Saturation Effects in Recorded Maneuver Data for the Test of Automated Driving

Lukas Hartjen*, Robin Philipp*, Fabian Schuldt* and Bernhard Friedrich †

Abstract: One key challenge for a scenario-based verification and validation approach of automated vehicles is the completeness of a scenario set for a specific operational design domain of the system. A possible solution is the use of real-world datasets to argue for the representativity of a given set of scenarios. In order to make these arguments, it is necessary to estimate the amount of data that has to be collected. This contribution investigates the occurrence of saturation effects in the data collection for the test of automated vehicles to estimate the representativity of a set of scenarios. To this end, we analyze a dataset collected in the city of Hamburg with respect to saturation effects in the observed maneuvers.

Keywords: Data Collection, Saturation, Test Driving, Validation

1 Introduction

In order to introduce highly automated driving (HAD) to the market, the safety verification and validation of these systems is a key requirement. Since distance-based validation approaches will likely not be feasible for economic reasons alone [1], scenario-based verification and validation of HAD-Systems are a focus of current research activities [2]. However, it is currently an unsolved question how a finite set of scenarios for the test of an automated vehicle should be constructed. Especially it is a challenge to argue for the completeness and representativity of a scenario set for a given operational design domain.

A possible approach to this argumentation for completeness and representativity is the exposure of the scenario set during real-world traffic recordings. The recognition of semantic traffic behavior based on vehicle and pedestrian maneuvers can facilitate the estimation of scenario exposure [3]. Due to the description through maneuvers, the traffic behavior of objects is abstracted from, becomes differentiable and thereby countable and collectable. However, it is necessary to know the amount of traffic data that has to be collected through real-world driving to make a representative statement about the exposure of different traffic scenarios. Therefore, this publication analyzes saturation effects during the data collection of vehicle maneuvers in urban traffic to estimate the data quantities that are needed for this task.

*Lukas Hartjen, Robin Philipp and Fabian Schuldt are with Volkswagen Group Innovation, Wolfsburg, Germany, lukas.hartjen@volkswagen.de

†Bernhard Friedrich is a Professor at the Institute of Transportation and Urban Engineering, Technische Universität Braunschweig, Brunswick, Germany

2 Related Work

This work is following a suggestion by Wachenfeld et al. [4], that test drives with automated vehicles are not only useful to test the system functionality directly. Rather, the data collected on these drives can also be applied to model traffic behavior and assemble a set of scenarios to test the automated vehicle in simulation or proving ground tests.

The assembly of a set of scenarios is also addressed by Hauer et al. [5]. They propose to model the search for new scenario types as a coupon collector's problem to estimate the likelihood of detecting previously unseen scenario types in new collected measurement data. These abstract scenario types correspond to the *Functional Scenarios* introduced by Menzel et al. [6]. In this paper, we investigate new maneuver combinations as a part of these *Functional Scenarios* to facilitate this and other approaches to evaluate the completeness of scenario catalogues.

Amersbach and Winner [7] propose a method to calculate the number of concrete scenarios required for the validation of automated vehicles. Based on the work of Langner et al. [8], they assume the share of unique scenarios to be one fifth of the overall amount of observed scenarios. Accordingly, they estimate the amount of scenarios required for the validation to be $n_{req} \approx 1.2 \cdot 10^{10}$. In our work, we investigate the occurence of new unique scenarios in measurement data by looking at the occurrence of abstract vehicle maneuvers. The findings could be incorporated into calculations such as the ones proposed by Amersbach and Winner [7].

Mauritz et al. [9] introduce a testing strategy for a lane change assistant based on an abstract semantic domain description. They demonstrate how it is possible to estimate the achieved test coverage over time using this semantic description. A transfer of this methodology to pedestrian traffic and the test of automated driving systems is described by Hartjen et al. [3], but did not include the semantic analysis of vehicle movements. While the approach proved helpful in identifying semantically redundant elements in recorded test drives, saturation behavior could not be observed due to the limited size of the analyzed data.

A framework to model driving maneuvers for the generation of test scenarios is presented by Krajewski et al. [10]. Their approach uses *Generative Adversarial Networks (GANs)* to synthesize new trajectories from a set of previously recorded maneuver instances. While this is an important addition to our work in this paper, the question of how much data should be collected is not addressed to the best of our knowledge.

