Embedded 2D LiDAR-Based Person Tracking for Safe Navigation in Assistive Autonomous Robots

Conference Paper

Author(s):
Plozza, Davide; Marty, Steven; Scherrer, Cyril; Schwartz , Simon; Zihlmann , Stefan; Magno, Michele

Publication date:
2025

Permanent link:
https://doi.org/https://doi.org/10.3929/ethz-c-000791089

Originally published in:
https://doi.org/10.1109/IWASI66786.2025.11122002

This article has been accepted for publication in the proceedings of the
2025 10th IEEE International Workshop on Advances in Sensors and Interfaces (IWASI)
DOI: 10.1109/IWASI66786.2025.11122002

Embedded 2D LiDAR-Based Person Tracking for Safe Navigation in Assistive Autonomous Robots

Davide Plozza*, Steven Marty*, Cyril Scherrer*, Simon Schwartz*, Stefan Zihlmann*, Michele Magno*
*Dept. of Information Technology and Electrical Engineering, ETH Zürich, Switzerland
{dplozza,martyste,cscherrer,sschwartz,szihlmann,magnom}@ethz.ch

***Abstract*—This paper presents a fully embedded real-time person tracking pipeline for assistive quadrupedal robots supporting safe navigation for visually impaired users. Our approach combines a deep learning-based 2D LiDAR person detector with a lightweight multi-object tracker and integrates it into a Guide Dog Robot (GDR) navigation framework. A novel detection post-processing scheme is proposed, reducing detector latency by 52.38% compared to state-of-the-art voting-based methods while preserving accuracy. The improved latency enables the entire pipeline to operate reliably at 20 Hz on a resource-constrained mobile robotic embedded platform based on the NVIDIA Jetson Xavier NX. The experimental setup shows that our system tracks dynamic obstacles and continuously localizes the user holding the robot's handle, enabling a dynamic safety footprint for proactive collision avoidance. Under the tested setup, the optimal configuration achieves a MOTA of 83.27% and a user tracking RMSE below 0.2 m on two custom datasets recorded with motion-capture ground truth. Real-world navigation experiments in indoor environments demonstrate effective collision prevention and smooth corrective maneuvers when the user drifts from the default following position. The modular design of the detection, tracking, and planning components ensures flexibility and ease of integration into other robotic platforms. This work contributes a scalable and efficient tracking and navigation solution for human-aware mobile robots operating in dynamic environments, supporting safer human-robot interaction in assistive contexts.**

***Index Terms*—2D LiDAR, people detection, people tracking, autonomous mobile robots, assistive robotics**

I. Introduction

In recent years, there has been an increasing amount of research and commercial solutions aimed at providing navigation-assistive technologies for visually impaired people, with the goal of providing an intuitive, comprehensive, and reliable substitute for vision [1], [2]. Concurrently, Autonomous Mobile Robots (AMR) have witnessed rapid advances in sensing and navigation capabilities, enabling new applications in human-centered environments [3].

These trends highlight an opportunity for robotic guide systems that integrate the autonomy and environmental perception capabilities of AMR to provide intuitive and robust navigation assistance for visually impaired individuals. Unlike smartphone-based apps or wearable devices, small-scale mobile platforms can provide more complete obstacle avoidance and improved physical human-robot interaction [2], [4].

One of the crucial and challenging tasks in this use-case is the capability of the robot to track the assisted individual, ensuring a comfortable and safe navigation experience by keeping the user in its designated space via feedback [4]–[6]. Recent literature has proposed various configurations using different sensors to detect and track humans in autonomous robots, with the most common being RGB(-D) cameras [7], [8], 2D LiDARs [9]–[11], and 3D LiDARs [12]. Compared to RGB(-D) and 3D LiDAR, 2D LiDAR sensors offer low power consumption and wide angular coverage, providing precise distance measurements and lightweight processing requirements that are ideal for embedded platforms and overall cost-effectiveness [13], [14].

In our prior work [15], we proposed a real-time 2D LiDAR-based person detection and tracking pipeline integrated into a navigation framework for low-latency obstacle avoidance. In this paper, we build upon that foundation by extending the system to a multi-LiDAR setup with 360° coverage and integrating it into a Guide Dog Robot (GDR) system for user tracking, enabling safe and dynamic navigation assistance.

