2015 IEEE 18th International Conference on Intelligent Transportation Systems

Testing of advanced driver assistance towards automated driving: A survey and taxonomy on existing approaches and open questions

Jan Erik Stellet[1], Marc René Zofka[2], Jan Schumacher[1], Thomas Schamm[2], Frank Niewels[1] and J. Marius Zöllner[2]

***Abstract*— In this work, we propose a novel taxonomy to partition the problem of testing advanced driver assistance systems (ADAS) into three basic dimensions. These dimensions are detailed and confirmed with recent research. Our framework permits the consideration of open research questions which have to be answered to pave the way for future highly automated driving. Despite the importance of this problem, a similarly comprehensive and structured survey has to the best of the authors' knowledge not been developed before.**

I. Introduction

A. Motivation

Advanced driver assistance systems (ADAS) feature an increasing degree of automation towards the goal of fully automated driving for safe and comfortable travel. This trend promises a reduction in the number and severity of traffic accidents, of traffic congestions as well as fuel consumption and thus leads to resource-saving mobility.

Therefore, intelligent vehicles have to perceive the environment, understand the current situation and plan and execute an appropriate behaviour. Due to the highly complex and dynamic workspace, algorithms for environment perception and cognition tasks have to cope with uncertainty in measurements and predictions.

However, with an increasing complexity of a system and the system's workspace, the effort for testing and evaluation rises. The safety and reliability of future ADAS have to be validated in high-dimensional and complex traffic situations, including traffic participants with different horizons of acting as well as interaction between different degrees of cooperation. Maurer and Stiller comment that:

> *"If testing and assessment methods cannot keep pace with this functional growth, they will become the bottleneck of the introduction of advanced DAS to the market."* [1]

Thus, new and efficient testing methods are required to pave the way for future ADAS.

B. Goal of this paper

In this paper, a novel taxonomy of testing methods for advanced driver assistance system is proposed. In our opinion, a structured reflection on recent accomplishments in research is necessary in order to highlight open questions related to future automated driving. Screening published testing approaches with respect to the methodological challenges, we have identified a framework of relevant axes. Thus, the current state of the art is compiled and structured.

[1]The authors are with Robert Bosch GmbH, Corporate Research, Vehicle Safety and Assistance Systems, 71272 Renningen, Germany

[2]The authors are with FZI – Research Center for Information Technology, 76131 Karlsruhe, Germany {zofka, schamm, zoellner}@fzi.de

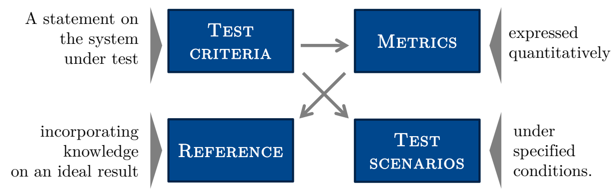


Fig. 1: Our taxonomy consists of three axes: 1) Test criteria and metrics (Sec. III), 2) referencing (Sec. IV) and 3) test scenarios (Sec. V).

II. Taxonomy

Our survey is based on the understanding that a test performs an evaluation of:

> A statement on the system-under-test *(test criteria)* that is expressed quantitavely *(metric)* under a set of specified conditions *(test scenario)* with the use of knowledge of an ideal result *(reference)*.

This understanding is based on the IEEE definition of testing [2] and implies three different axes, as visualized in Fig. 1.

The eventual goal is to validate whether a statement on the system is fulfilled. Thus, the derivation of technical test criteria and metrics from non-functional requirements (e.g. comfort, safety) represents the first axis in the framework and is discussed in Sec. III. The second axis, detailed in Sec. IV, is the reference system. This entity delivers ground truth information for quantitative attributes, necessary for an evaluation on an absolute scale. Thirdly, test scenarios represent specified conditions, under which the system-under-test is evaluated. Due to a complex, uncertain and unpredictable traffic environment, the definition and parametrisation of test scenarios for intelligent vehicles is a non-trivial task. This is detailed in Sec. V. Our taxonomy carves out open research questions, summarised in the conclusion in Sec. VI.

Although the axes will be discussed separately, there are of course dependencies. For example, in order to evaluate the effectiveness of a collision mitigation system, one has to define test scenarios that comprise accidents. This constraints

978-1-4673-6596-3/15 $31.00 © 2015 IEEE
DOI 10.1109/ITSC.2015.236

the possible ways of conducting the test to those, e.g. simulations, which can be safely executed.

III. Test Criteria and Metrics

The cornerstone of any ADAS test is a formal specification of the statement to be tested. We differentiate between three representations, that are 1) non-technical test criteria, 2) technical system level criteria and 3) sub-system metrics.

At the highest level, requirements on a system, e.g. from customers or legislations, specify the test goal. These are of course system-specific, but might include dimensions like comfort, reliability, robustness and safety [3]–[5].

We stress that it is often required to first translate such non-technical criteria to a technical, quantitative level, which is an important step of abstraction. The eventual test result thus reflects a technical view only on the demanded statement.

The derived technical test criteria are in general function-specific. For current ADAS though, one can often define system level criteria on a high level of abstraction in terms of a system response (correct, false) given an objective reason for intervention [6]. We argue that this concept, as it is visualised in Fig. 3, might be too simplistic for future continuously intervening automated driving functions. Hence, it might be necessary to define a generalisation to continuous scales measuring the level of objective danger in contrast to the appropriateness of the system's reaction.

Lastly, criteria and metrics can be defined on a sub-system level. That is because ADAS are typically not monolithic but modularised systems, comprising sensors for *environment perception*, algorithms for *object tracking* as well as *situation interpretation* and *action planning*. To select appropriate components and algorithms, metrics are required at this level.

At the *perception* layer, metrics quantify the completeness and accuracy of an observation [7], [8], e.g. for image-based sensors [9], [10], optical flow [11], stereo vision [12], [13], LIDAR sensors [14] or a grid-based representation [15].

Similarly, *object tracking* can be evaluated in terms of how complete and accurate the spatio-temporal estimates are [16]–[18].