An estimation of the necessary amount of Naturalistic Driving Data (NDD) for traffic modeling is presented by Wang et al. [11]. In their work, they apply the Kullback-Leibler (KL) Divergence [12] to measure the change in estimated probability density functions when using more data to compute them. While the focus of their work is the analysis of longitudinal driving behavior, the general methodology seems to be transferable to other applications as well. In this work, their presented methodology is transferred to analyze saturation effects in the collection of semantic traffic data for the test of automated vehicles. The usage of the KL-divergence measure is discussed in detail in the later part of this publication.

Despite an emerging consensus in the scientific community that scenario-based verification and validation approaches for automated driving will play an important role, it remains an unsolved challenge how to assemble a representative set of scenarios. In order to utilize real-world data collected during test drives, it is of interest to estimate the amount of data that has to be collected to make informed statements about the surrounding traffic of the automated vehicle. To the best of our knowledge, this area of research has not been widely investigated for the purpose of verifying automated vehicles. In this work, we therefore address two research questions that will be discussed in the following section.

3 Research Questions and Hypotheses

Based on the analyzed related work, the following research questions were identified to quantify the amount of measurement data that should be collected to create a representative set of scenarios for scenario-based testing of urban automated driving functions.

Research Question 1 *How much data is needed to achieve saturation with respect to the behavior of objects?*

In this contribution, the behavior of movable objects is defined as sequences of maneuvers [13]. To answer this research question, we will look at the parametric description of individual maneuvers on one side and at their occurrence in the form of maneuver sequences on the other side. A dataset is considered to be saturated in this respect when no previously unseen maneuver sequences are detected through additional data collection efforts. The amount of data necessary to observe these saturation effects is analyzed in this work.

Research Question 2 *Is it possible to identify commonly occurring Functional Scenarios by means of semantic classification?*

Since identifying all possible *Functional Scenarios* (cf. [6]) in recorded measurement data is likely going to be very time and cost expensive, it could be of great practical interest to determine common scenarios. These scenarios would be likely to occur during the field operation of an automated vehicle. Thereby, a first goal of the verification and validation process of the system could then be to establish the safe behavior of the system in these common *Functional Scenarios*, for example in regression tests of new software revisions. In this research question, it will be investigated if and how the classification of abstract maneuvers in recorded measurement data can aid the process of identifying common *Functional Scenarios*.

4 Methodology

To investigate the aforementioned research questions, data was collected on test drives with automated vehicles in the city of Hamburg. The driven route is shown in Figure 1. Our dataset consists of approximately 3 hours of urban driving, in which the vehicles travelled roughly 50.6 *km*. In this database, we classified 179 488 maneuvers that were executed by surrounding objects.

In an earlier publication [13], we introduced a catalog of vehicle maneuvers, as well as the different layers of semantic analysis for urban traffic that will be the foundation of of this work. The layers are shown in Figure 2. For the purpose of this work, classifiers

Figure 1: Routes driven for the collection of maneuver data in the city of Hamburg, Germany. Map data by OpenStreetMap [14]

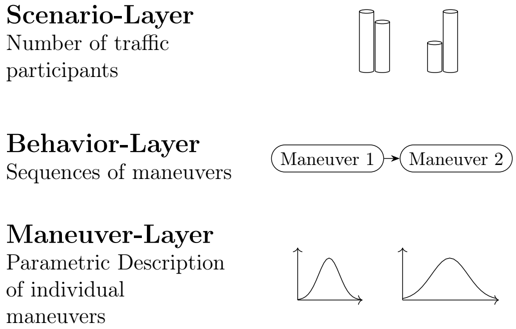

Figure 2: Layers of semantic analysis for traffic participants in urban traffic scenarios, extended from [13]

were implemented to detect all maneuvers of the catalog [13] in measurement data with the exception of *Follow Lane* and *Park* maneuvers. The following subsections describe the applied methodologies, as well as the preliminary results.

4.1 Saturation Analysis

With respect to *Research Question 1*, the collected data is analyzed on the *Maneuver Layer* and the *Behavior Layer* with respect to the occurrence of saturation effects.

4.1.1 Maneuver Layer

On the lowest layer of analysis, the execution of individual maneuvers is examined. An example is the velocity profile or the trajectory curvature during *Turn Left* maneuvers at intersections. For the purpose of this work, the analysis on the *Maneuver Layer* will be limited to investigate one exemplary maneuver parameter, the initial velocity v_0 observed in *Turn Left* maneuvers. Apart from being an illustrative example, this parameter could also be of practical interest to specify the initial state of a vehicle in a simulation scenario, even if the remaining movement of the object is determined by a traffic model.

