Reference Systems for Environmental Perception: Requirements, Validation and Metric-based Evaluation

Mohamed Brahmi, Karl-Heinz Siedersberger, Andreas Siegel and Markus Maurer

Abstract— **The performance of advanced driver assistance systems (ADAS) is strongly dependent on the quality of the environmental perception sensors and algorithms used. This quality needs therefore to be assessed in order to allow an exhaustive evaluation of the ADAS perception system. For the quantitative assessment of the perception sensors and algorithms, a reference system, that can provide ground-truth information on the environment of a vehicle with higher quality than the ADAS sensors, is required. This paper defines and discusses requirements that a system has to satisfy in order to be used as reference system. In addition, a typical example of reference systems for ADAS perception is evaluated in terms of satisfaction of these requirements and validation with help of other external systems. Furthermore, this paper describes a metric-based evaluation of object-based perception systems with help of a this reference system.**

I. INTRODUCTION

A. Motivation: Use of Reference Systems

The evaluation of the ADAS perception system can be made in a formative or in a summative way. Formative evaluation aims at the assessment of a system or a process during the design and development. The assessment results obtained are used as feedback to improve this system or process. Instead, a summative evaluation deals with the assessment of the end performance of a system or a process without feeding back the obtained results.

Reference systems can be used at different stages during the pre-development process of ADAS systems as shown in Fig. 1.

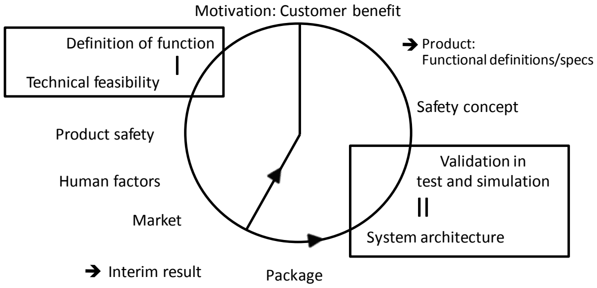

Fig. 1: Iterative ADAS development process [1]

1) Specification phase: During the specification phase of new ADAS functions, the requirements on the environmental perception are derived from requirements on the function. At this stage, the use of a reference system can help to verify whether the intended sensor system meets these requirements. Depending on the result of this verification, these specifications can be validated or changed without developing the whole system.

This measure can help to decrease development costs and time. These verifications allow the refinement of the specifications as well as the choice of the adequate perception sensors among different candidates for the intended ADAS system.

2) Development phase: For the development of perception algorithms, the detailed knowledge of the sensor behavior is crucial. Furthermore, sensor simulation models are essential for testing and optimizing the algorithms of ADAS systems. The development of such simulation models can be achieved by using reference systems that provide the ground-truth allowing the analysis of the sensor effects and properties such as separability, resolution and accuracy. These effects can be reproduced through these simulation models, as shown in [2], in order to provide the subsequent components of ADAS functions with artificial but realistic sensor measurements for test purposes. Once acquired, the knowledge about sensor effects leads to a correct interpretation of the sensor measurements in real time.

All these measures, as part of the formative evaluation, can help optimizing the perception systems of ADAS. At the end of this phase, final testing and benchmarking can be performed to assess the whole system in a summative evaluation.

B. Related Work

Different works used or developed a reference system for ADAS perception in several forms. In [3] and [4] a manual or semi-automated labeling was used for the generation of ground truth data. This results in inaccurate and incomplete reference data. [5] and [6] collect ground truth data with help of a mobile sensor setup. This data can only be used in specific applications (e.g., SLAM[1]). A similar reference system to the one used in this work is presented in [3] and considered as source of ground-truth data although its performance was not evaluated and validated. In summary, these works has not raised the issue of the requirements on a reference system for ADAS perception and its validation.

