Pedestrian-Robot Interactions on Autonomous Crowd Navigation: Dataset and Metrics

Diego Paez-Granados[1†], Yujie He[2], David Gonon[2], Lukas Huber[2], Aude Billard[2]

***Abstract*— Robot navigation in unstructured human environments remains challenging when crowd density limits visibility and scene understanding like in malls, markets and airports. In this abstract, we present a new dataset of pedestrian interactions with a service robot with egocentric sensor data from RGBD, 3D LIDAR, force sensinsg, and robot state. Moreover, we aim to highly the need to consider real-life navigation data in order to define proper benchmarking tools for autonomous navigation control and how to realize them from the robot's perspective. We provide over 110 trials with 266k frames in different crowd types and provide methods and code for comparing controllers within this navigation task.**

***Index Terms*— Mobile Service Robots, Human-Robot Interaction, People detection and tracking, autonomous navigation**

I. Introduction

Autonomous navigation in pedestrian areas is a prevalent topic for all service robot categories; delivery robots on the streets, cleaning robots in malls, and self-driving wheelchairs in airports. All bring high societal and economic value. However, interactions in highly populated areas remain challenging in perception, control and pedestrian interactions, where proxemics, self-localization, and disruptive questions remain open.

Collisions are one of the major safety concern as impacts with most service robots could lead to dangerous accidents even at "slow" speeds of 1.5 m/s (6 km/h) [1], moreover, likelihood of subsequent injuries from falls is even higher. At the same time, freezing the robot as safety measure in the middle of crowds could lead to other dangerous collisions with pedestrians stumbling on the robot, become a danger to itself and bystanders [2], [3].

Although several novel approaches are constantly investigated no proof or guarantees in natural crowds exists. Moreover, benchmarks are lacking for actual pedestrian-robot interactions. In our previous work [4], we focused on combining active compliance with DS-based obstacle avoidance [5] which provides a way to slide around obstacles while in contact and continue moving towards the goal. Effectively making the robot post-collision resilient as long as we follow safety design considerations for robot impacts through velocity limits and surface compliance as outlined in [1].

In this work, we present a dataset – **lasa-crowdbot**[1] in [6]– and the key benchmark points used for evaluating the controller described above on outdoor pedestrian interactions within natural crowds. Unlike existing datasets of indoor and outdoor robot navigation MOT20 [7], JRDB [8], and SCAND[9], we provide a detailed classification by type and density of pedestrian crowds, which we consider relevant for benchmark of autonomous social navigation.

The robot Qolo [10] (Fig.1), a personal mobility vehicle for people with lower-body impairments was equipped with a autonomous navigation control operating in shared-control [11] and full autonomy mode [4], [12]. We performed long-term testing at different locations in the city of Lausanne, Switzerland and evaluated multiple controllers through systematic testing with multiple crowd types and densities.

All source code for processing and analyzing interactions was made open source[2].

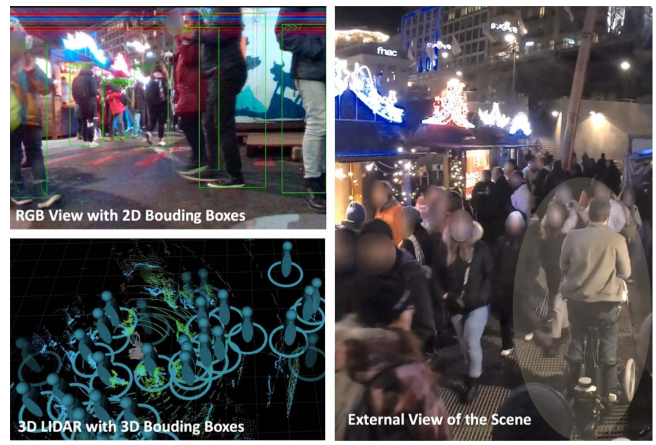

Fig. 1. Robot Qolo navigating on the crowds up to 1ppsq in density, the dataset includes annotated 3D BBx and 2D BBx on Lidar and RGBD data, respectively.

II. Dataset Description:

The dataset includes point clouds from a frontal and rear from two 16 lasers LIDAR (Velodyne VLP-16), a frontal facing RGBD camera (Real Sense D435), and Force/Torque sensor (Botasys Rokubi 2.0). The following are the main recorded sensing modalities:

- 2 x 3D point cloud (50 m, 20Hz)

*This work was funded in part by the EU H2020 project "Crowdbot" (779942).

