A Synthetic Dataset Generator for Automotive Overtaking Maneuver Detection

Luca Forneris[1[0009-0008-9595-3520]], Riccardo Berta[1[0000-0003-1937-3969]], Alessio Capello[1[0000-0003-4277-7283]], Marianna Cossu[1[0009-0008-4548-7648]], Matteo Fresta[1[0009-0000-7265-7501]], Fabio Tango[2[0000-0001-7283-4571]], and Francesco Bellotti[1]

[1] Department of Electrical, Electronic and Telecommunication Engineering (DITEN)-University of Genoa, Via Opera Pia 11a, 16145 Genova, Italy.
{luca.forneris, alessio.capello, marianna.cossu, matteo.fresta}@edu.unige.it,
{riccardo.berta, francesco.bellotti}@unige.it
[2] Centro Ricerche Fiat (CRF), Orbassano, Italy.
fabio.tango@crf.it

Abstract. This paper presents a novel tool for generating driving scenario datasets, that are a key asset to advance research and development in automated driving and driver assistance systems. The tool relies on the MATLAB Automated Driving Toolbox and focuses on the overtaking maneuver. It uses simulated vehicular data, without relying on camera-equipped real-world vehicles, thus providing a low-cost solution, while allowing to abstract the main action features, that are very important for the pre-training of machine learning models. The tool has been designed to target customization (in terms, e.g., of road curvature radii), in order to allow meeting specific requirements, while its interoperability (e.g., multiple-format export) supports integration with other development environments. A preliminary analysis of the first scenarios generated with the tool confirms the validity of the system under development.

Keywords: automated driving, driving scenario, synthetic datasets, vehicular data, driving scenario classification, driving scenario detection

1 Introduction

Automated driving systems require reliable and high-quality scenario datasets for testing and validation. Scenario-based approaches are increasingly important for verifying automated driving systems compared to traditional distance-based methods. Overtaking (OV) and Lane Change (LC) maneuvers are particularly crucial due to their complex interactions between the Ego Vehicle (EV) and the Leading Vehicle (LV). To address dataset challenges, this paper proposes a tool for generating synthetic OV maneuver scenarios, highlighting its versatility, variability, adaptability, and interoperability. By offering a comprehensive dataset, this tool advances research and development in automated driving and driver assistance systems, ultimately promoting safer and more efficient autonomous vehicles.

Existing literature lacks an analysis of OV maneuver classification that doesn't rely on camera-equipped vehicles but instead utilizes only vehicular data. Our dataset creation serves two main purposes: i) to develop an analysis tool for identifying and classifying the addressed scenario in an acquisition file, and ii) to form the training set for a scenario classifier (e.g., left/right LC and possibly OV maneuvers). This paper focuses on dataset design and implementation, providing significant opportunities for research and development in automated driving and driver assistance systems. By focusing on vehicular data like relative position, distance, angle, and velocity between vehicles, researchers can safely develop innovative and practical solutions for automated vehicles to perform LC and OV maneuver. This approach may reduce complexity and costs compared to camera-based systems that rely on sophisticated computer vision algorithms and hardware.

Section 2 reviews existing literature related to our study. In Section 3, we outline our systematic approach, data collection methods, experimental setup, and tools used, emphasizing our work's key strengths. Section 4 demonstrates the usefulness of our experimental results, while Section 5 summarizes key findings and contributions. We highlight the importance of our methodology and dataset generation tool, along with discussing future research directions.

2 Related Works

Wachenfeld et al. [1] propose a scenario-based approach for validating automated driving functions, as relying solely on distance-based validation would be impractical due to the extensive kilometers required. This scenario-based method is becoming increasingly important for verifying and validating automated driving systems [2], necessitating high-quality and reliable scenario datasets.

Geyer et al. [3] compare scenarios to dynamic storylines, incorporating anticipated driver actions without rigidly specifying every detail. Scenario implementation considers factors like driver free will, allowing for flexible and context-dependent realization. Accurate scenario detection and classification involve identifying and interpreting various road situations and events, such as Car Following (CF), Lane Changes (LC) and Overtaking (OV) maneuvers. This detection enables automated driving systems to make contextually appropriate decisions.

Scenario detection is crucial in real-time for determining vehicle control operational modes and in offline analysis for assessing system performance [4]. High-quality datasets from real-world sensors are essential for training and testing autonomous driving systems, but creating them can be costly and time-consuming. Virtual simulations, like the one proposed by Cossu et al. [5], offer a complementary solution to address dataset challenges. They allow the generation of diverse scenarios through CARLA simulator [6] based on user-specified parameters, providing customized and tailored content for research purposes.

