Scenario Based Testing of Automated Driving Systems: A Literature Survey

Conference Paper · October 2020

DOI: 10.46720/f2020-acm-096

6 authors, including:

Đemin Nalić
MQS Automotive GmbH

Tomislav Mihalj
Graz University of Technology

Maximilian Baeumler
Technische Universität Dresden

Matthias Lehmann
Technische Universität Dresden

F2020-ACM-096

SCENARIO BASED TESTING OF AUTOMATED DRIVING SYSTEMS: A LITERATURE SURVEY

Demin Nalic[1], Tomislav Mihalj[2], Maximilian Bäumler[3], Matthias Lehmann[4], Arno Eichberger[5], and Stefan Bernsteiner[6]

[1,2,5]*Technische Universität Graz*
Rechbauerstraße 12, 8010 Graz, Austria (E-mail: demin.nalic@tugraz.at, tomislav.mihalj@tugraz.at, arno.eichberger@tugraz.at)
[3,4]*Technische Universität Dresden*
George-Bähr-Straße 1b, 01069 Dresden, Germany (E-mail: maximilian.baeumler@tu-dresden.de, matthias.lehmann7@tu-dresden.de)
[6]*MAGNA Steyr Fahrzeugtechnik AG & CO KG*
Libenauerhauptstraße 317, 8041 Graz, Austria (E-mail: stefan.bernsteiner@magna.com)

ABSTRACT: Market introduction of automated driving features several motivations including road safety, driving comfort, energy efficiency and totally new transport systems. However, many challenges are blocking it, including performance of the perception system, safety validation, legal and ethical issues, Human-Machine interaction and others. Especially the safety validation of SAE Level 3-5 systems in complex environments with respect to road and weather conditions call for totally new approaches and processes. Scenario based methods for testing and validation of automated driving systems (ADS) in virtual test environments are gaining importance and becoming an essential component for verification and validation processes of ADS. The high system complexity and costs of real testing lead to an exponential increase of test efforts for real world testing. Using scenario and simulation based approaches this effort can be efficiently reduced with respect to costs and time. Research has shown that it is necessary to drive and test billions of kilometers to ensure safety of ADS which would not be rational considering the time and cost effort for real testing. The biggest challenges are the selection of a suitable simulation framework and the selection of relevant scenarios for the system under test. The literature reports different strategies and approaches for generating relevant scenarios for testing of ADS. All of them have their advantages and disadvantages related to the used environment, vehicle, traffic models and integration complexity. This paper presents a survey through different approaches and methods for scenario generation and evaluation for testing and validation of ADS. It reviews a total number of 86 different papers, most of them published recently. It proposes a terminology and classification scheme of the different methods for scenario generation but also for the related assessment criteria. The reader should get a thorough state of the art overview on scenario based verification and validation approaches of ADS.

KEY WORDS: Automated Driving, Scenario Generation, Assessment, Testing

1. Introduction

Today, as a consequence of digitalization and the related disruptive technologies, mobility is under-going revolutionary changes. The main aspects are alternative power trains targeting lower CO2 emission, new mobility concepts (e.g. shared mobility) and automated driving. The introduction of automated driving is mainly motivated by a desire to:

- Improve traffic safety with a target of zero fatalities;
- Improve occupant comfort and enable use of travel time for other activities;
- Improve CO2 emissions via predictive and cooperative driving;
- Develop sustainable mobility for elderly and dis-

abled people;
- Develop sustainable mobility that is compatible with growing urbanization.

Despite enthusiastic press releases, the development of automated driving at SAE 3 automation level and upwards faces many challenging issues, related to a complex cyber-physical system that could endanger human life. These issues include:

- Occupant acceptance including driver and passengers;
- Testing and validation of reliable automated driving;
- Reliable perception of the driving environment including scenario interpretation;
- Complexity of traffic scenarios and environmental conditions;
- Regulatory and ethical issues;
- Road capacities especially in mixed traffic;
- Irony of automation and automation surprises;
- Social acceptance;
- Cybersecurity including manipulation and misuse.

For automated driving systems (ADS) of all SAE levels, intensive testing and validation have to be carried out systematically in a persistent development process within its specifications. This process ranges from simulation, to test benches with different complexity, to on-road testing. The literature listed in this paper has been identified through a systematic review process, mainly in the subject databases Scopus (Elsevier B.V) and IEEE Xplore, and through a keyword-based Google Scholar search. With a focus on current literature, more than 900 search results were reviewed on the basis of title, abstract and full text, out of which about 10% were included in this paper. The paper is structured as follows: At first the terms used in the literature are defined in Section 2 , Section 3 deals with scenario based methods, Section 4 deals with metrics needed for the assessment and validation of ADS described Section 5.

2. Terminology

This section defines the most relevant terms and tries to give a short classification.

2.1. Basic definitions

The terms scenario, scenario-based approach and test scenario catalog are defined below.

Scenario

A scenario is "*the temporal development between several scenes in a sequence of scenes*" according to [3], based on the work of [4] and [5]. Thus, scenarios describe the chronological sequence of still images represented by scenes and can be enriched by actions and events (e.g. overtaking maneuvers).

Scenario-based approach

As [2] already pointed out in 2016, ADS can no longer be validated conventionally by test drives. This problem will be solved by the scenario-based approach, also called "database approach" [6]. The aim is to compress the test drive into scenarios and thus remove negligible sections without an action or event (e.g. monotonous straight driving) from the validation procedure. Sippl et al. demonstrated how scenario-based testing can already be used throughout the entire development process of an ADS [7].

Test scenario catalogues

Test scenario catalogues [8]-[9] or test scenario databases [10] represent a selected collection of test scenarios for the validation process.

