CENTER FOR CONNECTED AND AUTOMATED TRANSPORTATION

UMTRI-2023-2
March 2023

Development of an Integrated Augmented Reality Testing Environment and Implementation at the American Center for Mobility

Dr. Henry Liu

UMTRI

CENTER FOR CONNECTED AND AUTOMATED TRANSPORTATION

Report No. UMTRI-2023-2

March 2023

Project Start Date: January, 2021

Project End Date: December, 2021

Development of an Integrated Augmented Reality Testing Environment and Implementation at the American Center for Mobility

by

Henry Liu, Professor

University of Michigan

UMTRI | The University of Akron | Illinois Center for Transportation | Purdue University | Washtenaw Community College | Central State University

CENTER FOR CONNECTED AND AUTOMATED TRANSPORTATION

DISCLAIMER

Funding for this research was provided by the Center for Connected and Automated Transportation under Grant No. 69A3551747105 of the U.S. Department of Transportation, Office of the Assistant Secretary for Research and Technology (OST-R), University Transportation Centers Program. The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

Suggested APA Format Citation:
Liu, H.X (2022). Development of an Integrated Augmented Reality Testing Environment and Implementation at the American Center for Mobility. Final Report. UMTRI-2023-2.
DOI: 10.7302/7019

Contacts

For more information:

Dr. Henry X. Liu
University of Michigan
2350 Hayward, Ann Arbor, MI, 48109
Phone: (734) 647-4796
Email: henryliu@umich.edu

Center for Connected and Automated Transportation
University of Michigan Transportation Research Institute
2901 Baxter Road
Ann Arbor, MI 48152
umtri-ccat@umich.edu
ccat.umtri.umich.edu
(734) 763-2498

CENTER FOR CONNECTED AND AUTOMATED TRANSPORTATION

Technical Report Documentation Page

1. Report No. UMTRI-2023-2	**2. Government Accession No.**	**3. Recipient's Catalog No.**
4. Title and Subtitle Development of an Integrated Augmented Reality Testing Environment and Implementation at the American Center for Mobility DOI: 10.7302/7019		**5. Report Date** March 2023 **6. Performing Organization Code**
7. Author(s) Liu, Henry, Ph.D., https://orcid.org/0000-0002-3685-9920		**8. Performing Organization Report No.**
9. Performing Organization Name and Address		**10. Work Unit No.** **11. Contract or Grant No.** 69A3551747105
12. Sponsoring Agency Name and Address Center for Connected and Automated Transportation University of Michigan Transportation Research Institute 2901 Baxter Road Ann Arbor, MI 48109		**13. Type of Report and Period Covered** Final Report January 2021 – December 2021 **14. Sponsoring Agency Code**

15. Supplementary Notes
Conducted under the U.S. DOT Office of the Assistant Secretary for Research and Technology's (OST-R) University Transportation Centers (UTC) program.

16. Abstract
In this project, we developed an integrated solution for autonomous vehicle testing, in which the naturalistic driving environment (NDE) is combined with the augmented reality (AR) testing system. The integrated solution is implemented at American Center for Mobility (ACM). With the NDE, realistic virtual traffic flow can be generated in the testing environment and the real testing AV can interact with virtual background vehicles. The NDE is developed based on large-scale naturalistic driving data to ensure high fidelity. With AR techniques, the information between the simulation world and the physical world can be communicated and synchronized. Moreover, augmented images with virtual background vehicles can be rendered using the AR system. The developed system facilitates the testing processes of AV safety performance with low operational costs and high evaluation accuracy.

17. Key Words Autonomous vehicles, Testing and evaluation	**18. Distribution Statement** No restrictions.		
19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 29	**22. Price**

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

UMTRI · The University of Akron · Illinois Center for Transportation · Purdue University · Washtenaw Community College · Central State University

Table of Contents

List of Tables

List of Figures

Project Summary

This project develops an integrated solution for autonomous vehicle testing, in which the naturalistic driving environment (NDE) is combined with the augmented reality (AR) testing system. To add real vehicles to a test scenario, the time and monetary expense of coordinating and controlling real vehicles is very high. The developed system makes it possible to add virtual background traffic to the testing environment that the autonomous vehicle (AV) under test saw as "real". To achieve this, a high-fidelity NDE model is developed to generate virtual traffic flow with realistic human driving behaviors. The real AV under test will interact with simulated background vehicles so its safety performance can be evaluated and analyzed. The AR system communicates and synchronizes the information between the simulation and physical worlds and generates augmented images with virtual background vehicles. The proposed integrated solution is implemented at the American Center for Mobility (ACM), which is one of the world's premier test tracks for AVs located in Ypsilanti, Michigan. The developed system facilitates the testing processes of AV safety performance with low operational costs and high evaluation accuracy.

1. Introduction

Testing and evaluation is a critical step in the deployment of autonomous vehicles (AVs), which has received extensive attention from both the industry and academia in recent years [1-14]. Prevailing methods test AVs in the naturalistic driving environment (NDE), observe their performance, and make statistical comparisons to human driving performance [13]. Due to the rareness of safety-critical events, however, it has been pointed out that hundreds of millions and sometimes hundreds of billions of miles would be required in NDE to demonstrate the safety performance of AVs at the level of human-driven vehicles, which is intolerably inefficient for the real-world testing [1]. Therefore, testing AVs in NDE simulations has attracted increasing attention because of the advantages of controllability, repeatability, and efficiency [13,15].

The key to simulation testing is the trustworthiness of the testing results. As pointed out in various domains [15-18], the simulation-to-reality gap could hinder and even mislead the training and testing process of an agent. To fill this gap, existing studies have paid much attention to the fidelity of vehicle dynamics, sensor models, and photorealistic images based on techniques such as computer graphics, physics-based modeling, and data augmentation [15,19,20]. However, how to model the naturalistic behavior of human-driven vehicles with high fidelity still remains an open question. To answer this question, many AV companies tried to replay human driving behaviors according to the logged data collected from the real-world driving environment. However, as the human driving behaviors are pre-determined in the logged data, they cannot interact with AV models, which severely limits the scenarios that can be simulated. To address this issue, the human driving models developed in the transportation engineering field have been applied, such as the Intelligent Driving Model (IDM) [21] and MOBIL [22] models. However, although these models can interact with AV models, they were designed for traffic flow analysis purposes such as reproducing traffic oscillations and the fundamental diagram, which are not suitable for the AV simulation testing.

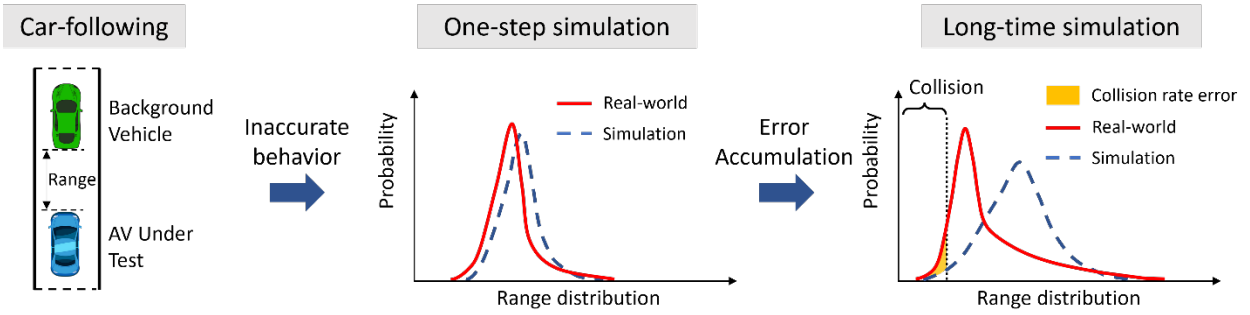


Figure 1 Illustration example that the distribution inconsistency might mislead AV development and deployment.