3 Methodology

We propose a synthetic OV maneuver scenario pipeline implemented using MATLAB [7] Automated Driving Toolbox [8]. The tool generates synthetic data extracted from simulations, focusing solely on vehicular data signals without using real sensors or cameras. Data is collected from vehicle dynamics, emulating sensors providing speed, yaw-rate, steering angle, etc. The current version allows the EV and LV to drive on a one-way road with two 3.5m wide lanes, simulating a common highway scenario. Future improvements may include different road types like two-way roads or intersections. The main features of our tool-chain are:

- Reproducibility: scenarios are completely deterministic, guaranteeing a complete control of the experiments.
- Variability: the case histories considered in the study were generated using real statistical distributions, ensuring a diverse and variable dataset. Moreover, the range of customizable parameters enables the creation of highly distinct scenarios.
- Versatility: the tool offers complete customization according to the user's specific requirements. Users have the possibility to parameterize multiple factors, enabling the creation of varied and/or highly specific scenarios. Moreover, the tool facilitates the realization of diverse situations, including intersections, traffic roundabouts, and more.
- Interoperability: the work environment facilitates the export of scenarios generated in multiple extensions, enabling their import into other simulation and development environments. This capability empowers users to make additional modifications to the scenarios as needed.

3.1 Specifications

As anticipated before, one of the strengths of this tool is the possibility to customize simulated scenarios seamlessly, as users can define all the meaningful values by editing a simple text file. In particular, the user can define EV and LV speeds, the duration of the LC maneuver, the curve direction (left or right) and radius. Those parameters are required to define the generate a LC. In accordance with the user's requirements, the specifications of the scenarios generator can be documented within an external file (e.g., in a JSON format). The versatility of this tool allows for the inclusion of additional parameters to accommodate various other requirements. The chosen parameters for varying the scenarios include:

- EV speed [km/h]
- LV speed [km/h]
- Lane change maneuver duration [s]
- Road curvature radii [m], 0 for straight roads
- Curve direction, left or right.

The software utilizes the set parameters to interpret and define the scenarios.

3.2 Creation of road network and vehicles

The road is generated, including straight and curved segments, with realistic infrastructure to represent a lifelike road layout. After road network generation, vehicles are placed according to predefined criteria and distribution to represent various traffic scenarios. To mimic real driver behavior, noise is added to the trajectories of EV and LV, simulating skids. An example of the road structure is depicted in Fig. 1.

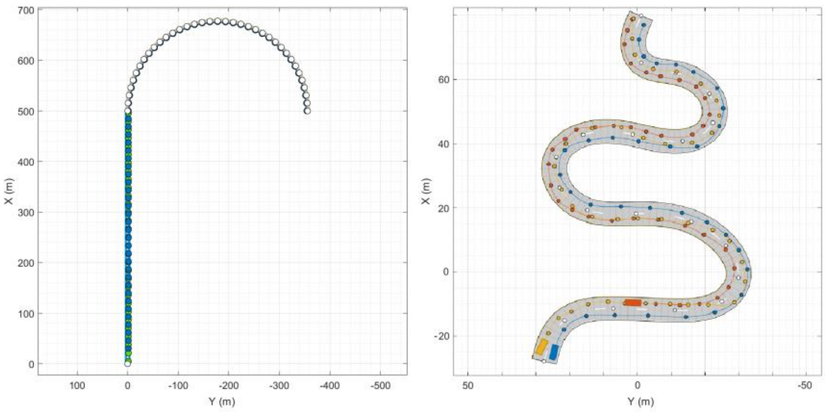

Fig. 1. Example of two road networks from the interface of Driving Scenario Designer application. On the left, the former portrays a road composed of a 500-meter straight section followed by a curve with a radius of 178 meters. Meanwhile, the latter illustrates a racetrack featuring three vehicles, each with a designated trajectory to follow.

3.3 Overtaking scenario instances

Two versions were established for each overtaking scenario instance. The first version comprises solely the EV driving along the road. The second version involves the EV following the LV without overtaking it. In the latter case, both vehicles maintain the same speed, and to ensure safety, a constant time headway of 1.3 seconds is maintained between them. In case of overtake, the difference between the speed of EV and LV is added to the list of parameters, so that the generated maneuvers can be as more realistic as possible.

3.4 Scenario Simulation and export

The scenario is simulated and data is stored at each timesteps, which can be defined a priori according to the user needs. Users should determine an appropriate timestep value striking a balance between simulation accuracy and computational burden. Once all pertinent vehicular data have been acquired, the data relevant to the newly simulated scenario is now ready to be stored and used. Scenarios may be saved in different formats, such as ASAM OPENDRIVE® or OPENSCENARIO®.

