2021-01-0175 Published 06 Apr 2021

# Infrastructure-Based Sensor Data Capture Systems for Measurement of Operational Safety Assessment (OSA) Metrics

**Niraj Altekar** The University of Arizona

**Steven Como and Duo Lu** Arizona State University

**Jeffrey Wishart** Exponent Inc.

**Donald Bruyere** The University of Arizona

**Faisal Saleem** Maricopa County Department of Transportation

**K. Larry Head** The University of Arizona

***Citation:*** Altekar, N., Como, S., Lu, D., Wishart, J. et al., "Infrastructure-Based Sensor Data Capture Systems for Measurement of Operational Safety Assessment (OSA) Metrics," SAE Technical Paper 2021-01-0175, 2021, doi:10.4271/2021-01-0175.

## Abstract

The operational safety of automated driving system (ADS)-equipped vehicles (AVs) needs to be quantified for an understanding of risk, requiring the measurement of parameters as they relate to AVs and human driven vehicles alike. In prior work by the Institute of Automated Mobility (IAM), operational safety metrics were introduced as part of an operational safety assessment (OSA) methodology that provide quantification of behavioral safety of AVs and human-driven vehicles as they interact with each other and other road users. To calculate OSA metrics, the data capture system must accurately and precisely determine position, velocity, acceleration, and geometrical relationships between various safety-critical traffic participants. The design of an infrastructure-based system that is intended to capture the data required for calculation of OSA metrics is addressed in this paper. The designed multi-modal sensor system includes a combination of traffic video cameras, vehicle-to-infrastructure (V2I) roadside units (RSUs), National Transportation Communications for Intelligent Transportation System Protocol (NTCIP)-compliant signal controllers streaming Signal Phase and Timing (SPAT) data, and Light Detection and Ranging (LIDAR) sensors. The system is contrasted with other design options to evaluate trade-offs between capability and cost. The designed data capture system was deployed at a SMARTDrive Program℠ Test Bed intersection in Anthem, AZ that has been developed by the University of Arizona Transportation Research Institute (TRI) in cooperation with the Maricopa County Department of Transportation (MCDOT). The intersection is equipped with a sensor system that includes a fiber optic data transfer backbone to support the data transfer to a server at the MCDOT Traffic Management Center. A measurement uncertainty (MU) analysis has been conducted using experimental data to better understand the performance and reliability of the proposed sensor system design. The data capture system will enable the development and validation of a methodology to continuously measure OSA metrics by gaining rich information through fusion of multi-modal data collected from available sources for safety assessment of the transportation system that includes AVs.

## Keywords

Traffic safety assessment, Data capture, Infrastructure, Vehicle-to-infrastructure, Sensors

## Introduction

The pace of advancements towards commercialization of automated driving system (ADS)-equipped vehicles (AVs) is rapidly increasing. Currently, human error is cited as the "critical reason" (meaning that human error was the "last event in the crash causal chain") for over 90% of crashes in the United States [1]. By gradually eliminating human involvement from the driving task, AVs have a potential to be safer than conventional human-driven vehicles [2]. However, the behavior of human-driven vehicles is more comprehensible to the public as compared to AVs that heavily depend on computer-based technologies which lack the desired transparency desired by the public [3]. The artificial intelligence based operational nature of AVs and minimal data

shared by AV developers make their safety assessment a challenging process. In prior work completed by the Institute of Automated Mobility (IAM), operational safety assessment (OSA) metrics were introduced to quantify the risk level associated with introduction of AVs on public roads [4]. The OSA metrics can be categorized as:

- Black Box metrics that do not require on-board vehicle data, including ADS data (although ADS and other on-board data can be used)
- Grey Box metrics that require limited access to ADS data (but not to proprietary AV developer information such as perception data)
- White Box metrics that require extensive access to ADS data (such as perception data)

The IAM has continued to develop the OSA metrics through validation and refinement via simulation [5] but also through real-world data capture via infrastructure-based sensors. Infrastructure-based sensors are capable of capturing data to support estimation of Black Box metrics.

Different types of infrastructure-based sensors have been utilized by past studies for collecting safety data. Saunier and Sayed [6] developed a method to extract vehicle trajectories from video data collected by infrastructure mounted cameras to identify potential conflicts by clustering conflicting vehicle trajectories. Atev *et al.* [7] developed a video-based system for real-time prediction of collisions at intersections. Due to their fixed nature, cameras mounted at intersections often face limitations regarding their field-of-view. Observability of all critical areas in the intersection should be considered when positioning these cameras. If multiple cameras are used, synchronization between different cameras is challenging. To address this challenge, Chen *et al.* [8] utilized the flexibility of unmanned aerial vehicles for capturing video data from a broader view, whereas Jackson *et al.* [9] proposed a flexible system based on movable cameras for road safety and behavioral analysis of road-users. A key challenge that video-based safety analysis techniques face is the loss of quality of extracted data in poor lighting conditions such as during night time or overcast weather [10]. Use of Light Detection and Ranging (LIDAR) sensors have been used by other studies to overcome this challenge. Studies in [10–13] utilized LIDAR sensors fixed at an intersection for data collection. Zyner *et al.* [14] presented a dataset collected by a vehicle equipped with LIDAR sensors that was parked near five different roundabouts.

The emergence of connected and automated vehicles has enabled the availability of high-resolution data for safety analysis. Basic Safety Messages (BSMs) broadcast by connected vehicles provide high-fidelity state information. Xie *et al.* [15] utilized BSM data from Safety Pilot Model Development (SPMD) for identifying high-risk locations on roadways. Liu and Khattak [16], and Zhang and Khattak [17] used the BSM data to capture extreme and risky driving events including extreme lane changes. Signal Phase and Timing (SPaT) data along with BSM data was used by Zhao and Zhang [18] for observing queueing dynamics at signalized intersections.

Analysis of existing literature revealed that different studies used the data from different types of sources for extracting safety-critical trajectory information about surrounding vehicles. This information is later used for extracting safety-critical events and their in-depth analysis. However, very few out of these studies validated or compared the extracted trajectories with another trustworthy source representing the ground truth. For calculating the metrics proposed in [4], information regarding position, velocity, acceleration, and geometric relationships between various safety critical traffic participants is necessary. The accuracy and fidelity of the extracted data are important for analysis using the proposed OSA metrics.