Our contributions are as follows:

- Extending an embedded 2D LiDAR-based person tracking pipeline with a novel post-processing method, enabling real-time operation at 20 Hz on the Jetson Xavier NX.
- Proposing a GDR by extending an autonomous navigation planner to incorporate dynamic user position and ensure collision avoidance during guidance tasks.
- Systematic benchmarking on custom datasets with motion-capture ground truth and real-life hallway and doorway experiments, demonstrating accurate user localization and effective collision prevention.

II. Related Work

Person detection from 2D LiDAR has evolved from hand-crafted features [16] to deep learning approaches such as DROW [10] and DR-SPAAM [9], the latter offering real-time capability with spatial attention. Despite their success, integration into multi-object tracking pipelines remains limited. End-to-end tracking frameworks [17], [18] are not yet common in 2D LiDAR applications, where trackers like SORT [19] and Norfair remain more practical.

Autonomous navigation is commonly divided into a global planner which determines an overall path using algorithms such as A*, and a local planner that adjusts the trajectory in response to sensor data. The ROS Navigation Stack [20] offers a widely adopted implementation of this structure. For local planning, the Timed-Elastic-Band (TEB) algorithm [21]

979-8-3315-6578-7/25/$31.00 © 2025 IEEE

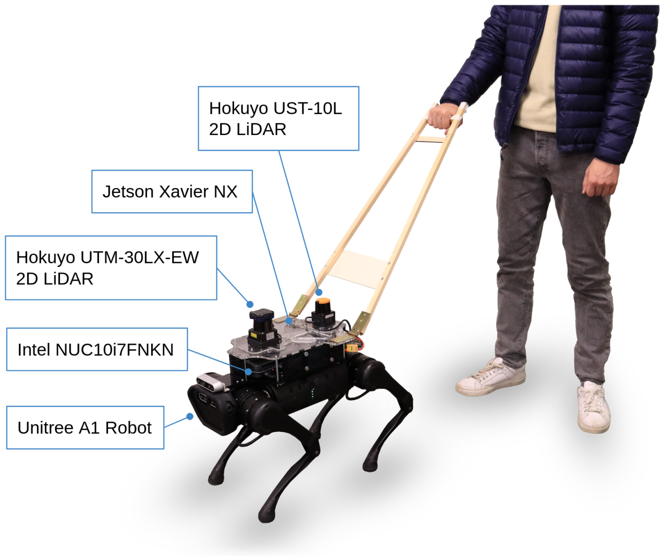

Fig. 1: Guide Dog Robot, composed of an Unitree A1 with the additional backpack system, with front and back-facing LiDARs.

is frequently used due to its fast optimization and ability to incorporate dynamic obstacle trajectories [22], [23].

In prior work [15], we developed a real-time embedded 2D LiDAR-based detection and tracking system running in real-time at $20\,\mathrm{Hz}$ on a Jetson Xavier NX, integrating the DR-SPAAM detector with the lightweight Norfair multi-object tracker [24]. Our approach achieved over 85% MOTA score in real-world datasets and successfully enabled close-range human avoidance using trajectory prediction.

The need for autonomous navigation support for blind and visually impaired people has motivated the development of intelligent assistive systems. Reviews by Manjari et al. [1] and Khan et al. [2] highlight a wide range of technologies, from smart canes to robotic platforms. Yet, many systems still lack portability, real-time performance, or validation in dynamic environments. Kim et al. [25] directly investigate blind user expectations for robotic guide dogs using quadrupeds, revealing preferences for rigid handles and smoother and intuitive motion planning. Hong et al. [4] highlight the importance of knowing the position of the user in GDR applications to improve safety and human-robot interaction. Thus, providing the robot with reliable person detection and tracking is essential.

Guerreiro et al. [26] use a simplified planning approach using the ROS Navigation Stack, assuming a fixed user position relative to the robot, which must be conservative to avoid collisions if the user drifts. Other methods actively track the user's position to capture user-robot dynamics for planning [5], [6], [25], [27], typically using rear-facing RGB-D cameras [5], [6] or a 2D LiDAR on a pan/tilt turret [27]. We leverage our proposed tracker with a 360° 2D LiDAR setup for both user tracking and dynamic obstacle avoidance, achieving efficient tracking without additional moving hardware. Instead of modeling complex human-robot interactions, we extend the planning strategy of Guerreiro et al. [26], enhancing safety by incorporating user tracking to anticipate and prevent collisions.

The proposed approach advances the state of the art and prior works by augmenting established tracking pipelines with real-time, on-device user localization coupled to a guide-robot navigation stack, thereby closing the perception-action loop and delivering safe and energy-efficient mobility assistance for visually impaired users in dynamic environments.