4 Experimental results

As a preliminary functional test, we defined a combination of 1000 parameter values. The corresponding baseline instances were generated and simulated, with a sample creation average time of 0.99 seconds. This process has been performed on a PC using 64-bit Windows 11 as operative system and equipped with an Intel core i7-12700H 2.30 GHz CPU, 16 GB RAM, and NVIDIA GeForce RTX 3060 GPU with 6 GB of memory. The generated instances have been verified by experts from the Hi-Drive project, especially to ensure physical consistency between values (e.g., speeds and radii of curvature). This demonstrates that by exploiting the proposed tool, users can effortlessly generate a dataset of OV maneuver scenarios by simply setting the desired values of some relevant parameters, with no need to write any script/code nor master a simulation environments.

5 Conclusions and future works

As efficient creation of driving scenario data is key to advance development of advanced driving assistance systems (ADAS) and automated driving functions (ADF), we have designed a tool, based on MATLAB Automated Driving Toolbox for implementing various instances of overtaking scenarios, in a variety of conditions. The relevant vehicular signals are recorded, so that can be later employed as part of machine learning datasets. This approach offers significant benefits, as it enables computationally efficient simulations and obviates the necessity of employing complex computer vision techniques to process camera-recorded data. The adaptability of the tool allows for extensive customization to suit specific user/company requirements. Authenticity of the scenario settings benefits from the incorporation in the tool of real maps from OpenStreetMap®. From serving as a diagnostic method for existing scenario classifiers with high accuracy to acting as a benchmark for comparing various classification methods. Furthermore, the flexibility of the tool goes beyond mere OV or LC scenarios. The combination of these attributes streamlines development, reducing computational complexity compared to numerous state-of-the-art tools, thereby unlocking a variety of potential applications.

A first already planned for future work will concern the implementation of other traffic vehicles beside EV and LV. On the other hand, we are interested also in involving of real users in data collection through the use of a steering wheel device connected to the tool. This approach aims to enhance the realism and reduce determinism in the data collection process, as a real driver exhibits non-deterministic behavior. Additionally, users can execute different driving styles (e.g., comfort, normal, aggressive), thus integrating the dataset with variety and human aspects (e.g., expertise, knowledge, emotions) into the dataset.

6 Acknowledgements

The authors would like to thank all partners within the Hi-Drive project for their co-operation and valuable contribution. This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 101006664. The sole responsibility of this publication lies with the authors. Neither the European Commission nor CINEA – in its capacity of Granting Authority – can be made responsible for any use that may be made of the infor-mation this document contains.

7 References

1. Wachenfeld, W., Winner, H.: The Release of Autonomous Vehicles. In: Maurer, M., Gerdes, J.C., Lenz, B., and Winner, H. (eds.) Autonomous Driving: Technical, Legal and Social Aspects. pp. 425–449. Springer, Berlin, Heidelberg (2016). https://doi.org/10.1007/978-3-662-48847-8_21.
2. Nalic, D., Mihalj, T., Eichberger, A., TU Dresden, GERMANY, Bäumler, M., Lehmann, M.: Scenario Based Testing of Automated Driving Systems: A Literature Survey. In: FISITA World Congress 2021 - Technical Programme. FISITA (2021). https://doi.org/10.46720/f2020-acm-096.
3. Geyer, S., Baltzer, M., Franz, B., Hakuli, S., Kauer, M., Kienle, M., Meier, S., Weißgerber, T., Bengler, K., Bruder, R., Flemisch, F., Winner, H.: Concept and development of a unified ontology for generating test and use-case catalogues for assisted and automated vehicle guidance. IET Intell. Transp. Syst. 8, 183–189 (2014). https://doi.org/10.1049/iet-its.2012.0188.
4. Bellotti, F., Osman, N., Arnold, E.H., Mozaffari, S., Innamaa, S., Louw, T., Torrao, G., Weber, H., Hiller, J., De Gloria, A., Dianati, M., Berta, R.: Managing Big Data for Addressing Research Questions in a Collaborative Project on Automated Driving Impact Assessment. Sensors. 20, 6773 (2020). https://doi.org/10.3390/s20236773.
5. Cossu, M., Berta, R., Capello, A., De Gloria, A., Lazzaroni, L., Bellotti, F.: Developing a Toolchain for Synthetic Driving Scenario Datasets. In: Berta, R. and De Gloria, A. (eds.) Applications in Electronics Pervading Industry, Environment and Society. pp. 222–228. Springer Nature Switzerland, Cham (2023). https://doi.org/10.1007/978-3-031-30333-3_29.
6. Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., Koltun, V.: CARLA: An Open Urban Driving Simulator. In: Proceedings of the 1st Annual Conference on Robot Learning. pp. 1–16. PMLR (2017).
7. Inc, T.M.: MATLAB version: 9.14.0 (R2023a), https://www.mathworks.com, (2023).
8. Inc, T.M.: Automated Driving Toolbox version: 3.7 (R2023a), https://www.mathworks.com, (2023).