2.2. Scenario Classification

Scenarios can be divided into many categories, which is why the following will attempt to classify and briefly describe them.

Level of detail

Menzel et al. divide scenarios into functional, logical and concrete scenarios according to their level of detail [11]. While functional scenarios (see Figure 1) can be described linguistically, logical scenarios already contain parameter spaces and concrete scenarios exact parameter values. This systematically reduces the degree of abstraction and increases the number of possible scenarios from functional to concrete scenarios.

Level of observation

The level of observation can be best described by a bird's-eye-view of the scenario and its content. Rodarius et al. distinguish between driving and traffic scenarios [12]. While driving scenarios contain just a few vehicles, traffic scenarios operate in a broader context, including a variety of driving scenarios. An examplary traffic scenario is driving in a highway section. Additionally, [13] introduces scenario classes, which contain similar scenarios like e.g. "urban scenarios". Thereby, a scenario can be assigned to different classes.

Level of information

Since a scene contains, among other things, dynamic elements, the scenery and information about actors and observers, [14] introduced five layers to describe the necessary information. According to [15] the five basic layers can be supplemented by a sixth layer digital information":

- **Layer 1:** Road level information (e.g. geometry)
- **Layer 2:** Traffic infrastructure (e.g. traffic signs)
- **Layer 3:** Temporaral modifications (e.g. construction sites)
- **Layer 4:** Objects (e.g. dynamic objects)

- **Layer 5:** Environment (e.g. weather)
- **Layer 6:** Digital information (e.g. V2X communication)

Level of freedom
Especially in simulative validation, e.g. performed with a driver-behavior-model-based stochastic traffic simulation [16], the behavior and reaction of all introduced actors and observers is important. Therefore, [17] introduces a distinction between static, dynamic and hybrid scenarios. In static scenarios all trajectories except those of the ADS under test are predefined. Dynamic scenarios are the opposite of static ones, where reactions to the ADS under test are desired (e.g. evasion). Hybrid scenarios do have predefined trajectories until a fixed time step, from which on reactions are allowed.

Level of risk
Scenarios can represent different risk levels in their overall course. Lehmann et al. define normal driving, critical and accident scenarios [9], whereas [18] labels the same categories as typical, critical and edge case scenarios. Hereby, normal driving scenarios have a high probability of occurence and a low risk. In contrast, accident scenarios are very risky, but have a low probability of occurence. Critical scenarios are in between, whereby a critical scenario can end up in an accident.

Level of representativeness
According to [8] and [9] scenario catalogues should be representative in order to achieve valid coverage. In this sense of representativeness, the test scenario catalogues should be able to replicate the real population (spatial, temporal, objective) of the considered road traffic as accurately as possible. However, a test scenario catalogue can only be representative if the containing test scenarios are also representative, which is why a distinction must be made between representative and non-representative test scenarios. The term representative scenario is also used in [19] to denote test scenarios that represent a subset of scenarios. Thus, representativeness in [8] and [9] is used in relation to a superordinate population, while [19] uses this for a scenario representant derived from a scenario catalogue.

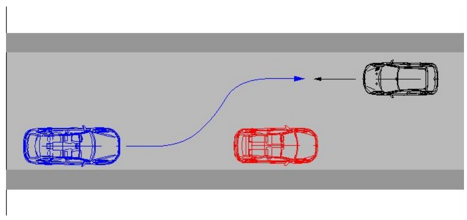

Figure 1. Exemplary functional evasive-scenario.

3. Scenario-Based Methods

This section introduces a scenario-based test concept for testing and validation of ADS ([20], [21], [22], [23]). The concept can be divided into acquisition phase, generation of logical scenarios, generation of concrete scenarios, test execution and evaluation. An acquisition phase is responsible for the collection and organization of knowledge and data obtained from different sources. It is followed by parametrization process also known as the generation of logical scenarios. Through this process, the parameter space is assigned to each parameter of a scenario. Once the logical scenarios are defined, the concrete scenarios can be generated. Those first steps are normally derived before the execution step, whereby the following issues shall be considered: Coverage of different types of scenarios [24] and coverage of variation of each scenario where assigned parameter spaces come into play.
The test execution and evaluation are usually iterative using concrete scenarios as input for the optimization or search algorithms to identify sufficient amount of relevant scenarios to fulfil research goals.

3.1. Generation of Concrete Scenarios

The test concept introduced earlier consists of two phases. The first one deals with the off-line approach of generating concrete scenarios. The second one deals with the on-line identification of relevant scenarios through test execution and evaluation. In this section, only the first stage is considered involving the generation of concrete scenarios.
Currently, there are two recognized approaches to generate logical scenarios, knowledge-driven and data-driven. The main difference between those two is that in the data-driven approach, the identified scenarios are the result of clustering of measurement data, while in the knowledge-driven approach experts are those that explicitly define functional and logical scenarios. It shall be clearly stated that neither of the approaches excludes expert's knowledge or support of data and measurements.
Those two approaches highly depend on the available data sources and it would be wrong to compare them in order to distinguish which approach is better. On the one hand, the knowledge-based scenarios could lack real-life representation if they are not inspired by measurements or field data. On the other hand, the data-driven approach could perform poorly in covering high diversity of scenarios if they are limited only on performed field tests and few databases. Maybe the biggest benefit of using a data-driven approach is its trustworthiness, while in

the case of a knowledge-driven approach it is its easy understandable and systematic concept of scenarios in a human-readable way that does not require extensive data analytic experience. Therefore combining those approaches and preparation of frameworks to handle both could be promising [25].