Quantifying the capabilities of *situation interpretation* and *action planning* algorithms becomes increasingly ambiguous though. On the one hand, recognised intentions of other traffic participants and predictions of their future course may be evaluated for being complete and accurate. Similarly, control actions can be assessed in terms of tracking error, control effort and robustness [3].

However, for future highly automated driving, this might be insufficient. Therefore, metrics are required to measure how human-like an autonomous vehicle can interact in a traffic situation [1], though without the inherent imperfections and misunderstandings caused by human drivers.

IV. Referencing

Test criteria, which aim at an evaluation of a system on an absolute scale[1], inherently require the definition of a reference frame. Therefore, *ground truth* information has to be obtained and we will review approaches to do so in the following. Commonly, this is related to perception and object tracking tasks but this is mainly a question of the focus of the test and technical feasibility.

[1]In contrast to a mere relative comparison between two systems.

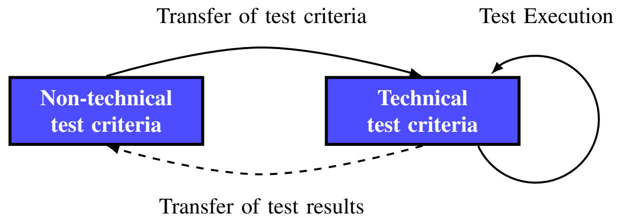


Fig. 2: Technical test criteria and metrics on system and sub-system level are usually derived from non-technical criteria.

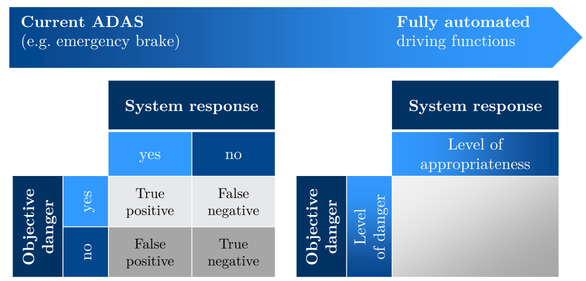


Fig. 3: Current ADAS are often evaluated on a binary scale but fully automated driving requires continuous measures.

Before we survey the different means for ground truth generation, it should be remarked that this concept is an abstraction of the real world and thus a postulated *axiomatic truth*. For example, vehicles which are detected by a perception sensor are typically abstracted to boxes with certain position and size. As it is illustrated in Fig. 4, a similar abstraction is performed for ground truth generation, e.g. by using a more precise sensor. Thus, misconceptions in this abstraction will propagate to the evaluation results.

We follow the usual classification, that there are two fundamental approaches for creating reference data, namely by measurement or simulation [18], [19]. The underlying difference is the origin of the reference information which can be empirical (data-based) or model-based.[2]

The ground truth property of the first approach has to be justified by the accuracy of the reference measurement, e.g. one order of magnitude better than the system-under-test. In the second approach, accurate and realistic models are assumed, so that the behaviour in a simulated environment represents the real system. Thus, the ground truth property relates to the congruence of these models with reality.

[2]Obviously, this differentiation is never really clear-cut. Using a reference sensor often requires to map two different physical measurement principles and thus models as well.

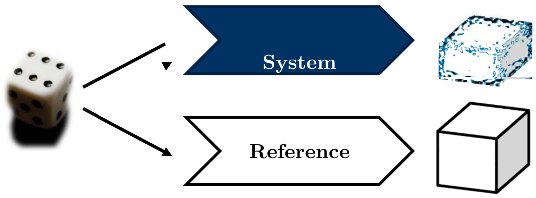


Fig. 4: Reference systems deliver ground truth information for the abstraction that is performed by a system and hence perform an abstraction of reality themselves.

Fig. 5: FZI's Cognitive Car: On-board reference sensors, like a high resolution laser scanner, facilitate referencing less accurate sensors.

A. Reference by measurement

Obtaining reference data by sensory measurements is the most intuitive approach at the perception layer. The basic principle is to run a second, significantly more accurate (in terms of spatial and temporal resolution) reference sensor in parallel to the perception system. Therefore, it is necessary that both systems share a similar measurement space or the mapping between these spaces is sufficiently simple.

One is then faced with a classical data association task, where reference measurements have to be assigned to all observations. For example, referencing a camera- with a laser scanner-based obstacle detection, see Fig. 5, requires associating all detected objects between both sensors. Performing this association task efficiently and unsupervised may become very challenging or impossible:

- Measurement spaces may be not perfectly congruent. For example, depth maps calculated from stereo video images and laser scanner point clouds exhibit different behaviour at transparent surfaces, e.g. windows [20]. Further post-processing steps are hence necessary.
- For certain attributes, e.g. image classification tasks, sufficiently reliable reference information cannot be provided by technical devices. Hence, manual labelling by human experts is required [21], which is a very costly and time-consuming task.
- A lower sampling rate of the reference sensor requires interpolation or limits the use to static scenes [20].

The key advantage of using reference sensors is that the vehicle can be operated in the real world. Therefore, realism with respect to environmental conditions, driver behaviour etc. is automatically ensured.

The central drawback though is that only one situation thread can be recorded in real-time and thus no alternative reactions of the system can be incorporated (open-loop testing). Moreover, a technical reference system may pose additional restrictions on the range of operation [8].

The following survey concerns reference measurements at the perception layer, i.e. the detection of static and dynamic objects in the vehicle's surroundings. Therefore, position and motion quantities are to be obtained with high accuracy. Two approaches can be distinguished, that are the use of on-board or external reference sensors.

1) On-board reference sensors: Having a reference sensor inside the ego vehicle allows a great flexibility, as the evaluation can be performed without any prerequisites on the test environment or other traffic participants. Many examples are reported for different features and we will list these according to the reference sensor that is used.

Laser scanner: Multi-layer laser scanners provide highly accurate range information. This can be used to validate the existence of object detections, e.g. for an image-based pedestrian detector [21]. Likewise, position and motion of dynamic objects can be obtained [22], [23].