The AV simulation testing brings brand new requirements for the NDE modeling. To evaluate AVs' safety performance quantitatively, the accident rate of AVs in NDE is usually utilized [2,7,8,13]. As the human driving behaviors significantly affect the response and performance of AVs, estimating the accident rate accurately requires the NDE model in simulation to be distributionally consistent with the real-world driving environment. Specifically, let X denote the variables that define the environment, such as the behaviors of human drivers. Then, testing an AV in NDE is essential to sample X from its underlying distribution, denoted as $X \sim P(X)$, to estimate its performance μ_E^κ by

$$\mu_E^\kappa := \mathbb{E}_X\left(\phi_E^\kappa(X)\right) \approx \frac{1}{n}\sum_{i=1}^{n}\phi_E^\kappa(X_i) \approx \frac{m}{n}, X_i \sim P(X), \qquad (1)$$

where E denotes the event of interest, κ denotes the AV agent under test, $\phi_E^\kappa(X)$ denotes the AV performance at the environment specified by X, n denotes the number of tests, and m denotes the number of event E occurred during tests. According to the Monte Carlo method [23], only with the accurate distribution $P(X)$ in the simulation, the estimation of μ_E^κ can be statistically accurate. Therefore, the AV simulation testing requires distributionally consistent NDE models, which are significantly different from those for traffic flow analysis.

To further illustrate the new requirements, we consider a simple car-following scenario, where the leading vehicle is a human-driven vehicle and the following vehicle is an AV, as shown in Figure 1. In the real-world driving environment, the behavior of the leading vehicle could have an underlying distribution, which leads to the range distribution between two vehicles after one timestep (see Figure 1, middle, red curve). If the NDE model cannot accurately represent the stochastic behaviors of the leading vehicle, there could exist inconsistency between the range distributions in simulation (Figure 1, middle, blue dashed line) and real-world. After several time steps, this inconsistency will be accumulated and amplified, which could lead to significant estimation errors of the accident rate (Figure 1, right, yellow area). In this example, the accident rate will be underestimated, which may mislead the further development and deployment. Therefore, a realistic NDE with distributionally consistent environment statistics is critical for AV testing, which is a challenging requirement. To the best of our knowledge, there is no existing method that can fulfill this requirement.

To fill this research gap, we propose a data-driven optimization-based NDE modeling framework in this project, which can ensure the distributional consistency of vehicle microscopic behaviors. The overall pipeline of the framework includes two major steps as shown in Figure 2. In the first data-driven step, we propose to directly construct human driving behavior models using empirical distributions from the largescale naturalistic driving data (NDD). Specifically, the driving

behaviors are modeled by two longitudinal models (i.e., free driving and car following) and four lateral models considering different driving situations, and each model is described as action distributions for different states. For example, the car-following behaviors can be modeled by the acceleration distribution of the following vehicle conditional on different self speeds, relative speeds, and relative distances with the preceding vehicle. These empirical behavior models serve as basic models. If the dataset is sufficiently accurate, diverse, and large, the empirical distributions can accurately characterize human driving behaviors in different situations. However, due to the limitation of data quantity and unavoidable data noise, the obtained empirical distributions may still have small inaccuracy compared with the ground truth. As illustrated in Figure 1, the small inaccuracy can be accumulated and compounded with the simulation, so the resultant simulation environment will deviate from the realistic environment.

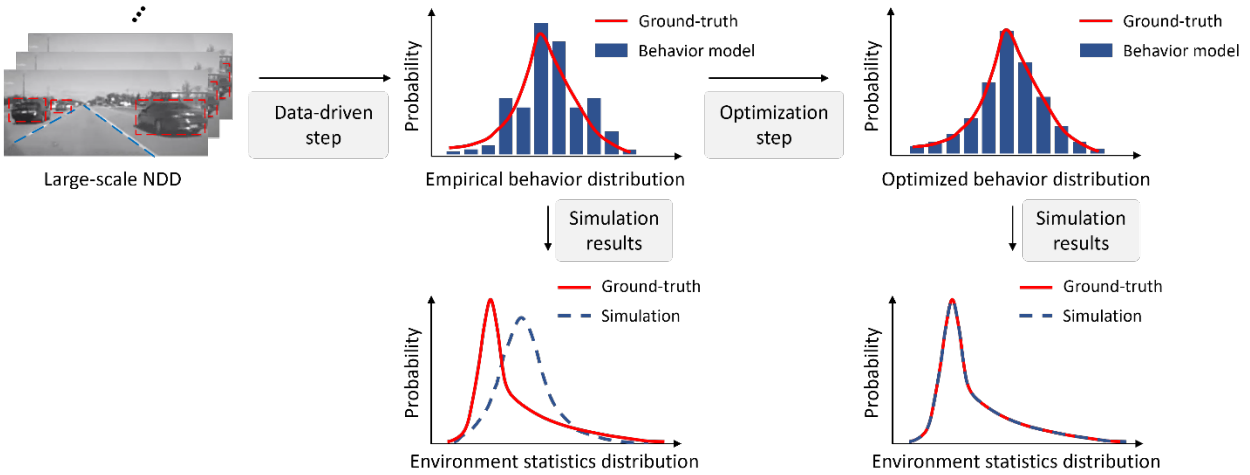


Figure 2 Proposed framework pipeline.

To further tackle the error accumulation issue, the second step of the framework is to refine empirical behavior models to minimize the accumulated errors using optimization methods. To achieve this goal, the key is to model the long-term effects of the error accumulation. In this study, by modeling the vehicle state evolution as a Markov chain, the long-term effects of the error accumulation could be modeled based on the stationary distribution of the Markov chain. Then, an optimization problem can be formulated to minimize the accumulated errors by adjusting the empirical behavior models. In this way, the error accumulation could be reduced, which results in the NDE model with more accurate distributions. This provides opportunities to conduct high-fidelity long-time simulation for full-length trip evaluation of AV's performance [24]. Using the large-scale real-world NDD collected by the University of Michigan, Ann Arbor, we validate the performance of the proposed method for a multilane highway driving environment. Compared with existing models, the proposed method demonstrates superior performance

regarding the distributional accuracy. To further validate the capability for AV testing, the generated NDE is also utilized to test the safety performance of an AV agent.

In summary, the main contributions of this project in generating NDE are threefold: first, the new requirements of NDE modeling are identified for AV testing purposes, which cannot be satisfied by most existing methods; second, a novel modeling framework is proposed to generate the NDE that is distributionally consistent with the real-world driving environment; third, the proposed method is validated using large-scale real-world NDD and the generated NDE is further validated by testing an AV model.

2. Data-driven NDE modeling

In this section, we propose a simple yet effective data-driven method for NDE modeling leveraging large-scale NDD. Specifically, six empirical behavior models are constructed including free-driving, car-following, and four lane-changing behaviors with different driving conditions, and then the NDE can be generated by combining the six empirical behavior models according to the driving condition at each time step (Section 2.1). To construct each empirical behavior model, the large-scale NDD is processed and utilized in Section 2.2. Then, a multi-lane highway driving environment is simulated to evaluate the performance of the empirical behavior models in Section 2.3, which validates the data-driven method and further motivates the robust modeling step in Section 3.

2.1 Empirical behavior models

To construct the NDE, both longitudinal and lateral behaviors of human drivers need to be modeled based on the vehicle's own state and its surrounding situations S. For example, human drivers need to decide their longitudinal accelerations and decide whether to take a lateral lane-change maneuver. In this study, six behavior models are proposed including free-driving, car-following, and four lane-changing behaviors with different driving conditions, as shown in Figure 3a. Specifically, the vehicle acceleration in the free-driving case is modeled that depends only on its current velocity, while the acceleration in the car-following case is modeled that depends on the velocity, range (relative position), and range rate (relative speed) of the subject vehicle and its preceding vehicle. To capture the lane-changing probability in different conditions, four lane-changing models are proposed, which output the lane-change probability of the subject vehicle at each moment. For example, in the cut-in lane change situation, the lane-changing probability depends on velocities and distances between the subject vehicle and the preceding vehicle in the current lane and the vehicle behind in the target lane. We note that more lane-changing models could be constructed in this framework by dividing the driving conditions into more categories if needed.