Depending on available data, the distribution of the parameters in logical scenarios can be defined as continuous or discrete which influences the selection of the method to derive concrete scenarios. Often combinatorial methods ([20], [21], [26], [27], [28]) are used to derive concrete scenarios from discrete domains while sampling methods ([22], [29]) are normally used in a case of Monte-Carlo simulation.

In the following sections, we will separately review the literature regarding knowledge-driven and data-driven approach.

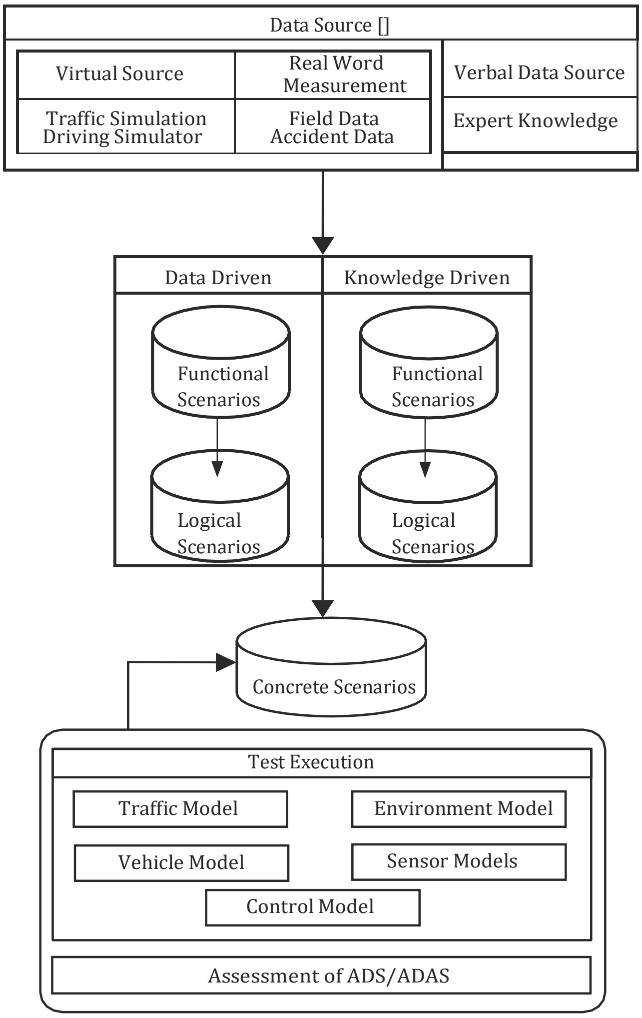

Figure 2. Scenario Based Approach.

3.1.1. Knowledge-Driven Approach

Derivation of concrete scenarios is divided into collection of knowledge, knowledge representation and parametrization. Sources for the knowledge collection are e.g. guidelines for road construction or traffic infrastructure, scenario catalogues, accident data, measurements from field tests and expert's knowledge [20]. Sources provide information to describe scenarios and support parametrization. A common practice to describe knowledge is to apply ontology. Conclusive definitions of the ontology can be found in the work of Guardino [30] but one often cited [31] describes the ontology as "a formal, explicit specification of a shared conceptualization". Ontology is a technique to organize and describe entities, their behaviour, relations and constraints of some domain in a formal way that is readable for both, humans and machines.

A common approach to represent the ontology is with two basic layers, terminological box (TBox) and assertional box (ABox) [32], [33]. TBox is an abstract layer that consists of objects, object properties and rules where objects are structured in hierarchical order. ABox is the concrete layer that contains data properties and specification of individuals.

Often, ontology is written using the Web Ontology Language (OWL), frequently based on the Resource Description Framework (RDF) model that expresses in the form subject-predicate-object, also known as triples. Typically, this is edited in software tools like Protégé, that already comes with the Semantic Web Rule Language (SWRL) to express rules and the reasoners like Pellet, Hermit, etc. [34] to infer logical consequences. Another method, used by researches for the knowledge representation is Unified Modeling Language (UML), the visual modelling language commonly used to describe and design software-based systems, business processes, etc. In this work, we will not cope with a comparison between OWL and UML, but readers can refer to [35] for more information.

The application of OWL can be seen in the work of Bagschik [36], [14]. He uses OWL together with the 5-layer model to represent knowledge and derive functional scenarios. Each layer within the layer model defines a certain group of scenario elements. The first layer defines the layout of the road, road markings and topology, the second layer includes traffic infrastructure added at the road-level. Temporary manipulations of the first two layers such as construction sites are added within the third layer. The fourth layer includes stationary and dynamic objects that are not part of traffic infrastructures like pedestrians or vehicles and the last layer deals with the environment. Such knowledge representation is a basis for automated generation of scenes and functional scenarios.

The application of UML class diagrams to depict ontology can be found in work of Klück [21] and Li [26]. To create an input model for combinatorial testing, they specify ontology as a tuple that contains sets of concepts, attributes, domains and relations between concepts, attributes and domains. Such defined tuple is a basis for automation to derive instances which are analogue to functional scenarios.

After defining fictional scenarios, the next step is parametrization and derivation of logical scenarios. The important thing to keep in mind when performing parametrization is the format in which the instances of logical scenarios will be transformed. Such a process is described by Menzel [37]. He divided the process into two steps: Detailing and format conversion. Detailing covers argumentation of functional scenarios with parameters and modelling relations and constraints between elements and parameters. While format conversion is responsible for transforming the representation of a road network in the syntax of the OpenDRIVE format and traffic participant and environmental conditions into OpenSCENARIO format. To assign parameters and and their distribution, classification trees are often used like in work of Li [26] or Schuldt [20] who was inspired by Grochtmann [38] and Thomason [39].