† is the corresponding author. [1] D.Paez-Granados is with the SCAI Lab, Swiss Federal School of Technology in Zurich - ETH Zurich and SPZ, Switzerland `dfpg@ieee.org`

[2] Y. He, D. Gonon, and A. Billard are with the LASA Laboratory, Swiss Federal School of Technology in Lausanne - EPFL, Switzerland {`yujie.he; david.gonon; lukas.huber; aude.billard`}`@epfl.ch`

[1]Dataset website: https://www.epfl.ch/labs/lasa/crowdbot-dataset/

[2]Pedestrian analysis tools can be found here: https://github.com/epfl-lasa/crowdbot-evaluation-tools

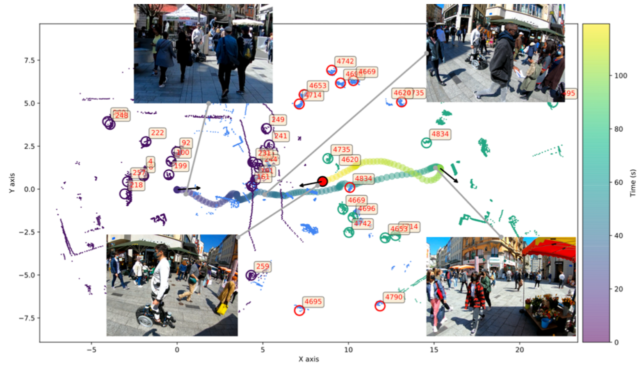

Fig. 2. Example of experimental setup scenario 1, a mixed influx of 6 streets during a farmer's market with up to 0.7 $ppsm$ in crowd density.

- 1x RGBD camera (10 m, 30fps)
- 1 x force sensor ($< 1kN$, 200Hz)
- Robot's state and controller output (100Hz)
- Pressure sensing (100Hz) (in shared-control mode)

As well, we provide the metadata of people detection and tracking from onboard real-time sensing (DrSPAAM detector [13]), and people class labelled from 3D point cloud (AB3DMOT [14]). Furthermore, several metrics were developed to be measured from an egocentric perspective and are given in the dataset: estimated crowd density at multiple radii, density variance, proximity to the robot, and path efficiency metrics (such as time to goal, path length, and virtual collisions).

The whole dataset comprises over 250k frames of data (over 5 hours of navigation in dense crowds), provided on ROS standard data type: rosbag. One recording of the dataset includes approximately 120s of data in a single rosbag format with all Qolo's sensors and state. Moreover, we include offline post-processed people tracking data over the 3D point cloud using our adapted tracker based on [14] exported in npy files for easy read and access.

Summarizing the data content is:

- Pedestrian motion information in the form of 3D point clouds around the robot, including all surrounding people and obstacles in a range of up to 50 m.
- Pedestrian's motion data from a forward-looking RGBD camera, with people, labelled and blurred.
- Output from 3 people detection layers and 1 integrated people tracker.
- Force/Torque information gathered by the contact sensors at the robot's bumper.
- Recordings of the navigation interface input given by the user/driver of the robot.
- Motion data was gathered from the robot inertia sensors and odometry sensors.
- Blurred video recordings of the scene from the robot's perspective.

III. Crowd Navigation Assessment

Previous works have focused on: collisions, success rate, and time to goal [15], [16], as main outcome metrics in pedestrian scenarios. Nonetheless, these metrics are not sufficiently detail to assess the actual pedestrian to robot interaction. Moreover, crowd type, density and homogeneity should be accounted to create a relevant benchmark tool. Therefore, we first categorize all interactions by density of the crowd. Here, we account for mean, max, and standard deviation of density measured at 2.5, 5 and 10 m radii from the robot, as depicted in Fig. 3.

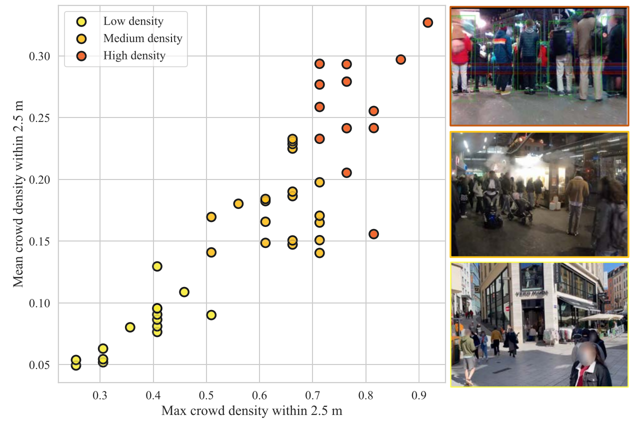

Fig. 3. Resulting density clusters within the current dataset.

