Benchmarking Sensing and Motion Planning Algorithms for Autonomous Driving

Yue Linn Chong[1], Huiyu Leong[2], Christina Dao Wen Lee[1], Marcelo H. Ang Jr.[1]

***Abstract*—Autonomous Driving (AD) requires sensing and motion-planning algorithms to perform well in the dynamic real-world environment. Sensing is required to detect obstacles and localize the vehicles, while motion planning computes a trajectory for the vehicle to follow. We propose a dual-layer approach for AD, with the lower layer guaranteeing safety and the intelligence layer to improve AD effectiveness. This paper examines existing metrics and datasets used for testing and evaluating algorithms on the intelligence layer and seeks to highlight gaps between current metrics and real-world AD performance. Generally, there is insufficient emphasis on developing and verifying comprehensive datasets or test cases that can evaluate algorithms' performances in challenging real-world scenarios. Moreover, metrics for sensing such as mean average precision (mAP) and Final Displacement Error (FDE) do not capture the impact of erroneous algorithms on AD. Motion planning metrics could also be improved upon by considering using qualitative methods of evaluation which include more critical features for AD.**

***Index Terms*—autonomous vehicles, benchmarks, sensing, perception, planning, controls**

I. Introduction

In autonomous driving (AD), the algorithm stack can be divided into two layers. The lower layer guarantees safety and cannot afford errors. This safety layer must accurately detect free space and avoid collisions. The higher layer includes algorithms that improve the effectiveness of AD. This intelligence layer may be overridden by the safety layer to guarantee safety. Modules on the safety layer require perfect performance for the AD to be sufficiently safe for on-road driving. Due to the simpler nature of these algorithms, their evaluations are straightforward. However, the evaluation of the intelligence layer is non-trivial, so they are the focus of this paper.

These AD algorithms can generally be divided into two categories, namely sensing and motion planning. Sensing algorithms aim to understand the environment the AD is being conducted in. Motion planning algorithms aim to plan a reference trajectory for AD with input from the sensing algorithms and other apriori data and execute the driving commands to track the reference trajectory.

There are many available metrics proposed to evaluate the performance of the algorithms, but not all may be aligned with the AD performance. The metrics should be reflective of the algorithm's future real-world behavior, though those cases are yet to be known. Ideally, the metric is succinct and intuitive for ease of risk propagation calculation [1]. It should also be agnostic to the variations of algorithms in a task and avoid user-defined thresholds [1]. Fortunately, all categories consider computational time as a criterion for AD, and so it is not discussed further in this paper.

[1] Yue Linn Chong, Christina Dao Wen Lee, Marcelo H. Ang Jr., are with Department of Mechanical Engineering, National University of Singapore, Singapore. {mpechon, mpecldw, mpeangh}@nus.edu.sg

[2] Huiyu Leong hleong@alumni.stanford.edu

II. Evaluating Sensing Algorithms

Sensing algorithms can be divided into two broad objectives: (1) to map and localize the ego vehicle and (2) to detect, track, and predict objects. These objectives are discussed separately in II-A and II-B respectively.

A. Evaluation Mapping and Localization Algorithms

Mapping algorithms represent the environment in a map such that the ego vehicle can be localized and a path can be planned from the start to the goal position. The evaluation of mapping algorithms as stand-alone is challenging as the type of the map generated depends on the method used, which leads to method-specific metrics that are not generalizable. Moreover, ground truth maps are not easily obtained. Alternatively, since mapping and localization use the same core methods, they can be evaluated together using the relative localization error. Since AD performance only depends on localization error, this metric is a feasible approach. That is, localization error is a metric of map quality.

However, these current metrics do not measure the risks due to algorithm errors to the AD stack as a whole unless the dataset used with the metric is sufficiently diverse. For example, some scenarios are known to be challenging, such as scenes with crowded dynamic objects or with sparse features should also be included. Ideally, datasets should also differentiate specific applications with different challenging elements such as urban, rural or highway driving. The datasets must also have a sufficiently diverse sensor suite to encompass the wide variety of sensors being used for different methods with 2D and 3D LiDARs, and mono and stereo cameras. However, there is currently no such dataset to the authors' knowledge.

B. Evaluating Classification, Tracking and Prediction Algorithms

Classification algorithms locate and classify objects. Tracking algorithms track the classified objects over time, which generates the past trajectories of each object used in trajectory prediction algorithms to predict future trajectories.

The current metric for classification is the mean average precision (mAP). Object tracking algorithms are evaluated using CLEAR-MOT [1] metrics or by measuring the number of mostly-tracked, partly-tracked and mostly-lost objects.

