3D object tracking for self-driving vehicles using synthetically-generated data

Diogo Mendonça
diogomendonca1421@ua.pt
DETI, University of Aveiro
Aveiro, Portugal

Miguel Drummond
mvd@av.it.pt
Instituto de Telecomunicações, University of Aveiro
Aveiro, Portugal

Petia Georgieva
petia@ua.pt
IEETA and DETI, University of Aveiro
Aveiro, Portugal

Abstract

In this work, we propose a strategy to generate synthetic data for 3D object tracking using the Precise-Synthetic Image and LiDAR (PreSIL) source code [7].

We tested the highest ranked LiDAR-only tracking algorithm with the generated data and compared with the baseline performance in the KITTI object tracking benchmark in order to verify the veracity of our data.

This is the first publicly available 3D object tracking dataset using an idealized next generation LiDAR sensor, permitting the study of LiDAR object tracking now aided by radial velocity.

1 Introduction

Autonomous vehicles, or self-driving cars is a topic of active research and development, expected to play an increasingly important role in the transportation landscape in the coming years, driven by the desire to improve the safety, efficiency, and convenience. These vehicles use highly advanced software algorithms to perceive and navigate through an environment without the need for a human driver. In order to develop these algorithms, high-quality and rich datasets are required to train, evaluate and benchmark different implementations.

Real-world datasets like the KITTI Vision Benchmark [6], nuScenes [1] and the Waymo Open Dataset [3] are the most commonly used as the experimental basis in the research of automotive tasks like object segmentation or 3D object detection, despite displaying significant drawbacks such as expensive equipment and labelling costs.

These drawbacks can be solved by recurring to synthetic datasets, that allow for the inexpensive generation of data at scale and high customization of driving scenarios that are too difficult or dangerous to replicate in a real-world environment. These datasets are reliant on a simulation platform, like CARLA [5], that allows researchers to create diverse and realistic scenarios with accurate 3D object models and test different accurate virtual sensors, such as cameras, RADAR and LiDAR.

In this work, we intend to verify the feasibility of the development of a 3D object tracking model using a synthetic dataset generated using the Precise Synthetic Image and LiDAR (PreSIL) source code [7]. This was achieved by testing a state-of-the-art algorithm with said dataset and comparing the resulting metrics with the baseline values in the KITTI object tracking benchmark.

2 Methodology

Starting with dataset generation, our goal was to replicate the driving environments and sequence numbers and sizes of the KITTI object tracking dataset with PreSIL [7]. This data generation code leverages the realistic and detailed 3D environments and objects of GTA V and the game's native functions and plugins, such as the Script Hook V library [2], to simulate LiDAR and camera data and create a KITTI-like format dataset with automated annotation, as can be seen in Fig. 1.

Despite providing an identical vertical and horizontal angular resolution and frame rate as the Velodyne HDL-64E sensor used in the capture of the KITTI dataset, PreSIL limits the virtual LiDAR's FoV (Field of View) to the FoV of the camera (90°), leading to a similar but cropped synthetic point cloud counterpart. Another encountered limitation was the lack of control over the environment, with no ability to modify the GTA V world or control the cars' movement (due to the reliance on native AI pathing functions).

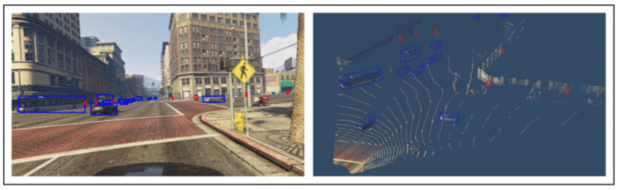

Figure 1: Annotated image and point cloud pair in PreSIL generated dataset. Reproduced from: [7].

The final generated dataset contains 20 training and 19 testing sequences, with over 16k frames and 80k annotations in multiple different driving scenarios such as interstate roads, highways or urban traffic.