In order to describe the observed execution of the maneuvers in the database in a uniform manner, the object movements are modeled by Beziér curves of third order [13]. Three curves describe the spatial movement as well as the observed velocity profile. Figure 3 depicts the first parameter of the Beziér curve modeling the velocity v_0 during classified *Turn Left* maneuvers in the analyzed database. After the trajectory of each recorded maneuver of the chosen type has been converted to this local coordinate system, the Bézier points are calculated. By combining the Bézier representations of all observed maneuver instances, the underlying probability density function *(pdf)* for v_0 is estimated using kernel density estimation *(kde)* with Scott's rule [15] for bandwidth selection.

Following Wang et al. [11], saturation effects regarding the amount of collected maneuver samples are investigated by repeating the process of *kde* for different sample sizes. To this end, 125 instances of *Turn Left* maneuvers are sampled to 30 different sets of

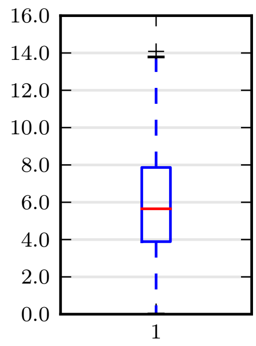

Figure 3: Recorded initial velocities during 125 *Turn Left* maneuvers

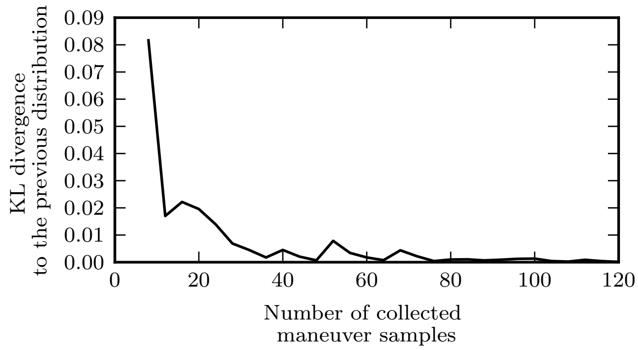

Figure 4: Kullback-Leibler divergence to the previous distribution of initial velocities during *Turn Left* maneuvers over the amount of maneuver samples used for kernel density estimation

equidistant size. The obtained *pdfs* are then compared using the Kullback-Leibler [12] divergence to the next smaller sample set. The results are shown in Figure 4.

4.1.2 Behavior Layer Analysis

To analyse saturation effects on the *Behavior Layer*, the amount of observed unique maneuver sequences is plotted over the cumulated number of recorded objects. Since multiple maneuvers can be executed at each point in time, we introduce the concept of a *Paralell Maneuver Combination (PMC)*. A *PMC* captures all the maneuvers that an object is executing at a defined point in time (cf. Figure 5). A *Parallel Maneuver*

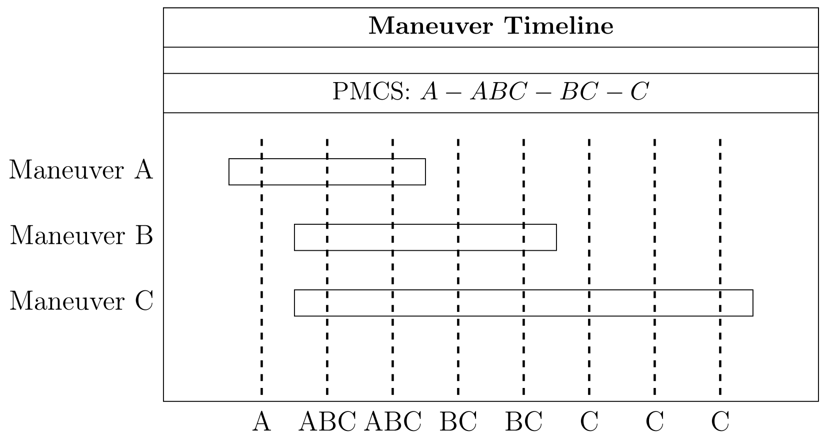

Figure 5: Abstract example for the notation of an object's *Parallel Maneuver Combination Sequence (PMCS)* from its maneuver timeline