M. Brahmi and M. Maurer are with the Faculty of Electrical Engineering, Department of Vehicle Electronics, Technische Universitt Braunschweig, Germany `brahmi@ifr.ing.tu-bs.de`, `maurer@ifr.ing.tu-bs.de`

K-H. Siedersberger is with the Advanced Driver Assistance Development Group, AUDI AG, Germany `Karl-Heinz.Siedersberger@audi.de`

A. Siegel is with Audi Electronic Venture, Germany `andreas.siegel@audi.de`

---

[1] Simultaneous Localization And Mapping

The purpose of this paper is to define requirements that a system has to fulfill in order to be used as a reference system for ADAS environmental perception. Furthermore, a specific reference system is discussed in terms of satisfaction of these requirements and performance evaluation and a metric-based evaluation of a perception system with help of this system is presented.

II. REQUIREMENTS ON A REFERENCE SYSTEM

The requirements defined in this work can be divided in qualitative and quantitative requirements. The satisfaction of the qualitative requirements can be answered by Yes or No. Whereas, the quantitative requirements are described through quantities and values that must be specified.

A. Qualitative Requirements

1) Mobility: Since the system under test is operating under different conditions and in different environments, the reference system also has to be usable in different conditions and environments. To be qualified as mobile, a system must be portable and self mobile.

- Portability: a portable system is movable and should not be fixed to a predefined place.
- Self mobility: is the systems ability to move: thereby the ability to accompany the system under test.

Therefore, the reference system must be mobile in order to cover different aspects when evaluating the ADAS perception system.

2) Self-esteem: The reference system should be able to provide its own metrics of uncertainty that can be associated with the measurement results to verify its quality. These metrics, such as the standard deviation, must furthermore be realistic and be in agreement with the real measurement quality of the reference system.

B. Quantitative Requirements

1) Reliability: In order to obtain reference data, the reference system must be reliable. Accordingly, the reference data obtained from the system must guarantee a minimum required quality during a given specified period of time. The required period of time can be derived from the duration of a reference measurement where both reference and sensor data are collected.

2) Field of View: The reference system should be able to cover the entire Field of View (FOV) of the sensors under test even for sensor fusion configurations where the FOV is expanded. The needed FOV to be covered by the reference system can then be derived from the FOV of the system under test according to (1).

$$\textit{FOV-Volume}_{\textit{Ref}} \geq \textit{FOV-Volume}_{\textit{Sensor}} \tag{1}$$

3) Accuracy: One of the most crucial aspects of a reference system is its measurement accuracy. The requirements on the accuracy of the reference system can be derived from the required or expected accuracy of the sensor system. Generally, the accuracy of reference systems should be one order of magnitude higher than the accuracy of the system under test.

Since modern ADAS sensor systems are becoming more and more accurate, this requirement leads to a trade-off between the feasibility and required accuracy. Therefore, a slightly more accurate system than the system under test can be used as a reference system if its reliability is guaranteed and more accurate systems are unavailable.

These accuracy requirements on the reference system can be derived, if requirements on directly measured quantities are defined, by applying the rule mentioned above. Many other functions, however, have requirements on the accuracy of derived quantities which cannot be measured in a direct way. Through a measurement function in form of

$$\delta = f\left(\epsilon_1, \epsilon_2, \ldots, \epsilon_n\right) \tag{2}$$

the derived quantity δ can be obtained from the directly measured quantities $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$. The uncertainty of the derived quantity $\Delta\delta$ can then be obtained from the uncertainties of the directly measured quantities $\Delta\epsilon_1, \Delta\epsilon_2, \ldots, \Delta\epsilon_n$ by means of the linear law of error propagation according to (3), under the assumption that $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are uncorrelated.

$$\Delta\delta = \frac{\partial f}{\partial \epsilon_1}\Delta\epsilon_1 + \frac{\partial f}{\partial \epsilon_2}\Delta\epsilon_2 + \ldots + \frac{\partial f}{\partial \epsilon_n}\Delta\epsilon_n \tag{3}$$

Hence, the requirement on directly measurable quantities can be defined.

4) Timing:

a) Time delay: In order to be considered as ground truth, the reference data must be correctly time stamped. This can be achieved by satisfying the following relationship:

$$|T_{meas} - Timestamp| \leq \epsilon_{max} \tag{4}$$

where T_{meas} denotes the real measurement time and ϵ_{max} the maximal tolerable time delay. Since every system is affected by latency problems, the timestamp of the reference data must take into account this latency by compensating it internally or providing it to the measurement system. Hence, the sensor measurement data can be correctly associated and temporally aligned with the reference data for a proper comparison and evaluation.

b) Measurement rate: The reference system provides discrete measurements of the quantities in view. These measurements are mostly not synchronized with those of the sensor system and have therefore to be temporally aligned. This allows a reliable comparison between the reference data and the sensor data.

