A METHOD FOR THE CHARACTERIZATION OF PERCEPTION SENSORS

Loren Stowe, Greg Beale, Matt Palmer
Virginia Tech Transportation Institute (VTTI)
United States

Steven Huggins
Global Center for Automotive Performance Simulation (GCAPS)
United States

Sebastian Silvani
National Highway Traffic Safety Administration (NHTSA)
United States

Paper Number 23-0160

ABSTRACT

Vehicle perception systems for both advanced driver assistance systems (ADAS) and automated driving systems (ADS) rely on a plurality of sensors and sensor modalities to "see" the surrounding environment. Each sensor type has its own inherent strengths and weaknesses (e.g., cameras perceive color but need ambient light, radar is insensitive to light but does not perceive color). The goal of this study was to develop systematic and adaptable tests and analysis methods that would allow the performance characterization for a variety of sensors and sensor types. Three common sensor modalities (radar, LiDAR, and camera) were selected for demonstrating the application of the methodology.

Three test maneuver templates were developed to exercise the relative motion of target objects within the sensor's field of view (FOV). These allowed a broad set of conditions to be configured that corresponded to ones that might occur during real-world driving. These were combined with external conditions (e.g., simulated rain, variable ambient light, sensor degradations) to identify compounding factors that may affect sensor performance. Sensor characteristics and test factor sensitivities were then calculated across different metrics, including distance accuracy and maximum detection range, to demonstrate the process and efficacy of the method in characterizing perception sensor performance.

INTRODUCTION

Safe operation of a vehicle is dependent on an accurate awareness of the driving environment. ADS, as well as lower levels of automation found in ADAS, rely on varied sensor modalities and technologies to gain this awareness. Currently, a review of the Voluntary Safety Self-Assessment (VSSA) [1] reports published by ADS developers indicate that radar, camera and LiDAR sensors are common sensors being used in ADS perception systems. Though each original equipment manufacturer (OEM) may develop unique algorithms for object perception and classification, perception starts with and acts on the output from sensors.

While the sensor manufacturer provides performance specifications for their products, these are typically based on standardized test specifications that are performed in a highly controlled environment. These results are not necessarily indicative of the performance realized in the final operating environment. Other studies [1] have demonstrated that the lab-based methods for defining spec sheet performance can be inadequate at capturing the applied performance of the sensor (Figure 1).

[1] These can be accessed at https://www.nhtsa.gov/automated-driving-systems/voluntary-safety-self-assessment.

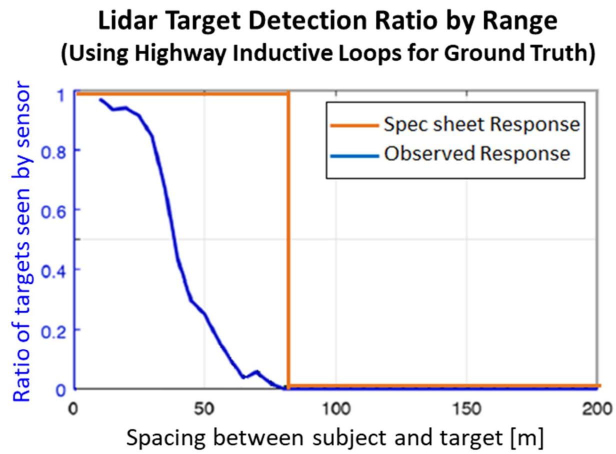


Figure 1. Actual LiDAR target detection ratio versus expected from specification sheet [1].

RESEARCH QUESTION / OBJECTIVES

The objective of this study was to develop test and analysis methods that can be used to characterize the performance of a variety of perception sensors in a variety of operational conditions. This was accomplished by

- Surveying test procedures and measures used by industry
- Identifying potential gaps in current practices for characterizing performance in operational conditions
- Developing a method for efficiently testing and characterizing a variety of sensors under a range of conditions

METHOD

The study sought to develop flexible and efficient characterization methods that could apply to a range of sensors with a variety of inherent sensitivities (e.g., camera to light, LiDAR to precipitation) and data outputs (e.g., object information, point cloud) that were designed for different functions (e.g., forward safety applications, parking assist). To facilitate this, the study explored functional testing to ensure a more direct link between test results and sensor performance in the applicable operational design domain (ODD). The development of the methodology consisted of three primary objectives.

1. Develop repeatable and adaptable tests based on the expected operational conditions such as relative motion with objects of interest and environmental factors.
2. Characterize sensor performance with a common set of metrics defined as a function of the test inputs to facilitate comparisons of sensors and sensor modalities.
3. Capture key sensor performance traits including the following:
    a. Nonlinear responses
    b. Temporal effects
    c. Interactions
    d. Degradation effects

Test Templates

To support these objectives, three test templates were defined to allow a wide range of configurations that could be used to capture primarily longitudinal, lateral, and field of view assessment. For the dynamic tests, ground truth measurements were obtained using differential global position satellite (DGPS) receivers on the subject vehicle (SV) and the target vehicle (TV). This functionality is indicated as the vehicle reference system (VRS) in the images below. For the static objects (e.g., signs, lane markings), the positions were previously mapped with DGPS.