III. Methods

Our autonomous guidance pipeline follows a modular approach, and consists of the following components: people detection, tracking, and autonomous navigation. The pipeline is augmented to specifically account for user tracking and ensuring collision avoidance of both robot and user.

A. System

As seen in Fig. 1, the GDR consists of a Unitree A1 quadruped equipped with a custom-built backpack housing two LiDAR sensors, an Intel NUC10i7FNKN with a six-core 10th-generation Intel i7 CPU, and an NVIDIA Jetson Xavier NX for onboard real-time processing. A handle attached to the back allows a blind user to be guided by the robot.

The sensing system includes a Hokuyo UTM-30LX-EW and a Hokuyo UST-10LX 2D LiDAR, both mounted at approximately $45\,\mathrm{cm}$ from the ground, providing a full $360°$ field of view. Both LiDARs operate at $20\,\mathrm{Hz}$ with a $0.25°$ angular resolution. The front-facing UTM-30LX-EW features a $270°$ scan angle, while the rear-facing UST-10LX extends coverage to complete the 360° field of view. Their scans are merged into a single full-circle scan, yielding 1440 points.

The local planner computes velocity commands, which are executed through Unitree's built-in walking controller, which also provides an odometry estimation at $50\,\mathrm{Hz}$. The entire tracking and navigation framework is implemented in ROS [28].

B. Detection

The proposed system employs the novel 2D-LiDAR human detection model DR-SPAAM [9], also employed in [15], using an adapted version of its open-source implementation [29]. This paper proposes a faster post-processing scheme enabling faster real-time operation on a full 360° scan. The DR-SPAAM detector processes 2D LiDAR scans by first generating *cutout windows* around each LiDAR point. Each cutout captures neighboring points within a fixed region, normalized and re-sampled to 56 points. These cutouts are processed by a 1D convolutional network, which performs both classification to identify if a cutout contains a person and regression to estimate the relative offset to the person's center. To enhance detection accuracy and efficiency, DR-SPAAM introduces a spatial attention mechanism and an auto-regressive model, allowing the network to aggregate temporal information across scans without the need for explicit alignment for ego-motion.

Two post-processing schemes are available: a more accurate voting scheme (*vote*) used in [9] which casts individual cutout predictions into a spatial grid followed by non-maximum suppression (NMS), and a faster method (*nms*) available in the open-source implementation [29] which applies NMS directly on raw detections, keeping only the highest-confidence ones within a 0.5 m radius. While the second method is faster, it generates a high number of false positives that the voting scheme can effectively reduce. To address this, we propose an alternative post-processing step (*nms**) that clusters detections using NMS (with a 0.5 m radius), but then averages both the positions and confidence values of all detections within each cluster to produce a centroid, yielding similar behavior to the voting scheme without the additional computational overhead. The post-processing is not used during training, allowing flexibility for selecting and adjusting it at deployment.

The detector design enables successful deployment on various robots and sensors without requiring retraining for a specific sensor, as demonstrated by the CROWDBOT project [30], which confirms its independence from angular resolution and distance. Therefore, we decided against creating a new training dataset for our LiDAR setup and instead trained the model using the JRDB dataset [31], which features a 360° field of view with a 0.32° angular resolution, closely resembling our 0.25° and making it a better fit than the DROW dataset [32]. The ground-truth labels are extracted from the dataset as in [33].

To further improve processing efficiency, we examine the impact of window stride on inference speed, which differs from the frame stride studied by the DR-SPAAM authors. Window stride involves generating cutout windows only at every nth LiDAR point.

The model is deployed on a Jetson Xavier NX platform, which is less powerful than the Jetson Xavier AGX used in [9].

C. Tracking

The proposed system leverages the tracker designed in our previous work [15], based on Norfair [24], a lightweight, real-time multi-object tracking library utilizing a Kalman filter with a constant velocity model and data association via the Hungarian algorithm. Track initiation and deletion are managed using count of matched frames: a new track is confirmed once its counter exceeds C_{init}, and it is removed after C_{del} consecutive unmatched updates. These parameters are tuned to balance fast detection with minimal false positives. To ensure consistency, all detections are transformed into the robot's odometry frame before tracking.

D. Guide Dog Robot Navigation and User Tracking

In the GDR setup, a visually impaired user is guided by holding a handle positioned behind the robot. The user's position relative to the robot is tracked to ensure safe and comfortable navigation. The estimated user track is determined by selecting the closest detected person track to the default user position of $0.6\,\mathrm{m}$ directly behind the robot center. If no detections are found within a specified $1.0\,\mathrm{m}$ radius around this default position, the user's position is set to the default.