3.1.2. Data-Driven Approach

In a similar way as in knowledge-driven approach it is possible to divide the data-driven approach in the collection of data and data processing to derive logical scenarios and define parameter ranges. Pütz [6] grouped the data sources in a virtual, real and verbal category. The virtual category consists of traffic simulation and driving simulator data while the real category contains field data, proving ground tests and accident data. He also included expert knowledge as a source which is used to help interpreting complex scenarios that can be challenging for sensors. The data sources are subject to the data processing chain to form a few databases including logical scenarios, parameter spaces and test specifications. The data processing chain is conducted in seven steps starting from raw data transformation and fusion [40] over enrichment with additional signals that cannot be found in the recorded data such as time-to-collision (TTC) and finally concluded with clustering.

Examples how to use a simulation as source is given in [41] and [42]. In [41], Nalic calibrated a traffic flow model using measured data of an official highway test section in Austria and demonstrated a framework and an approach for testing a Highway Chauffeur function. On the other hand, Jenkins used a simple traffic simulation model to generate collision scenarios. His focus is to train and validate the Recurrent Neural Network on accident data generated from traffic simulation.

Nitsche [43] evaluated junction scenarios obtained from accident data which were collected in the UK between 1999 and 2010. The collected data are processed at two levels. The first level deals with a reduced number of attributes to describe main collision parameters and is achieved with the k-medoids clustering algorithm, which shows better performance on a smaller set of attributes. In level two, the association rule mining is used to add additional attributes. This separation in two levels helps to increase the efficiency of the clustering algorithm that performs better on a smaller set and helps to ease the job for the association rule mining due to a smaller dataset that needs to be interpreted. The results of clustering are described in natural language and associated with the functional scenarios while results of association rule add parameter ranges, therefore they are the representation of logical scenarios.

Another popular example of a data-driven approach is Waymo [44]. They are reproducing real driving data into parametrized simulations to analyse different variations. Such analysis helps to test new manoeuvres and refine old ones. With the help of simulation, they run up to 8 000 000 miles per day using 25 000 virtual self-driving vehicles. Gelder et al. [29] present a methodology based on real-life scenarios, extracting braking events from a dataset of 60 hours of driving. The recorded scenarios are parametrized and parameter values are presented as probability density functions (PDF) fitted with the help of Kernel Density Estimation (KDE). Those parametrized scenarios are used to generate concrete scenarios for Monte-Carlo simulations.

4. Safety Metrics

In section 3 various techniques for the generation and detection of scenarios are analyzed and reviewed. According to Figure 2 the output of the scenarios detection methods are concrete scenarios which are relevant for testing procedures of ADS. An essential part of the testing procedure is the assessment of the executed scenarios. The biggest challenge hereby is to find appropriate metrics which can be used for the safety assessment of ADS. In [45] Mahmud has performed a general review of possible safety metrics for automotive systems and classified them in four groups, temporal, distance and deceleration based indicators and other safety indicators. Very common and widely used for the assessment of concrete scenarios and ADS are temporal safety metrics and distance related metrics as shown in Tab. 1 and 2 respectively. In this case, they are additionally classified according to their underlying method. Hereby, **S** characterizes metrics, which are "stationary" extrapolating and thus assuming a stationary scenario development. **D**, which stands for dynamic extrapolation, indicates metrics considering the dynamic development of a scenario, e.g. the acceleration / deceleration of participants. Finally, **A** marks metrics, which are mainly based on specific assumptions to be made about the behaviour of the systems / participants in combination with a predominantly dynamic basic characteristic. These may, for example, be assumptions about planned trajectories or certain deceleration values to avoid accidents.

Temporal Metric	Description	Method	Reference
TTC	Time-to-Collision	S	[47], [48]
MTTC	Modified TTC	D	[49], [50]
WTTC	Worst TTC	A	[51]
ETTC	Extended TTC	D	[52]
TTB	Time-to-Brake	D	[46],[53]
TTK	Time-to-Kickdown	D	[54]
TTS	Time-to-Steer	D	[54]
TA	Time-to-Accident	S	[55], [56]
TTCE	Time-to-Closest-Encounter	A	[57]
THW	Time Headway	S	[58], [59]

Table 1. Temporal metrics used for assessment of the safety performance of ADS (S = stationary; D = dynamic; A = assumption-based).

Distance Metric	Description	Method	Reference
DCE	Distance-of-Closest-Encounter	S	[57]
MTC	Margin to Collision	D	[60]
PSD	Proportion of stopping distance	D	[61], [62]

Table 2. Distance metrics used for assessment of the safety performance of ADS (S = stationary; D = dynamic; A = assumption-based).

A famous representative of temporal metrics is, for example, the time-to-collision (TTC) with various forms and combinations as presented in [66], [53], [47], [46] and [67]. While easy to calculate and interpret, the TTC as a stationary metric is not able to handle the acceleration and deceleration behaviour of the participants. As a result, it makes sense to combine different metrics of different basic methods for a comprehensive assessment. In [46] Junietz differentiates between deterministic and probabilistic metrics and combines different temporal metrics for the criticality assessment of detected scenarios. The deterministic metrics are equivalent with the temporal metrics of [45]. Probabilistic metrics are used for trajectory planning and prediction methods and are presented in [68]. For the identification of critical scenarios Hallerbach uses in [69] temporal metrics in various forms and introduces a traffic metric which is used to identify critical scenarios for cooperative and automated vehicles in traffic and to evaluate the influence of ADS on traffic quality. Distance related metrics are mainly based on the prediction of a certain minimal distance to the target vehicle with a parallel prediction of the time to reach the target. This is shown in [63] and shown in the Mazda and Honda Algorithms in [64] and [65]. The Mazda and Honda Algorithms additionally consider the required deceleration to avoid a collision.