High resolution laser scanner: In contrast to the former, lidar sensors with dozens of scan layers provide a dense 3-D image. This can be employed as reference for comprehensive world representations, e.g. built from dense stereo video images [20]. Creating reference measurements for objects and static environment [24] as well as semi-dense 3-D maps [25] are also reported.

Tri-focal optical sensor: An interesting approach presented in [13] is to use a third optical sensor as reference for a stereo camera setup (prediction error evaluation).

Reference from the system-under-test: Lastly, specialised approaches are known which avoid the need for additional sensors altogether by intelligent use of model assumption. Concerning stereo imaging, [26] discusses a pixel-wise reference for depth measurements on the road surface by exploiting a planar road assumption. Similarly, [27] proposes to check the existence and estimated velocity of stereo vision features by comparison to the vehicle's future driving corridor, which is of course collision-free. Although often impractical, one can also improve the accuracy in static scenes by simply averaging over multiple measurements [28].

2) External reference sensors: Given that techniques for vehicle localisation with a centimetre level accuracy exist, these can be used to provide ground truth information for sensors that measure relative distances.

To this end, both observer and observed vehicle are equipped with, e.g. differential GPS, and their absolute positions transformed to relative distances [8], [29], [30]. The concept is equally applicable to other objects such as a remotely controlled moving base [3], full-scale robotic vehicles [27] or artificial pedestrian dummies [31].

Apart from using satellite-based global positioning, cameras mounted above a proving ground can be used to identify and measure the position of vehicles below [32]. Using laser scanners and video sensors, a similar concept is applied to create a dataset of traffic at a public intersection in [33].

B. Reference by simulation

Synthetic ground truth data solves many of the aforementioned problems in the technical realisation of reference measurements. A simulation model generates all underlying input signals of the algorithm or system-under-test and maps these to the ideal output signal.

For example, consider a camera-based algorithm for the detection and distance measurement of vehicles. Here, the ideal result is the correct distance if a vehicle is present in the simulated scene. To test the algorithm, one has to synthetically generate its input data, i.e. images. The validity of all conclusions drawn from the algorithm's assessment is hence linked to the degree of realism of the synthetic images (e.g., ambient lighting, shadows, reflections etc.).

One central advantage is that simulations provide an omniscient observer model, which means that the past and all future evolutions are known in every time step. Thereby, ADAS components can be tested in closed-loop by regarding the exertion of influence [34].

Another relevant aspect is the ability to easily combine multiple simulation models and thus reuse existing work. This is one of the reasons why many approaches focus on just one specific but complex component of a holistic simulation environment, e.g. models of vision-based sensors.

The central issue with simulations is that all conclusions are only valid with respect to the underlying models. Thus, an important question is how the realism of a simulation environment can be evaluated and quantified. Contributions on model validation have been reported for vehicle dynamics [35], synthetic imaging [36] and human behaviour in driving simulators [37]. Still, given the variety of sensors and other parts of a simulation environment, e.g. other traffic participants, further research in this area could yield great value.

In the following, works on holistic simulation environments, as visualised in Fig. 6, will be surveyed first. Secondly, dedicated models of environment sensors as the central component of ADAS are reviewed.

1) Simulation environments: A simulation environment comprising multiple idealised sensor models, vehicle dynamics and traffic situations is developed in [38]–[40] within the European research project DECOS. From an industrial perspective, a similar objective is followed with the VIRTUAL TEST DRIVE [41]. Another commercial product is introduced in [42]. The development of an open-source solution is presented in [43]. In [44], [45], a simulation architecture is described and evaluated on the platooning behaviour of automatically controlled electrical carts. Aiming at the evaluation of a preventive pedestrian protection system, [6] presents a simulation environment with probabilistic models of driver's and pedestrian's behaviour as well as injury severity in accident situations.

2) Simulation of perception sensors: Two levels of abstraction can be distinguished for the modelling of environment perception sensors. Depending on whether an early indication on the feasibility of a new function or an in-depth evaluation are strived for, either idealised high-level or accurate low-level models can be employed [46].

High-level sensor models: In a generic approach, a sensor's ability to detect objects and environment features within its field of view with a specific measurement uncertainty is modelled [40], [47], [48]. Because an effect-based model is created, the adaptation to different sensors is straightforward. However, the physical measurement principles and hence sensor-specific effects are not taken into account.

Low-level sensor models: The goal is to accurately model the physical measurement principle (cause-based), which leads to sensor-specific models:

- **Video sensors:** A virtual camera sensor is presented in [49] and with fisheye lenses in [50]. In [51], a model is developed and compared to real images with a focus on colour correctness.
- **Stereo vision:** An evaluation of stereo imaging algorithms on synthetic and real images is given in [52].
- **Radar:** Realistic modelling of radar sensor measurements in vehicular applications is studied in [53], [54]. To this end, multiple reflections per object and the angular resolution are considered.

C. Combination of measurements and simulations

Given the strengths and weaknesses of the aforementioned concepts, it is a natural further step to combine the advantages of reference measurements with simulations [30]. Our taxonomy distinguishes between two general concepts:

1) X-in-the-loop: As it is visualised in Fig. 7a, the overall complex of vehicle, environment and driver is divided blockwise (sequentially) to real world or virtual data. Examples for this partitioning are:

- **Hardware in the loop (HiL)**: Synthetic environment models, e.g. renderings of 3-D environments, are projected on a screen and recorded by real video sensor hardware [42], [55].
- **Driver in the loop (DiL)**: A human driver takes a seat in an artificial vehicle with projections of a virtual environment and vehicle-like controls. An immersive virtual driving experience can be generated but at the cost of expensive, immovable simulators [56]. Hence, a simple PC-based solution has been proposed in [57].
- **Vehicle hardware in the loop (VEHiL)**: A real vehicle is fixed on a chassis dynamometer and equipped with real sensors. A robotic moving base acts as a controllable target [3].
- **Vehicle in the loop (ViL)**: In contrast to the former approach, a human driver is additionally included in the simulation loop. In order to add virtual objects to both human and the system's perception, these are projected into the driver's gaze using a HeadUp display [58], [59] or virtual reality glasses [60]. Because the potential