Object type	#Occurrences
Pedestrian	40214
Car	32105
Misc	6137
Van	1152
Truck	1102
Cyclist	696
Person_sitting	254

Table 1: Object occurrences in the synthetic dataset.

In order to evaluate the quality of the dataset for 3D object tracking tasks, the highest ranked LiDAR-only tracking algorithm in the KITTI object tracking benchmark was selected, PC3T (Prediction Confidence 3D Tracker) [10].

Figure 2: Visualization of object tracking in the PreSIL-generated dataset using PC3T.

This algorithm follows the tracking-by-detection paradigm, meaning that it is reliant on the output of a 3D object detection algorithm to predict the future trajectories of traffic participants (in this study only cars where considered relevant). As can be seen in Fig. 3, this algorithm can be split into 4 parts:

(1) Detect objects from point clouds using an external 3D object detector,

(2) Estimate possible current states of tracked objects using a motion model and assign a prediction confidence score,

(3) Associate the predicted states with the detected states,

(4) Update the matched pairs and set unmatched detected states as new tracked states.

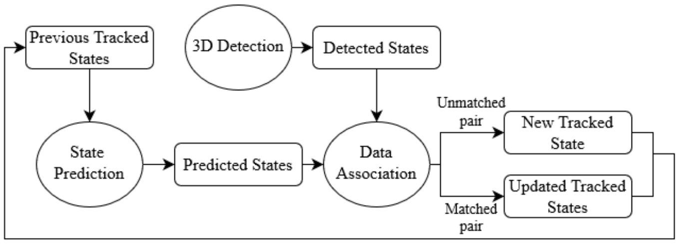

Figure 3: Flow diagram of the design of PC3T.

The SECOND-IoU 3D object detector was chosen to function as the external 3D object detector for PC3T due to displaying great performance in the KITTI object detection and object orientation estimation benchmark and being used to benchmark PC3T in its original paper, but also due to its availability in OpenPCDet [9], a toolbox developed by OpenMMLab that is the codebase of state-of-the-art 3D object detection models and provides an user-friendly and less error-prone interface for user utilization and customization.

With this toolbox, we were able to train the SECOND-IoU detector in a PreSIL-generated 3D object detection dataset created by Leandro Alexandrino, composed of 15k distinct frames [4]. This allowed the inference in our dataset, generating the detections of the annotated objects with an AP_R40 score for Easy, Medium and Hard difficulties of 95.85%, 90.78% and 90.41%, respectively.

3 Results

When evaluating this dataset for 3D object tracking, the metric used as reference was the same as in the KITTI 3D object tracing benchmark, HOTA (Higher Order Tracking Accuracy) [8]. This metric is very complete, due to its capability to explicitly measure and combine two important metrics in object tracking, detection (DetA) and association (AssA) accuracy. DetA measures the alignment between the set of all predicted detections and the set of all ground-truth detections, while AssA measures how well a tracker links detections over time into the same identities.

As illustrated in figure 4, the PC3T algorithm performs slightly worse in our dataset than in KITTI. This might due to some problems encountered during an visual examination of the sequences:

- PC3T struggles to maintain the identity of a specific car model, leading to a lower AssA score in sequences where that car is present. In frames where its identity is changed, this car is considered as a "new" car (never seen before), due to its predicted states not being assigned to a detected state, leading to a new tracked state.
- The virtual LiDAR sensor struggles to detect far-away objects in frames with road inclination, due to the static vertical direction of the LiDAR, always perpendicular to the z-axis. This creates an angle mismatch between the road surface and the LiDAR's direction when on a hill, that, combined with the LiDAR's limited vertical FoV (28.6°), can lead to non-detected objects.