For navigation, as in [15], we utilize the ROS Navigation Stack [20], configuring it with AMCL for localization, A* as the global planner, and TEB as the local planner.

TABLE I: Configurations

Config	Window Stride	Conf. Thresh.	C_{init}	C_{del}
C1	1	0.85	10	15
C2	10	0.85	10	15
C3	10	0.8	5	15

TABLE II: Benchmark Results For Each Dataset.

DS	Tracker config.	Valid	ID sw.	Miss	FP	MOTA	MOTP [m]
GR1	*C1 vote*	6264	1	28	**497**	**91.64%**	0.15
	C1 nms	6271	**0**	21	4656	25.68%	0.15
	*C1 nms**	6264	1	28	740	87.78%	0.15
	C2 vote	6262	1	30	958	84.28%	0.15
	C2 nms	6268	1	24	4703	24.87%	0.15
	*C2 nms**	6262	1	30	875	85.60%	0.15
	C3 vote	6279	1	13	2472	60.50%	0.15
	C3 nms	**6283**	1	**9**	5631	10.36%	0.15
	*C3 nms**	6279	1	13	1848	70.41%	0.15
GR2	*C1 vote*	18441	33	1665	453	89.32%	0.19
	C1 nms	18998	24	1117	13788	25.87%	0.19
	*C1 nms**	18474	33	1632	468	89.41%	0.19
	C2 vote	18582	28	1529	313	90.71%	0.19
	C2 nms	19037	**19**	1083	9534	47.19%	0.19
	*C2 nms**	18546	28	1565	**166**	**91.27%**	0.19
	C3 vote	18853	29	1257	2084	83.27%	0.19
	C3 nms	**19149**	23	**967**	11688	37.05%	0.19
	*C3 nms**	18846	30	1263	1271	87.27%	0.19

Additionally, some adjustments are made to accommodate the human-robot configuration. As in [26], the human-robot pair is treated as a single entity with a shifted center towards the user and an enlarged footprint encompassing both the robot and user. The planner thus provides collision avoidance assuming the user is in the default position directly behind the robot, but if the user deviates too far, a collision could be possible. Thus a dynamic footprint is maintained, which includes a $0.5\,\mathrm{m}$ padding around the estimated user position, allowing real-time collision detection for immediate future trajectories.

In the event of a potential collision, the robot stops and initiates a corrective maneuver by gently rotating, providing physical feedback through the handle to nudge the user back to the default position. Once the planner no longer detects a future collision, normal navigation resumes.

IV. Experimental Results

We assess tracking accuracy and GDR system performance using custom real-world datasets and integrated experiments. Our tracker is evaluated in three configurations, as detailed in Table I: *C1* prioritizes overall tracking accuracy, *C2* is optimized for real-time performance, and *C3* minimizes track initiation time. Additionally, each configuration can be coupled with each of the three post-processing methods *vote*, *nms*, and *nms**.

A. Datasets

To assess tracking accuracy in a GDR setting, we record two datasets. Ground truth position and orientation of both the robot and participants are recorded at $100\,\mathrm{Hz}$ using a motion capture system equipped with six Vicon Vero 2.2 cameras. The robot setup, as detailed in Section III-A, provides the sensor and odometry data for evaluation. Experiments are conducted within a $4\,\mathrm{m}\mathrm{x}4\,\mathrm{m}$ space, where movement occurs in an environment featuring moderate clutter from chairs and tables.

The datasets are as follows:

1) *Guide Robot 1 (GR1)*: A single person follows the robot while holding a handle. The robot moves semi-randomly, with speeds up to $1.0\,\mathrm{m\,s^{-1}}$ (forward/backward) and $\pm 1.5\,\mathrm{rad\,s^{-1}}$ (rotation). Duration: 5 minutes and 18 seconds.
2) *Guide Robot 2 (GR2)*: A person follows the robot using the handle, acting as a user, while two additional people move randomly around it. The robot moves semi-randomly with speeds up to $1.0\,\mathrm{m\,s^{-1}}$ (forward/backward) and $\pm 1.5\,\mathrm{rad\,s^{-1}}$ (rotation). Duration: 5 minutes and 39 seconds.

B. MOT Benchmark

Tracking performance is assessed using CLEAR MOT metrics [34] as in [15].