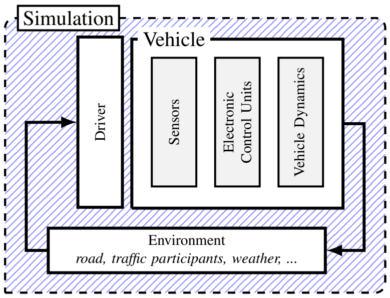
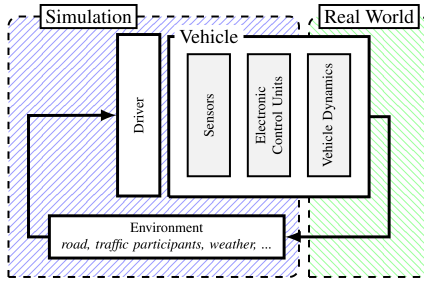
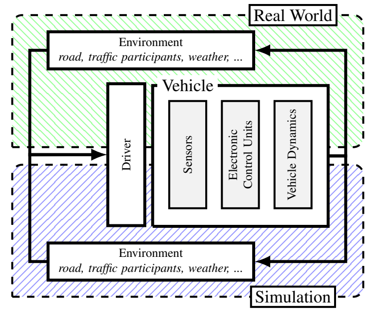


Fig. 6: An ADAS simulation includes models of the environment, the human driver, sensors and the vehicle dynamics.



(a) X-in-the-loop: Hardware components are connected to the virtual environment.



(b) Measured and simulated environmental aspects are augmented and aligned in order to test ADAS on both worlds.

Fig. 7: Different concepts of combining measurements and simulations.

driver reactions stem from synthetic objects, a proving ground with sufficient free space is required.

2) Augmentation of measurement data: For complex measurement principles, e.g. vision-based, simulating all influences with sufficient realism can be infeasible. One approach to alleviate this problem is to augment real recordings with additional synthetic elements. For example, recorded video sequences from test drives can serve as realistic image background. Then, the simulated appearance of objects, e.g. pedestrians [61], [62] or vehicles [63] is added to this data. Effectively, real and virtual data are used in parallel, as illustrated in Fig. 7b. A comparison between real and augmented images is presented in [64], where a pedestrian classifier is learned from both datasets and the eventual performance is evaluated.

A different direction is pursued in [65], where image data is first collected using 360° cameras and laser scanners. From these raw recordings, synthetic images with an arbitrary viewing angle and camera field of view can be rendered in order to emulate different sensor configurations.

D. Conclusion

Retrieving ground truth information from measurements is advantageous in order to capture real world realism (e.g. environmental influences and human reactions) but often limited in practice due to repeated effort per measurement. The ground truth assumption can be justified by comparing the measurement accuracy of the reference sensor and the system-under-test. Generating reference data with simulations offers the potential for reproducible closed-loop testing. The main drawback is the high initial effort to build the simulation environment. How to quantify the achieved degree of realism has not yet been finally answered.

Generally, a further significant benefit of virtual test environments is the ability to cover a variety of test scenarios, especially those which involve dangerous accident situations.

V. Test scenarios

In order to compare different (meta-) models for the structuring and derivation of test scenarios, the term *traffic situation* is generically defined to integrate different works. Thereby, in our framework a *test scenario* is defined as the sum of all relevant conditions in the workspace, under which a system is operated during the test. This concerns dimensions like road type, current weather, static environment and especially other traffic participants. The fact that intelligent vehicles have to cope with different grades of automation increases the space of behaviours to be considered further. Our taxonomy implies two general research questions in order to structure the compilation of appropriate test scenarios for ADAS to automated driving:

1) How to obtain test scenario definitions?
2) What is encompassed by the test scenario?

In each case, two different concepts of approaching these research questions have been identified. They are visualised in Fig. 8 and will be detailed in the following.

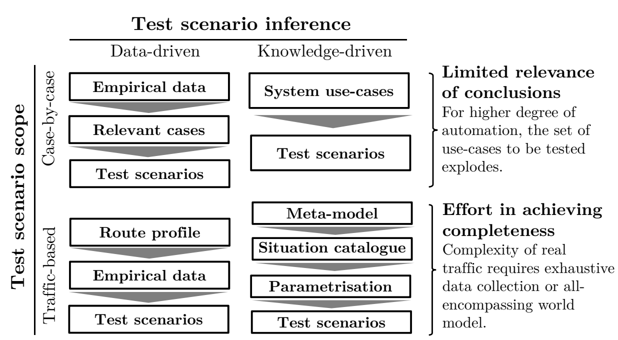


Fig. 8: Our framework identifies two complementary axes of defining and deriving test scenarios.

A. Test scenario inference

The first approach for the definition of a test scenario is called *knowledge-driven*: Thereby, holistic considerations lead to the definition of dimensions, which have to be taken into account to build up a test scenario for the system-under-test. In order to do so, known use-case catalogues or meta-models can be used. In this approach, implicit information is made explicitly available. This is similar to the concept of deduction, where no extension of the deductive hull takes place and thus, the resulting test scenarios are restricted to the specified meta-models or use-cases.

In [66] an ontology for generating test and use-case catalogues for cooperative vehicle guidance is presented. A fundamental terminology is given in order to describe the traffic situation dimensions from the ego vehicle's point of view. In [67] a meta-model consisting of four layers is presented: the road course, adaptions like building sites, the traffic situation and the environmental conditions like snow or rain. The model presented in [68] consists of three dimensions, considering driver, environment as well as the vehicle. These patterns are then refined with attributes and can be instantiated in order to create a test scenario. A hierarchical tree is presented in [69], which splits up a situation into two sub-trees for the driver within its vehicle as well as the locality. Another example for a knowledge-driven derivation is given in [70] for a parking assist function. In [71] promising design considerations for a testing language are presented, which formalises different aspects of a test scenario, including the cyber-physical system, traffic participants as well as the definition of evaluation agents.