5. Assessment and Validation

This section deals with the methods and requirements for validating the functionality of ADS.
For system assessment and validation, a searchable scenario database is required that contains all described scenarios and allows a targeted scenario extraction that corresponds to the system focus. This theoretical structure and the practical implementation can be found in projects, such as SePIA or Pegasus [9] [70].
The main areas, which need scenarios for the assessment and validation, are the function development itself, the homologation of functions for legal authorisation in the market, rating organizations like Euro NCAP and in the field observation institutes e.g. periodical technical inspection. In each of these areas, the system operation according to the specifications has to be ensured.
The assessment environment ranges from pure simulation e.g. [71] via XiL-methods [72] to real-world drives, which could take place in real traffic (e.g. field operational tests - FOTs) or on a test site. Over all levels, the demands on the maturity level of the systems are constantly increasing. The controllability of the possible scenario parameters is always reproducible, except for the movement in real traffic. In contrast, the scenario space and the level of detail is potentially most extensive in real road traffic [73].
A main request for every assessment is the right scenario selection. The scenarios can be derived mostly from the system focus itself and expert knowledge [74] or by other methods e.g. risk assessment [75]. Systems that aim at a higher degree of automation must be able to handle a large scenario space, which can only be addressed efficiently with simulations [73]. These should be validated, but up to now there are almost no model validations of simulations, which are

necessary to make virtual assessment methods reliable and usable [76]. The test scenarios must focus on the challenging tasks for ADS, such as critical traffic scenarios or environmental influences [77].

The system assessment is divided into the areas of information gathering, action derivation and execution. On the system level, this corresponds to sensor and information processing, behaviour planning and actuator activation. For the ADS assessment, all areas must be taken into account [77].

For the assessment pass and fail criteria must be defined. These must be able to provide the safe functional area of the ADS. For the individual case consideration of scenarios, the criteria can be usual criticality metrics, such as the TTC as described in section 4.1. The pass or fail values can be specific values or ranges or can be derived from statistical distributions. Latter one can describe, for example, the impact of ADS on the reduction or the severity and occurrence of accidents and are normally one result by simulations.

A further criterion is the impact of the systems on the occurrence of accidents and the question whether new or unknown scenarios are caused by them. This prediction can be addressed with simulation as described in [78].

6. Conclusion

This review paper gives an overview of current approaches and methods for scenario generation techniques. Based on a general taxonomy for defining test-scenarios, the survey mainly shows how concrete scenarios can be generated both systematically knowledge-driven through expert knowledge and data-driven through the evaluation of mass driving- and accident-data. The safety metrics, which are divided into temporal and distance metrics, offer first possibilities for an evaluation of the generated scenarios, e.g. concerning their criticality.

Scenario-based-testing can provide one major pillar for the validation of automated driving systems. Compressing the real- and test drives into exemplary scenarios promises to reduce the validation effort tremendously. About the further development of scenario-based testing, three major challenges need to be solved in the future: It must be possible to 1) represent the real traffic situation as completely and realistically as possible for the corresponding use cases, 2) evaluate the scenarios as objectively and comprehensively as possible and 3) standardize and harmonize the existing scenario generation approaches for a neutral validation process.

References

[1] W. Wachenfeld and H. Winner (2016). The Release of Autonomous Vehicles. Berlin, Heidelberg: Springer Verlag. pp. 425–449.

[2] N. Kalra, S. M. Paddock (2016). "Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability?. In: Transportation Research Part A, Vol. 94, pp. 182-193.

[3] ISO/PAS 21448:2019(E) (2019). "Road vehicles - Safety of the intended functionality". ISO

[4] S. Geyer et al. (2014). "Concept and development of a unified onotlogy for generating test and use-case catalogues for assisted and automated vehicle guidance". In: IET Intelligent Transportation Systems, Vol. 8, Iss. 3, pp.193-189.

[5] S. Ulbrich et al. (2015). "Defining and Substantiating the Terms Scene, Situation, and Scenario for Automated Driving". IEEE 18th International Conference on Intelligent Transportation Systems, Las Palmas, pp. 982-988.

[6] A. Putz et al. (2017). "System validation of highly automated vehicles with a database of relevant traffic scenarios". In: situations, Vol. 1, pp. 19–22.

[7] C. Sippl et al. (2019). "Scenario-Based Systems Engineering: An Approach Towards Automated Driving Function Development". IEEE International Systems Conference (SysCon).

[8] C. Amersbach, H. Winner (2019). "Defining Required and Feasible Test Coverage for Scenario-Based Validation of Highly Automated Vehicles". IEEE 22nd Intelligent Transportation Systems Conference (ITSC).

[9] M. Lehmann, M. Bäumler, D. Hamelow, G. Prokop (2019). "Use of a criticality metric for assessment of critical traffic situations as part of SePIA". In: 19. Internationales Stuttgarter Symposium, Springer Vieweg

[10] Pegasus Project (2020). [Online] https://www.pegasusprojekt.de/.

[11] T. Menzel, G. Bagschik and M. Maurer (2018). "Scenarios for Development, Test and Validation of Automated Vehicles," IEEE Intelligent Vehicles Symposium (IV), Changshu, pp. 1821-1827.

[12] C. Rodarius et al. (2015). "Deliverable D7.1 // Test and Evaluation Plan AdaptIVe". [Online].