The range sweep test (Figure 2) varied the distance between the SV and TV by adjusting the relative speed between the two vehicles. The SV and TV started at the minimum following distance at the baseline test speed. The SV decreased its speed by a defined amount below the nominal speed to increase the headway for the first half of the trial. When the sensors' maximum detection range was reached, the SV increased its speed above the baseline speed to reduce the headway distance, down to a predefined minimum safe following distance. The baseline speed, relative speed, lane offset, and road configuration can be adjusted to evaluate the effect of these variables on sensor performance.

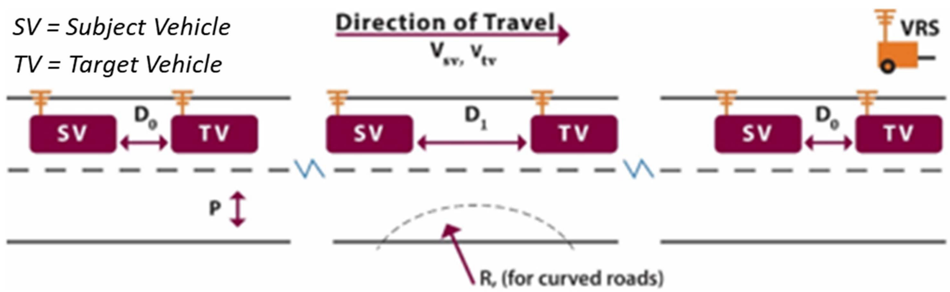


Figure 2. Schematic of Range Sweep Test Template

The lane slalom trials were designed to sweep the target laterally in a sensor's horizontal FOV. In a lane slalom trial, the SV repeated s-turns while approaching the stationary TV, which was parked one or more lanes over from the SV's path (Figure 3*)*. The frequency (cycles per second) and amplitude of the s-turns was varied to create increase the rate the TV moved in the FOV and the distance it moved within the FOV. Expressed as a percentage of lane width, the amplitude of these s-turns ranged from 50 percent of lane width to 200 percent or two-lane widths. The TV offset and road geometry can also be varied.

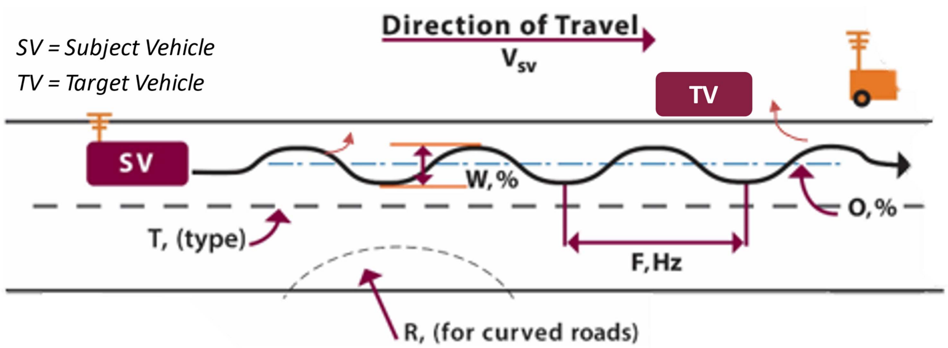


Figure 3. Schematic of Lane Slalom Test Template

The turning circle test was designed to test a sensor's ability to follow a turning vehicle at several viewing angles and range locations (Figure 4). The SV was stationary while it observed the TV traveling in a circle at a constant radius (R_{TV}). The SV's distance from the TV's circle and speed of the TV can be altered. In addition, the orientation of the SV relative to the circle can be changed to test the performance in different regions of the sensor FOV including the edge.