This paper presents a design and development of an infrastructure-based system that captures the data required for computing the OSA metrics. The system collects data from fixed video cameras and a V2I communication system. For validation of extracted information, separate high-accuracy data sources are employed for collecting the data representing the ground truth, including an on-board differential Global Positioning System (GPS) unit, and an aerial drone, LIDAR sensors.

## Measurement Uncertainty

Accuracy, precision, and resolution are all components of the measurement uncertainty (MU) of a data capture system. These terms are often used interchangeably and thus incorrectly. First, it is important to understand the difference between these three concepts as they will be discussed in this paper. According to the Joint Committee for Guides in Metrology (JCGM) [19], the accuracy of a measurement is defined as the "closeness of the agreement between the result of a measurement and a true value of the measurand." In contrast to accuracy, precision relates to the repeatability of measurements suggesting that values could be far from accurate, but the measurements could be considered precise if they are close to one another. Finally, resolution is the smallest increment of the value that is being measured. Accuracy, precision, and resolution are just a few out of numerous parameters that can be considered in determination of the MU. In the context of this work, it is important to consider the MU of data capture systems and how it impacts the collected data.

In order to determine the MU for a system, it is necessary to have a source of ground truth, or some known measurement, to compare with measured values. As it will be discussed in detail later in the paper, the data capture system considers a variety of sensor modalities, several of which are considered high-fidelity measurement techniques (e.g., LIDAR); however, the one data source with the most transparent, yet minimal uncertainty, in use is the OxTS RT3000 differential GPS unit. The differential GPS unit has a documented accuracy within 10 cm in the test configuration and is used to understand the MU for the accompanying sensor modalities. The differential GPS, LIDAR, drone footage, and added video cameras acted as temporary installments to the sensor suite for the validation of data.

**TABLE 1** MU for data capture sensor modalities.

| Sensor Modality | MU | Direct Measure |
|---|---|---|
| RT3000 Differential GPS | ±10 cm | GPS Position |
| LIDAR | ±2 cm | Light-based Position (static) |
| Telematics Unit | ±3 m | GPS Position |
| Drone Footage Tracking | Algorithm Dependent | Visual Position |
| Traffic Camera Tracking | Algorithm Dependent | Visual Position |

The primary variables needed to calculate the proposed OSA metrics from [4] are the position, velocity, and acceleration of the relevant vehicles. As such, different sensor modalities will propagate different levels of uncertainty through data capture (e.g., Radio Detection and Ranging (RADAR) allows for the direct measurement of velocity resulting in an uncertainty only as great as that of the RADAR; however, if GPS position is being measured to then derive velocity, the error magnifies as it is propagated and further again if the GPS position is used to derive acceleration). An MU analysis was conducted to determine the potential error which would propagate through the OSA metric calculations to understand the level of accuracy needed to produce useful metric values.

To better understand the uncertainty in the collected data, the generalized MU is documented for each sensor modality in Table 1.

Next, the propagation of uncertainty was studied for the key parameters needed to calculate the metrics including position, velocity, and acceleration. Since the RT3000 differential GPS was used as ground truth due to its low MU and high accuracy, the key parameters were evaluated in the context of this modality. The uncertainty in position is a direct measure of the accuracy documented for the GPS unit. When utilizing the RT3000 differential GPS unit with the dual-antenna configuration connected to the base station, the MU is ±2 cm; however, as the subject testing was completed offsite, the Omni-Star subscription was utilized for ±10 cm MU and the data acquisition rate was 100 Hz. The RT3000 uses an inertial navigation unit to measure accelerations with an MU of ± 10 mm/s$^2$ as well as integrate velocity data with an MU of ± 0.08 kph. The GPS correction is applied at a frequency of 10 Hz to reduce drift associated with the inertial measurements.

The importance of understanding MU in data capturing is apparent as the results demonstrate the ability of the different sensor modalities to measure key vehicle parameters. Depending on the needs for perception systems, different levels of uncertainty may be acceptable for the system. The focus of this study was to evaluate these different sensor modalities in the context of the OSA metrics introduced in [4]. The RT differential GPS unit was implemented as a near-ground truth source with the described uncertainties for vehicle trajectory parameters allowing for a direct comparison with the other modalities. As such, the other modalities were compared against the baseline of the RT unit to determine the MU for each.

MU analysis was applied to a specific case of an OSA metric previously proposed in [4], the minimum safe distance violation (MSDV), in order to understand how the uncertainty of the various sensor modalities propagates through the metric formulation. The equation (1) and equation (2) adopted from [4] describe the conditions when MSDV occurs.

$$SDV' = \begin{cases} 1 & if\ d^{lat} < d^{lat}_{min} \wedge\ d^{long} < d^{long^*}_{min} \\ 0 & else \end{cases} \tag{1}$$

$$MSDV = \begin{cases} 1 & if\ \ MSDV' = 1 \wedge Originated\ by\ ego\ vehicle \\ 0 & else \end{cases} \tag{2}$$

where $d^{long^*}_{min}$ is either $d^{long,same}_{min}$ or $d^{long,opp}_{min}$, depending on whether the entities are driving in the same or opposite direction, and is the minimum safe longitudinal distance based on maximum longitudinal acceleration/deceleration for each involved vehicle and a given response time; $d^{lat}_{min}$ is the minimum safe lateral distance based on maximum lateral acceleration and a given response time.

$$d^{long,same}_{min} = \left[ \begin{array}{l} v_1^{long}\rho_1 + \frac{1}{2} a^{long}_{1,max,accel}\ \rho_1^2 \\ + \frac{\left(v_1^{long} + \rho_1 a^{long}_{1,max,accel}\right)^2}{2a^{long}_{1,min,decel}} \\ - \frac{\left(v_2^{long}\right)^2}{2a^{long}_{2,max,decel}} \end{array} \right]_+ \tag{3}$$

Equation (3) provides the mathematical formulation for $d^{long,same}_{min}$ and is adopted from Section 3.1 of [20]. In this equation, one vehicle (subscript 1) is following behind another vehicle (subscript 2) and both are moving in the same direction.