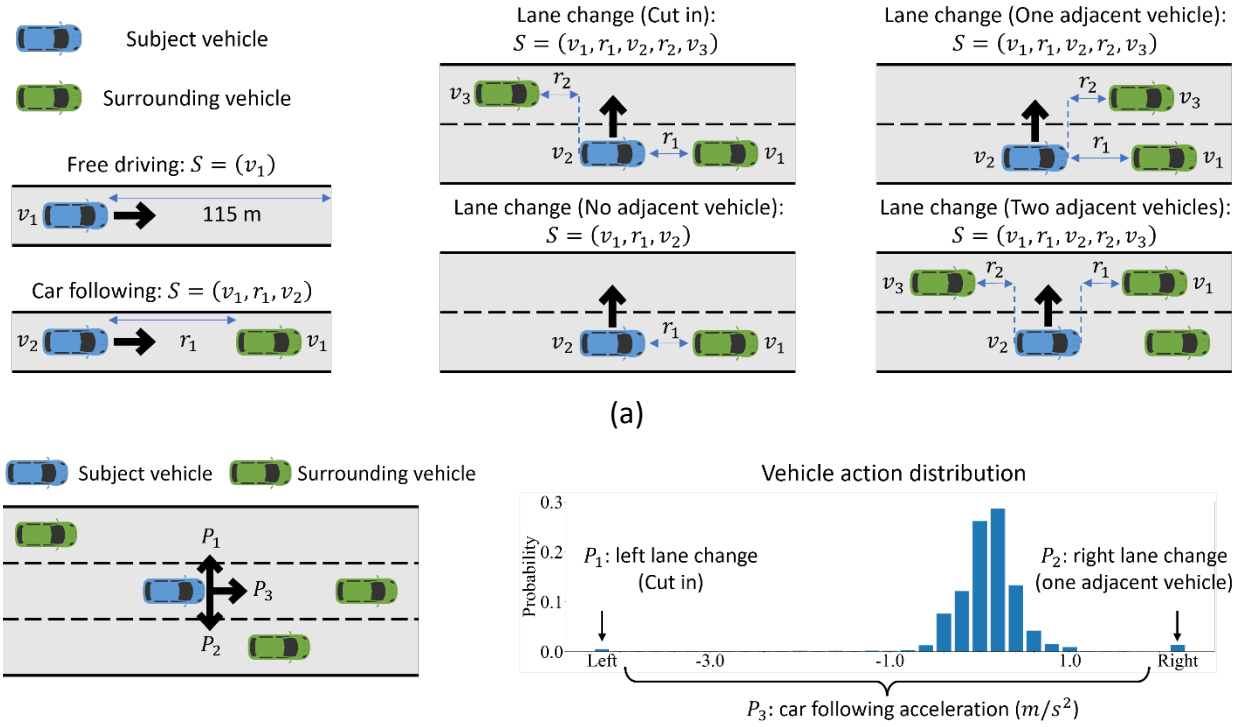


Figure 3 Illustration of empirical behavior models. a, longitudinal and lateral behavior models in this study. **b,** illustration example of how to simulate a vehicle's action at a decision moment**.**

After constructing the six behavior models, the NDE can be generated by combining the behavior models according to the driving condition at each time step. Taking Figure 3b (left) as an example, the subject vehicle can take left lane change, keep car-following, or take right lane change at the moment. The left lane change behavior can be categorized as a cut-in behavior, where the lane-changing probability P_1 can be obtained from the cut-in behavior model. The right lane change behavior can be categorized as a lane change with one adjacent vehicle, where the lane-changing probability P_2 can also be obtained by the corresponding behavior model. Moreover, the longitudinal acceleration probability P_3 can be obtained by the car-following behavior model. After normalization, we can obtain the action distribution of the vehicle as shown in Figure 3b (right). Then, the subject vehicle's action will be sampled from this distribution and used to update its state to the next time step. To simplify the modeling process, longitudinal acceleration is assumed zero if the vehicle is making a lane change behavior. Also, if there is no vehicle in front, the ego vehicle will not take lane-changing behavior. By repeating this process for all vehicles and time steps, the NDE can be generated.

The remaining question is how to construct the six behavior models with distributional accuracy.

In this study, we propose to directly estimate empirical behavior distributions as the behavior models by leveraging large-scale NDD. As the NDD records all the information needed for the human driving behaviors, accurate empirical behavior models could be constructed if using a sufficient amount of data with perfect quality. Although the actual data is usually limited by the data quality and quantity, these empirical behavior models could provide a good foundation and can be further improved as discussed in Section 3. Specifically, for each behavior model, we obtain the empirical probability $P(a \mid S)$ for all the vehicle actions $a \in \mathcal{A}$ at all discretized states $S \in \mathcal{S}$, where $\mathcal{A}$ denotes the action space, and $\mathcal{S}$ denotes the state space. Let $F(S) = \left[P(a_1 \mid S), \cdots, F\left(a_{|\mathcal{A}|} \mid S\right) \right]$ denote the probability mass function under a certain state S, then the empirical behavior model can be denoted by

$$F = \left[F(S_1), \cdots, F\left(S_{|\mathcal{S}|}\right) \right] \in \mathbb{R}^{|\mathcal{S}| \times |\mathcal{A}|}. \qquad (2)$$

Next, we will introduce how to process the NDD and construct F for all the six behavior models.

2.2 Naturalistic driving dataset processing

To construct empirical behavior models, we utilized large-scale NDD from the Integrated Vehicle Based Safety System (IVBSS) dataset [25] and the Safety Pilot Model Deployment (SPMD) dataset [26] at the University of Michigan Transportation Research Institute (UMTRI). In the IVBSS program, 108 drivers ranging from 20 to 70 years old were recruited. Each participant drove the IVBSS vehicle equipped with the data acquisition system (DAS) for 6 weeks. The relative distance and speed with the leading vehicle are recorded by radar at 10 Hz. The SPMD program covered over 34.9 million travel miles and included 98 vehicles equipped with the DAS and Mobileye to record human naturalistic driving behaviors. The data were also recorded at 10 Hz with positions, speeds, and accelerations of ego-vehicles, relative speeds with surrounding vehicles, and both longitudinal and lateral distances between vehicles and lane markings. We queried partial datasets with the following criteria: (1) vehicle was traveling at a speed between 20 m/s and 40 m/s; (2) dry surface condition; (3) daylight condition. The resulting dataset includes approximately 8,200 driving hours data.

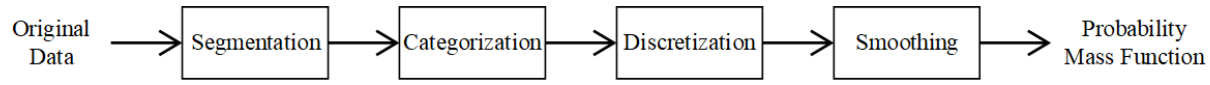
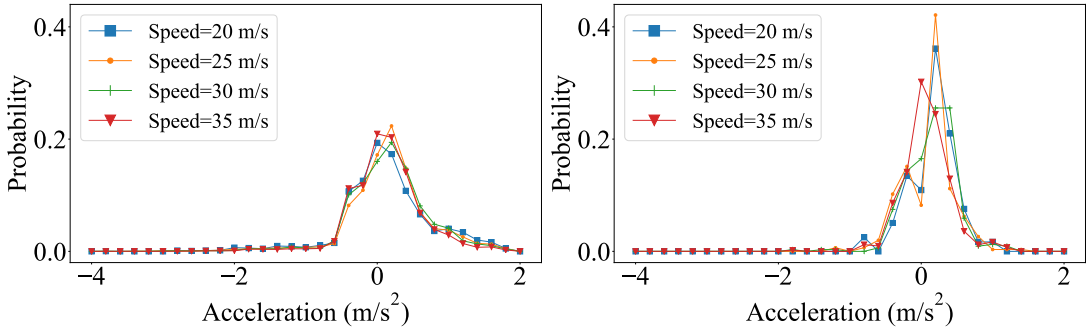


Figure 4 Data processing flow chart.

The data processing consists of four steps including segmentation, categorization, discretization, and smoothing, as shown in Figure 4. Specifically, the original data were first segmented into trajectories and then categorized into specific groups based on the six driving situations defined in empirical behavior models. Then, a smoothing technique was applied to the discretized action

distribution and finally, we could obtain the probability mass functions for each group, which constituted the six empirical behavior models. Figure 5 demonstrates examples of constructed empirical behavior models. Specifically, Figure 5a and Figure 5b show examples of the vehicle longitudinal acceleration distributions in free-driving and car-following situations, respectively. For the car-following case, Figure 5b indicates acceleration distribution when the ego-vehicle and its preceding vehicle have the same speed and their range is 30 meters. We can find that for both free-driving and car-following cases, the mean of acceleration is around zero, which is consistent with the intuition. Compared with the car-following situation, the probability of acceleration greater than zero is generally higher in the free-driving case, which is reasonable as well.