We propose a set of metrics which highlight different characteristics of the robot performance in social navigation:

1) Controller performance: In order to observe the controller against high-level planning algorithm these metrics allows to assess different patterns in the crowd compatibility. We compute the controller drive contribution, the agreement, and the fluency, and compare them among different crowd types in reference to non-crowded scenarios.
2) Pedestrian interactions: In a simple interaction metric we use minimal mean distance to pedestrians, virtual collisions with robot boundaries, and real collisions were selected as main egocentric metrics to estimate the level of agreement with the social navigation.
3) Path efficiency: with the goal of comparing the overall system performance we selected the relative time to goal, relative path length and relative jerk w.r.t. a non-crowded baseline recording. So that, all metrics are valid regardless of distance travelled and crowd density.

IV. Results

The current dataset includes two type of scenarios, the first, running the robot over 5.0 km in sets of 20 m round trips, for a total of 95 recordings. Figure 3 shows an example around low and mid crowd densities. The second type of scenario was successfully recorded 15 times with higher densities approximately reaching 1 $ppsm$ from on-board measurements.

Given the type of scenario and crowd density we clustered the data into 3 types of: sparse, flows, and mixed traffic, and divide it with low- (< 0.15 $ppsm$), mid- (< 0.65 $ppsm$), and high- (< 1 $ppsm$) pedestrian densities. We concluded that such granularity level into the scenarios construction and

TABLE I

CROWD NAVIGATION DATASET COMPARISON

Name [Year]	# Sequences	Indoor/Outdoor	# Frame	# 2D BBx	# 3D BBx	# 2D tracks	Crowd Density	Crowd Type
MOT20[7] (2020)	8	Indoor & Outdoor	13k	1.6M	-	3.6k	No	No
JRDB[8] (2020)	54	Indoor & Outdoor	28k	2.4M	1.8M	3.5k	No	No
SCAND[9] (2022)	138	Indoor & Outdoor	-	-	-	-	No	No
CROWDBOT (2022)	110	Outdoor	266k	$\sim 4M^*$	$\sim 4M^*$	TBC*	Yes	Yes

* Under verification with multiple tracking methods.

classification of crowd data is required in order to appropriately compare the results of interacting around unstructured human environments.

Compared with existing datasets, we present crowd types and clusters by crowd density, something non-existing in other robot navigation datasets (Table I). Moreover, we offer a large dataset with 110 trials exceeding 266k frames in 2D and 3D data with labelled bounding boxes. Further details of the compatibility with the specific crowd navigation scenarios and tested controller are described in detail in our work in [17]. Subsequent work on our dataset will quantify the uncertainty on bounding boxes both in 2D and 3D data, as well as, validate the interaction types with pedestrians through the 2D tracks that were measured.

ACKNOWLEDGEMENT

The experiments were approved by the human research ethical committee of EPFL (Approval No: HREC-032-2019), and with the approval and cooperation of the office for mobility, the police and the office for parks, and the public domain of the city of Lausanne. Approval numbers: 395128 and 416008. *Disclaimer:* D.P. and K.S. hold the patents of the robot Qolo and shares in the company Qolo Inc.

REFERENCES

[1] D. Paez-Granados and A. Billard, "Crash test-based assessment of injury risks for adults and children when colliding with personal mobility devices and service robots," *Nature Scientific Reports*, vol. 12, pp. 1–13, 3 2022. [Online]. Available: https://doi.org/10.1038/s41598-022-09349-9

[2] P. Salvini, D. Paez-Granados, and A. Billard, "On the Safety of Mobile Robots Serving in Public Spaces: Identifying gaps in en ISO 13482:2014 and calling for a new standard," *ACM Transactions on Human-Robot Interaction*, vol. 10, no. 3, pp. 1–29, 2021. [Online]. Available: https://doi.org/10.1145/3442678