Consequently, the measurement rate of the reference system must be sufficient to detect small changes of the quantity in view and to reconstruct it. Generally, the measurement frequency of the reference system must satisfy at least the Shannon-Nyquist sampling theorem:

$$f_{ref} \geq 2 \times f_{veh} \tag{5}$$

where f_{veh} denotes the highest frequency component in the frequency spectrum of the considered vehicle signal. This condition is necessary but not sufficient to reconstruct the measured signal. Depending on the interpolation method between the samples, the signal can or cannot be correctly reconstructed. Generally, the interpolation error depends both on the sampling rate and on the used interpolation method. Using a linear interpolation method, it can be expected that the interpolation error decreases with an increasing sampling rate.

Besides, the signals of vehicle dynamics are band-limited signals due to limited acceleration and braking abilities of the vehicle. Spectral analysis of selected signals of vehicle dynamics, that describe the motion of center of gravity[2], shows that the highest frequency component is lower than 10 Hz (Fig. 2).

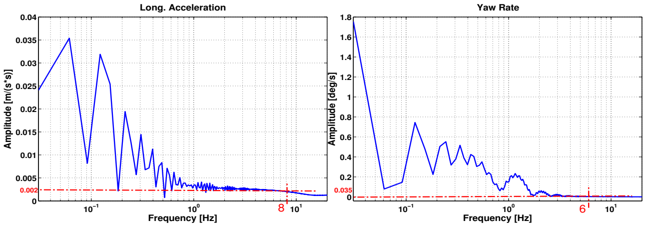

Fig. 2: Spectral analysis of the acceleration and yaw rate signals. The thresholds are set at $0,002\frac{m}{s^2}$ (left) and $0,035\frac{deg}{s}$ (right)

III. VALIDATION OF A REFERENCE SYSTEM

In this section, validation methods of reference systems are described. These methods are applied on a reference system used during our development activities of perception systems for ADAS.

A. Description

A **D**ifferential GPS aided **I**nertial **N**avigation **S**ystem (INS/DGPS) is used as reference system. It consists of an **I**nertial **M**easurement **U**nit(IMU) combined with a DGPS receiver on a navigation computer allowing the estimation of the position, velocity and orientation using the measured accelerations and rotation rates as well as the DGPS measurements. This combination helps to overcome the problems of the the low measurement rate and the poor availability of GPS. Using RTK[3]-DGPS techniques, the measurement of the absolute position can be achieved with 2 cm accuracy (horizontal RMS).

This system provides absolute measurements of the vehicle which is carrying it. However, ADAS sensors and perception algorithms generally provide relative measurements of the objects in the surrounding scene. If many vehicles are equipped with this INS/DGPS system and if they can communicate with each other, a localization of the ego-vehicle relative to target-vehicles in the scene can be obtained. Besides the dynamic objects, static objects can also be referenced. Therefore, their static DGPS position is measured off-line and their relative positions to the ego-vehicle are calculated by the computing unit at each step. This data can then be used as reference data for the ADAS sensors mounted on the ego-vehicle as shown in Fig. 3.

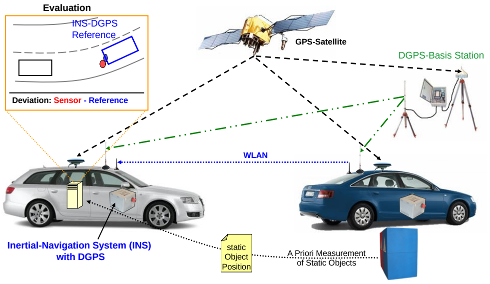

Fig. 3: Working principle of the INS/DGPS based reference system [7]

The referencing process consists of an on-line simultaneous collection of the measurements of both the reference and the ADAS perception system and a subsequent data post-processing for the evaluation of the ADAS perception system. An overview on this process is given in Fig. 4