$$\begin{aligned} d^{long,opp}_{min} = & \frac{2v_1^{long} + \rho_1 a^{long}_{1,max,accel}}{2}\rho_1 \\ & + \frac{\left(v_1^{long} + \rho_1 a^{long}_{1,max,accel}\right)^2}{2a^{long}_{1,min,decel,correct}} \\ & + \frac{2\left|v_2^{long}\right| + \rho_2 a^{long}_{2max,accel}}{2}\rho_2 \\ & + \frac{\left(\left|v_2^{long}\right| + \rho_2 a^{long}_{2max,accel}\right)^2}{2a^{long}_{2min,decel}} \end{aligned} \tag{4}$$

Equation (4) provides the mathematical formulation for $d^{long,opp}_{min}$ and is adopted from Section 3.3 of [20]. In this equation, one vehicle (subscript 1) and another vehicle (subscript 2) and both are moving in the opposing direction towards each other.

$$d^{lat}_{min} = \mu + \left[ \begin{array}{l} \frac{2v_1^{lat} + \rho_1\ a^{lat}_{1,max,accel}}{2}\rho_1 \\ + \frac{\left(v_1^{lat} + \rho_1\ a^{lat}_{1,max,accel}\right)^2}{2\ a^{lat}_{1,min,decel}} \\ - \left( \begin{array}{l} \frac{2v_2^{lat} - \rho_2 a^{lat}_{2,max,accel}}{2}\rho_2 \\ - \frac{\left(v_2^{lat} - \rho_2\ a^{lat}_{2,max,accel}\right)^2}{2\ a^{lat}_{2,min,decel}} \end{array} \right) \end{array} \right]_+ \tag{5}$$

Equation (5) provides the formulation for $d_{min}^{lat}$, for adjacent vehicles and is adopted from Section 3.4 of [20]. In this equation, one vehicle (subscript 1) is to the left of another vehicle (subscript 2) and $d_{min}^{lat}$ is the distance between the right side of the first vehicle and the left side of the other vehicle, and μ is the lateral fluctuation margin [m].

Since the MSDV metric incorporates the relative position and velocity of the ego and other vehicle, this calculation was performed with an assumed error consistent with the capabilities of the RT differential GPS unit. In calculating the MSDV threshold for the ego vehicle traveling behind another vehicle, velocities of both vehicles are required. Based on the RT differential GPS unit velocity MU of 0.08 kph, the variability in the calculated MSDV threshold would range from ±0.01 m to ±0.07 m, assuming the vehicles are traveling the same speed up to a differential of 104 kph. Larger speed differentials result in larger errors in the MSDV calculation due to the resulting impacts on closing speed (i.e., if the vehicles are traveling at the same speed, they have a closing speed of 0 kph and therefore will experience a minimum error due to uncertainty in the measure of velocity; however vehicles traveling at different speeds will induce larger error resulting from uncertainty in velocity measurements for each of the vehicles relative to one another). In addition to the uncertainty in the MSDV calculation, the relative distance between the vehicles also needs to be considered. The additional uncertainty associated with the localization of the two vehicles is ±0.1 m multiplied by two for the localization of each vehicle resulting in ±0.2 m. In total, the uncertainty in determining an MSDV would be ±0.21 m to ±0.27 m given the uncertainty associated with the differential GPS of ±10 cm for localization and 0.08 kph for velocity.

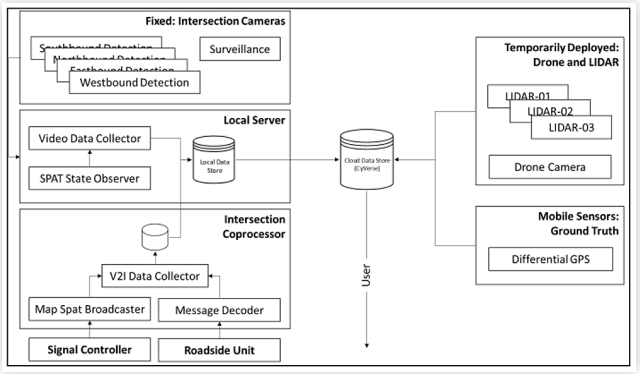
Although differential GPS can accurately determine position for vehicles, on-board vehicle data present challenges due to the lack of sharing during vehicle testing. Furthermore, the need to process data from each vehicle independently on the roadway is an inefficient technique to evaluate large quantities of traffic participants. For these reasons, infrastructure-based sensors are more effective data capture modes in the context of this work.

# System Architecture

Based on the capabilities of different sensors and accuracy requirements an infrastructure-based sensor data capture system was designed and deployed at the intersection of W Daisy Mountain Drive and N Gavilan Peak Parkway in Anthem, AZ, for validation and implementation. Figure 1 shows the system-level architecture of the designed data capture system.

The system consists of a multi-modal sensor suite that includes a combination of fixed traffic detection video cameras, an actuated traffic signal controller, V2I communication unit and intersection coprocessor, temporarily deployed drone with a video sensor and LIDAR sensors, and a vehicle equipped with mobile sensors to collect the data representing the ground truth.

**FIGURE 1** System architecture.






The data storage system, which is supported by the fiber optic network backbone, is composed of the data storage space on the intersection coprocessor, a local data store located at the transportation management center, and a cloud-based data store – CyVerse that facilitates the distribution of the data to end-users.

# Fixed Sensors

The term "fixed sensors" relates to video-camera based sensors that are permanently installed for traffic signal detection at the intersection. These cameras are passive sensors in that they use the ambient light available for imaging the traffic scene. They have the advantage of producing imagery familiar and easily interpretable to those viewing the video playback. The study intersection is equipped with two types of cameras: detector cameras, a surveillance camera.

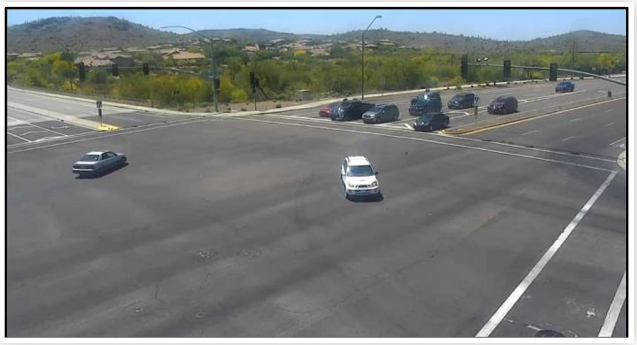
**Detector Cameras** The study intersection uses a video-based vehicle detection system for facilitating the actuated traffic signal control. Detector cameras capture video frames at a frequency of 30 frames per second and stream them on the internal network following Real-time Streaming Protocol (RTSP). The study intersection is equipped with four detector cameras. Each camera is associated with a unique inbound approach of the intersection. The *Video Data Collector* hosted on the local server collects and timestamps video frames from all four detector cameras at the intersection. It also retrieves the signal phase state information available from the *SPAT State Observer* and stores the current state information of applicable signal phases. The state information of signal phases is then associated with each frame. An example of single video frames from each detector camera overlaid with applicable SPaT information is shown in Figure 2.