[3] ——, "Safety Concerns Emerging from Robots Navigating in Crowded Pedestrian Areas," *International Journal of Social Robotics (SORO)*, vol. 14, no. 2, pp. 441–462, 2022. [Online]. Available: https://doi.org/10.1007/s12369-021-00796-4

[4] D. Paez-Granados, V. Gupta, and A. Billard, "Unfreezing social navigation : Dynamical systems based compliance for contact control in robot navigation," in *IEEE International Conference on Robotics and Automation (ICRA-2022)*, vol. 1, 5 2022, pp. 8368–8374. [Online]. Available: https://doi.org/10.1109/ICRA46639.2022.9811772

[5] L. Huber, A. Billard, and J.-J. Slotine, "Avoidance of Convex and Concave Obstacles With Convergence Ensured Through Contraction," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1462–1469, 2019. [Online]. Available: https://doi.org/10.1109/lra.2019.2893676

[6] D. Paez-Granados, Y. He, D. Gonon, L. Huber, and A. Billard, "3d point cloud and rgbd of pedestrians in robot crowd navigation: Detection and tracking," *IEEE DataPort*, 12 2021. [Online]. Available: https://doi.org/10.21227/ak77-d722

[7] P. Dendorfer, A. Osep, A. Milan, K. Schindler, D. Cremers, I. Reid, S. Roth, and L. Leal-Taixé, "Motchallenge: A benchmark for single-camera multiple target tracking," *International Journal of Computer Vision*, vol. 129, pp. 845–881, 4 2021.

[8] R. Martin-Martin, M. Patel, H. Rezatofighi, A. Shenoi, J. Gwak, E. Frankel, A. Sadeghian, and S. Savarese, "Jrdb: A dataset and benchmark of egocentric robot visual perception of humans in built environments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021. [Online]. Available: 10.1109/TPAMI.2021.3070543

[9] H. Karnan, A. Nair, X. Xiao, G. Warnell, S. Pirk, A. Toshev, J. Hart, J. Biswas, and P. Stone, "Socially compliant navigation dataset (scand): A large-scale dataset of demonstrations for social navigation," 3 2022. [Online]. Available: http://arxiv.org/abs/2203.15041

[10] D. Paez-Granados, H. Kadone, M. Hassan, Y. Chen, and K. Suzuki, "Personal mobility with synchronous trunk-knee passive exoskeleton: Optimizing human-robot energy transfer," *IEEE/ASME Transactions on Mechatronics*, vol. 1, pp. 1–12, 2022. [Online]. Available: https://doi.org/10.1109/TMECH.2021.3135453

[11] D. J. Gonon, D. Paez-Granados, and A. Billard, "Reactive Navigation in Crowds for Non-holonomic Robots with Convex Bounding Shape," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 4728–4735, 2021. [Online]. Available: https://doi.org/10.1109/LRA.2021.3068660

[12] L. Huber, J.-J. Slotine, and A. Billard, "Avoiding dense and dynamic obstacles in enclosed spaces: Application to moving in crowds," *IEEE Transactions on Robotics*, pp. 1–10, 2022.

[13] D. Jia and B. Leibe, "Person-MinkUNet: 3D Person Detection with LiDAR Point Cloud," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2021.

[14] X. Weng, J. Wang, D. Held, and K. Kitani, "3D Multi-Object Tracking: A Baseline and New Evaluation Metrics," *IROS*, 2020.

[15] C. Chen, Y. Liu, S. Kreiss, and A. Alahi, "Crowd-robot interaction: Crowd-aware robot navigation with attention-based deep reinforcement learning," in *IEEE International Conference on Robotics and Automation*, 2019.

[16] Y. Kobayashi, T. Sugimoto, K. Tanaka, Y. Shimomura, F. J. Arjonilla Garcia, C. H. Kim, H. Yabushita, and T. Toda, "Robot Navigation Based on Predicting of Human Interaction and its Reproducible Evaluation in a Densely Crowded Environment," *International Journal of Social Robotics*, 2021.

[17] D. Paez-Granados, Y. He, D. Gonon, D. Jia, B. Leibe, K. Suzuki, and A. Billard, "Pedestrian-robot interaction on crowd navigation: Reactive control methods and evaluation," in *IEEE Conference International on Intelligent Robots and Systems (IROS)*, 10 2022, pp. 1–8. [Online]. Available: https://www.research-collection.ethz.ch/handle/20.500.11850/561754