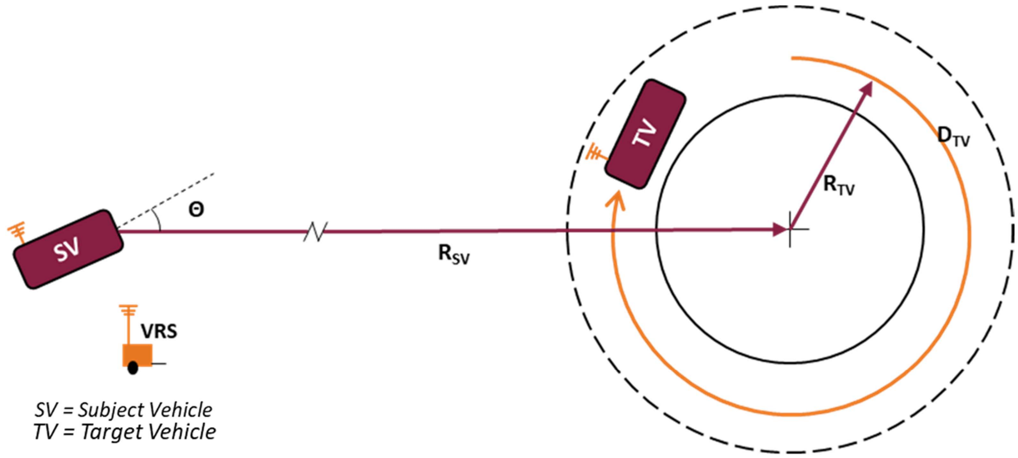


Figure 4. Schematic of Turning Circle Test Template

Table 1 provides a list of the parameters associated with the three templates and an example set of values used during the study. Bold values indicate values used for baseline conditions.

Table 1. Example Parameter Definitions for Test Templates

Variable	Range Sweep Test	Slalom Test	Turning Circle
Distance (D)	*40 -> 2 -> 40 (m)*		
SV Velocity (Vsv)	*Vtv +/- dD*	10, **15**, 20 (mph)	
TV Velocity (Vtv)	10, **15**, 20 (mph)		5, 10, **15** mph
Rate of Change of D (dD)	3, **5**, 10 (mph)		
Position (Lane) of TV (P)	-100%, **0**, +100%		
Radius of Curvature (R)	**0**, 15 (m)	**0**, 15 (m)	
Type (T)		*Solid, Dashed, Straight, Curved, White, Yellow, Ped X-ing, Stop Bar*	
Amplitude/Width (W)		**50**, 100, 150 (%)	
Frequency (F)		**0.4**, 0.75, 1 (Hz)	
Offset (O)		**0**, 25, 50 (%)	
Turning Radius (Rtv)			30, **40**, 50 m
Viewing Range (Rsv)			15, 25, **35**, 45 m
Viewing Angle (Θ)			-30, **0**, 30 deg

Workflow for Test Methodology

The test maneuvers provided a set of parameters that could be varied. Additional elements were included (e.g., rain, sensor degradations) to provide a characterization that could be more representative of the real-world performance. This, along with the different sensor modalities, creates the potential for a large test space. To help make the test space more manageable and relevant for a given sensor and/or application, the following process (Figure 5) was developed to help guide the development of a test plan, resulting in a final test matrix.

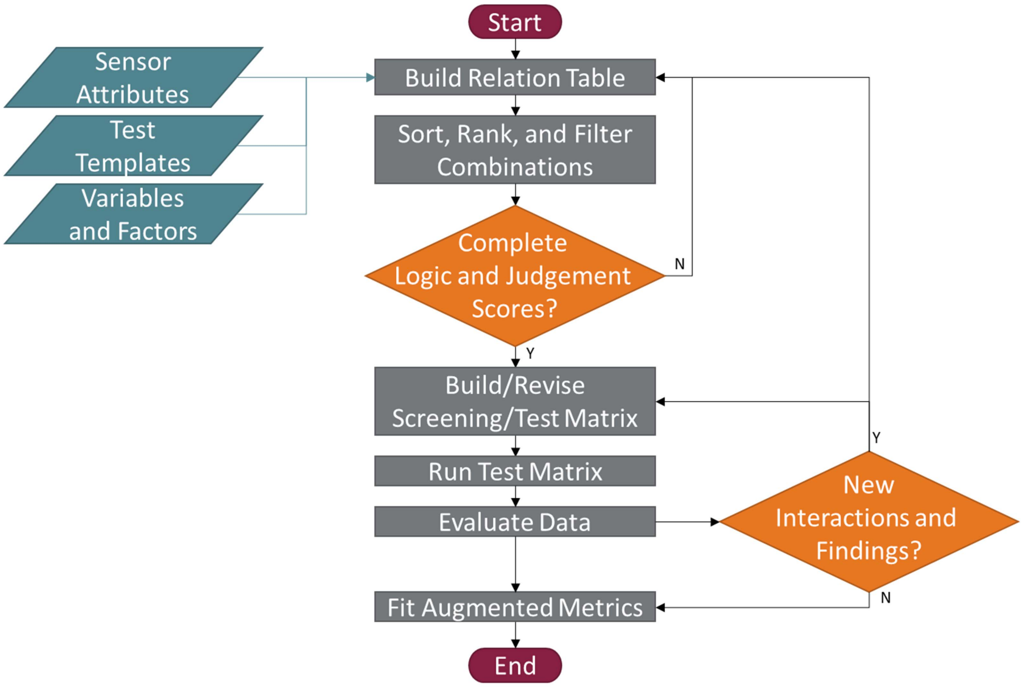


Figure 5. Process flow for the development and execution of the sensor characterization test plan.