The MOTA score (Equation 1) quantifies tracking accuracy by penalizing identity switches (ID_k), missed detections (Miss_k), and false positives (FP_k), normalized by total ground truths (g_k) per scan k. Errors are reported individually, with tracks matched using a $0.75\,\mathrm{m}$ threshold.

$$MOTA = 1 - \frac{\sum_k (ID_k + Miss_k + FP_k)}{\sum_k g_k} \tag{1}$$

MOTP (Equation 2) evaluates precision as the average Euclidean distance (d_k^i) between matched estimates and ground truths across all correct matches (c_k), where lower values indicate better precision.

$$MOTP = \frac{\sum_{i,k} d_k^i}{\sum_k c_k} \tag{2}$$

Ground truth annotations are extracted from motion capture data, considering only detections within the LiDAR field of view for fair evaluation.

Table II shows the benchmark results, with CLEAR MOT metrics and the raw counts of each error type. Configurations *C1* and *C2* yield similar scores across all datasets, indicating that the stride parameter can be increased without sacrificing accuracy. Only in *GR1* does a slight reduction in accuracy occur due to more false positives. Testing stride values above 10 revealed rapid performance degradation, identifying 10 as the optimal trade-off. *C3* as expected trades fewer misses for a higher count of false positives across both datasets. All configurations show similar MOTP, which stays below $0.2\,\mathrm{m}$ even in the most challenging configuration.

Naive non-maximum suppression implemented in *nms* leads to excessive false positives. Although raising the confidence threshold nearly to 1 reduces false positives, it significantly lowers true positives. In contrast, our proposed *nms** post-processing closely matches the performance of the original voting scheme. It is even more effective in filtering more false positives, resulting in higher MOTA scores, especially in configuration *C3*.

TABLE III: User Tracking Benchmark Results.

Dataset	Tracker	ATE [m]	RMSE [m]
GD1	*C2 voting*	0.154	0.181
	*C2 nms**	0.154	0.182
	C3 voting	0.153	0.182
	*C3 nms**	**0.152**	**0.179**
GD2	*C2 voting*	0.146	0.181
	*C2 nms**	0.147	0.188
	C3 voting	**0.145**	0.179
	*C3 nms**	**0.145**	**0.178**

C. User Tracking Benchmark

In addition to evaluating the general tracking performance, we specifically benchmark the system's ability to accurately track the user positioned behind the robot, essential for the GDR application. This benchmark leverages the positional differences between the ground truth and the estimated user positions, obtained as described in Section III-D.

We adopt the metrics commonly used in single-object tracking: Average Tracking Error (ATE) and Root Mean Square Error (RMSE).

$$ATE = \frac{1}{N}\sum_{j=1}^{N} d_k \tag{3}$$

The ATE, shown in Equation (3), is similar to the Multi-Object Tracking Precision (MOTP), in which it measures the mean Euclidean distance (d_j) between the estimated user positions and the ground truth positions across all scans (j) containing ground truth. The RMSE is computed similarly between the ground truth and estimated position.

The results in Table III demonstrate the system's ability to accurately track the user behind the robot in both the *GR1* and *GR2* datasets, validating its effectiveness for human guidance applications even under dynamic conditions. All configurations achieve low ATE and RMSE values, with configuration *C3* using *nms** consistently yielding the best results, achieving an overall RMSE of $17.85\,\mathrm{cm}$.

D. Computation Time

The computation times for each configuration of our tracker were benchmarked on a Jetson Xavier NX. We measured the detector inference time (T^i_{det}), tracker update time (T^i_{track}), and total latency time (T^i_{lat}). The results for different configurations are summarized in Table IV.

The configuration *C1* is not capable of real-time execution at $20\,\mathrm{Hz}$ for either LiDAR setup. Configurations *C2* and *C3* with voting post-processing show worst-case times exceeding $50\,\mathrm{ms}$, whereas with *nms**, all times are within the real-time limit, indicating that only the proposed *nms** can achieve reliable real-time performance in this setting.

Interestingly, *nms** exhibits slower performance than *vote* when paired with *C1*, likely due to increased computation times

TABLE IV: Computation Times.