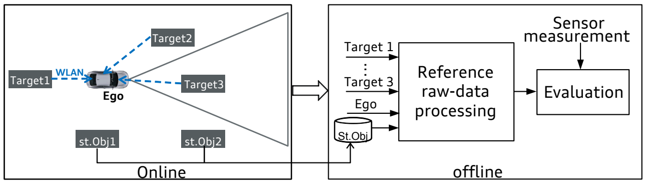

Fig. 4: Referencing process: first, on-line measurement are collected on a common car-PC, then ground-truth data is generated to evaluate the perception system

B. Validation of the Reference System

The reference data is the measurement data collected from the reference system and is considered as the ground-truth data to compare sensor measurements to. Therefore, the reference system used must be thoroughly verified and validated, in order to be qualified as a trustful source of reference data.

1) Satisfaction of the requirements: In order to validate the used reference system, it should be discussed to which extent the requirements derived above are satisfied by this system.

a) Mobility: this system is portable and self mobile and can therefore be mounted in and moved by test vehicles carrying the perception system under test. The differential correction data needed for RTK can be provided by local stations installed on test sites or by commercial base stations installed in different location and transmitting over GSM/GPRS.

b) Field Of View: due to the wide availability of GPS, the only restriction of the FOV is the range of the wireless communication (e.g., 250m), which is generally wider than the ranges of the ADAS sensors. Hence, the FOV, which can

---

[2] The shown signals have the highest dynamic among analyzed position, velocity and acceleration signals

[3] Real Time Kinematic

be covered by the presented reference system, is sufficient for evaluating ADAS perception sensors.

c) Reliability and self-esteem: The reference system can be qualified as reliable only if during a reference measurement the collected reference data can guarantee a minimum required quality. The information about this quality is given by the provided self-esteem. By driving repeatedly the same trajectory and with the help of a high accuracy laser sensor, detecting and extracting edges of buildings, the self-esteem of this system was evaluated by comparing the measurement uncertainty to the real measurements quality. Unlike other similar systems, this reference system provides realistic self-esteem information [7].

d) Time delay: This reference system timestamps its data using a synchronization mechanism based on the UTC time and an accurate synchronization signal provided by the GPS Card. As shown in Fig. 5, the time delay of this signal and the corresponding synchronization message is lower than 11 μs. All the measurements of this system are time stamped relatively to this signal which is transmitted to the host system with a very low delay through a high priority CAN[4]-message.

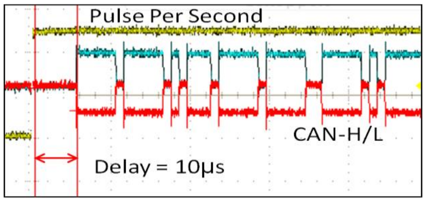

Fig. 5: Time delay of the synchronization message

e) Measurement rate: the INS/DGPS system used operates with a minimum frequency of 100 Hz providing new measurement data every 10 ms. Due to the limited motion dynamic of passenger cars, this update rate satisfies by far the Shannon-Nyquist Sampling Theorem and even rule of thumb (6).

$$f_{ref} \geq 10 \times f_{veh} \tag{6}$$

Additionally, the errors caused by a linear interpolation between two reference measurement timestamps were estimated and evaluated. Therefore, a 1 MHz simulation signal modeling realistic vehicle dynamics is sampled with $f_{Ref} = 100\,Hz$. Fictive measurements of a sensor system are then temporally aligned with the discrete reference data with a linear interpolation of the reference measurements exactly at the sensor measurements timestamps. Therefore, a sensor system with a sampling rate of

$$T_{CycleSens} := 40ms + \tau \tag{7}$$

is analyzed, where τ denotes the variable part of the sensor measurement cycle which is modeled by a white Gaussian noise $\tau = N(2, 4)$.