Sensor attributes help define the measures of interest and the test conditions. The intended application for the sensor and the environment in which it may operate are important when defining the test template parameters and test conditions. For example, a radar and a camera intended for forward safety applications will likely have different test conditions (e.g., lighting) that are useful for characterizing each sensor's performance efficiently and accurately. Similarly, the same sensor modalities intended for parking assist applications will have different test conditions and potentially different measures of interest i.e., object position and pose accuracy may be more relevant than max range and object velocity.

A key part of the framework is the test templates introduced above that can be configured to specify a variety of maneuvers, relative motions, and road definitions The framework considers general sensor attributes as well as test variables and factors to define the test template parameters and the conditions which these maneuvers will be executed.

In addition to the sensor considerations, the intended operational conditions help identify test variables and factors of interest. While the influence of variables and factors are treated the same in the analysis, they were kept separate to distinguish between test maneuvers and the conditions in which the maneuvers were carried out. Test variables were used in the test templates to define the test maneuvers (e.g., vehicle velocity, following distance, radius of curvature, lane position). Test factors define the conditions the tests are performed (e.g., ambient light, rain rate, sensor surface degradation). Taken together, variables and factors are referred to as parameters.

Once the sensor attributes, test maneuvers, and parameters of interest are defined, they are organized into a relation table to help prioritization the test conditions. Since the number of test conditions has the potential to be very large, this step facilitates a structured down selection process to improve testing efficiency. It also provides an intermediate step in defining the test matrix. Figure 6 shows an example of an excerpt from the relation table and subsequent test matrix.

Index	Attribute	Test	Variable or Factor	Logical Check	Time Added	Development Required	Triviality	Camera	LiDAR	Radar	Rank
2297	Lane Width Accuracy	Lane Slalom	Mount Displacement	1	5	5	4	1	1	0	95%
2858	Pose Accuracy	Reveal	Degradation #1	1	5	5	5	1	1	0	90%
2323	SV Lane Position Accuracy	Intersection	Subject Vehicle Velocity	1	4	4	5	1	1	0	90%
391	Lane Position Accuracy	Ranging Sweep	Mount Displacement	1	4	5	5	1	1	0	90%
2256	Noise Succeptability	Intersection	Degradation #3	1	5	4	4	1	1	1	90%
2380	Object Resolvability	Lane Slalom	Degradation #2	1	4	5	4	1	1	1	90%
2398	Lane Position Accuracy	Intersection	Degradation #4	1	5	5	4	1	1	0	90%
2444	Pose Accuracy	Intersection	Ta[illegible]ehicle Velocity	1	3	4	5	1	1	0	85%

File Code	Test	Location	Variant	Speed	Offset	Rate
2.2.1.01	Ranging Sweep	Highway	Baseline	45 mph	0%	5 mph
2.2.2.01	Ranging Sweep	Highway	Velocity Sensitivity	5.0 mph	0%	5 mph
2.2.2.02	Ranging Sweep	Highway	Velocity Sensitivity	10.0 mph	0%	5 mph
2.2.2.03	Ranging Sweep	Highway	Velocity Sensitivity	15.0 mph	0%	5 mph
2.2.2.04	Ranging Sweep	Highway	Velocity Sensitivity	20.0 mph	0%	5 mph
2.2.2.05	Ranging Sweep	Highway	Velocity Sensitivity	25.0 mph	0%	5 mph
2.2.2.06	Ranging Sweep	Highway	Velocity Sensitivity	30.0 mph	0%	5 mph

Figure 6. Excerpt from a relation table (top) and test matrix (bottom).

Construction of the final test matrix was an iterative process. The test configurations identified in the test plan framework were used to populate a relation table that provided a means to qualitatively assess different test parameter combinations and prioritize which configurations to include in the initial testing. The first round of data collection employed a screening matrix consisting of two to four levels for a given set of parameters to help identify interactions, sensor sensitivities, and areas that may need additional levels. Figure 7 provides and examples of how the hypothetical or expected response can be used to identify initial parameter values. The left graph shows that the expected position accuracy will have a non-linear relationship to range and rain intensity and there will be an interaction between these two variables. The right graph provides a sampling strategy for the initial values for the screening matrix (levels shown in orange). Results from the initial test are then used to identify areas that require additional levels of rain intensity and range (shown in maroon) to be able to characterize the performance of the sensor more accurately.