A *data-driven* approach on the other hand generalises implicit information in empirical data in order to obtain critical dimensions. In [72], for example, the GIDAS[3] database is used to infer representative, critical traffic situations for pedestrian protection systems. Similarly, [31] combines accident data from different country-specific databases. In [73] the traffic situation space is partitioned using accident patterns from the database of the German Insurance Association [74]. Test scenarios for Autonomous Emergency Braking (AEB) are defined in [75] based on an accidentology study. These have been adapted to the EURONCAP testing procedure [76] and their use in simulations is discussed [77].

B. Test scenario scope

The scope of the test scenarios is directly linked to the conclusion that should be drawn from the test. That is, the performance of a system-under-test can be evaluated in specific singular use-cases *case-by-case* or concerning its overall impact in the field. The importance of the latter category, denoted as *traffic-based*, is that unwanted negative side-effects are revealed as well [6].

As an example for *case-by-case* evaluation, a collision prevention system can be evaluated within delimited accident situations for crossing pedestrians [31], [72], [75]. This enables the evaluation of an ADAS regarding its effectiveness in a specific use-case. Similarly, [70] considers only the geometry of parking spaces due to the limited functionality of a parking assistant. In [38] situation templates are used, that are parametrised by a single criticality factor.

In a *traffic-based* evaluation, generic traffic scenes are considered. Typically, this applies to public road tests or field operational tests like the Bertha-Benz drive [78]. Presented in [79], [80] urban and inter-continental public road tests are performed without any prior assumption on the passed test cases. Thereby, varying road infrastructures, weather and illumination conditions, and traffic patterns are included.

Reference sensor recordings from real test drives can be replayed in simulations [81], however the behaviour of traffic participants is implicitly contained, which restricts to open-loop testing. In order to identify relevant scenarios from recorded data, model-based footprints on different scales [82], [83], system exposure metrics [84] or machine learning, e.g. of convoy merging situations [85], can be used.

[3]German In-Depth Accident Study

C. Implications on future ADAS

Our taxonomy raises several questions, which have to be discussed for the creation of test scenarios for future ADAS.

Systems with higher degree of automation have to cope with more complex traffic situations. Thus, evaluating the system's performance in a *case-by-case* fashion becomes tedious or leads to incomplete conclusions only.

It is therefore essential to consider efficient means for *traffic-based* evaluation. The effort in inference of traffic situations from recordings is accompanied by the challenge of achieving completeness when creating an *all-encompassing traffic situation world model*. Growing and learning ontologies [66], based on an open-world assumption, may be the key to this challenge.

VI. CONCLUSION

The main contribution of this work is the proposal of several dimensions that allow comparing and contrasting testing methods for advanced driver assistance system. We demonstrate how a significant number of relevant works fits into this framework and thus confirm its usefulness.

Despite the abundance of promising approaches that have been identified and classified, a number of research questions remain unanswered to date. These can be summarised as:

1) How to define test criteria to measure the performance and safety of automated vehicles and transfer these to quantitative metrics on system and sub-system level?
2) How to overcome the challenge of accurately modelling complex measurement principles, e.g. in vision-based sensors?
3) How to overcome the dilemma of testing the entire complexity of real-world traffic?

Future works could tackle the problem with a twofold approach, that is a reduction of the problem space and a faster test execution, e.g. using simulations or analytical models.

On the one hand, efficient inference of test scenarios could be achieved by an intelligent combination of data- and knowledge-driven approaches. On the other hand, once the challenges in different test scenarios are understood and related to the individual components and algorithms of the system, testing might be executed individually, using the most efficient and accurate reference available at this layer.

References

[1] K. Bengler, K. Dietmayer, B. Farber, M. Maurer, C. Stiller, and H. Winner, "Three decades of driver assistance systems: Review and future perspectives," *Intelligent Transportation Systems Magazine, IEEE*, vol. 6, pp. 6–22, winter 2014.

[2] IEEE, "610.12-1990 - IEEE standard glossary of software engineering terminology," 1990.

[3] O. J. Gietelink, *Design and validation of advanced driver assistance systems*. PhD thesis, Delft University of Technology, 2007.

[4] S. Ebel, U. Wilhelm, A. Grimm, and U. Sailer, "Wie sicher ist sicher genug? Anforderungen an die funktionale Unzulänglichkeit von Fahrerassistenzsystemen in Anlehnung an das gesellschaftlich akzeptierte Risiko," in *6. Workshop Fahrerassistenzsysteme*, 2009.

[5] S. Fabris, J. D. Miller, and J. Luo, "Validation of an aeb system," in *3rd International Symposium on Road Vehicles Functional Safety Standards and Its Application*, (Xi'an, China), 2014.

[6] T. Helmer, *Development of a Methodology for the Evaluation of Active Safety using the Example of Preventive Pedestrian Protection*. PhD thesis, Technische Universität, Berlin, 2014.

[7] V. Mariano, J. Min, J.-H. Park, R. Kasturi, D. Mihalcik, H. Li, D. Doermann, and T. Drayer, "Performance evaluation of object detection algorithms," in *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, vol. 3, pp. 965–969, 2002.

[8] M. Brahmi, "Reference systems for environmental perception," in *Automotive Systems Engineering* (M. Maurer and H. Winner, eds.), pp. 205–221, Springer Berlin Heidelberg, 2013.

[9] K. Smith, R. Schweiger, W. Ritter, and J.-E. Kallhammer, "Development and evaluation of a performance metric for image-based driver assistance systems," in *Intelligent Vehicles Symposium (IV), 2011 IEEE*, pp. 381–386, 2011.

[10] J. Fritsch, T. Kuhnl, and A. Geiger, "A new performance measure and evaluation benchmark for road detection algorithms," in *Intelligent Transportation Systems - (ITSC), 16th International IEEE Conference on*, pp. 1693–1700, 2013.