[13] H. Elrofai, J.P. Paardekooper, E.d. Gelder, S. Kalisvaart and O.o.d. Camp (2018). "StreetWise - Scenario-Based Safety Validation Of Connected And Automated Driving", [Online]. Available: https://www.persistent-identifier.nl/urn:nbn:nl:ui:24-uuid:2b155e03-5c51-4c9f-8908-3fa4c34b3897

[14] G. Bagschik, T. Menzel and M. Maurer (2018). "Ontology based Scene Creation for the Development

of Automated Vehicles". IEEE Intelligent Vehicles Symposium (IV), Changshu, pp. 1813-1820.
[15] J. Sauerbier, J. Bock, H. Weber, L. Eckstein (2019). "Definition of Scenarios for Safety Validation of Automated Driving Functions". ATZ worldwide. Vol. 121, No. 1, pp.42-45.
[16] M. Mai (2017). "Fahrerverhaltensmodellierung für die prospektive, stochastische Wirksamkeitsbewertung von Fahrerassistenzsystemen der Aktiven Fahrzeugsicherheit". Göttingen. Cuvillier Verlag
[17] E. Rocklage, H. Kraft, A. Karatas and J. Seewig (2017). "Automated scenario generation for regression testing of autonomous vehicles". IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, pp. 476-483.
[18] OICA (2019). "Future Certification of Automated/Autonomous Driving Systems". [Online]. Available: http://www.unece.org/ fileadmin/DAM/trans/doc/2019/wp29grva/ GRVA-02-09e.pdf
[19] H. Watanabe, T. Maly, J. Wallner, T. Dirndorfer, M. Mai, G. Prokop (2019). "Methodology of Scenario Clustering for Predictive Safety Functions". In: 9. Tagung Automatisiertes Fahren.
[20] Schuldt, Fabian et al. (2013). "Effiziente systematische Testgenerierung für Fahrerassistenzsysteme in virtuellen Umgebungen". In: Automatisierungssysteme, Assistenzsysteme und Eingebettete Systeme Für Transportmittel.
[21] F. Klueck, Y. Li, M. Nica, J. Tao and F. Wotawa (2018). "Using Ontologies for Test Suites Generation for Automated and Autonomous Driving Functions". IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW), Memphis, TN, pp. 118-123.
[22] P. Nitsche, R. H. Welsh, A. Genser and P. D. Thomas (2018). "A novel, modular validation framework for collision avoidance of automated vehicles at road junctions". 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, pp. 90-97.
[23] Rogic, B., Bernsteiner, S., Samiee, S., Eichberger, A., Payerl, C. (2016). "Konzeptionelle virtuelle Absicherung von automatisierten Fahrfunktionen anhand eines SAE Level 3 Fahrstreifenwechselassistenten". Poster session presented at VDI/VW-Gemeinschaftstagung, Wolfsburg, Germany.
[24] F. Hauer, T. Schmidt, B. Holzmüller and A. Pretschner (2019). "Did We Test All Scenarios for Automated and Autonomous Driving Systems?". IEEE Intelligent Transportation Systems Conference (ITSC), Auckland, New Zealand, pp. 2950-2955.
[25] A. Erdogan, E. Kaplan, A. Leitner and M. Nager (2018). "Parametrized End-to-End Scenario Generation Architecture for Autonomous Vehicles". 6th International Conference on Control Engineering & Information Technology (CEIT), Istanbul, Turkey, pp. 1-6.
[26] Y. Li, J. Tao, and F. Wotawa (2020). "Ontology-based test generation for automated and autonomous driving functions". Information and Software Technology, vol. 117, p. 106200.
[27] Rocklage, Elias, et al. (2017). "Automated scenario generation for regression testing of autonomous vehicles". IEEE 20th International Conference on Intelligent Transportation Systems (ITSC).
[28] Tuncali, Cumhur Erkan, et al. (2018). "Simulation-based adversarial test generation for autonomous vehicles with machine learning components". IEEE Intelligent Vehicles Symposium (IV).
[29] E. de Gelder and J. Paardekooper (2017). "Assessment of Automated Driving Systems using real-life scenarios". IEEE Intelligent Vehicles Symposium (IV), Los Angeles, CA, pp. 589-594.
[30] N. Guarino, D. Oberle, and S. Staab (2009). "What Is an Ontology?". Berlin, Heidelberg: Springer Verlag, pp. 1–17.
[31] R. Studer, V. R. Benjamins, and D. Fensel (1998). "Knowledge engineering: Principles and methods". Data & Knowledge Engineering, Vol. 25, No. 1, pp. 161 – 197.
[32] M. Hülsen, J. M. Zöllner and C. Weiss (2011). "Traffic intersection situation description ontology for advanced driver assistance". IEEE Intelligent Vehicles Symposium (IV), Baden-Baden, pp. 993-999.
[33] A. Armand, D. Filliat and J. Ibañez-Guzman (2014). "Ontology-based context awareness for driving assistance systems". IEEE Intelligent Vehicles Symposium Proceedings, Dearborn, MI, pp. 227-233.
[34] S. Abburu (2012). "A survey on ontology reasoners and comparison". In: International Journal of Computer Applications, Vol. 57, No. 17.
[35] C. Atkinson and K. Kiko (2005). "A detailed comparison of uml and owl". [Online]. Available: https://madoc.bib.uni-mannheim.de/1898/
[36] G. Bagschik, T. Menzel, C. Korner, and M. Maurer (2018). "Wissensbasierte Szenariengenerierung für Betriebsszenarien auf deutschen Autobahnen". In: Workshop Fahrerassistenzsysteme und automatisiertes Fahren, Vol. 12.
[37] T. Menzel, G. Bagschik, L. Isensee, A. Schomburg and M. Maurer (2019). "From Functional to Logical Scenarios: Detailing a Keyword-Based Scenario Description for Execution in a Simulation Environment". IEEE Intelligent Vehicles Symposium (IV), Paris, France, pp. 2383-2390.
[38] Grochtmann, Matthias and Klaus Grimm (1993). "Classification Trees for Partition Testing." Softw. Test. Verification Reliab. 3, pp. 63-82.