Combination Sequence (PMCS) is subsequently defined as the chronological development of an object's *PMC* state. A *PMCS* is constructed by identifying the *PMC* at each discrete point in time and then appending it to the sequence if it differs from the previous *PMC*. An abstract example for the construction of a *PMCS* can be found in Figure 5. In the beginning, the object executes only the abstract maneuver *A*. Next, it simultaneously

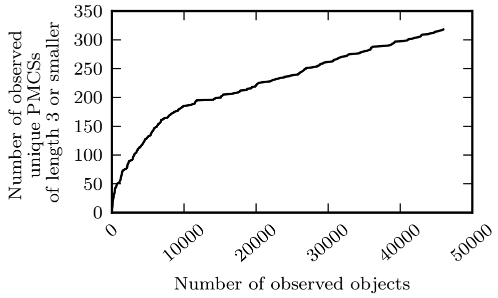

Figure 6: Amount of observed unique *Parallel Maneuver Combination Sequences (PMCSs)* of length 3 or smaller over the total number of observed objects

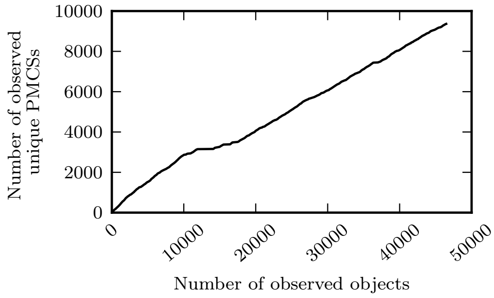

Figure 7: Amount of observed unique *Parallel Maneuver Combination Sequences (PMCSs)* over the total number of observed objects

starts to execute the maneuvers B and C, changing its current *PMC* from A to ABC and adding a new entry to its *PMCS*. After the execution of maneuver A terminates, the current *PMC* is now BC, adding another entry to the *PMCS* which is followed by C after the execution of maneuver B terminates as well. Note that repeated *PMC* samples are not duplicated in the *PMCS* since it reflects only changes in the object's current maneuver constellation. This constitutes an abstraction by neglecting the temporal extension of individual *PMC* states in the sequence.

Once the *PMCS* has been constructed for every object in the recorded database, the occurrence of new unique sequences is analyzed regarding saturation effects. For every *PMCS*, it is checked whether or not an identical *PMCS* was already observed for another object in the database. If not, it is added to the set of known sequences and the respective count is incremented by one. Figure 6 shows the development of the number of observed *PMCS* of length three or shorter over the time driven, while Figure 7 shows the development for all sequences.

4.2 Common Scenarios

To answer *Research Question 2*, common patterns of maneuver sequences are identified and their frequencies of occurrence are compared. This way, commonly occurring elements of scenario layer 4 [16] can be found and subsequently be turned into *Functional Scenarios*[6] for simulation or proving ground tests. To this end, Figure 8 shows the 15 *Parallel Maneuver Combination Sequences* with the highest number of occurrences in the database.

5 Discussion

In the first research question of this paper, the goal is formulated to estimate the amount of recorded data after which repetitions in the observed maneuver data make the collection of more data inefficient. When looking at the analysis of the *Maneuver Layer*, a qualitative evaluation of the results in Figure 4 shows the occurrence of saturation behavior with

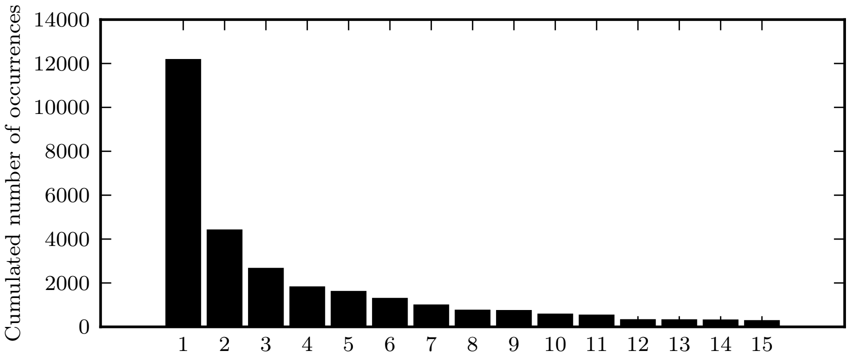

Figure 8: 15 *Parallel Maneuver Combination Sequences (PMCS)* with the highest number of occurrences in the database

respect to the distribution of initial velocities v_0 during *Turn Left* maneuvers. As shown by Wang et al. [11], the detection of saturation can be quantified by introducing a threshold parameter ϵ. The dataset is then considered saturated if none of the remaining KL-divergence values exceeds ϵ. Therefore a discussion should be held if and for what value of ϵ this kind of saturation effect should be incorporated into the overall validation process.