Tracker	Detector T^i_{det} worst/avg [ms]	Tracker T^i_{track} worst/avg [ms]	Total time T^i_{lat} worst/avg [ms]
C1 vote	258.63 / 6.158	33.13 / 12.24	272.36 / 238.36
*C1 nms**	288.75 / 275.90	35.21 / 14.64	311.52 / 290.54
C2 vote	53.38 / 37.66	21.81 / 5.88	64.18 / 43.54
*C2 nms**	**38.48 / 26.84**	22.83 / **5.60**	**51.98 / 32.47**
C3 vote	59.69 / 38.01	23.08 / 6.42	68.83/44.43
*C3 nms**	41.79 / 27.18	**21.52** / 6.16	52.35 / 33.34

TABLE V: Post-processing Computation Times for *C3*.

FOV	*vote* worst/avg [ms]	*nms* worst/avg [ms]	*nms** worst/avg [ms]
270°	14.88/14.13	**5.90/5.01**	6.48/6.11
360°	23.23/20.41	**10.98/9.74**	11.65/10.61

scaling with the number of input detections, whereas *vote* is less sensitive to this.

Additionally, the post-processing computation times for configuration *C3* are shown in Table V. The proposed *nms** is only slightly slower than *nms*, but is on average 52.38% faster than *vote*.

E. User collision avoidance experiments

To validate the effectiveness of the dynamic footprint and feedback mechanism in preventing user collisions in our proposed GDR, two experiments were conducted in a hallway and door navigation scenario, where the robot and user must pass through a $1\,\mathrm{m}$ wide doorway, as shown in Fig. 2. The *C3 nms** configuration was used.

In the first experiment, with dynamic footprint collision detection disabled, the navigation system failed to prevent a collision as the user drifted too close to the door frame. In contrast, with the dynamic footprint and feedback mechanism enabled, and running the tracker in the *C3 nms** configuration, the robot successfully avoided collision by detecting it early, stopping, and performing corrective maneuvers.

V. Discussion

Our experiments show that all tested configurations of the tracking pipeline achieve robust and accurate person tracking in cluttered, close-range indoor environments, as required in assistive scenarios. Notably, this performance is achieved without retraining, despite the detector being trained on JRDB with a different LiDAR setup, highlighting the generalization capability of DR-SPAAM and its resolution and distance independence [9].

Our proposed post-processing method *nms** achieves similar or better results compared to the original voting scheme in [9], while reducing the average computation time by 52.38%. The results also show that decreasing the *window stride* to 10 has minimal impact on accuracy. These changes enable reliable real-time operation at $20\,\mathrm{Hz}$ on the Jetson Xavier NX with a 360° dual-LiDAR setup, demonstrating the method's scalability.

Configuration *C3* achieves a low track initiation time and fewer overall misses while maintaining a manageable number

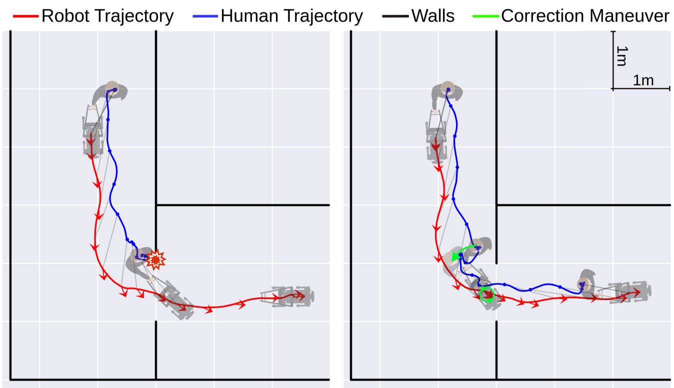

Fig. 2: Collision avoidance in a door navigation task with Vicon-tracked robot and user trajectories. *Left*: Static footprint results in a collision with the wall (red star). *Right*: Dynamic footprint with feedback prevents collision as the robot stops, provides corrective feedback (green arrows), and then continues navigation.

of false positives and achieving an overall MOTA score of 83.27% when paired with *nms**. This makes it our preferred choice for people avoidance applications. Additionally, the *C3 nms** configuration is validated by the user tracking benchmark, achieving an overall RMSE of $17.85\,\mathrm{cm}$. This demonstrates its reliability in maintaining accurate user localization even in dynamic environments.

The collision avoidance experiments show the effectiveness of the proposed GDR system in utilizing the accurate user tracker to detect and prevent collisions with the environment, ensuring user safety.

Some limitations remain. The linear motion model may underperform in response to sudden or unpredictable user behavior, suggesting the need for more expressive prediction models. Furthermore, while our GDR system ensures safe navigation without explicitly planning for complex human-robot dynamics, the navigation efficiency may be hindered by frequent stops and corrective maneuvers.