[11] P. Marquez-Valle, D. Gil, and A. Hernandez-Sabate, "Evaluation of the capabilities of confidence measures for assessing optical flow quality," in *Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on*, pp. 624–631, 2013.

[12] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vision*, vol. 47, no. 1-3, pp. 7–42, 2002.

[13] S. Morales, Y. Woo, R. Klette, and T. Vaudrey, "A study on stereo and motion data accuracy for a moving platform," in *Advances in Robotics*, vol. 5744 of *Lecture Notes in Computer Science*, pp. 292–300, Springer Berlin Heidelberg, 2009.

[14] A. Kapp, "Quality measures for lidar signal processing," in *Intelligent Vehicles Symposium, 2006 IEEE*, pp. 163–168, 2006.

[15] B. Grabe, T. Ike, and M. Hoetter, "Evaluation method of grid based representation from sensor data," in *Intelligent Vehicles Symposium, 2009 IEEE*, pp. 1245–1250, 2009.

[16] M. Stampfle, D. Holz, and J. Becker, "Performance evaluation of automotive sensor data fusion," in *Intelligent Transportation Systems, 2005. Proceedings. 2005 IEEE*, pp. 50–55, 2005.

[17] B. Ristic, B.-N. Vo, D. Clark, and B.-T. Vo, "A metric for performance evaluation of multi-target tracking algorithms," *Signal Processing, IEEE Transactions on*, vol. 59, no. 7, pp. 3452–3457, 2011.

[18] R. Schubert, H. Klöden, G. Wanielik, and S. Kalberer, "Performance evaluation of multiple target tracking in the absence of reference data," in *Information Fusion (FUSION), 13th Conference on*, pp. 1–7, 2010.

[19] S. N. R. Meister, *On Creating Reference Data for Performance Analysis in Image Processing*. PhD thesis, University of Heidelberg, 2014.

[20] D. Pfeiffer, S. Morales, A. Barth, and U. Franke, "Ground truth evaluation of the stixel representation using laser scanners.," in *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*, pp. 1091–1097, 2010.

[21] T. Tatschke, F.-J. Farber, E. Fuchs, L. Walchshausl, and R. Lindl, "Semi-autonomous reference data generation for perception performance evaluation," in *Information Fusion, 2007 10th International Conference on*, pp. 1–7, 2007.

[22] F. de Ponte Müller, L. Navajas, and T. Strang, "Characterization of a laser scanner sensor for the use as a reference system in vehicular relative positioning," in *Communication Technologies for Vehicles*, vol. 7865 of *Lecture Notes in Computer Science*, pp. 146–158, Springer Berlin Heidelberg, 2013.

[23] D. M. Kirchhöfer, C. Uhlmann, L. Beste, B. Reiß, and T. Brandmeier, "Automated object labelling for advanced driver assistance systems using a camera- and a lidar-system," in *12th International Symposium and Exhibition on Sophisticated Car Occupant Safety Systems*, 2014.

[24] J. Klandt, M. Radimirsch, A. Kirchner, and H. Philipps, "Referenzierung von Fahrerassistenzsystemen. Mess-und Auswertesysteme fuer die Umfelderfassung," *VDI-Berichte*, no. 2166, 2012.

[25] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, pp. 3354–3361, 2012.

[26] Z. Liu and R. Klette, "Approximated ground truth for stereo and motion analysis on real-world sequences," in *Advances in Image and Video Technology*, vol. 5414 of *Lecture Notes in Computer Science*, pp. 874–885, Springer Berlin Heidelberg, 2009.

[27] N. Schneider, S. Gehrig, D. Pfeiffer, and K. Banitsas, "An evaluation framework for stereo-based driver assistance," in *Real-World Scene Analysis 2011*, Lecture Notes in Computer Science, p. 2751, 2012.

[28] R. Nair, S. Meister, M. Lambers, M. Balda, H. Hofmann, A. Kolb, D. Kondermann, and B. Jähne, "Ground truth for evaluating time of flight imaging," in *Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications*, vol. 8200 of *Lecture Notes in Computer Science*, pp. 52–74, Springer Berlin Heidelberg, 2013.

[29] S. Zeng, "Performance evaluation of automotive radars using carrier-phase differential GPS," *Instrumentation and Measurement, IEEE Transactions on*, vol. 59, pp. 2732–2741, Oct 2010.

[30] B. Strasser, A. Siegel, K.-H. Siedersberger, M. Maurer, and H. Bubb, "Vernetzung von Test-und Simulationsmethoden für die Entwicklung von Fahrerassistenzsystemen (FAS)," in *4. Tagung Aktive Sicherheit durch Fahrerassistenz*, 2010.

[31] P. Lemmen, J. Stoll, U. Bergelt, P. Seiniger, *et al.*, "Evaluation of pedestrian targets for use in automomous emergengy brake system testing-a report from the harmonistion platform 2 dealing with test equipment," *23nd Conference on the Enhancement of the Safety of Vehicles (ESV)*, 2013.

[32] T. Kubertschak, M. Wittenzellner, and M. Maehlisch, "Kamerabasiertes Referenzsystem für Fahrerassistenzsysteme," in *Forum Bildverarbeitung 2014*, pp. 261–272, KIT Scientific Publishing, 2014.

[33] E. Strigel, D. Meissner, F. Seeliger, B. Wilking, and K. Dietmayer, "The ko-per intersection laserscanner and video dataset," in *Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on*, pp. 1900–1901, 2014.

[34] V. Punzo and B. Ciuffo, "Integration of driving and traffic simulation: Issues and first solutions," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 12, no. 2, pp. 354–363, 2011.

[35] U. Baake, K. Wüst, M. Maurer, and A. Lutz, "Testing and simulation-based validation of ESP systems for vans," *ATZ worldwide*, vol. 116, no. 2, pp. 30–35, 2014.

[36] M. Nentwig, *Untersuchungen zur Anwendung von computergenerierten Kamerabildern für die Entwicklung und den Test von Fahrerassistenzsystemen*. PhD thesis, Friedrich-Alexander-Universität, Erlangen-Nürnberg, 2013.