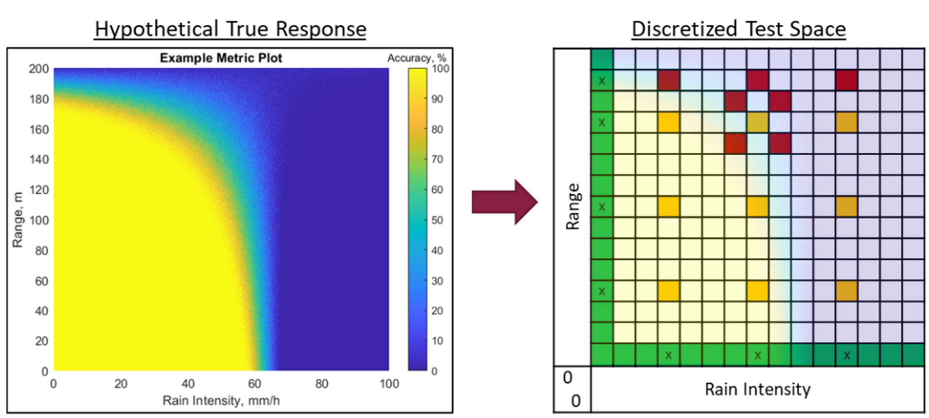


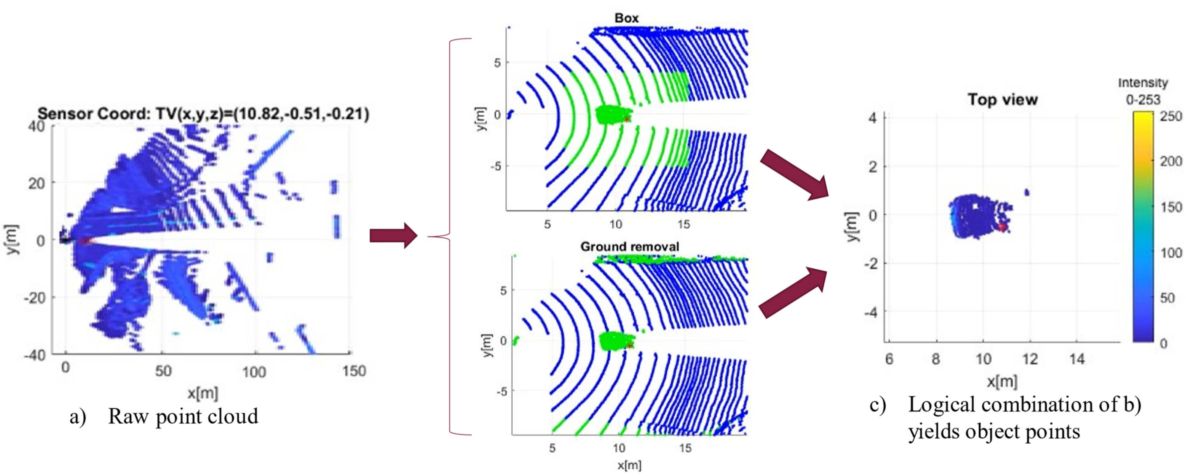
Figure 7. Example of how a screening matrix can be used to select relevant parameter levels.

Data Processing

Figure 7 also provides a visual representation of what data points might be needed to create a relationship for the multi-way dependencies. By fitting an equation to the surface response, sensor accuracy as a function of both range and rain intensity can be found. This can be extended to higher dimensions as well.

As noted above, one of the objectives for the test plan was to facilitate comparison of different sensors and sensor types. However, not all sensors provide the same native output. Many sensors have on-board processing that allows them to provide a list of objects with relevant measures. Consequently, data from sensors that provided less processed data e.g., point cloud from LiDAR or pixel information from a camera, had to be processed to extract object information. The goal of the study was not an evaluation of different object identification algorithms. Therefore, the team used a priori knowledge of the location of the object of interest from the DGPS as a basis for distinguishing which points in the sensors' output were associated with the TV.

Error! Reference source not found. shows an example of the steps used to identify the TV. The top figures provide a bird's eye view of the point cloud. Subplot (a) shows the full FOV for this LiDAR. Subplot (b) is zoomed into the region around the SV and TV. In both, the SV is located at (0, 0) and the TV, represented by the red dot, is located at (10, 0). The first frame (a) shows the density of information provided by the LiDAR. Subplot *b* (right) is zoomed closer to the TV position (note the different axis scale) and highlights the steps used to extract the points associated with the TV. First a bounding box was created based on the DGPS position (shown in green). An algorithm was used to identify the points in this region that were associated with the ground plane. Any points in the point cloud not in this plane were then separated. The ground plane points are shown in blue in the lower image of (b) and all other points are shown in green. The final step extracts the points around the TV position from the DGPS that are not part of the ground plane. The resulting points are shown in subplot (c) where the color indicates the intensity of the returned signal with the rear of the vehicle having a stronger return than the other surfaces. Using this approach allowed the team to have object data from the LiDAR sensors structured in a manner consistent with radars and other object level sensors.



a) Raw point cloud

b) DGPS centered bounding box and rejected ground plane

c) Logical combination of b) yields object points

Figure 8. Identification of the target vehicle using the DGPS-seeded bounding box.

DATA SOURCES

The sensor types for this research were limited to perception sensors (rather than localization sensors), including various types of radar, LiDAR, and cameras intended for use in ADS-equipped vehicles. The method developed and applied in this research effort would also apply to other perception sensing technology. Though localization sensors are an important part of ADS operation, GPS and inertial sensors were not included in the testing portion of this research.