**Surveillance Camera** Intersections account for a large percentage of conflicts and crashes that occur [21]. As can be seen in Figure 2, most areas of the intersections box are observable by at least one detector camera. However, there are some blind-spots which are not covered by the field-of-view (FOV) of any detector camera mounted at the intersection. To

**FIGURE 2** Video frames from detector cameras overlaid with SPaT information.




**FIGURE 3** Example video frame from the surveillance camera.





capture events occurring in these areas, video data from a surveillance camera that is mounted near the northeast corner of the study intersection are also collected. Unlike detector cameras, the surveillance camera allows pan, tilt, and zoom controls so that any possible blind spots in the intersection box can be observed. An example video frame from the surveillance camera is shown in Figure 3.

## Temporarily Deployed Sensors

The term "temporarily deployed sensors" relates to sensors that are temporarily installed or used near the study intersection during planned data capture events. For such events, LIDAR sensors and a drone-mounted camera are utilized for capturing additional data.

**LIDAR Sensors** As seen in Figure 2, each detector camera at the study intersection is monocular and directed towards a unique approach to the intersection. This legacy configuration, while serving its purpose for traffic detection, can make automated tracking of vehicles through the intersection difficult since the algorithm must track vehicles through several camera views. Another disadvantage of this camera configuration is the fact that distance from the intersection is difficult to measure. This adds to difficulties in estimating a vehicle's position, particularly distance from the intersection. While camera technology has a low-cost advantage and its imagery readily available, the reliance on ambient light makes imagery obtained at night and in severe weather conditions that limit visibility problematic. This motivated the IAM team to consider active sensors that provide their own energy and do not rely on the ambient light.

A LIDAR sensor sends out an electromagnetic pulse of energy and measures how long it takes to return to the sensor. The advantage of LIDAR is the fact that lasers are used to provide the source of the light pulse. Since lasers are highly directional, angular resolution is very fine. Scanning the laser in azimuth and elevation allows the sensor to form a three-dimensional (3D) representation of the scene, providing accurate measurements of distance to the detected object and an ability to view the scene from any perspective during post-event analysis. LIDAR sensors can also provide accurate estimates of the outside boundary of a vehicle, which can be critical when trying to determine some of the OSA metrics such as MSDV. However, self-shading is a type of occlusion that occurs when a single LIDAR is used to illuminate a vehicle from one side of the intersection which introduces uncertainty of the vehicle true boundary. Large vehicles can also completely obscure smaller vehicles, leaving them undetected by the LIDAR. This problem can be resolved by using multiple LIDAR units illuminating the traffic scene from multiple observation points (although it must be said that this solution is applicable to other sensor modalities as well).

Considering the advantages and limitations of using LIDAR sensors the data capture system employs three Luminar® LIDAR sensors temporarily deployed during data capture events, each on a different corner of the intersection, as indicated in Figure 4.

Each LIDAR unit has a raster scan pattern that covers 30 degrees in elevation and 120 degrees in azimuth so that a single LIDAR unit can cover two lanes of the traffic. The LIDAR units are connected by Ethernet to a single processor so that the detection data could be combined into a single

**FIGURE 4** LIDAR sensor mounting locations near the study intersection.

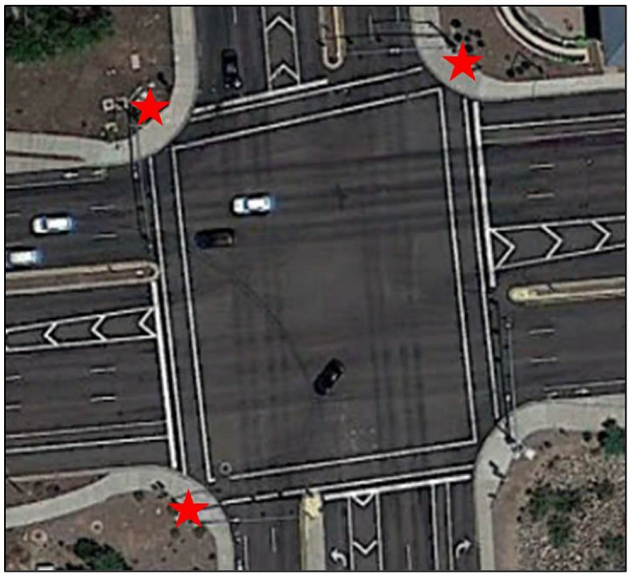
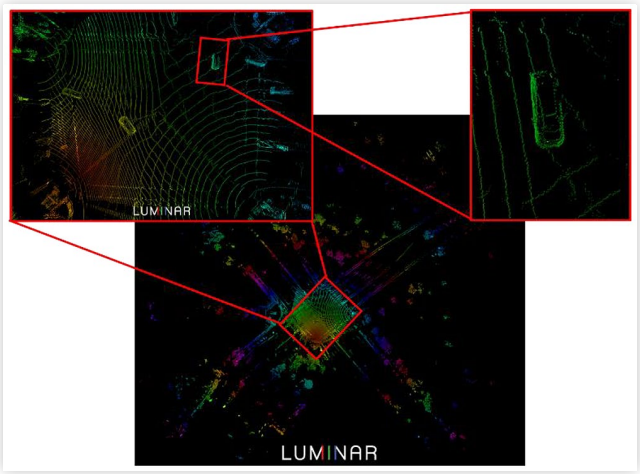


**FIGURE 5** Multiple LIDAR units linked together to form a single image.






point cloud image as shown in Figure 5. This configuration allows illumination of vehicles from both sides to reduce shading and provides more accurate estimates of the external boundaries of vehicles within the scene with a maximum detection range of more than 200 m for all approaching lanes.