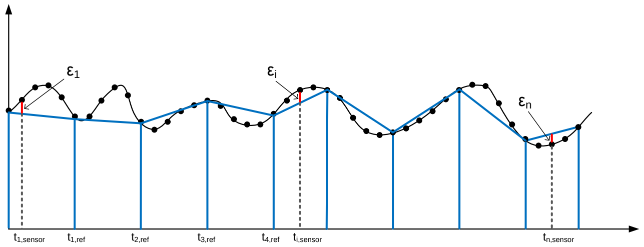

Fig. 6: Estimation of the interpolation error: for each sensor timestamp $t_{j,Sen}$, the closest reference timestamps $t_{i,ref}$ and $t_{i+1,ref}$ around $t_{j,Sen}$ are found for the linear interpolation

As illustrated in Fig. 6, at each step, the corresponding reference measurement is calculated by a linear interpolation between the two closest reference measurement timestamps $t_{i,ref}$ and $t_{i+1,ref}$ around the sensor measurement timestamp $t_{j,Sen}$. The corresponding reference measurement $ref_{i,interp}$ is calculated according to (8)

$$Ref_{i,interp} = ref_i + \frac{ref_{i+1} - ref_i}{t_{i+1,ref} - t_{i,ref}}(t_{j,Sen} - t_{i,ref}) \tag{8}$$

Hence, the so made approximation errors can be estimated as $\epsilon_i = Sim_i - Ref_{i,interp}$. The root-mean-square error, as criterion of the interpolation quality of the complete measurement sequence is obtained according to (9).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\epsilon_i^2} \tag{9}$$

The simulation results for selected signals are summarized in Table I and show that the errors caused by the linear interpolation are negligible. These obtained results prove that the measurement rate of the reference system is sufficient to reconstruct the real vehicle signals.

TABLE I: Estimated Interpolation Errors

Signal	Pos X	Pos Y	lon. Acc	hor. Vel	Yawrate
RMSE	0.1cm	0.01cm	0.02m/s^2	$10^{-5}m/s$	$10^{-4}rad/s$

2) Validation by means of external systems: In order to verify and validate the accuracy of the used reference system, independent and more accurate external systems are needed. The validation can be made either in static or dynamic situations.

a) Static Validation: Generally, INS/DGPS systems can, at best, perform absolute position measurements with the RTK-DGPS accuracy of 2 cm (horizontal RMS). Therefore, a system with a slightly better accuracy (e.g., 1cm) but a high reliability can be used for validating this reference system under predetermined conditions. For this purpose, a high end GNSS receiver is used for evaluating the static measurement accuracy of the reference system. The verification principle is shown in Fig. 7.

This receiver performs long term RTK-DGPS measurements of the position with predetermined required accuracy

---

[4] Control Area Network

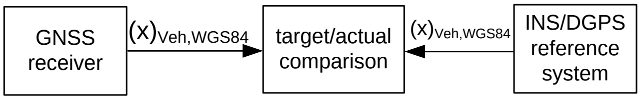

Fig. 7: Validation of the static position measurement

(e.g., 1 cm), so that only very accurate measurements are processed and accepted. This same position is then measured with the reference system allowing its static performance assessment (Fig. 8). The resulting measurement deviations between both systems are in the confidence range of 2 cm. Furthermore, this verification is performed prior to reference measurements in order to exclude decalibration errors.

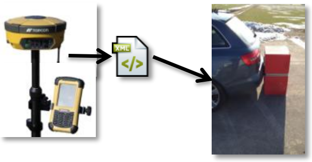

Fig. 8: Validation of the static position measurement: experimental setup

b) Dynamic Validation: Dynamic situations are more relevant and challenging for the ADAS perception systems. Therefore, the performance of the reference system must be assessed and validated in dynamic situations. For this purpose a stereo-video based validation system was developed. It consists of a stereo camera system which detects and localizes the vehicle that carries the reference system, with help of special landmarks mounted on the roof of the vehicle, as shown in Fig. 9.

The exact position of these landmarks, relative to the vehicle coordinate system, $(\vec{x})^{lm}_{veh}$ is measured with help of high accuracy tachymeter. In addition, the DGPS position of fixed landmarks $(\vec{x})^{lm}_{WGS84}$ is measured with help of the system used in III-B.2.a. The detection and localization of the vehicle relative to the camera system is triggered by the light barrier. This event is timestamped according to the UTC time allowing a precise temporal alignment for the subsequent processing. The vehicle position relative to the camera coordinate system $(\vec{x})^{veh}_{cam}$ is then obtained and the ground-truth GPS pose of the vehicle $(\vec{x})^{veh}_{WGS84}$ is deduced and on-line transmitted to the reference system for a target/actual comparison. Thus, the on-line dynamic validation of the reference system during or prior to reference measurements can be achieved.