The sensors used in this research included four radar, four LiDAR, and two camera sensors. Table 2 lists the sensors, including the type, reference name, a brief description of the sensor, and the native output from the sensor. One of the 77-GHz radars produced two datasets: firstly, a short range (SR) set with a wide field of view (FOV) without any vertical information, and secondly a long range (LR) set with a narrow FOV and vertical location information. As noted previously, DGPS was included in both the SV and TV to provide ground truth measurements.

Table 2. Sensors Included in Research

Sensor Type	Anonymized Name	Brief Description	Sensor Output
Camera + processor	Object Camera	Mono camera with on-board processing for object and lane tracking	Identified Objects
Camera	Machine Vision	Mono camera, Color	Pixels (photos)
LiDAR	32	32 beam,360 degrees FOV	Point Cloud
LiDAR	128	128 beam,360 degrees FOV	Point Cloud
LiDAR	Solid State	Solid state LiDAR, forward FOV, multiple beams	Point Cloud
LiDAR	Budget	Mechanical LiDAR, forward FOV, single beam	Point Cloud
Radar	SR 24 GHz	Short range radar operating at 24 GHz.	Bulk Objects
Radar	SR 77 GHz	Short range radar operating at 77 GHz.	Point Cloud
Radar	LR 77 GHz	Long-range radar operating at 77 GHz with two scan formats: 1. Wide near field and 2. Narrow far field	Point Cloud

A t-slot frame was used to mount the sensors to the SV. The frame provided a means to mount and align a variety of sensors and simultaneous collection data from all sensors. This ensured the test conditions were the same for each sensor, improving collection efficiency.

The data were collected asynchronously with their own time stamp to utilize each sensor's maximum sample rate. This flexibility required additional steps during data processing but allowed the team to collect data from a variety of sensors simultaneously during each trial.

RESULTS

The research resulted in several key findings. First, it confirmed that the sensor performance observed in representative driving conditions differed from the performance listed in sensor specification sheets (Table 3). Similarly, degraded conditions tended to negatively impact sensor output as can be seen in the last two columns.

Table 3. Comparison of Measured Maximum Range to Specification Sheet

	Sensor	Specification Sheet (m)	Baseline (m)	Percent Change	In Rain (m)	Percent Change
	Object Camera	250	133.3	**-47%**	116.3	**-53%**
LiDAR	32-Beam	100	37.5	**-63%**	29.2	**-71%**
LiDAR	Solid State	600	145.0	**-76%**	113.7	**-81%**
Radar	SR 77 GHz	64	66.4	**4%**	66.2	**3%**
Radar	SR 24 GHz	95	98.0	**3%**	24.8	**-74%**
Radar	LR 77 GHz (wide)	100	102.2	**2%**	100.5	**1%**
Radar	LR 77 GHz (far)	250	203.2	**-19%**	203.2	**-19%**

The effect of degradation on performance were often non-linear and dependent on the type of degradation and the sensor. Rain resulted in a decrease in the detection range by at least ten percent for the nine sensors tested. While lighting primarily affects camera performance, on-coming headlights at dusk did result in a measurable difference in distance accuracy for one LiDAR. Surface degradations, which attenuate or distort the signal, tended to reduce the performance of the sensor. However, the effect was dependent on the sensor and the degradation (both type and level).

The following plots further explore the effect of rain. The baseline ranging sweep test was performed along with four lane offset conditions from one half lane (50%) to two full lanes (200%). Rain towers were positioned along the roadway and the vehicles executed the test maneuver with and without rain.

Camera performance was particularly susceptible to rain conditions (Figure 9). In addition to attenuating the signal, water droplets distort the light. In addition, while the lateral offset did not have a strong influence on distance error without rain, there was an influence in the presence of rain.

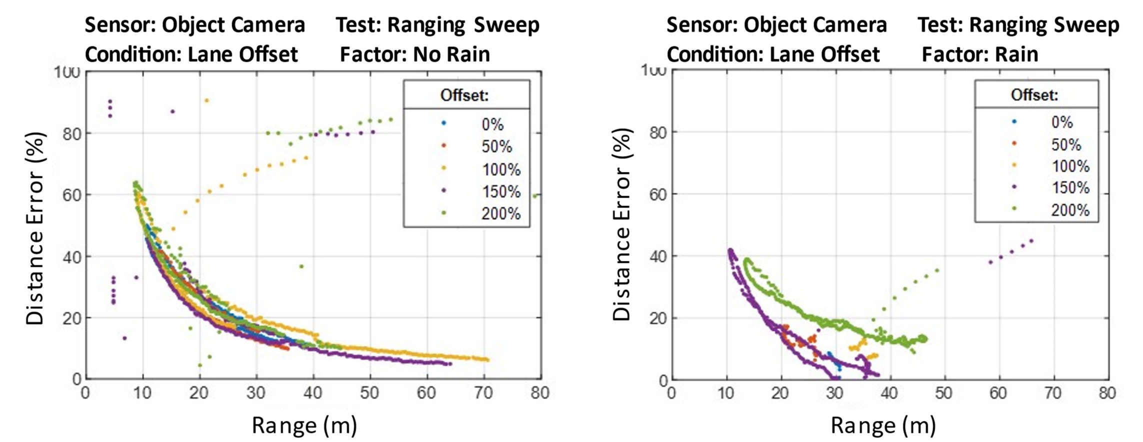


Figure 9. Object camera detection accuracy plots without rain (left) and with rain (right).