(a) (b)

Figure 5 Examples of empirical behavior models. a, Free driving. **b,** Car-following ($r_1 = 30m, v_1 = v_2$).

2.3 Performance evaluation of empirical behavior models

In this subsection, the performance of the NDE constructed by the six empirical behavior models is evaluated in a three-lane highway simulation, as illustrated in Figure 3b (left). The Hellinger distance [27] is used to quantitatively measure the dissimilarity between the simulated distribution and the true distribution. The Hellinger distance ranges from 0 to 1, and the smaller the measurement, the better the model performance. To demonstrate the performance, the background vehicles velocity and range distributions, which are important for the AV testing, are investigated as shown in Figure 6. Results show that although the distributions can roughly capture the trends of the real-world distributions, there still exists significant distributional inconsistency, particularly for the vehicle velocity. This inconsistency is caused by the error accumulation of the empirical behavior models, where the small model errors caused by the limited data quality and quantity are accumulated and amplified along with the simulation steps, as illustrated in Figure 1. To address this issue, the robust NDE modeling step is developed in the next section to further improve the empirical models.

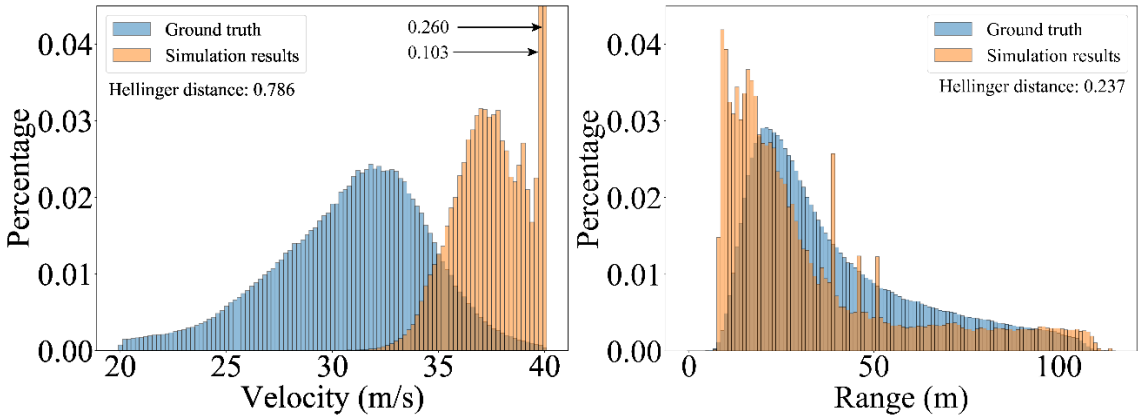


Figure 6 Velocity and range distributions of the NDE using empirical behavior models.

3. Robust NDE modeling

In this section, the robust NDE modeling step is proposed to refine empirical behavior models to minimize the accumulated errors using optimization methods. To achieve this goal, the key is to model the long-term effects of the error accumulation. Specifically, by modeling the vehicle state evolution as a Markov chain, the long-term effects of the error accumulation can be characterized by the stationary distribution of the Markov chain. Then, an optimization problem is formulated to minimize the accumulated errors by adjusting the empirical behavior models, which results in the NDE model with more accurate distributions. In the following paragraphs, we first propose the optimization framework in Section 3.1 and then apply the framework to the longitudinal behavior models in Section 3.2.

3.1 Optimization framework

In order to solve the error accumulation problem, we need to measure the accumulated error of the NDE generated by the empirical behavior models. One possible way is to simulate the NDE, collect the data, and obtain the simulated NDE distribution. However, the computational burden of this method is very heavy since a large number of simulations are needed to obtain an accurate estimation. To overcome this problem, we propose to measure the NDE distribution by analyzing the stationary distribution of the NDE Markov chain. By using this analytical method, the NDE stationary distribution serves as an accurate approximation of the simulated environment, which reflects the performance of the empirical behavior models. By fitting the simulated stationary distribution with real-world ground truth, we can improve the NDE accuracy and solve the error accumulation issue.

Following this idea, the optimization framework can be formulated as in Figure 7. The decision variable is the vehicle behavior model F as shown in Equation 2, and the objective is to minimize the adjustment to the empirical behavior model $F^{\star}$ while ensuring the accuracy of the

stationary distribution. To achieve this objective, there are generally four sets of constraints in the optimization formulation. The first set of constraints is the standard definition of stationary distribution, which indicate that the state will always follow its stationary distribution after reaching the steady state. The second set of constraints describes the relationship between the behavior model and Markov chain state transition probability. The stochastic vehicle behavior model outputs actions for the next time step and therefore determines the state transition process. The third set of constraints is to match the stationary distribution of the simulation with the real-world ground-truth distribution, which is the key to reducing the accumulated errors. As a result, the simulated environment can be guaranteed to fit the desired real-world statistics (e.g., velocity and range distributions) even after a long simulation time horizon. The last set of constraints denotes other standard requirements, such as non-negative constraints of probability mass functions, normalization of stationary distribution, etc.

$$\min_{F} \quad \text{distance}(F, F^{\star})$$

$s.t.$ 1. Definition of the stationary distribution.

2. Relation between behavior model and Markov chain transition probability.

3. Relation between stationary distribution and ground-truth distribution.

4. Other constraints.

Figure 7 Overall formulation of the proposed robust modeling step.

3.2 Optimization of longitudinal behavior models

In this section, we apply the proposed framework to optimize the two empirical longitudinal behavior models as a proof of concept, while keeping the four empirical lateral behavior models unchanged. As the velocities cannot be well modeled by the empirical models as shown in Figure 6, we choose the velocity distribution as the optimization target.

For the free-driving behavior model, we define the discretized speed as the state of the Markov chain. It is easy to find that the Markov chain is finite, irreducible, and aperiodic, so there exists a unique positive stationary distribution π [28] satisfying

$$\pi^T \boldsymbol{P} = \pi^T, \quad (3)$$

$$\sum_{S \in \mathcal{S}} \pi_S = 1, \quad (4)$$

$$\pi \succeq 0, \quad (5)$$

where $\mathbf{P}$ is the state transition probability matrix. As the vehicle longitudinal acceleration depends only on its current speed in the free-driving situation, the state transition probability matrix $\mathbf{P}$ is essentially a function of the behavior model F in Equation 2 as

$$\boldsymbol{P}\left(S_i, S_j\right) = G\left(F\right), \forall S_i, S_j \in \mathcal{S}, \left(6\right)$$

where $\mathrm{G}\left(\cdot\right)$ is a linear mapping from the longitudinal acceleration to the state transition. For example, if the current speed falls in the state S_i, the next speed after the transition is S_j, the time resolution is Δt, and the probability of taking acceleration a that satisfies the $S_j = S_i + a \cdot \Delta t$ is $p(a \mid S_i)$, then $\mathbf{P}\left(\mathrm{S_i}, \mathrm{S_j}\right) = p(a \mid S_i)$. Moreover, as the goal of the optimization is to match the vehicle stationary speed distribution with the real-world speed distribution in the free-driving situation, we have

$$\pi = \pi^{\star}, \left(7\right)$$

where $\pi^{\star}$ is the ground truth of the speed distribution in free-driving situations that is obtained from the large-scale NDD.