[39] Thomason M.G., Gonzalez R.C. (1985). "Data Structures and Databases in Digital Scene Analysis". In: Tou J.T. (eds) Advances in Information Systems Science. Springer, Boston, MA.

[40] M. Bäumler, L. Dziuba-Kaiser, Z. Yin, M. Lehmann, G. Prokop (2020). "Use information you have never observed together: Data fusion as a major step towards realistic test scenarios", 23. IEEE Intelligent Transportation Systems Conference (ITSC).

[41] D. Nalic, A. Eichberger, G. Hanzl, M. Fellendorf and B. Rogic (2019). "Development of a Co-Simulation Framework for Systematic Generation of Scenarios for Testing and Validation of Automated Driving Systems*". IEEE Intelligent Transportation Systems Conference (ITSC), Auckland, New Zealand, pp. 1895-1901.

[42] I. R. Jenkins, L. O. Gee, A. Knauss, H. Yin and J. Schroeder (2018). "Accident Scenario Generation with Recurrent Neural Networks". 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, pp. 3340-3345.

[43] Nitsche, Philippe & Thomas, Pete & Stuetz, Rainer & Welsh, Ruth. (2017). "Pre-crash scenarios at road junctions: A clustering method for car crash data". In: Accident analysis and prevention, 107, pp. 137-151.

[44] L. Waymo, "On the road to fully self-driving," 2018. [Online]. Available: https://waymo.com/safety/

[45] Mahmud, S. M. & Ferreira, Luis & Hoque, Md & Hojati, Ahmad. (2017). "Application of proximal surrogate indicators for safety evaluation: A review of recent developments and research needs". IATSS Research, 41.

[46] Junietz, Philipp & Schneider, Jan & Winner, Hermann. (2017). Metrik zur Bewertung der Kritikalität von Verkehrssituationen und -szenarien. In: 11. Workshop Fahrerassistenzsysteme.

[47] C. Roesener, F. Fahrenkrog, A. Uhlig and L. Eckstein (2016). "A scenario-based assessment approach for automated driving by using time series classification of human-driving behaviour". 19th International Conference on Intelligent Transportation Systems (ITSC), Rio de Janeiro, pp. 1360-1365.

[48] Rogic, B., Nalic, D., Eichberger, A. and Bernsteiner, S. (2020). "A Novel Approach to Integrate Human-in-the-Loop Testing in the Development Chain of Automated Driving: The Example of Automated Lane Change". 1-8. Paper presented at IFAC 2020 World Congress..

[49] Yang, Hong (2012). "Simulation-based evaluation of traffic safety performance using surrogate safety measures". Doctoral dissertation. Rutgers University-Graduate School-New Brunswick.

[50] Yang, Hong, Kaan Ozbay and Bekir Bartin (2010). "Application of simulation-based traffic conflict analysis for highway safety evaluation". Proceedings of the 12th WCTR, Lisbon, Portugal, 4.

[51] W. Wachenfeld, P. Junietz, R. Wenzel and H. Winner (2016). "The worst-time-to-collision metric for situation identification". IEEE Intelligent Vehicles Symposium (IV), Gothenburg, pp. 729-734.

[52] Minderhoud MM, Bovy PH (2001). "Extended time-to-collision measures for road traffic safety assessment". In: Accident Analysis and Prevention. 33 (1), pp. 89-97.

[53] J. Hillenbrand, K. Kroschel and V. Schmid (2005). "Situation assessment algorithm for a collision prevention assistant". IEEE Proceedings. Intelligent Vehicles Symposium. Las Vegas, NV, USA, pp. 459-465.

[54] J. Hillenbrand, A. M. Spieker and K. Kroschel (2006). "A Multilevel Collision Mitigation Approach - Its Situation Assessment, Decision Making, and Performance Tradeoffs". In: IEEE Transactions on Intelligent Transportation Systems, Vol. 7, No. 4, pp. 528-540.

[55] Zhou, Huanyun and Huang, Fei. (2013). "Development of Traffic Safety Evaluation Method based on Simulated Conflicts at Signalized Intersections". Procedia - Social and Behavioral Sciences.

[56] A. Svensson (1998). "A Method for Analysing the Traffic Process in a Safety Perspective". Lund University.

[57] J. Eggert, "Predictive risk estimation for intelligent ADAS functions," 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), Qingdao, 2014, pp. 711-718.

[58] K. Kircher (2003). "A comparison of headway and time to collision as safety indicators". In: Accident analysis and prevention, Vol. 35, pp. 427-33.

[59] Michelle L. Arnold and Ron VanHouten (2020). "Increasing Following Headway in a Driving Simulator and Transfer to Real World Driving". Journal of Organizational Behavior Management.

[60] Kitajima, S., et al. (2009). "Estimation of Driver's Dangerous States of Rear-end collision based on Driver Video Recorder Data and Ordinary Driving Data". Proceedings of Automotive Engineers of Japan. Vol. 97, pp. 89-97.

[61] Guido, Giuseppe, et al. (2011). "Comparing safety performance measures obtained from video capture data". In: Journal of Transportation Engineering, 137 (7), pp. 481-491.