LiDAR performance also deteriorated with increased rain (Figure 10). Water has similar detrimental effects for LiDAR and camera. Rain attenuates and blocks the laser pulses to and from the target. This was observed with both the solid-state and 32-beam LiDAR. The LiDAR does not display the sensitivity to lane offset in the presence of rain as was observed for the camera.

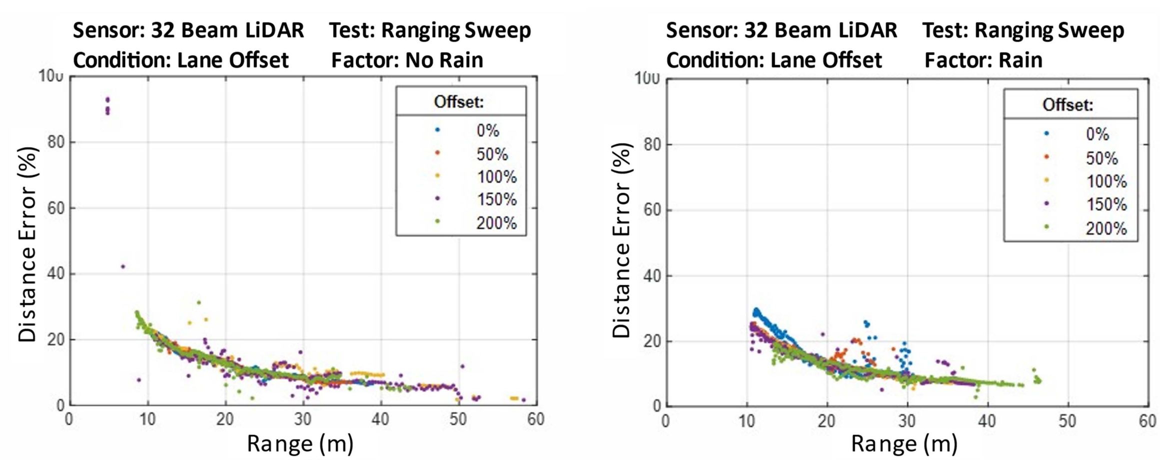


Figure 10. Thirty-two beam LiDAR detection accuracy plots without rain (left) and with rain (right).

As noted earlier, the 24 GHz sensor was the only radar that was significantly affected during the rain trials (Figure 11). Because the 24 GHz radar reports objects, not point clouds, the object tracking software within the object radars could be more susceptible to reduced detection range than those radars that just report point clouds.

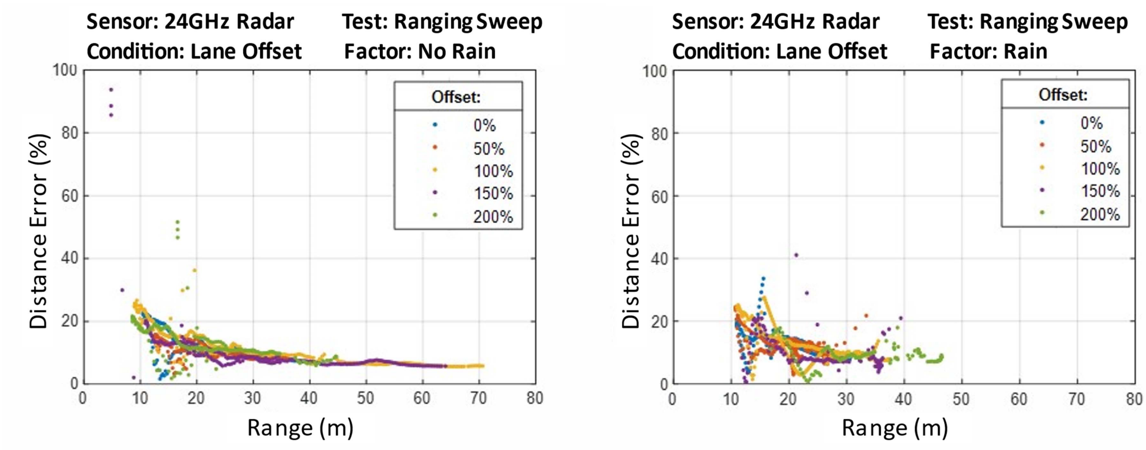


Figure 11.SR 24 GHz radar detection accuracy plots without rain (left) and with rain (right).