Finally, the optimization problem can be formulated as below:

$$\min_{F} \quad \left\| F - F^{\star} \right\|_{Frob} \qquad \left(8\right)$$

$$s.t. \quad \pi^T \boldsymbol{P} = \pi^T, \qquad \left(9\right)$$

$$G\left(F\right) = \boldsymbol{P}\left(S_i, S_j\right), \forall S_i, S_j \in \mathcal{S}, \qquad \left(10\right)$$

$$\pi = \pi^{\star}, \quad \left(11\right)$$

$$\sum_{a \in \mathcal{A}} F\left(a|S\right) = 1, \ \forall S \in \mathcal{S}, \qquad \left(12\right)$$

$$\sum_{S_j \in \mathcal{S}} \boldsymbol{P}\left(S_i, S_j\right) = 1, \ \forall S_i \in \mathcal{S}, \qquad \left(13\right)$$

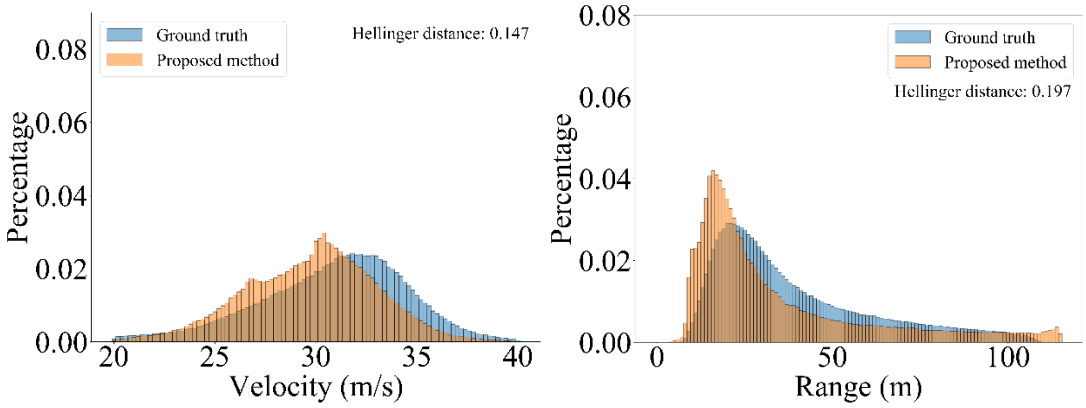
$$\sum\nolimits_{S \in \mathcal{S}} \pi_S = 1, \forall S \in \mathcal{S}, \qquad \left(14\right)$$

$$F, \boldsymbol{P}, \pi \succeq 0. \qquad \left(15\right)$$

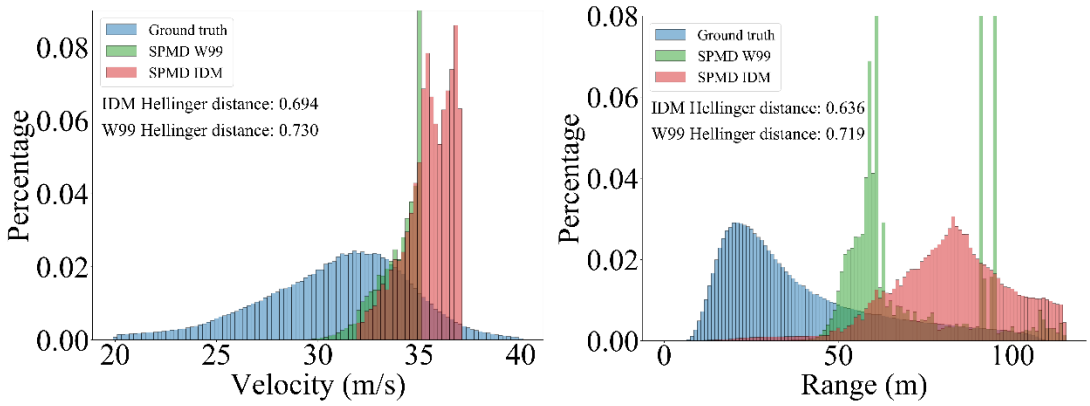
The Frobenius norm $\left\| \cdot \right\|_{Frob}$ is adopted to measure the distance between the optimized free-driving model F and the empirical free-driving model $F^{\star}$. Comparing with the constraints

discussed in Figure 7, Equation 9 denotes the definition of the stationary distribution, Equation 10 denotes the relation between behavior model and transition probability, Equation 11 denotes the distributional consistency between the stationary distribution and the ground-truth, and Equations 12-15 denote other constraints including the normalization requirements for the acceleration probability mass function, state transition probability matrix, and stationary distribution, respectively, and the non-negative requirements. It can be found out that this is a linear programming problem that can be solved efficiently using commercial solvers, for example, Gurobi [29].

For the car-following situation, the vehicle state is composed of the speed of the subject vehicle (v), range (r), and range rate (rr) with the preceding vehicle. The state transition in the car-following situation depends not only on the subject vehicle action but also preceding vehicle action, which makes the optimization problem more complex. To solve this issue, we optimize the steady-state situation of the car-following model, which is a necessary condition regardless of the evolving process to the stationary distribution. As the preceding vehicle has reached the steady-state, the ego-vehicle state transition relies only upon its own action. Then, the optimization problem can be formulated as the same as Equations 8-15, where decision variables are the probability mass functions of the car-following accelerations (F in Equation 2), and Equation 9 is a three-dimensional joint state distribution (i.e., v, r and rr). It is also a linear programming problem that can be solved efficiently.



(a)



(b)

Figure 8 Velocity and range distributions. a, proposed method**. b,** SUMO simulator**.**

4. Simulation performance evaluation

In this section, the performance of the proposed NDE modeling framework is evaluated.

4.1 Distributional consistency

As the distributional consistency is critical for the generated NDE, we first evaluate whether the proposed NDE can generate accurate velocity and range distributions, compared with the existing NDE baseline (i.e., SUMO [30]) and the empirical models constructed in Section 2. Specifically, two existing car-following models, the IDM [21] and the Wiedemann 99 model [31], and SUMO LC2013 lane-changing model [32] are selected as the existing NDE baseline, which are widely applied in existing traffic simulators. For fair comparisons, the model parameters are calibrated with the SPMD NDD described in Section 2.2 using the calibration method developed in [33-34].

Figure 8 shows the results of the proposed NDE model and existing NDE models. It can be found that the proposed NDE model can significantly better reproduce the real-world velocity and range distributions than existing ones. Specifically, both the IDM and Wiedemann 99 models are concentrated in a small interval of velocity and range, while the real-world distributions range among a much wider interval. It is reasonable as these existing models are designed for accident-free purposes and therefore might be more conservative and deterministic.

Table 1 Quantitative performance evaluation for different methods.

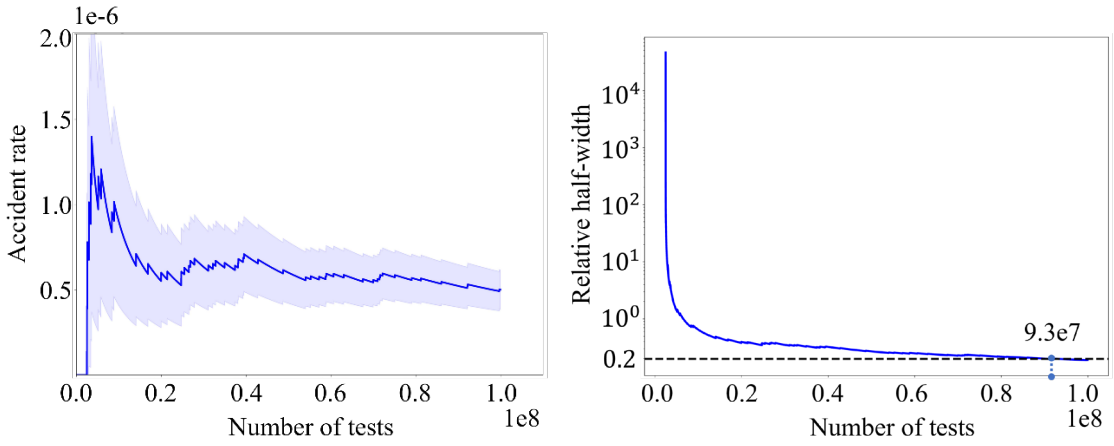
Method \ Metric	Velocity	Range
Empirical behavior models	0.786	0.237
SUMO (SPMD IDM)	0.755	0.659
SUMO (SPMD W99)	0.629	0.691
Proposed method	**0.147**	**0.197**

Compared with the results before the optimization as shown in Figure 6, the accuracy of the simulated velocity distribution is also significantly improved, which validates the effectiveness of the robust modeling step. To further quantify the performance improvement, the Hellinger distances of all these models are calculated as listed in Table 1. Results show that the empirical behavior models can achieve a better performance than the existing NDE models (similar performance in velocity but significantly better performance in range), and the optimized models can achieve the best performance. These results validate the effectiveness of both the data-driven and optimization steps and show that the proposed method can generate the NDE consistent with the real-world driving environment.