[62] A. Vittorio et al. (2012). "A new microsimulation model for the evaluation of traffic safety performances".

[63] A. Polychronopoulos, M. Tsogas, A. Amditis, U. Scheunert, L. Andreone and F. Tango, "Dynamic

situation and threat assessment for collision warning systems: the EUCLIDE approach," IEEE Intelligent Vehicles Symposium, 2004, Parma, Italy, 2004, pp. 636-641.
[64] Ayumu Doi, Tetsuro Butsuen, Tadayuki Niibe, Takeshi Takagi, Yasunori Yamamoto, & Hirofumi Seni (1994). Development of a rear-end collision avoidance system with automatic brake control JSAE Review, 15(4), 335 - 340.
[65] Fujita, K. Akuzawa und M. Sato: Radar brake system. In: Proceedings of the 1995 Annual Meeting of ITS America, S. 95-101, March 1995.
[66] Paardekooper, J. et al. (2019). "Automatic Identification of Critical Scenarios in a Public Dataset of 6000 km of Public-Road Driving". In 26th International Technical Conference on the Enhanced Safety of Vehicles (ESV). Mira Smart.
[67] P. Junietz, F. Bonakdar, B. Klamann and H. Winner (2018). "Criticality Metric for the Safety Validation of Automated Driving using Model Predictive Trajectory Optimization". In: 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, pp. 60-65.
[68] Lefèvre, S., Vasquez, D. & Laugier, C. (2014). "A survey on motion prediction and risk assessment for intelligent vehicles". Robomech J 1, 1.
[69] S. Hallerbach et al. (2018). "Simulation-Based Identification of Critical Scenarios for Cooperative and Automated Vehicles". SAE Technical Papers.
[70] Zlocki, A., Eckstein, L., Fahrenkrog, F. (2015). "Evaluation and sign-off methodology for automated vehicle systems based on relevant driving situations". Transportation Research Record: Journal of the Transportation Research Board, 2489(1), pp. 123–129.
[71] Kitajima S., Shimono K., Tajima J., Antona-Makoshi J., and Uchida N. (2019). "Multi-agent traffic simulations to estimate the impact of automated technologies on safety". In: Traffic injury prevention, Vol. 20, No. sup1, pp. 58–64.
[72] Schuldt, F., Menzel, T., Maurer, M. (2015). "Eine Methode für die Zuordnung von Testfällen für automatisierte Fahrfunktionen auf X-in-the-Loop Verfahren im modularen virtuellen Testbaukasten". In C. Stiller (Ed.), 10. Workshop Fahrerassistenzsysteme (pp. 171–182). Uni-DAS e.V.
[73] SAE On-Road Automated Vehicle Standards Committee (2018). "Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles". SAE International: Warrendale, PA, USA
[74] Stellet, J. E., Zofka, M. R., Schumacher, J., Schamm, T., Niewels, F., Zollner, J. M. (2015). "Testing of advanced driver assistance towards automated driving: A survey and taxonomy on existing approaches and open questions". In: IEEE 18th International Conference on Intelligent Transportation Systems, pp. 1455–1462.
[75] Pierson A., Schwarting W., Karaman S., and Rus D. (2019). "Learning Risk Level Set Parameters from Data Sets for Safer Driving". In: IEEE Intelligent Vehicles Symposium (IV), pp. 273–280.
[76] Riedmaier S., Ponn T., Ludwig D., Schick B. and Diermeyer F. (2020). "Survey on Scenario-Based Safety Assessment of Automated Vehicles". In IEEE Access, Vol. 8, pp. 87456-87477.
[77] Batsch, F., Kanarachos, S., Cheah, M., Ponticelli, R., Blundell, M. (2020). "A taxonomy of validation strategies to ensure the safe operation of highly automated vehicles". In: Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, 0 (0), 1–20.
[78] Roesener C., Harth M., Weber H., Josten J., and Eckstein L. (2018). "Modelling Human Driver Performance for Safety Assessment of Road Vehicle Automation". In 2018 21st International Conference on Intelligent Transportation Systems (ITSC), pp. 735–741.
[79] A. Saswat et al. (2013). "An orchestrated survey on automated software test case generation". In: Journal of Systems and Software, Vol. 86, pp. 1978-2001.
[80] Mitra, Sovan (2006). "A White Paper on Scenario Generation for Stochastic Programming". Optirisk systems: white paper series, ref. No. OPT004.
[81] Kuhn, D. & Bryce, Renee & Duan, Feng & Ghandehari, Laleh & Lei, yu & Kacker, Raghu. (2015). "Combinatorial Testing. Theory and Practice". In: Advances in Computers, Vol. 99, pp. 1-66.
[82] Kuhn, D. et al. (2009). "Combinatorial Software Testing". In: Computer, Vol. 42.
[83] Xia, Qin & Duan, Jianli & Gao, Feng & Chen, Tao & Yang, Cai. (2017). "Automatic Generation Method of Test Scenario for ADAS Based on Complexity". No. 2017-01-1992. SAE Technical Paper.
[84] Winner, H., Geyer, S., & Sefati, M. (2013). "Maße für den Sicherheitsgewinn von Fahrerassistenzsystemen". Maßstäbe des sicheren Fahrens, 6.
[85] Benmimoun, Mohamed & Eckstein, Lutz. (2014). "Detection of Critical Driving Situations for Naturalistic Driving Studies by Means of an Automated Process". In: Journal of Intelligent Transportation and Urban Planning. pp. 11-21.

Acknowledgement

This work is founded by the Austrian Federal Ministry of Transport, Innovation and Technology as part of the FFG Program "EFREtop".