While saturation behavior was observable for the investigated example of initial velocities during *Turn Left* maneuvers after around 80 samples, this does not necessarily apply to other parameters or other maneuvers for that matter. The introduced methodology could be applied to a wider range of parameters to research similar effects for these maneuver properties.

Analysis on the *Behavior Level* showed no significant saturation effects in the data quantities that were investigated in this paper. In fact, a linear regression for the number of unique *PMCS* n_{PMCS} over the amount of observed objects $n_{objects}$ yields $R^2 \approx 0.99$, indicating a nearly linear relationship between the two variables. This absence of saturation effects in the observed behavior could have multiple causes. On the one hand, it is possible that the quantity of analyzed data was simply not sufficient. Since we are not looking at entire scenarios, but rather at individual object behavior as an element of scenario layer 4 [17], it is difficult to compare the obtained results to estimations of the amount of required scenarios such as by Amersbach and Winner [7] to assess plausibility. Another possibility would be that the concept of *PMCS* is not well suited to compare object behavior. This cannot be ruled out but is viewed to be unlikely since it introduces a significant abstraction from the actual concrete behavior. Therefore an overestimation of the actual traffic complexity should not take place in our current opinion. Nevertheless, our approach manages to quantify the variability of the underlying dataset by differentiating the sequences of abstract maneuvers from each other. Thereby, we hope to further contribute to the understanding of large traffic datasets with this approach, which could not only be applicable for the verification and validation of automated vehicles, but also in training data selection for machine learning algorithms.

Another important factor influencing the saturation results are the classification algorithms for the individual maneuvers. Since recognizing them reliably in noisy measurement data is not a trivial task, the output can generally include false positive as well as false negative detections. The classification algorithms written in the context of this publication are considered to be experimental software that could be subject to bugs. Therefore, the results should be treated with care, as future improvements could also influence the statistical findings discussed in this paper. After manual inspection, we do however believe the general trend of our results to remain, even if the implementation of the classification algorithms could change in the future.

Looking at the second research question concerning the identification of commonly occurring scenarios, only a partial answer can be given. By looking at the occurrence of different *PMCS* in Figure 8, common elements of scenario layer 4, the movable objects, can be identified. For example, a common sequence that was identified this way was the combination of a *Cross Junction* maneuver with a simultaneous *Keep Velocity* maneuver. The most common sequence was a single *Standstill* maneuver. This could be an indication to increase testing efforts with respect to this behavior, since there seems to be a high exposure to it, at least in the dataset that was analyzed in this work.

These analyzed sequences of maneuvers only represent one element of scenario layer 4, the movable objects. They can however be used as building blocks in the construction of test scenarios. For example, a strategy could be to construct scenarios that include one of the most occurring *PMCS* to test the automated vehicle's reaction to this common behavior.

6 Conclusion and Future Work

In this paper, we have investigated the occurrence of saturation effects in collected measurement data for the verification and validation of automated driving. While our analysis of individual maneuver parameters indicated saturation behavior for the chosen exemplary maneuver parameter, more research is necessary to determine the applicability of the obtained quantitative statements for other parameters and a wider range of maneuvers. The investigation should also be extended to include parameter correlations to more accurately reflect the complex nature of urban traffic. On the *Behavior Layer*, saturation of the observed maneuver sequences could not be determined. Future work will likely focus on investigating larger quantities of data. It was, however, possible to determine commonly occurring elements of *Functional Scenarios* for the analyzed test drives in the form of *Parallel Maneuver Combination Sequences (PMCS)*. These could be a valid starting point to further refine and construct a data-driven set of scenarios for the test of urban automated vehicles.

Finally, the effect of route choice on the qualitative and quantitative structure of the classified maneuver data will be analyzed in the future to further facilitate the practical application of the proposed methodology in the verification and validation process of automated vehicles.

Acknowledgment

We would like to thank Sven Chlosta, Stefan Diddens, Ting Fu, Hansang Lee, Asis Das, Hedan Qian and Zhijing Zhu for their help in implementing the proposed concepts.