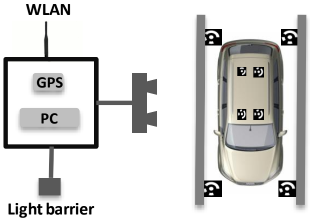

Fig. 9: Experimental setup of the on-line validation system

IV. METRIC-BASED EVALUATION OF PERCEPTION SYSTEMS

A. Pre-Processing

This section describes a metric-based evaluation of an object detection perception system with help of the reference system presented in Sec. III-A. In [8], a performance metric for image-based ADAS was developed. In this work, similar metrics are defined and implemented that, however, are not limited to image-based environmental perception. This metric-based evaluation is sensor independent and can even be applied on multi-sensor perception systems.

As shown in Fig. 4, sensor measurements are collected and the corresponding reference measurements are generated in form of lists of detected objects. The evaluation is based on a comparison between the reference and sensor object lists, where each object is described by a state vector $\vec{X}$. First of all, in a pre-processing, the sensor and reference measurement data must be temporally aligned (according to Fig. 6) and transformed into a common coordinate system. Then, a distance-based association is performed in order to assign the sensor measurements to the corresponding reference measurements. Thereby, a deviation state vector $\vec{\epsilon}_{ij}$, between the reference object i and the sensor object j, is calculated and an association indicator a_{ij} is obtained with help of a weighting matrix $w = diag[w_1, w_2, \ldots, w_n]$, that allows different weighting of measurement quantities.

$$\begin{aligned} \vec{\epsilon}_{ij} &= \vec{X}_{ref_i} - \vec{X}_{sens_j} \\ a_{ij} &= \sqrt{\vec{\epsilon}_{ij} \cdot w \cdot \vec{\epsilon}_{ij}} = \sqrt{w_1 \epsilon_1^2 + \ldots + w_n \epsilon_n^2} \end{aligned} \tag{10}$$

The resulting association indicator is then compared to a predefined threshold to decide whether the association was successful or not. Depending on the results of this association, different metrics are calculated.

B. Definition and Calculation of the Metrics

The result of the association is illustrated in the so-called Identification Graph (Fig. 10) and constitutes the basis for the further processing. Each reference object is represented with a black bar, whose start indicates the appearance of the reference object in the sensor FOV. The IDs of the sensor objects associated with their reference objects are color coded.

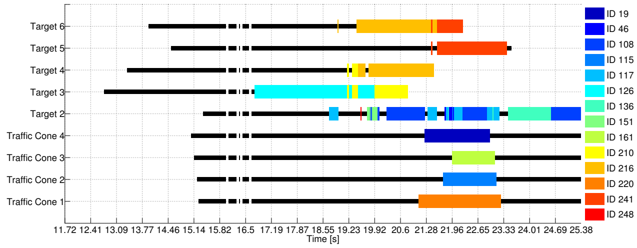

Fig. 10: Identification Graph

1) Multiple Track MT: This is given when a single reference object is detected by the perception system as two (ore more) sensor objects. The frequency of occurrence, can be taken for the assessment of the quality of the object building.

2) Multiple Object MO: This occurs when two or more reference objects are detected by the sensor system as only one object. This reflects the sensor separation abilities of the perception system.

3) True Positive TP $\rightarrow$ Localization Error LE: : In best case, a reference object is detected as a single sensor object. In this case, the localization error is analyzed by comparing some measurement quantities (such as $d_x, d_y, vel_{rel}, a_{rel}$). A statistical analysis is performed by calculating the precision and trueness of the measurements as a criterion for the accuracy.

4) False Positive FP: A big issue of perception systems is the detection of ghost-objects. Such a case is given when a sensor object is not assigned to any reference object.