The eye-safe lasers used can detect vehicles surrounding the intersection even in adverse weather. This provides ample observation time to the sensor system so tracking of traffic behavior can begin as the traffic approaches the intersection. A Track File Manager (TFM) is created to provide a more stable vehicle track file and more accurate state information for the vehicle as it passes through the intersection.

Post-processing of the LIDAR data provides an opportunity to back propagate the most accurate estimate of the vehicle's position through all observations associated with a particular vehicle, since the vehicle's outside boundaries never change and the best measurement of vehicle boundary occurs in the intersection with multiple sensors. By building accurate 3D models, LIDAR imagery provides a means of playing back a traffic event from any perspective, as illustrated in Figure 6 and Figure 7.

**Drone-Mounted Camera** FOV- and occlusion-related limitations of fixed video cameras motivated the IAM team to utilize a drone-mounted video camera to investigate the advantage of a ground-facing camera from a height of approximately 115 m, providing a single bird's eye view of the intersection, also referred to as a plan view. A DJI Phantom 4 Pro drone was utilized to hover over the study intersection to capture video footage of test vehicles and other traffic that passed through the intersection. These data can be analyzed using vehicle tracking algorithms under development to calculate the OSA metrics.

The main advantage of drone video footage is the removal of occlusions due to vehicles being hidden behind others. Furthermore, the near perpendicular orientation relative to the vehicle travel direction provides an optimal viewing angle for planar determination of vehicle motion which becomes distorted at shallower capture angles due to the camera perspective. One disadvantage of drone video capture is the limitation of continuous filming as the limited battery capacity results in 20–30 min of flying time. In total, approximately 70 min of drone footage was captured at a resolution of 1080p at 30 frames per second (fps) during the latest data capture event to provide a traffic dataset for the study intersection.

**FIGURE 6** LIDAR view 1.

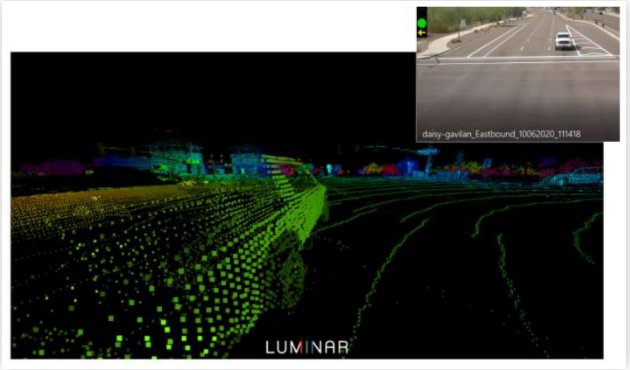





**FIGURE 7** LIDAR view 2.




## V2I Communication Units

Data sources enabled by V2I communication technology are SPaT data generated by traffic-actuated signal controllers, intersection map messages, and BSMs generated by surrounding connected vehicles.

**Signal Phase and Timing (SPaT)** SPaT data that represents the state of an intersection traffic signal controller are a data element defined in the SAE J2735 standard [22]. SPaT data describe the current state of all active traffic- and pedestrian-signal indications including the minimum and maximum expected times until their current states change. The actuated traffic signal controller generates SPaT data at a frequency of 10 Hz and forwards the data to the intersection coprocessor. A software component, called the *Map Spat Broadcaster*, listens for the SPaT data from the controller and decodes and processes it into a JavaScript Object Notation (JSON) string that can be logged. In addition to decoding the information contained in the SPaT data, for each phase, the *Map Spat Broadcaster* calculates the time elapsed since the initiation of the current state. The data are sent to the *V2I Data*

*Collector* hosted on the intersection co-processor for storing in the appropriate format, to the *SPAT State Observer* hosted on the local server for integration with data from other sources, and to the Roadside Unit (RSU) for broadcasting the information over the air. In the context of safety analysis, time synchronized SPaT data can help in establishing the relationship between occurrences of metric violations and critical states of signal phases.

**Intersection Map** Intersection Map is another data element defined in the SAE J2735 standard [22]. It is used to convey information about the geographic extents of an intersection or roadway, connections between various intersection lanes, lane-signal phase association, and other road information such as local speed limits. The intersection map data are developed using the USDOT Connected Vehicle ISD Message Creator [23]. A Map-Engine library developed by California PATH [24] is used to process the intersection map data and to locate vehicles on the intersection map, which helps in establishing the geometric relationship between various safety-critical traffic participants. An example visualization of content of the intersection map message is shown in Figure 8.

**Basic Safety Messages (BSMs)** A BSM is another data element defined in the SAE J2735 standard [22]. The BSMs include a vehicle's position, speed, heading, acceleration, and other state information at the instant of message broadcast. Vehicles equipped with a Connected Vehicle On-board Unit (OBU) broadcast BSM data at a frequency of 10 Hz. BSMs broadcasted by connected vehicles surrounding an intersection are received by the intersection RSU and are forwarded to the Wireless Message Decoder hosted on the intersection co-processor. The Wireless Message Decoder decodes the received BSMs and forwards the information to the V2I Data Collector for storing in the appropriate format. Figure 9 shows the information extracted from BSMs that is filtered using the Map-Engine library. Currently, MCDOT operates a small fleet of 12 vehicles that are equipped with OBUs that are deployed in the test bed. Note that one of MCDOT's connected vehicles left the road as it approached the intersection from the west in order to provide study support and does not indicate an errant vehicle path.