Optical degradations can affect LiDAR intensity values and number of points. Adding the sandblasted degradations (which impacts transmission and scattering of light) to the solid-state LiDAR in the ranging sweep trials decreased the maximum and average intensity values (Figure 12). The high-level sandblasted degradation resulted in very few returned points. While this level of surface wear may not be realized, it does provide a limit condition for characterization testing.

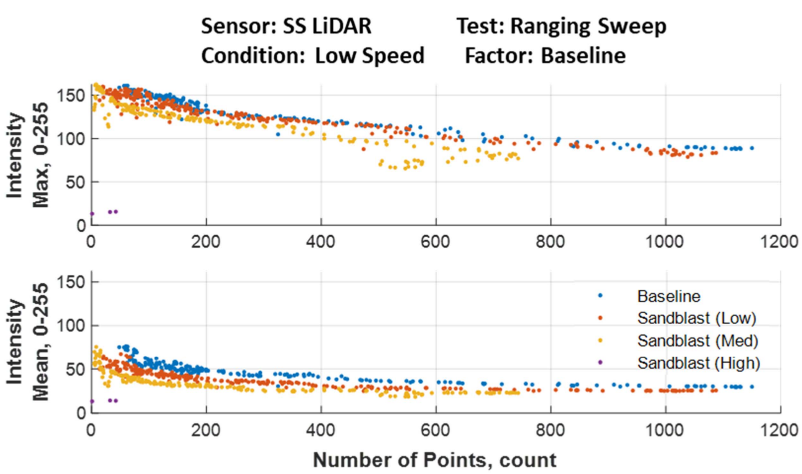


Figure 12. Intensity values for the solid-state LiDAR with three levels of sandblasted degradations.

DISCUSSION AND LIMITATIONS

Standardized tests are designed to minimize noise and subsequently test in what could be considered best-case conditions with ideal targets. However, this may not provide a good indication of the performance in the driving environment. The approach presented can be applied across a variety of sensors to provide metrics to compare different sensors and sensor types for a common set of conditions. The method developed provides a framework to guide testing and analysis. The use of the method will require the practitioner to select factors and variables that are relevant to the sensors and applications of interest. This could include additional test conditions, external factors, and metrics.

As noted, this research focused on sensor evaluation rather than perception software. The methods used for extracting object information is not indicative of how it would be accomplished in system integrated into a vehicle. The method employed used the DGPS signal and knowledge regarding the size of the target to identify points associated with the target object. Consequently, the analysis could identify individual returns associated with the object of interest. Since the DGPS is doing the job of the perception system in identifying the target, it is unlikely that the system, when integrated into a vehicle, will be able to detect objects with this limited of data. Without an understanding of the constraints of the object detection algorithm to be used, this could lead to an unrealistic expectation of system performance based on the characterization results. For example, if an algorithm needs a minimum number of points distributed across the vehicle to perceive it in the point cloud, the maximum detection range may be lower than what the characterization data shows. Therefore, it is important to understand the method and its assumptions in the context of how the sensor may be used and the associated data processing. Though not discussed, this also holds true for evaluating lane line detection performance for sensors that do not natively output this information.

Another consideration in the use of this method is that it, by design, is unlike a standardized test which is tightly defined to help ensure repeatability and reproducibility. This method requires careful documentation and a means to account for external factors that could affect repeatability (e.g., sunlight) or that may vary and effect reproducibility (e.g., the use of different target vehicles). The quality of the ground truth measurement system could also influence the results. The DGPS system used for this study could measure the location to within 10 cm, or approximately the width of a lane line. This may be adequate for many applications. The performance of a different ground truth system may or may not be sufficient for another party's system requirements.

CONCLUSIONS

The research team developed a methodology that provides a structured way to characterize the performance of sensors and sensor systems. The method defines a set of generalized test templates, a means to select the most relevant test factors, and an approach for analysis that includes multi-factor metric fitting to facilitate sensor performance characterization in the expected operational conditions. This approach targets representing the final performance rather than the current sensor data provided based on laboratory tests.

However, the test space can become large when considering the different parameters and combinations. The method outlined provides a framework to aid in the identification and selection of key test configurations. The flexibility associated with the method also makes it applicable to sensor systems that provide object level data. This flexibility does come at a cost. Unlike a standardized test where the steps and conditions are tightly defined, successful application of developed method benefits from knowledge and understanding of the sensors, their potential applications, operational conditions, and relevant measures and metrics.

It is also important to apply the results appropriately. While the results from this method may not necessarily reflect the final performance of the integrated perception, it does provide a means to generate results that more closely reflect how a sensor would likely perform in the automotive environment.

REFERENCES

[1] Wu, M., & Coifman, B. (2019, 10/01). Quantifying what goes unseen in instrumented and autonomous vehicle perception sensor data – A case study. *Transportation Research Part C: Emerging Technologies, 107*, 105-119. https://doi.org/10.1016/j.trc.2019.07.024.