5) False Negative FN: On the opposite side to FP, FN occurs when a reference object is not detected by the perception system although it is in its FOV. The occurrences of this cases are then calculated at each cycle and summarized for the measurement time.

Based on these metrics, other extended metrics, reflecting the tracking capabilities of the perception system, are derived.

6) Coverage CV: Is defined as the quotient of the TP occurrences to the number of objects that must have been detected by the perception system.

$$CV = \frac{TP}{TP + FN} \tag{11}$$

7) Object Purity OP: This describes the capability to track a reference object with the same identifier (best ID : bID) over time and is defined for each reference object i as

$$OP_i = \frac{n_i}{TP} \tag{12}$$

where n_i denotes the number of frames in which the reference object i is detected with bID.

8) Time To First Detection: Is the time until a reference object is detected by the sensor system and can be interpreted as an estimation of the time of response of the sensor system. An overview on these metrics is given in Fig. 11.

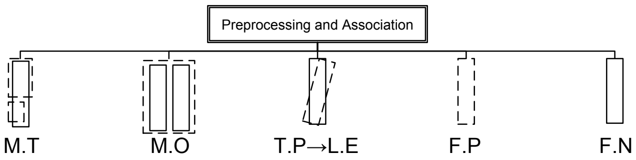

Fig. 11: Results of the association; — reference , - - - sensor

These metrics are automatically computed and reported for the chosen measurement scenario where the measurement environment is completely known and where the positions of and information on all reference objects are known during the measurement time.

C. Conclusion and future Work

The importance of reference systems for the ADAS environmental perception was raised in this paper. These reference systems allow the evaluation of the ADAS perception system in a formative way, towards its optimization and in a summative way for benchmarking and testing purposes. Therefore, these systems must meet some requirements in order to be qualified as reference systems.

This paper presents and describes a set of requirements that must be fulfilled. These requirements are derived from practical experience with different reference systems for the ADAS environmental perception. Furthermore, this paper describes ideas and methods for evaluating and validating a reference system so that it can be used as a trustful source of ground-truth reference data.

For the evaluation of the perception sensor systems with help of reference systems, different metrics and performance characteristics were developed. This metric based performance evaluation of the ADAS perception system and the corresponding performance characteristics are computed and generated automatically to allow an objective assessment of the sensor system performance.

In the future, the validation methods of the reference system will be extended to allow the performance assessment of the reference system in multi-object and dynamic situations. Ideally, systems based on other independent measurement principles (as GPS) are needed in order to allow an exhaustive evaluation of the reference system.

In order to make an objective comparison of different sensor systems or versions, an overall-score will be derived from the obtained metrics, which allows the assessment of a perception system with a minimal and compact set of criteria.

REFERENCES

[1] M. Maurer and H. Winner, *Automotive Systems Engineering*. Springer-Verlag GmbH, 2013.

[2] M. Brahmi, “Entwicklung, Implementierung und Validierung eines Sensormodells fuer zukuenftige Radarsensorik,” Master’s thesis, Technical University Munich, Munich, Germany, 2010.

[3] R. Lindl, “Tracking von Verkehrsteilnehmern im Kontext von Multisensorsystemen.” Ph.D. dissertation, Technical University Munich, 2009. [Online]. Available: http://mediatum2.ub.tum.de/doc/667321/document.pdf

[4] R. Kastner, T. Kuhnl, J. Fritsch, and C. Goerick, “Detection and motion estimation of moving objects based on 3d-warping,” in *Intelligent Vehicles Symposium (IV), 2011 IEEE*, june 2011, pp. 48 –53.

[5] A. Takeuchi, M. Shneier, T. Hong, T. Chang, and G. Cheok, “Ground truth and benchmarks for performance evaluation,” in *SPIE Aerosense Symposium*, 2003.

[6] J.-L. Blanco, F.-A. Moreno, and J. Gonzalez, “A collection of outdoor robotic datasets with centimeter-accuracy ground truth,” *Autonomous Robots*, vol. 27, pp. 327–351, 2009. [Online]. Available: http://dx.doi.org/10.1007/s10514-009-9138-7

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

[8] 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*, 2011, pp. 381–386.