**FIGURE 8** Visualization of intersection map developed using USDOT ISD Message Creator tool.

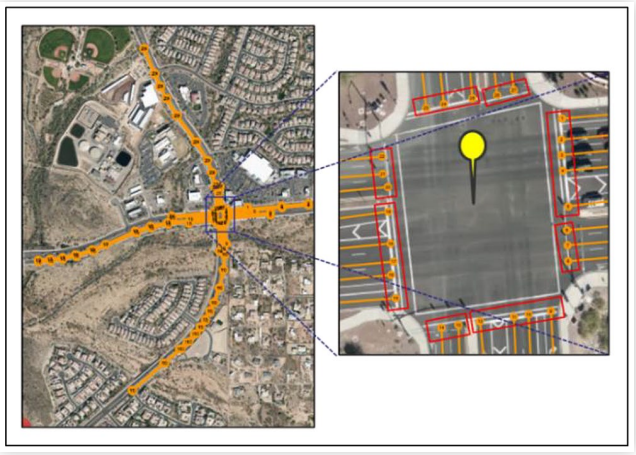
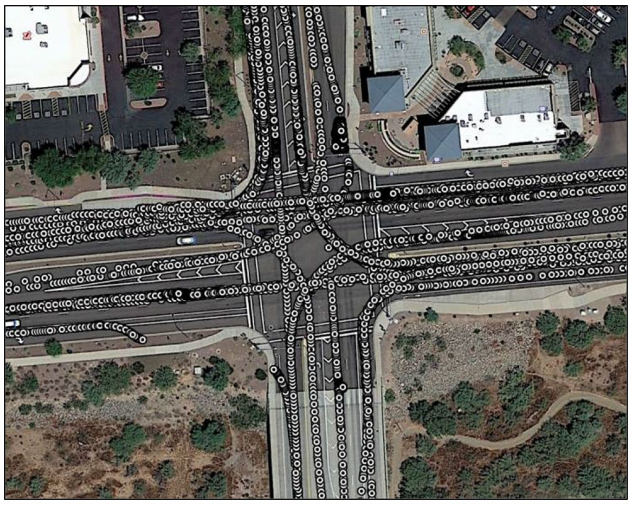




**FIGURE 9** Information extracted from BSMs allow to form high-resolution trajectories of vehicles that are on the intersection MAP.





## Mobile Sensors for Establishing Ground Truth

Although the different sensors utilized to capture data at the subject intersection all introduce useful and interesting data, perhaps the most important data collected at the intersection were the ground truth data captured using an RT3000 differential GPS unit. This GPS unit utilized a dual-antenna setup equipped with Omni-Star satellite subscription for positioning accuracy within ±10 cm with an acquisition rate of 100 Hz. The high accuracy and resolution of the differential GPS provides the data needed to validate the calculated vehicle trajectories from lower-fidelity sensors and tracking algorithms utilizing near-ground truth data.

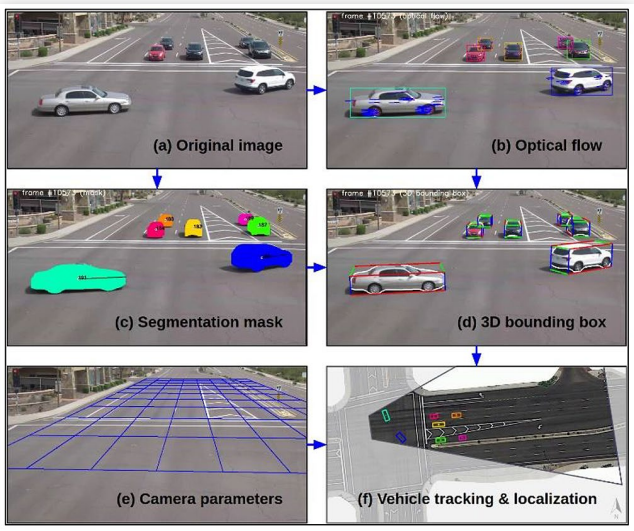
## Data Processing and Validation

The IAM team ran a pipeline of computer vision algorithms with the following steps to track and localize vehicles on the videos obtained by cameras mounted on the road infrastructure, as shown in Figure 10.

The algorithm pipeline is defined by the following steps:

1. Detection: A deep neural network is used to detect vehicles and obtain the segmentation mask on each video frame. The optical flow is used to estimate the vehicle motion and associate detected object instances on adjacent frames.

**FIGURE 10** Vehicle tracking with cameras on the road infrastructure.



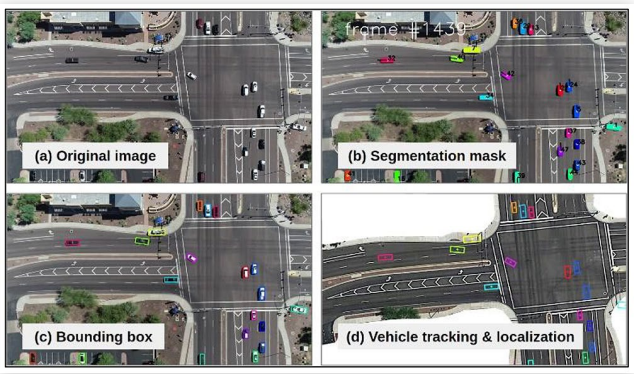



2. Calculation of 3D bounding box: The vehicle motion and mask is used to calculate the 3D bounding box according to the camera perspective geometry. This requires calibration of the camera relative to the ground and assumes the camera does not move between frames, although minor shifting of the camera can be corrected through the use of software.
3. Vehicle localization and tracking: The calibrated camera parameters, 3D bounding boxes, and estimated vehicle motions on the image are used to compute the 3D location and the speed in the world coordinate system. Then, a Kalman filter with rigid body kinematics is used to smooth the vehicle states.

The vehicle localization and tracking results contain the position, speed, and other properties of each vehicle on every frame. These results can be used for further analysis such as OSA metrics calculation or visualization on a map.

Similarly, the team ran a pipeline of computer vision algorithms to track vehicles in the drone videos, as shown in Figure 11. The steps are similar to those used for infrastructure-based cameras on road infrastructures, but with a few differences. First, the drone images are stabilized to a reference image (typically the first one in the video), since a drone may move slightly in the air when recording the video. Second, instead of 3D bounding boxes, oriented 2D bounding boxes are calculated from segmentation masks, which are essentially minimum area rectangles covering the masks. Third, another neural network model was trained to determine the orientation of each vehicle on the image. After the drone video is processed, sets of vehicle states similar to those in the tracking results from infrastructure-based cameras were obtained. Hence, the drone video results can also be used for the calculation of the proposed OSA metrics.

**FIGURE 11** Vehicle tracking with drone videos.






Vehicle localization and speed measurement results from both the infrastructure-based cameras and the drone videos were validated by driving a test vehicle at the test site. As shown in Figure 12, the northbound approach of the intersection was used during testing. The test vehicle was instrumented with an RT3000 differential GPS receiver which can achieve centimeter-level localization accuracy. Meanwhile, a drone (DJI Pro) flying at approximately 130 m above the ground was used to record the videos from a bird's-eye view. The GPS data, videos from infrastructure-based cameras, and videos from the drone were synchronized with a programmable flashlight connected to the GPS data logger, which was triggered as the test vehicle was passing through the intersection such that the flash could be observed in the videos. In this section, GPS data are considered as the reference measurements, which are compared with the vehicle tracking results.