In addition to the velocity and range distributions, we also calculated the lane-changing statistic of the proposed NDE to further examine its lateral behavior performance. From the simulation results, the average travel distance for one lane change is 4.86 kilometers. In the real-world driving environment, the same statistic is 4.45 kilometers per lane change on the highway [35]. Therefore, the proposed NDE can also reproduce a reasonable number of lane changes as in the real-world driving environment, which can further demonstrate the fidelity of the proposed NDE.

4.2 AV testing using the proposed NDE

To further demonstrate the capability of the proposed NDE model for AV testing, we test the safety performance of an AV model utilizing the proposed NDE and SUMO (SPMD IDM) environments, respectively. Following previous studies on AV testing [2,7,8,13], the AV accident rate in the NDE is chosen as the measurement, and the Monte Carlo method is applied to estimate the accident rate as shown in Equation 1. Specifically, the IDM and MOBIL models [13] are used as the AV model, one simulation test is conducted for a constant driving distance (400 m) of the AV, and the testing result (accident or not) of each simulation is utilized to calculate the accident rate.



(a) (b)

Figure 9 AV testing and evaluation results using the proposed NDE. a, estimation results of the accident rate (shaded area denotes the 90% confidence interval). b, accident type distribution.

During the simulation tests in the SUMO environment, there is no accident occurred so the estimated AV accident rate is zero. This is because the SUMO environment is designed for accident-free simulations, and therefore, it cannot effectively evaluate the AV safety performance. During the simulation tests in the proposed NDE, however, the estimated accident rate of the AV is 5.5×10^{-5} accidents per simulation, where 276 accidents are generated in the AV testing process. The estimation results with the number of tests can be found in Figure 9a, where the shaded area denotes the 90% confidence interval, and the accident type distribution is shown in Figure 9b based on the definitions of the National Highway Traffic Safety Administration [36]. The results show that the proposed NDE can successfully generate diverse safety-critical situations to evaluate the AV safety performance. With more distributionally accurate NDE models, the safety performance of AVs could be evaluated more effectively.

5. Implementation at American Center for Mobility (ACM)

5.1 Overall introduction

With over 500 acres of variable road systems and customized test environments, American Center for Mobility (ACM) is one of the world's largest test tracks for AVs located in Ypsilanti, Michigan. It includes a 4km long freeway loop, 2.4km long urban arterial, a 700-feet long curved tunnel, six lanes' boulevards, etc. In this project, we utilized the 4km highway loop as the testing environment.

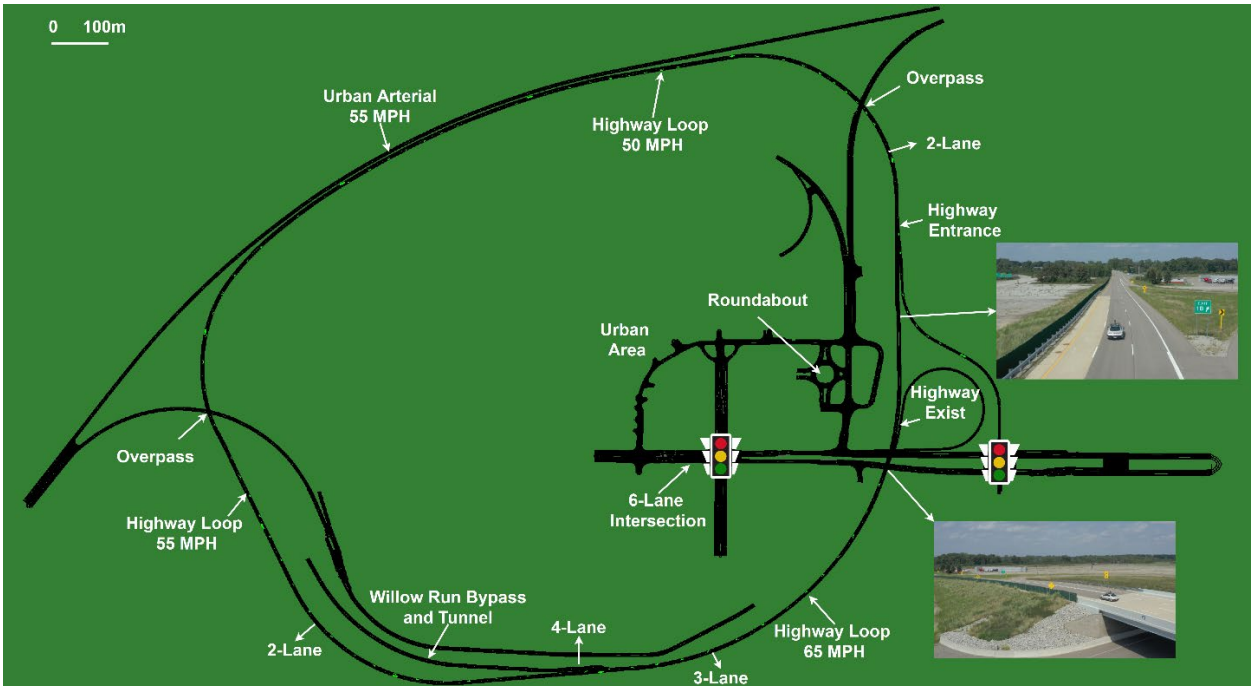


Figure 10 Illustration of the highway test track at the ACM.

To build a high-fidelity simulation environment of ACM, the proposed NDE modeling method is utilized to develop human car-following and lane-changing models. Then, we utilized both the C++ and TRACI interfaces [37] to substitute the default driving models in SUMO [30] simulator. The illustration figure of the generated ACM simulation environment is shown in Figure 10. Then, the AR system [38, 39] is integrated with the NDE simulator, which will synchronize the simulation world and the physical test tracks regarding the information of background vehicles (BVs), autonomous vehicles, traffic signals, high-definition maps, etc., through the TRACI interface. The AR system will send all virtual world information to the real AV so that the AV under test can interact with virtual traffic flows. The detailed system framework is discussed in the next section.

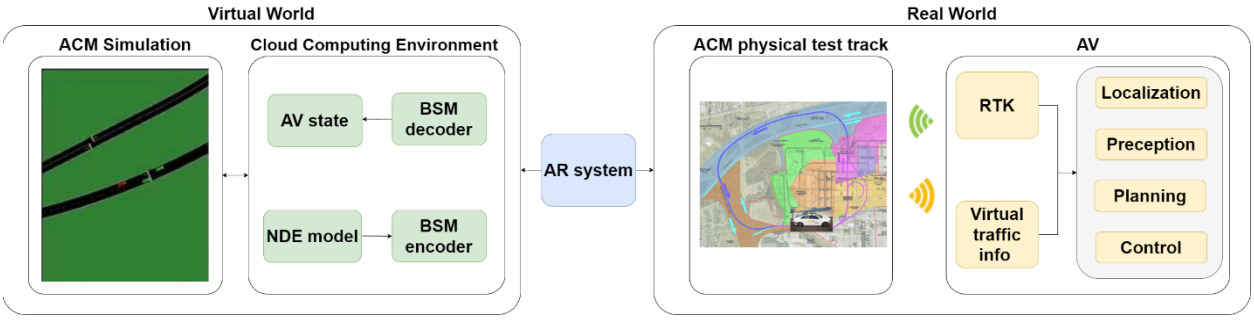


Figure 11 ACM implementation system framework.

5.2 System framework

The system framework is shown in Figure 11. On the left-hand side, it is the virtual world of the testing environment. Virtual background traffic is generated and controlled by our developed NDE models. On the right-hand side, it is the real world in the ACM physical test track. The AR system will communicate and synchronize information between the virtual simulation world and the real world. The virtual BVs information will be broadcasted as Basic Safety Message (BSM) and this information will be received by the AV and input into its software stack. Therefore, the AV under test needs to properly interact with its surrounding virtual BVs and we can evaluate its safety performance based on the testing results. At the same time, the AV state information will also be communicated back to the virtual world so that virtual BVs can react to the AV in the physical test track. As a result, we build the digital twin of ACM environments to enable the testing processes in such closed testing facilities.