[37] I. Zöller, N. Mautes, W. Ren, and B. Abendroth, "Knowledge base for valid driving simulation," *ATZ worldwide*, vol. 117, no. 3, pp. 46–49, 2015.

[38] Christian Schmidt, *Hardware-in-the-Loop gestützte Entwicklungsplattform für Fahrerassistenzsysteme: Analyse und Generierung kritischer Verkehrsszenarien*. PhD thesis, Universität Kassel, 2010.

[39] Carsten Schmidt, *Hardware-in-the-Loop-gestützte Entwicklungsplattform für Fahrerassistenzsysteme: Modellierung und Visualisierung des Fahrzeugumfeldes*. PhD thesis, Universität Kassel, 2011.

[40] D. Tellmann, *Hardware-in-the-Loop-gestützte Entwicklungsplattform für Fahrerassistenzsysteme: Modelle der Umfeldsensorik und angepasste Fahrermodelle*. PhD thesis, Universität Kassel, 2011.

[41] K. von Neumann-Cosel, *Virtual Test Drive: Simulation umfeldbasierter Fahrzeugfunktionen*. PhD thesis, Technische Universität München, 2014.

[42] S. Noth, J. Edelbrunner, and I. Iossifidis, "An integrated architecture for the development and assessment of adas," in *Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on*, pp. 347–354, 2012.

[43] K. Swanson, A. Brown, S. Brennan, and C. LaJambe, "Extending driving simulator capabilities toward hardware-in-the-loop testbeds and remote vehicle interfaces," in *Intelligent Vehicles Symposium (IV), 2013 IEEE*, pp. 122–127, 2013.

[44] F. Gechter, J.-M. Contet, S. Galland, O. Lamotte, and A. Koukam, "Virtual intelligent vehicle urban simulator: Application to vehicle platoon evaluation," *Simulation Modelling Practice and Theory*, vol. 24, no. 0, pp. 103–114, 2012.

[45] F. Gechter, B. Dafflon, P. Gruer, and A. Koukam, "Towards a hybrid real/virtual simulation of autonomous vehicles for critical scenarios.," *The Sixth International Conference on Advances in System Simulation (SIMUL 2014)*, 2014.

[46] R. Schubert, N. Mattern, and R. Bours, "Evaluating automated vehicle systems using probabilistic sensor simulations," in *Proceedings of ITS European Congress*, 2014.

[47] W. Deng, J. Dai, Q. Zhao, B. Litkouhi, N. Moshchuk, and R. Nisonger, "Modeling of range sensing and object detection, for vehicle active safety," in *Intelligent Transportation Systems (ITSC), 2009. 12th International IEEE Conference on*, pp. 1–6, 2009.

[48] M. Munz, M. Mahlisch, J. Dickmann, and K. Dietmayer, "Probabilistic modeling of sensor properties in generic fusion systems for modern driver assistance systems," in *Intelligent Vehicles Symposium (IV), 2010 IEEE*, pp. 760–765, 2010.

[49] D. Gruyer, M. Grapinet, and P. De Souza, "Modeling and validation of a new generic virtual optical sensor for adas prototyping," in *Intelligent Vehicles Symposium (IV), 2012 IEEE*, pp. 969–974, 2012.

[50] B. Schick and S. Schmidt, "Evaluation of video-based driver assistance systems with sensor data fusion by using virtual test driving," in *Proceedings of the FISITA 2012 World Automotive Congress*, vol. 196 of *Lecture Notes in Electrical Engineering*, pp. 1363–1375, 2013.

[51] D. Hospach, S. Mueller, O. Bringmann, J. Gerlach, and W. Rosenstiel, "Simulation and evaluation of sensor characteristics in vision based advanced driver assistance systems," in *Intelligent Transportation Systems (ITSC), IEEE 17th International Conference on*, pp. 2610–2615, 2014.

[52] T. Vaudrey, C. Rabe, R. Klette, and J. Milburn, "Differences between stereo and motion behaviour on synthetic and real-world stereo sequences," in *Image and Vision Computing New Zealand (IVCNZ), 23rd International Conference*, pp. 1–6, 2008.

[53] L. Hammarstrand, M. Lundgren, and L. Svensson, "Adaptive radar sensor model for tracking structured extended objects," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 48, no. 3, pp. 1975–1995, 2012.

[54] D. Svensson, M. Ulmke, and L. Hammarstrand, "Multitarget sensor resolution model and joint probabilistic data association," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 48, no. 4, pp. 3418–3434, 2012.

[55] F. Schmidt, *Funktionale Absicherung kamerabasierter Aktiver Fahrerassistenzsysteme durch Hardware-in-the-Loop-Tests*. PhD thesis, Universität Kaiserslautern, 2012.

[56] H. Amelunxen and C. Schmidt, "Driver-in-the-loop simulation for advanced driver assistance systems," in *AVEC 14*, pp. 750–756, 2014.

[57] K. Abdelgawad, B. Hassan, M. Grafe, and I. Gräßler, "A modular architecture of a PC-based driving simulator for advanced driver assistance systems development," in *Proceedings of the IEEE 15th International Workshop on Research and Education in Mechatronics REM*, 2014.

[58] T. Bock, M. Maurer, and G. Farber, "Validation of the vehicle in the loop (vil); a milestone for the simulation of driver assistance systems," in *2007 IEEE Intelligent Vehicles Symposium*, pp. 612–617, 2007.

[59] B. Blissing, F. Bruzelius, and J. Ölvander, "Augmented and mixed reality as a tool for evaluation of vehicle active safety systems," *International Conference Road Safety and Simulation*, 2013.

[60] S. Schwab, T. Leichsenring, M. R. Zofka, and T. Bär, "Consistent test method for assistance systems," *ATZ worldwide*, vol. 116, no. 9, pp. 38–43, 2014.

[61] J. Nilsson, *Handbook of Augmented Reality*, ch. Using Augmentation Techniques for Performance Evaluation in Automotive Safety, pp. 631–649. Springer, 2011.