The test vehicle passes through the area covered by the infrastructure-based camera in approximately 12 s, which allows the camera to generate 338 individual measurements on 338 video frames. Corresponding measurements from the GPS data and the tracking results with the drone video were selected. The trajectory of the test vehicle is shown in Figure 13, with four selected video images at different locations. Overall, the vehicle localization results from all three different types of measurement methods largely agree with each other. Table 2, Figure 14, and Figure 15 show the localization and speed measurement results, using the differential GPS data as a reference since it has the highest accuracy. In general, for both camera-based methods, the average location errors are below 1 m and the speed measurement errors are, on average, lower than 1 m/s, which means the localization and tracking results are promising with respect to meeting the accuracy requirements for further tasks such as computing the OSA metrics.

**FIGURE 12** An overview of vehicle localization and speed measurement result validation.

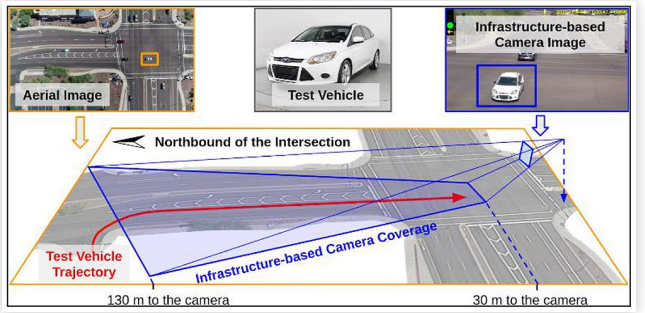


**FIGURE 13** Trajectories from GPS, the infrastructure-based cameras, and the drone video.

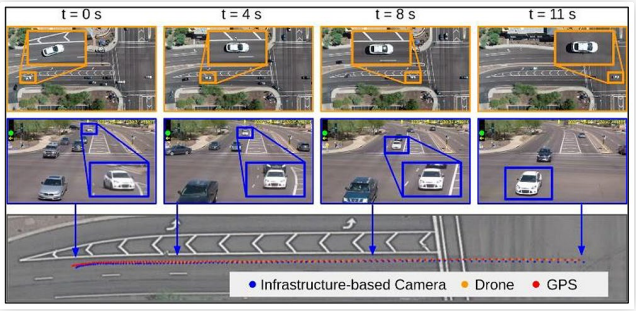





**TABLE 2** Localization and speed measurement results using the GPS data as the reference.



| Method | Average Localization Error | Maximum Localization Error | Average Velocity Error | Maximum Velocity Error |
|---|---|---|---|---|
| Infrastructure-based camera | 0.60 m | 1.24 m | 0.86 m/s | 3.50 m/s |
| Drone video | 0.48 m | 0.889 m | 0.14 m/s | 0.35 m/s |

**FIGURE 14** Vehicle localization errors for the infrastructure-based camera and the drone video, using the GPS data as the reference.

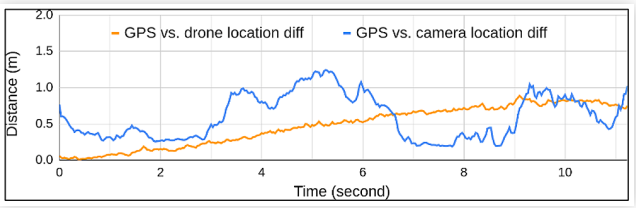
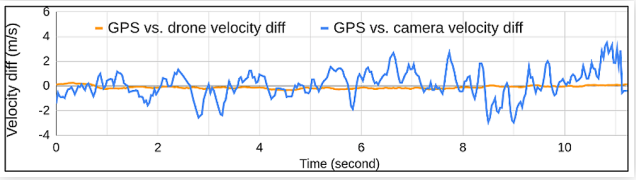





**FIGURE 15** Vehicle speed measurement errors for the infrastructure-based camera and the drone video, using the GPS data as the reference. The speed measurement differences are very small between the GPS data and the drone video results.






The localization results from the infrastructure-based cameras have larger errors, especially concentrated in the camera pointing direction. Also, the speed measurements from the infrastructure-based cameras are noisy mainly due to the uncertainty of the calculated 3D bounding box at certain perspective angles because the vehicle motion on the image is not significant when it is moving towards the camera. These errors can potentially be reduced by a more robust vehicle 3D pose estimator, a more advanced filtering technique, and using prior knowledge of vehicle shapes.

There are a few causes for localization and speed measurement errors for the camera-based methods. First, the ground plane is assumed to be perfectly flat when calibrating the camera using feature correspondences between the image and the map; thus, errors are to be expected at locations where the ground surface has uneven heights. Second, when comparing results with the GPS data, coordinates in latitude and longitude must be transformed to the local XYZ coordinates using an anchor point (typically the differential GPS base station). Due to the small uncertainty of the coordinates of the anchor point on the geodesic ellipsoid, there can be location errors introduced in this transformation. Third, the computer vision algorithms assume the vehicles are roughly "boxy" for calculating the 3D bounding box boundaries, and errors can be made for "non-boxy" vehicles. Fourth, the computer vision algorithm has difficulties in estimating the "depth" in the 3D world along the camera pointing direction due to the inherent property of camera projection, which may lead to inaccuracy in the estimated vehicle dimension and location.

## Case Studies

Two case studies on operational safety assessment using the results from the data capture system are presented in this section. In the first case, as shown in Figure 16, two vehicles are driving in the same lane at relatively high speeds, while their separation distance is small. Using the tracking results of the sensor package, the location and speed of both vehicles can be obtained, and hence, whether an MSDV has occurred can be determined. In the second case, as shown in Figure 17, a vehicle is changing lanes at a location that forbids lane changing. Using the tracking results of the data capture system, the location and trajectory of the vehicle can be obtained to check such traffic law violations (another OSA metric) by comparing the trajectory to annotated semantic information of lanes on the map. The OSA metrics can

**FIGURE 16** A case study on determining an MSDV using the results of the sensor package.




**FIGURE 17** A case study on determining traffic law violation using the results of the sensor package.

be calculated automatically by analyzing the traffic monitoring videos obtained from the data capture system paired with the algorithms, quantitative operational safety assessment at a specific segment of the road can be conducted to evaluate the performance of an AV.