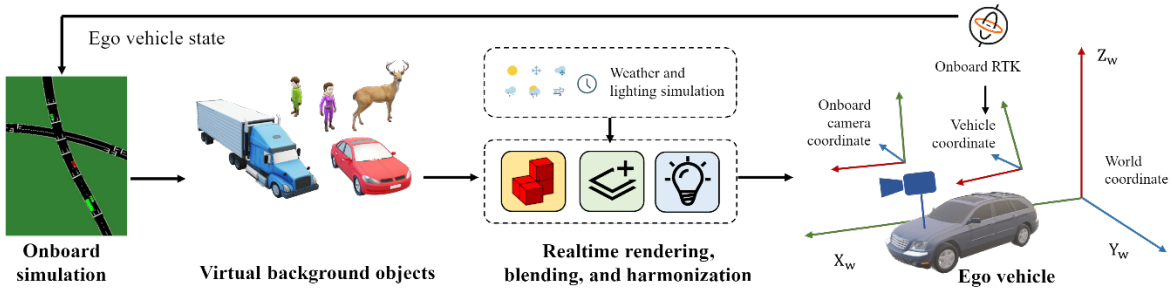


Figure 12 Flowchart of the augmented image rendering.

Moreover, the developed AR system can also render and blend virtual objects (e.g., vehicles) onto the camera view of the AV to generate augmented images (as shown in the lower right part of Figure 13). The illustration figure of the augmented image rendering is shown in Figure 12. A two-stage transformation is applied to project virtual objects to the onboard camera, where the first step is to transform the object from world coordinate to the ego-vehicle (i.e., AV) coordinate, and the second step is from ego-vehicle coordinate to the onboard camera coordinate. During the first step, the ego vehicle pose and location are obtained from the real-time signal of the onboard high-precision RTK. In the second step, the projection is based on the pre-calibrated camera intrinsic and extrinsic. We also perform relighting on the rendered layer to harmonize the visual quality of the blending result. The augmented view is generated based on a linear blending with the rendered foreground layer, camera's background layer, and the rendered alpha matte. On top of the blending result, a weather-control layer is further added to simulate different weather conditions, e.g., rain, snow, and fog.

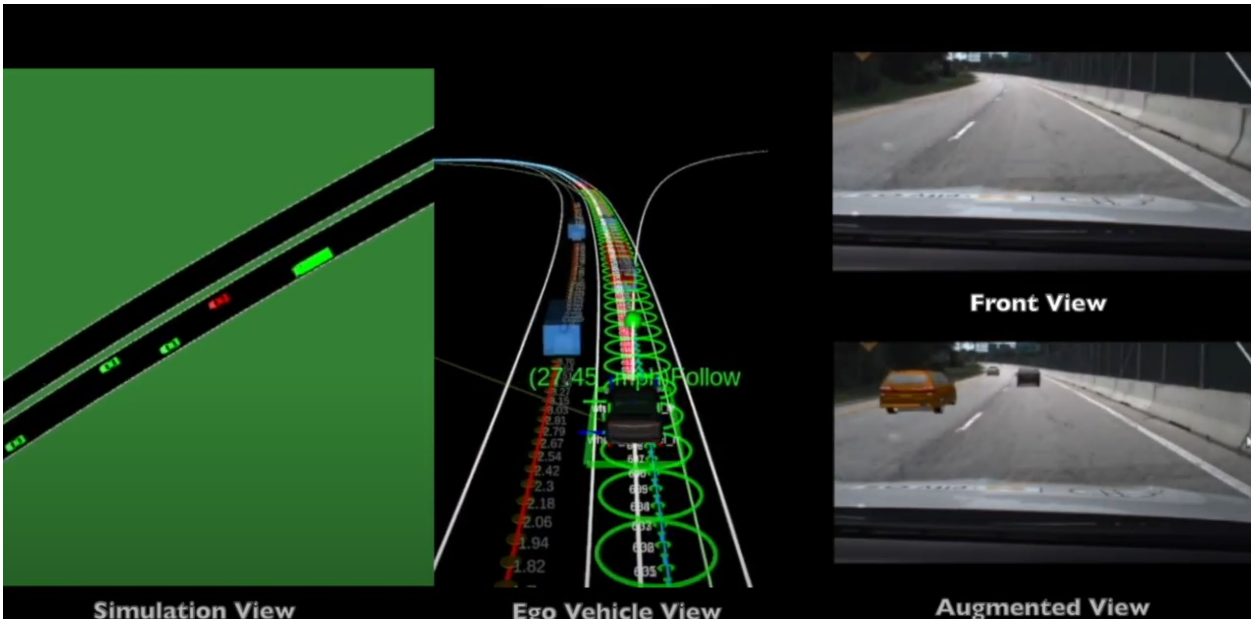


Figure 13 Illustration of the real-time visualization of the testing process.

5.3 Testing demo at ACM

To demonstrate the implemented system, a testing demo is shown in this section. We used a Lincoln MKZ hybrid vehicle as the AV under test, which is equipped with the open-source Autoware [40] system as the AV control agent. A real-time visualization view of the implemented system during the testing process is shown in Figure 13. On the leftmost view, it demonstrates the simulation environment where green vehicles denote simulated virtual traffic and the red vehicle denotes the AV under test. On the middle view, it shows the ego-vehicle view, where Autoware can obtain the surrounding BVs' information and detect them. On the rightmost view, it shows the original front-view camera (upper right) and the augmented image with simulated traffic (lower right). From the results, we can find that the simulated world is well-synchronized with the physical world and the AV under test will be evaluated in realistic traffic conditions.

6. Findings

In this project, we first propose a distributionally consistent NDE modeling framework for AV testing purposes. Then we implement the proposed NDE methodology with the AR system at ACM to build an integrated AV testing framework. To generate a high-fidelity NDE, we first propose to directly generate empirical behavior models using large-scale NDD. These empirical behavior models can serve as basic models since they will converge in probability to the real behavior models when the data size and quality are sufficiently large and accurate. To address the error accumulation issue and guarantee the accuracy of the NDE throughout the simulation, a robust modeling step based on the Markov process is proposed, which optimizes the empirical

models by matching simulated stationary distributions with the ground truth. The proposed method is validated for the multilane highway driving environment using large-scale real-world NDD. The vehicle speed and range distributions generated by the proposed NDE are consistent with the empirical ground truth, which are important for AV testing, and the lane-changing statistic is also examined to be consistent with the ground truth. Moreover, the generated NDE is utilized to test the safety performance of an AV agent, which further validates the effectiveness of the proposed NDE. From the ACM implementation results, we can find that the proposed system can successfully test the AV more safely and efficiently in closed test facilities.

7. Recommendations

One direction for future research is to further increase the fidelity of the proposed NDE. Homogeneous vehicles are considered in this study where all background vehicles share the same behavior models and vehicle classes. In the real-world driving environment, there are different vehicle classes, such as sedan, truck, heavy vehicle, etc. Moreover, within a specific vehicle class, they may have different behavior models, for example, some drivers are more aggressive, and some drivers are more conservative. We leave these for future research.

8. Outputs, Outcome, and Impacts

The proposed integrated solution for autonomous vehicle testing has significant advantages and real-world implementation impacts. By incorporating the NDE model with AR techniques, we can generate a high-fidelity testing environment to evaluate AV safety performance in closed testing facilities. With the proposed NDE model, realistic human driving behaviors can be simulated that will significantly influence the testing accuracy of AV performance. With the AR system, we can test physical AVs more safely and efficiently. To sum up, the proposed system is able to increase the efficiency of AV performance testing, reduce operational costs, and accelerate product validation.

The following outputs were generated during the performance of this project:

- Yan, X., Feng, S., Sun, H., & Liu, H. X. (2021). Distributionally consistent simulation of naturalistic driving environment for autonomous vehicle testing. arXiv preprint arXiv:2101.02828.

References

[1] N. Kalra and S. M. Paddock, "Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability?" Transportation Research Part A: Policy and Practice, vol. 94, pp. 182–193, 2016.

[2] D. Zhao, H. Lam, H. Peng, S. Bao, D. J. LeBlanc, K. Nobukawa, and C. S. Pan, "Accelerated evaluation of automated vehicles safety in lane-change scenarios based on importance sampling techniques," IEEE transactions on intelligent transportation systems, vol. 18, no. 3, pp. 595–607, 2016.