[62] J. Nilsson, *Computational verification methods for automotive safety systems*. PhD thesis, Chalmers University of Technology, 2014.

[63] M. R. Zofka, R. Kohlhaas, T. Schamm, and J. M. Zöllner, "Semivirtual simulations for the evaluation of vision-based adas," in *IEEE Intelligent Vehicles Symposium*, pp. 121–126, 2014.

[64] J. Nilsson, P. Andersson, I.-H. Gu, and J. Fredriksson, "Pedestrian detection using augmented training data," in *Pattern Recognition (ICPR), 2014 22nd International Conference on*, pp. 4548–4553, 2014.

[65] A. Oko, T. Sato, H. Kume, T. Machida, and N. Yokoya, "Evaluation of image processing algorithms on vehicle safety system based on free-viewpoint image rendering," in *Intelligent Vehicles Symposium Proceedings, 2014 IEEE*, pp. 706–711, 2014.

[66] S. Geyer, M. Baltzer, B. Franz, S. Hakuli, *et al.*, "Concept and development of a unified ontology for generating test and use-case catalogues for assisted and automated vehicle guidance," *Intelligent Transport Systems, IET*, vol. 8, no. 3, pp. 183–189, 2014.

[67] F. Schuldt, F. Saust, B. Lichte, and M. Maurer, "Effiziente systematische Testgenerierung für Fahrerassistenzsysteme in virtuellen Umgebungen.," *AAET2013 - Automatisierungssysteme, Assistenzsysteme und eingebettete Systeme für Transportmittel*, 2013.

[68] C. Domsch and H. Negele, "Einsatz von Referenzfahrsituation bei der Entwicklung von Fahrerassistenzsystemen," in *3. Tagung Aktive Sicherheit durch Fahrerassistenz*, 2008.

[69] J. Schwarz, *Response 3 - Code of Practice for development, validation and market introduction of advanced driver assistance systems*. VDI Berichte, VDI, 1960 ed., 2006.

[70] J. Wegener and O. Bühler, "Evaluation of different fitness functions for the evolutionary testing of an autonomous parking system," in *In Proceedings of GECCO*, pp. 1400–1412, Springer-Verlag, 2004.

[71] C. Berger, "Design considerations for a cyber-physical testing language on the example of autonomous driving," in *Proceedings of the 2012 Workshop on Domain-specific Modeling*, pp. 49–54, 2012.

[72] A. Ebner, T. Helmer, R. R. Samaha, and P. Scullion, "Identifying and analyzing reference scenarios for the development and evaluation of active safety: Application to preventive pedestrian safety," *International Journal of Intelligent Transportation Systems Research*, vol. 9, no. 3, pp. 128–138, 2011.

[73] F. Saust, T. C. Müller, J. M. Wille, and M. Maurer, "Entwicklungsbegleitendes Simulations- und Testkonzept für autonome Fahrzeuge in städtischen Umgebungen," in *AAET - Automatisierungssysteme, Assistenzsysteme und eingebettete Systeme für Transportmittel*, pp. 87–107, GZVB e.V, 2009.

[74] German Insurers Accident Research, "Unfalltypenkatalog der Unfallforschung der Versicherer (UDV): UNKA."

[75] W. Hulshof, I. Knight, A. Edwards, M. Avery, and C. Grover, "Autonomous emergency braking test results," *Proceedings of the 23rd International Technical Conference on the Enhanced Safety of Vehicles (ESV)*, 2013.

[76] R. Schram, A. Williams, and M. van Ratingen, "Implementation of autonomous emergency braking (aeb), the next step in euro ncap's safety assessment," *Proceedings of the 23rd International Technical Conference on the Enhanced Safety of Vehicles (ESV)*, 2013.

[77] C. Berger, D. Block, S. Heeren, C. Hons, S. Kuhnel, A. Leschke, D. Plotnikov, and B. Rumpe, "Simulations on consumer tests: A systematic evaluation approach in an industrial case study," in *Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on*, pp. 1474–1480, 2014.

[78] J. Ziegler, P. Bender, M. Schreiber, H. Lategahn, *et al.*, "Making bertha drive—an autonomous journey on a historic route," *IEEE Intelligent Transportation Systems Magazine*, vol. 6, no. 2, pp. 8–20, 2014.

[79] A. Broggi, M. Buzzoni, S. Debattisti, P. Grisleri, M. C. Laghi, P. Medici, and P. Versari, "Extensive tests of autonomous driving technologies," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 14, no. 3, pp. 1403–1415, 2013.

[80] M. Distner, M. Bengtsson, T. Broberg, and L. Jakobsson, "City safety—a system addressing rear-end collisions at low speeds," in *Proc. 21st International Technical Conference on the Enhanced Safety of Vehicles.*, 2009.

[81] U. Lages, M. Spencer, and R. Katz, "Automatic scenario generation based on laserscanner reference data and advanced offline processing," in *2013 IEEE Intelligent Vehicles Symposium (IV)*, pp. 153–155, 2013.

[82] M. Benmimoun, F. Fahrenkrog, A. Zlocki, and L. Eckstein, "Detection and classification of critical incidents by means of vehicle data," *ATZ worldwide*, vol. 114, no. 10, pp. 60–66, 2012.

[83] C. Berger, "Scenario pattern matching in large sensor recordings with simulation models for cyber-physical systems," *Proceedings of the 46th Summer Computer Simulation Conference 2014*, 2014.

[84] P. Glauner, A. Blumenstock, and M. Hauneis, "Exposition zur Raffung der Felderprobung für Fahrerassistenzsysteme," in *9. Workshop Fahrerassistenzsysteme*, 2014.

[85] M. Reichel, M. Botsch, R. Rauschecker, K.-H. Siedersberger, and M. Maurer, "Situation aspect modelling and classification using the scenario based random forest algorithm for convoy merging situations," in *2010 13th International IEEE Conference on Intelligent Transportation Systems - (ITSC 2010)*, pp. 360–366, 2010.