However, these metrics do not capture the resulting risks for AD application. Specifically, these metrics average over each object equally, ie. critical objects, such as objects near the ego, are treated equally as less critical objects. For prediction algorithms, the common metrics are the Average Distance Error (ADE) and the Final Displacement Error (FDE). However, these metrics do not take into account errors in the predicted yaw. Moreover, trajectory prediction, especially for pedestrian and cyclists, use ground truths past trajectories as input during the evaluation, while the actual AD systems use imperfect inputs past object trajectories from the object tracking algorithm.

Most object classification, tracking and prediction algorithms are data-based algorithms, requiring a dataset for algorithm training. With several datasets freely available, there lacks an evaluation between datasets to ensure the datasets are sufficiently diverse and well distributed. For example, in tracking and prediction datasets, there is a large percentage of stopped versus moving target vehicles. Since stationary objects are easier to predict and track, a lower overall error is obtained.

III. Evaluating Motion Planning Algorithms

Motion planning algorithms can be divided into two broad categories: (1) planning, and (2) controls. These categories are separately evaluated in III-A, III-B, while challenges of evaluation common to both categories are discussed in III-C.

A. Evaluating Planning Algorithms

The objective of planning algorithms is to plan a feasible route for the ego vehicle given a goal position. Current path-planning metrics include [2]: (1) collision avoidance, (2) path length, and (3) feasibility based on vehicle dynamics in different scenarios. These metrics are often tested across a limited set of scenarios, thus may not reflect the performance of the algorithms in real AD applications. The evaluation of path planning algorithms may be improved by extending the number of test scenarios using a given controller and obtaining the statistical average of the numeric indicators to quantify the performance of each path planning algorithm. Moreover, there are additional qualitative characteristics to consider in evaluating the safety of planned trajectories. This includes comparing the AD with that of a human driver with regards to compliance with traffic rules as well as rider comfort. Furthermore, vehicles in a pedestrian environment such as wheelchairs in shopping malls require planning algorithms to be evaluated in such environments where traffic rules are not present, but social rules apply instead. The challenge, therefore, lies in designing a comprehensive set of scenarios to test for these. The compliance with guidelines for proper AV behavior such as the TR68 [3] should also be measured as a form of metric for path planning algorithms.

B. Evaluating Controls Algorithms

The objective of control algorithms is to calculate the required speed and steering input for the vehicle to track the reference trajectory. There are several quantitative indicators of controller performance [4]: (1) lateral cross-track error, (2) yaw error, (3) longitudinal speed error, and (4) steering angle These are root-mean-square errors measured over paths of varying curvatures. However, there are other characteristics of controls for AD which are not captured, thereby requiring qualitative characteristics to supplement the evaluation.

An integral measure of controller robustness is its ability to stabilize external disturbances. Inaccurate localization could result in noticeable disturbances in errors. The stability of the controller under such situations is paramount for safe lane-keeping in urban driving but may not be measured with the above indicators. Moreover, the algorithm's reliance on a dynamically feasible trajectory should be considered. Algorithms without vehicle dynamics constraints imposed could have larger lateral errors at higher speeds, resulting in unsafe road behavior. There may also be situations where the controller is unable to solve to track the path. Hence, we would require metrics that take into account these characteristics to compare control algorithms.

C. Common Challenges

It is challenging to evaluate each motion planning algorithm solely using quantitative indicators due to their different qualitative characteristics. Instead, it may be useful to note the characteristics of each algorithm and develop an evaluation criteria specific to the use case of the AD. The metrics in the evaluation criteria may consist of the aforementioned indicators discussed, and application-specific performance requirements of the algorithms.

IV. Conclusion

We discuss benchmarks available for sensing and motion planning algorithms. Some metrics for sensing algorithms are not reflective of the risks propagated to the AD application. Motion planning metrics could also be improved upon by considering more critical features for AD. Some metrics lack a comprehensive dataset or set of test cases, which reflects the absence of verification that these are currently sufficient to be representative of real-world AD.

Acknowledgment

This research is supported by the National Research Foundation, Prime Ministers Office, Singapore under its LTA Urban Mobility Grand Challenge Program, ST Kinetics Autonomous Bus Trial.

References

[1] 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.

[2] D. G. Bautista, M. Arzamendia, J. Pérez, V. Milanes, and F. Nashashibi, "A review of motion planning techniques for automated vehicles," *IEEE Transactions on Intelligent Transportation Systems*, 2015.

[3] S. S. Council, *Technical Reference for autonomous vehicles - Part 1 : Basic behaviour*. Singapore: Enterprise Singapore, 2019, vol. TR 68 : Part 1 : 2019.

[4] N. H. Amer and Z. A. K. Hairi Zamzuri, Khisbullah Hudha, "Modelling and control strategies in path tracking control for autonomous ground vehicles: A review of state of the art and challenges," *Journal of Intelligent Robotic Systems*, 2016.