on time-dependent social interaction spaces: An application to assistive robots in caregiving centers," *Robotics and Autonomous Systems*, vol. 145, p. 103873, 2021.

[14] 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.

[15] D. Brščić, T. Kanda, T. Ikeda, and T. Miyashita, "Person tracking in large public spaces using 3-d range sensors," *IEEE Transactions on Human-Machine Systems*, vol. 43, no. 6, pp. 522–534, 2013.

[16] S. Molina, G. Cielniak, T. Krajník, and T. Duckett, "Modelling and predicting rhythmic flow patterns in dynamic environments," in *Towards Autonomous Robotic Systems: 19th Annual Conference, TAROS 2018, Bristol, UK July 25-27, 2018, Proceedings 19*. Springer, 2018, pp. 135–146.

[17] Z. Yan, T. Duckett, and N. Bellotto, "Online learning for human classification in 3d lidar-based tracking," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 864–871.

[18] T. Vintr, J. Blaha, M. Rektoris, J. Ulrich, T. Rouček, G. Broughton, Z. Yan, and T. Krajník, "Toward benchmarking of long-term spatio-temporal maps of pedestrian flows for human-aware navigation," *Frontiers in robotics and AI*, vol. 9, 2022.

[19] C. S. Swaminathan, T. P. Kucner, M. Magnusson, L. Palmieri, S. Molina, A. Mannucci, F. Pecora, and A. J. Lilienthal, "Benchmarking the utility of maps of dynamics for human-aware motion planning." *Frontiers in Robotics and AI*, vol. 9, 2022.

[20] L. Castri, S. Mghames, M. Hanheide, and N. Bellotto, "Causal discovery of dynamic models for predicting human spatial interactions," in *Social Robotics: 14th International Conference, ICSR 2022, Florence, Italy, December 13–16, 2022, Proceedings, Part I*. Springer, 2023, pp. 154–164.

[21] M. Stuede and M. Schappler, "Non-parametric modeling of spatio-temporal human activity based on mobile robot observations," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 126–133.

[22] T. Krajník, J. P. Fentanes, C. Dondrup, J. Santos, M. Hanheide, and T. Duckett. (2013) Strands public datasets. Accessed: 2022-11-28. [Online]. Available: https://lcas.lincoln.ac.uk/nextcloud/shared/datasets/index.html

[23] T. Vintr, K. Eyisoy, V. Vintrová, Z. Yan, Y. Ruichek, and T. Krajník, "Spatiotemporal models of human activity for robotic patrolling," in *Modelling and Simulation for Autonomous Systems: 5th International Conference, MESAS 2018, Prague, Czech Republic, October 17–19, 2018, Revised Selected papers*. Springer, 2019, pp. 54–64.

[24] 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.

[25] T. Schreiter, T. R. de Almeida, Y. Zhu, E. G. Maestro, L. Morillo-Mendez, A. Rudenko, T. P. Kucner, O. M. Mozos, M. Magnusson, L. Palmieri *et al.*, "The magni human motion dataset: Accurate, complex, multi-modal, natural, semantically-rich and contextualized," *arXiv preprint arXiv:2208.14925*, 2022.

[26] T. Yang, Y. Li, C. Zhao, D. Yao, G. Chen, L. Sun, T. Krajnik, and Z. Yan, "3D ToF LiDAR in mobile robotics: A review," *CoRR*, vol. abs/2202.11025, 2022. [Online]. Available: http://arxiv.org/abs/2202.11025