References

[1] H. Winner, W. Wachenfeld, and P. Junietz, "(How) can safety of automated driving be validated?" 2016, presentation at Virtual Vehicle Symposium Graz. [Online]. Available: https://www.fzd.tu-darmstadt.de/media/fachgebiet_fzd/publikationen_3/2016_5/2016_Wi_Wf_Ju_ViV-Symposium_Graz.pdf

[2] "Pegasus Research Project," https://www.pegasusprojekt.de/en/home, accessed: 2019-04-05.

[3] L. Hartjen, F. Schuldt, and B. Friedrich, "Semantic classification of pedestrian traffic scenarios for the validation of automated driving," in *2019 22nd IEEE International Conference on Intelligent Transportation Systems (ITSC)*. Auckland, New Zealand: IEEE, 2019, pp. 3696–3701.

[4] W. Wachenfeld and H. Winner, "Die Freigabe des autonomen Fahrens," in *Autonomes Fahren*, M. Maurer, J. C. Gerdes, B. Lenz, and H. Winner, Eds. Springer Vieweg, 2015, pp. 439–464.

[5] F. Hauer, T. Schmidt, B. Holzmüller, and A. Pretschner, "Did we test all scenarios for automated and autonomous driving systems?" in *2019 22nd IEEE International Conference on Intelligent Transportation Systems (ITSC)*. Auckland, New Zealand: IEEE, 2019, pp. 2950–2955.

[6] T. Menzel, G. Bagschik, and M. Maurer, "Scenarios for development, test and validation of automated vehicles," in *2018 IEEE Intelligent Vehicles Symposium (IV)*. Changshu, Suzhou, China: IEEE, 2018, pp. 1821–1827.

[7] C. Amersbach and H. Winner, "Defining required and feasible test coverage for scenario-based validation of highly automated vehicles," in *2019 22nd IEEE International Conference on Intelligent Transportation Systems (ITSC)*. Auckland, New Zealand: IEEE, 2019, pp. 425–430.

[8] J. Langner, J. Bach, L. Ries, S. Otten, M. Holzapfel, and E. Sax, "Estimating the uniqueness of test scenarios derived from recorded real-world-driving-data using autoencoders," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, Changshu, Suzhou, China, 2018, pp. 1860–1866.

[9] M. Mauritz, F. Howar, and A. Rausch, "Assuring the safety of advanced driver assistance systems through a combination of simulation and runtime monitoring," in *Leveraging Applications of Formal Methods, Verification and Validation: Discussion, Dissemination, Applications*. Corfu, Greece: Springer International Publishing, 2016, pp. 672–687.

[10] R. Krajewski, T. Moers, D. Nerger, and L. Eckstein, "Data-driven maneuver modeling using generative adversarial networks and variational autoencoders for safety validation of highly automated vehicles," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, Maui, Hawaii, USA, 2018, pp. 2383–2390.

[11] W. Wang, C. Liu, and D. Zhao, "How much data are enough? a statistical approach with case study on longitudinal driving behavior," *IEEE Transactions on Intelligent Vehicles*, vol. 2, no. 2, pp. 85–98, June 2017.

[12] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[13] L. Hartjen, R. Philipp, F. Schuldt, F. Howar, and B. Friedrich, "Classification of driving maneuvers in urban traffic for parametrization of test scenarios," accepted for publication in 2019 Tagung Automatisiertes Fahren, Munich, Germany, 2019.

[14] OpenStreetMap contributors, "Planet dump retrieved from https://planet.osm.org ," https://www.openstreetmap.org, 2019.

[15] D. W. Scott, "On optimal and data-based histograms," *Biometrika*, vol. 66, no. 3, pp. 605–610, 12 1979. [Online]. Available: https://doi.org/10.1093/biomet/66.3.605

[16] F. Schuldt, "Ein Beitrag für den methodischen Test von automatisierten Fahrfunktionen mit Hilfe von virtuellen Umgebungen - english title: Towards testing of automated driving functions in virtual driving environments," Ph.D. dissertation, Technische Universität Braunschweig, 2017.

[17] G. Bagschik, T. Menzel, and M. Maurer, "Ontology based scene creation for the development of automated vehicles," in *2018 IEEE Intelligent Vehicles Symposium (IV), Changshu, Suzhou, China*, 2018, pp. 1813–1820.