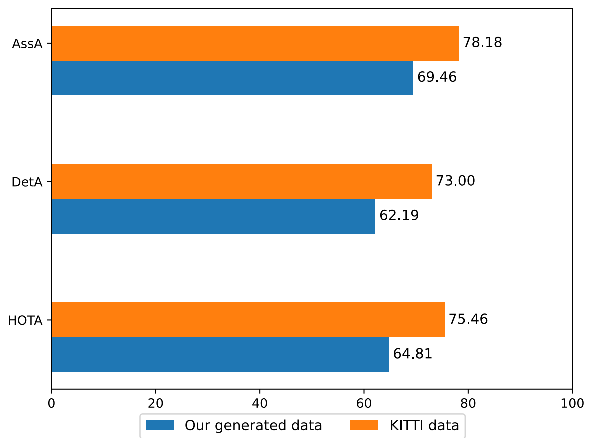

Figure 4: HOTA evaluation.

4 Conclusion

In this paper, we generated a synthetic dataset using PreSIL suited for the task of 3D object tracking, with 20 training and 19 testing sequences, over 16k frames and 80k annotations and multiple different driving scenarios.

With OpenPCDet, we were able to conclude that, using our dataset, it is possible to generate accurate detections of the dataset's annotated objects, with an AP_R40 score of over 90% on Easy, Medium and Hard detection difficulties.

When testing the PC3T algorithm in our dataset, we were able to obtain only slightly worse HOTA metrics than the ones set with the KITTI object tracking dataset, verifying its feasibility as the basis of the development of a new object tracking algorithm.

5 Future Work

The final objective is to design a "proof-of concept" object tracking and collision detection model using a synthetic dataset with an next-generation coherent LiDAR sensor and comparing said model with state-of-the-art solutions with LiDAR. This new sensor might provide an advantage on tracking tasks due to being capable of measuring the frequency difference between the emitted and received signals and, taking advantage of the Doppler effect, calculate the relative angular velocity of the captured object, much like a RADAR.

This is very much an exploratory research, because, as far as we know, there is no other publicly available 3D object tracking dataset using this next-generation sensor.

6 Acknowledgments

This work is funded by FCT/MCTES through national funds and when applicable co-funded EU funds under the project UIDB/50008/2020-UIDP /50008/2020.

References

[1] nuScenes, August 2023. URL https://www.nuscenes.org. [Online; accessed 27. Aug. 2023].

[2] Script Hook V + Native Trainer, August 2023. URL https://pt.gta5-mods.com/tools/script-hook-v. [Online; accessed 27. Aug. 2023].

[3] Open Dataset – Waymo, August 2023. URL https://waymo.com/open. [Online; accessed 27. Aug. 2023].

[4] Leandro Alexandrino, Miguel Drummond, Petia Georgieva, and Hadi Zahir. Synthetic automotive LiDAR dataset with radial velocity additional feature - (x,y,z,v). *Zenodo*, October 2022. doi: 10.5281/zenodo.7276691.

[5] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. CARLA: An open urban driving simulator. In *Proceedings of the 1st Annual Conference on Robot Learning*, pages 1–16, 2017.

[6] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[7] Braden Hurl, Krzysztof Czarnecki, and Steven Waslander. Precise synthetic image and lidar (presil) dataset for autonomous vehicle perception. In *2019 IEEE Intelligent Vehicles Symposium (IV)*, pages 2522–2529. IEEE, 2019.

[8] Jonathon Luiten, Aljosa Osep, Patrick Dendorfer, Philip Torr, Andreas Geiger, Laura Leal-Taixé, and Bastian Leibe. Hota: A higher order metric for evaluating multi-object tracking. *International journal of computer vision*, 129:548–578, 2021.

[9] OpenPCDet Development Team. Openpcdet: An open-source toolbox for 3d object detection from point clouds. https://github.com/open-mmlab/OpenPCDet, 2020.

[10] Hai Wu, Wenkai Han, Chenglu Wen, Xin Li, and Cheng Wang. 3d multi-object tracking in point clouds based on prediction confidence-guided data association. *IEEE Transactions on Intelligent Transportation Systems*, 23(6):5668–5677, 2021.