[3] H. Hungar, F. Köster, and J. Mazzega, "Test specifications for highly automated driving functions: Highway pilot," 2017.

[4] M. Koren, S. Alsaif, R. Lee, and M. J. Kochenderfer, "Adaptive stress testing for autonomous vehicles," in 2018 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2018, pp. 1–7.

[5] M. O'Kelly, A. Sinha, H. Namkoong, R. Tedrake, and J. C. Duchi, "Scalable end-to-end autonomous vehicle testing via rare-event simulation," Advances in Neural Information Processing Systems, vol. 31, pp. 9827–9838, 2018.

[6] L. Li, X. Wang, K. Wang, Y. Lin, J. Xin, L. Chen, L. Xu, B. Tian, Y. Ai, J. Wang, D. Cao, Y. Liu, C. Wang, N. Zheng, and F.-Y. Wang, "Parallel testing of vehicle intelligence via virtual-real interaction," Science Robotics, 2019.

[7] S. Feng, Y. Feng, C. Yu, Y. Zhang, and H. X. Liu, "Testing scenario library generation for connected and automated vehicles, part I: Methodology," IEEE Transactions on Intelligent Transportation Systems, 2020.

[8] S. Feng, Y. Feng, H. Sun, S. Bao, Y. Zhang, and H. X. Liu, "Testing scenario library generation for connected and automated vehicles, part II: Case studies," IEEE Transactions on Intelligent Transportation Systems, 2020.

[9] S. Feng, Y. Feng, H. Sun, Y. Zhang, and H. X. Liu, "Testing scenario library generation for connected and automated vehicles: An adaptive framework," IEEE Transactions on Intelligent Transportation Systems, 2020.

[10] S. Feng, Y. Feng, X. Yan, S. Shen, S. Xu, and H. X. Liu, "Safety assessment of highly automated driving systems in test tracks: a new framework," Accident Analysis & Prevention, vol. 144, p. 105664, 2020.

[11] M. Klischat and M. Althoff, "Synthesizing traffic scenarios from formal specifications for testing automated vehicles," in Proc. of the IEEE Intelligent Vehicles Symposium, 2020.

[12] L. Waymo, "Waymo safety report," https://storage.googleapis.com/ sdc-prod/v1/safety-report/2020-09-waymo-safety-report.pdf, 2020.

[13] S. Feng, X. Yan, H. Sun, Y. Feng, and H. X. Liu, "Intelligent driving intelligence test for autonomous vehicles with naturalistic and adversarial environment," Nature communications, vol. 12, no. 1, pp. 1–14, 2021.

[14] H. Sun, S. Feng, X. Yan, and H. X. Liu, "Corner case generation and analysis for safety assessment of autonomous vehicles," Transportation research record, vol. 2675, no. 11, pp. 587–600, 2021.

[15] W. Li, C. Pan, R. Zhang, J. Ren, Y. Ma, J. Fang, F. Yan, Q. Geng, X. Huang, H. Gong et al., "AADS: Augmented autonomous driving simulation using data-driven algorithms," Science Robotics, 2019.

[16] X. B. Peng, M. Andrychowicz, W. Zaremba, and P. Abbeel, "Sim-to-real transfer of robotic control with dynamics randomization," in 2018 IEEE international conference on robotics and automation (ICRA). IEEE, 2018, pp. 3803–3810.

[17] S. James, P. Wohlhart, M. Kalakrishnan, D. Kalashnikov, A. Irpan, J. Ibarz, S. Levine, R. Hadsell, and K. Bousmalis, "Sim-to-real via sim-to-sim: Data-efficient robotic grasping via randomized-to-canonical adaptation networks," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 12 627–12 637.

[18] W. Zhao, J. P. Queralta, and T. Westerlund, "Sim-to-real transfer in deep reinforcement learning for robotics: a survey," in 2020 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE, 2020, pp. 737–744.

[19] S. Manivasagam, S. Wang, K. Wong, W. Zeng, M. Sazanovich, S. Tan, B. Yang, W.-C. Ma, and R. Urtasun, "Lidarsim: Realistic lidar simulation by leveraging the real world," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 11 167– 11 176.

[20] Y. Chen, F. Rong, S. Duggal, S. Wang, X. Yan, S. Manivasagam, S. Xue, E. Yumer, and R. Urtasun, "Geosim: Realistic video simulation via geometry-aware composition for self-driving," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 7230–7240.

[21] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," Physical review E, vol. 62, no. 2, p. 1805, 2000.

[22] A. Kesting, M. Treiber, and D. Helbing, "General lane-changing model MOBIL for car-following models," Transportation Research Record, vol. 1999, no. 1, pp. 86–94, 2007.

[23] A. B. Owen, Monte Carlo theory, methods and examples, 2013.

[24] L. Waymo, "Simulation city: Introducing waymo's most advanced simulation system yet for autonomous driving," https://blog.waymo.com/ 2021/06/SimulationCity.html, 2021.

[25] J. Sayer, D. LeBlanc, S. Bogard, D. Funkhouser, S. Bao, M. L. Buonarosa, A. Blankespoor et al., "Integrated vehicle-based safety systems field operational test: Final program report," United States. Joint Program Office for Intelligent Transportation Systems, Tech. Rep., 2011.

[26] D. Bezzina and J. Sayer, "Safety pilot model deployment: Test conductor team report," Report No. DOT HS, vol. 812, no. 171, p. 18, 2014.

[27] Wikipedia contributors, "Hellinger distance," 2022. [Online]. Available: https://en.wikipedia.org/wiki/Hellinger_distance

[28] G. Grimmett and D. Stirzaker, Probability and random processes. Oxford university press, 2020.

[29] Gurobi Optimization, LLC, "Gurobi Optimizer Reference Manual," 2022. [Online]. Available: https://www.gurobi.com

[30] P. A. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, and E. WieBner, "Microscopic traffic simulation using SUMO," in 2018 21st International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2018, pp. 2575–2582.

[31] VISSIM, "Vissim 5.40-01, User Manual," 2012, planung Transport Verkehr AG, Karlsruhe, Germany.

[32] J. Erdmann, "Sumo lane-changing model," in Modeling Mobility with Open Data. Springer, 2015, pp. 105–123.

[33] B. Hammit, R. James, and M. Ahmed, "A case for online traffic simulation: Systematic procedure to calibrate car-following models using vehicle data," in 2018 21st International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2018, pp. 3785–3790.

[34] B. Hammit, "A case for online traffic simulation: Systematic procedure to calibrate car-following models using vehicle data," https://github.com/ bhammit1/Analysis, 2018.

[35] S. E. Lee, E. C. Olsen, W. W. Wierwille et al., "A comprehensive examination of naturalistic lane-changes," United States. Department of Transportation. National Highway Traffic Safety Administration, Tech. Rep., 2004.

[36] National Highway Traffic Safety Administration, "FARS/CRSS coding and validation manual," Report No. DOT HS, vol. 813, no. 251, 2022. [66] E. de Gelder, J. Manders, C. Grappiolo, J.-P. Paardekooper, O. O. den Camp, and B. D. Schutter, "Real-world scenario mining for the assessment of automated vehicles," 2020.

[37] TRACI: https://sumo.dlr.de/docs/TraCI.html.

[38] Feng, Y., Yu, C., Xu, S., Liu, H.X. and Peng, H., 2018, June. An augmented reality environment for connected and automated vehicle testing and evaluation. In 2018 IEEE Intelligent Vehicles Symposium (IV) (pp. 1549-1554). IEEE.

[39] Feng, S., Feng, Y., Yan, X., Shen, S., Xu, S., & Liu, H. X. Safety assessment of highly automated driving systems in test tracks: a new framework. Accid. Anal. Prev. 144, 105664 (2020).

[40] S. Kato, S. Tokunaga, Y. Maruyama, S. Maeda, M. Hirabayashi, Y. Kitsukawa, A. Monrroy, T. Ando, Y. Fujii, and T. Azumi. Autoware on Board: Enabling Autonomous Vehicles with Embedded Systems. IEEE International Conference on Cyber-Physical Systems (ICCPS) 287-296 (2018).