VI. Conclusion

This paper presents an autonomous GDR equipped with a fully embedded 2D LiDAR-based tracking pipeline designed to enhance obstacle avoidance and user safety. Integrated into an autonomous navigation framework and deployed on a quadrupedal robot, the tracker achieves consistent real-time performance at $20\,\mathrm{Hz}$ on an embedded GPU. The optimized *C3 nms** configuration is validated on custom datasets, achieving a MOTA score of 83.27% and an RMSE below $0.2\,\mathrm{m}$ for user tracking. Additionally, a novel post-processing method for the detector is proposed, reducing computational overhead by 52.38% compared to the original method while maintaining accuracy.

Real-world experiments demonstrate that the proposed GDR system reliably prevents collisions and ensures safe user guidance, combining accurate tracking with efficient computation. These results validate the effectiveness of our real-time tracker for assistive navigation, highlighting its scalability and modularity for integration into autonomous robotic systems.

Future work will include benchmarking against alternative person-tracking methods such as [35], and conducting broader real-world evaluations to better assess safety and generalization.

Acknowledgment

This work was supported by the ETH Future Computing Laboratory (EFCL).

References

[1] K. Manjari, M. Verma, and G. Singal, "A survey on assistive technology for visually impaired," *Internet of Things*, vol. 11, p. 100188, 2020.

[2] S. Khan, S. Nazir, and H. U. Khan, "Analysis of navigation assistants for blind and visually impaired people: A systematic review," *IEEE access*, vol. 9, pp. 26 712–26 734, 2021.

[3] M. Rubagotti, I. Tusseyeva, S. Baltabayeva, D. Summers, and A. Sandygulova, "Perceived safety in physical human–robot interaction—a survey," *Robotics and Autonomous Systems*, vol. 151, p. 104047, 2022.

[4] B. Hong, Z. Lin, X. Chen, J. Hou, S. Lv, and Z. Gao, "Development and application of key technologies for guide dog robot: A systematic literature review," *Robotics and Autonomous Systems*, vol. 154, p. 104104, 2022.

[5] Y. Chen, Z. Xu, Z. Jian, G. Tang, L. Yang, A. Xiao, X. Wang, and B. Liang, "Quadruped guidance robot for the visually impaired: A comfort-based approach," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 12 078–12 084.

[6] A. Xiao, W. Tong, L. Yang, J. Zeng, Z. Li, and K. Sreenath, "Robotic guide dog: Leading a human with leash-guided hybrid physical interaction," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 11 470–11 476.

[7] N. S. Punn, S. K. Sonbhadra, S. Agarwal, and G. Rai, "Monitoring covid-19 social distancing with person detection and tracking via fine-tuned yolo v3 and deepsort techniques," 2021.

[8] Z. Xu, X. Zhan, Y. Xiu, C. Suzuki, and K. Shimada, "Onboard dynamic-object detection and tracking for autonomous robot navigation with rgb-d camera," *arXiv preprint arXiv:2303.00132*, 2023.

[9] D. Jia, A. Hermans, and B. Leibe, "Dr-spaam: A spatial-attention and auto-regressive model for person detection in 2d range data," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 10 270–10 277.

[10] L. Beyer, A. Hermans, T. Linder, K. O. Arras, and B. Leibe, "Deep person detection in 2d range data," *arXiv preprint arXiv:1804.02463*, 2018.

[11] E.-J. Jung, J. H. Lee, B.-J. Yi, J. Park, S.-T. Noh *et al.*, "Development of a laser-range-finder-based human tracking and control algorithm for a marathoner service robot," *IEEE/ASME transactions on mechatronics*, vol. 19, no. 6, pp. 1963–1976, 2013.

[12] J. Gómez, O. Aycard, and J. Baber, "Efficient detection and tracking of human using 3d lidar sensor," *Sensors*, vol. 23, no. 10, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/10/4720

[13] M. Hasan, J. Hanawa, R. Goto, R. Suzuki, H. Fukuda, Y. Kuno, and Y. Kobayashi, "Lidar-based detection, tracking, and property estimation: A contemporary review," *Neurocomputing*, vol. 506, pp. 393–405, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231222009365

[14] D. Jia, A. Hermans, and B. Leibe, "2d vs. 3d lidar-based person detection on mobile robots," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 3604–3611.

[15] D. Plozza, S. Marty, C. Scherrer, S. Schwartz, S. Zihlmann, and M. Magno, "Autonomous navigation in dynamic human environments with an embedded 2d lidar-based person tracker," in *2024 IEEE Sensors Applications Symposium (SAS)*, 2024, pp. 1–6.

[16] K. O. Arras, O. M. Mozos, and W. Burgard, "Using boosted features for the detection of people in 2d range data," in *Proceedings 2007 IEEE international conference on robotics and automation*. IEEE, 2007, pp. 3402–3407.

[17] Z. Wang, L. Zheng, Y. Liu, Y. Li, and S. Wang, "Towards real-time multi-object tracking," in *European conference on computer vision*. Springer, 2020, pp. 107–122.

[18] F. Zeng, B. Dong, Y. Zhang, T. Wang, X. Zhang, and Y. Wei, "Motr: End-to-end multiple-object tracking with transformer," in *European Conference on Computer Vision*. Springer, 2022, pp. 659–675.

[19] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *2016 IEEE international conference on image processing (ICIP)*. IEEE, 2016, pp. 3464–3468.

[20] E. Marder-Eppstein, "navigation," 2022, [Online; accessed 07-November-2023]. [Online]. Available: http://wiki.ros.org/navigation

[21] C. Rösmann, F. Hoffmann, and T. Bertram, "Integrated online trajectory planning and optimization in distinctive topologies," *Robotics and Autonomous Systems*, vol. 88, pp. 142–153, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0921889016300495

[22] M. Pittner, M. Hiller, F. Particke, L. Patino-Studencki, and J. Thielecke, "Systematic analysis of global and local planners for optimal trajectory planning," in *ISR 2018; 50th International Symposium on Robotics*. VDE, 2018, pp. 1–4.

[23] C. Z. Looi and D. W. K. Ng, "A study on the effect of parameters for ros motion planer and navigation system for indoor robot," *International Journal of Electrical and Computer Engineering Research*, vol. 1, no. 1, pp. 29–36, 2021.

[24] J. Alori, A. Descoins, javier, F. Lezama, KotaYuhara, D. Fernández, A. Castro, fatih, David, F. Kurucz, B. Ríos, shafu.eth, K. Nar, D. Huh, and Moises, "tryolabs/norfair: v2.2.0," Jan. 2023. [Online]. Available: https://doi.org/10.5281/zenodo.7504727

[25] J. T. Kim, W. Yu, Y. Kothari, J. Tan, G. Turk, and S. Ha, "Transforming a quadruped into a guide robot for the visually impaired: Formalizing wayfinding, interaction modeling, and safety mechanism," *arXiv preprint arXiv:2306.14055*, 2023.

[26] J. Guerreiro, D. Sato, S. Asakawa, H. Dong, K. M. Kitani, and C. Asakawa, "Cabot: Designing and evaluating an autonomous navigation robot for blind people," in *Proceedings of the 21st international ACM SIGACCESS conference on computers and accessibility*, 2019, pp. 68–82.

[27] Z. Li and R. Hollis, "Toward a ballbot for physically leading people: A human-centered approach," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 4827–4833.

[28] M. Quigley, B. Gerkey, K. Conley, J. Faust, T. Foote, J. Leibs, E. Berger, R. Wheeler, and A. Ng, "Ros: an open-source robot operating system," in *Proc. of the IEEE Intl. Conf. on Robotics and Automation (ICRA) Workshop on Open Source Robotics*, Kobe, Japan, May 2009.

[29] Visual Computing Institute Aachen, "2d_lidar_person_detection," 2020, [Online; accessed 07-November-2023]. [Online]. Available: https://github.com/VisualComputingInstitute/2D_lidar_person_detection

[30] CROWDBOT, "Crowdbot data sets / softwares," 2021, [Online; accessed 07-November-2023]. [Online]. Available: https://crowdbot.eu/data-sets-softwares/

[31] 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.

[32] L. Beyer, A. Hermans, and B. Leibe, "Drow: Real-time deep learning-based wheelchair detection in 2-d range data," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 585–592, 2016.

[33] D. Jia, M. Steinweg, A. Hermans, and B. Leibe, "Self-supervised person detection in 2d range data using a calibrated camera," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 13 301–13 307.

[34] K. Bernardin and R. Stiefelhagen, "Evaluating multiple object tracking performance: the clear mot metrics," *EURASIP Journal on Image and Video Processing*, vol. 2008, pp. 1–10, 2008.

[35] A. Leigh, J. Pineau, N. Olmedo, and H. Zhang, "Person tracking and following with 2d laser scanners," in *2015 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2015, pp. 726–733.