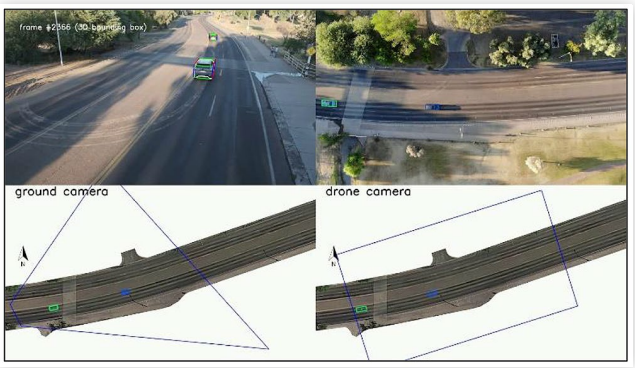
# Discussion

This paper presents a sensor system developed by the IAM team to capture the data required for computing the OSA metrics that were proposed in [4]. Primary data required for computing the metrics are the position of the vehicle and its higher derivatives including velocity and acceleration. Considering capabilities and limitations of each available sensor modality, the IAM team utilized a combination of active and passive sensors and capabilities of V2I communications for capturing the required data. The optical sensor modalities in some ways complement each other towards extracting the required information. various Table 3 summarizes the comparison of different optical sensor modalities based on the respective capabilities and limitations.

Although the current work is largely done at an instrumented intersection, the described data capture system can be expanded to other intersections, roundabouts, highway ramps, or other places with potential traffic safety concerns. Some preliminary experimental results from the data captured from another site are shown in Figure 18.

Through the data obtained from optical sensors such as video cameras and LIDAR sensors, vehicle position can be directly identified, and higher order derivatives need to be derived based on the changes in the vehicle position over time. However, the noise in the position measurement is magnified when used to derive the velocity or acceleration. RADAR sensors can help in mitigating this limitation as it can exploit the use of Doppler effect to directly measure the velocity of moving object. This makes RADAR a perfect complement to a conventional camera mounted with their FOVs pointed into traffic approaching the intersection. The complement of using a RADAR sensor in conjunction with a camera can also provide an all-weather capability since RADAR sensors can penetrate rain, snow, fog or other low visibility conditions with minimal degradation to its performance. Additional benefits of using RADAR sensors will be explored since the data capture system will need to operate in the similar environmental conditions.

**FIGURE 18** Preliminary experimental results at an additional site in AZ.






Two other sensor modalities that are in consideration by the IAM team but have not yet been incorporated in the data capture system included stereoscopic cameras and infrared cameras. By mounting two cameras together at a fixed distance, a better estimate of an object's distance and a 3D image of the viewed scene can be realized. These types of systems, referred to as stereoscopic cameras, have had significant success in the application of robotic computer vision where an indication of depth is necessary for robots designed for grasping objects. While this technology has the potential

**TABLE 3** Comparison of infrastructure-based sensors.

| Sensor Modality | Benefits | Limitations |
|---|---|---|
| Infrastructure-based camera | • Unlimited operational time<br>• Video can be obtained easily<br>• Convenience in automated analysis software integration.<br>• No moving parts, long lifetime | • Limited range and accuracy due to the viewing angle<br>• Measurement errors or failures with partially or totally occluded vehicles<br>• The camera requires installation work, and it is not flexible for adjustment<br>• Performance relies on the quality of ambient light |
| Drone Video | • Larger coverage, bird's-eye view<br>• No vehicle occlusion issue<br>• Flexibility in choosing the viewpoint and monitored sites | • Limited operational time (typically 20 to 30 minutes per flight)<br>• The drone requires a skilled human pilot and extra effort to obtain the video<br>• There are safety concerns about flying the drone over the traffic<br>• Performance relies on the quality of ambient light |
| Infrastructure-based LIDAR | • Unlimited operational time<br>• Point cloud can be obtained without extensive post-processing effort<br>• Accurate 3D information on vehicle location and shape<br>• Performance is unaffected by the ambient light | • High cost<br>• Most LIDAR sensors available on the market at present have mechanically rotating components with a limited lifetime<br>• It generally requires a different suite of software that is not well developed and not openly available |

of being an inexpensive alternative to LIDAR for 3D imaging, additional research is needed to provide a working model for the traffic management use case. Infrared cameras process a part of the electro-magnetic spectrum that is directly tied to an object's temperature. At night, infrared cameras have an advantage over conventional cameras since infrared cameras can better distinguish warm objects, like cars and pedestrians, and the cool evening background. During the day, hot pavement serves as a good background against cars and pedestrians which are cooler. However, as the background transitions from warmer temperatures to cooler temperatures, infrared cameras can have a hard time distinguishing our objects of interest and the background if the infrared camera's minimally detectable temperature difference is large with respect to the camera's own thermal noise level.

## Conclusions and Future Work

This paper presents the development of a data capture system that exploits the capabilities of different sensor modalities for operational safety assessment, including evaluation of the navigation of an AV in a given scenario using the proposed OSA metrics. To accurately compute the OSA metrics proposed in the earlier work of the IAM team, analysis of measurement uncertainty was performed and the effect of measurement uncertainties on the accuracy of computed OSA metrics was studied. Based on the required measurement accuracy, capabilities of different active and passive sensors were explored to develop the data capture system. The information extracted from various sensors in the developed data capture system was validated against the high-accuracy differential GPS data representing the ground truth. Finally, two case studies were presented to demonstrate the computation of OSA metrics using the information extracted from the collected data.

In the next phase of this work, capabilities of additional infrastructure-based sensor modalities will be explored that can provide the information required for computation of OSA metrics in varying environmental and operational conditions. Additionally, the use of V2I communications data to elucidate geometric relationships between safety-critical traffic participants and relationships between signal state dynamics and high occurrences of OSA metrics violations will be explored.

It is possible that a single data capture system design will not be optimal in every situation, and thus several designs including varying permutations of sensors may be developed. This will allow for OSA metrics to be measured at additional types of intersections and roadways. The eventual goal is to develop a suite of robust data capture systems that can provide the means of measuring OSA metrics, thereby allowing the operational safety of both AVs and human-driven vehicles to be evaluated in order to improve the road traffic safety. Future work will also include exploring the use of such data capture systems in additional locations (both in AZ and elsewhere) to compare the calculation of OSA metrics in varying environments.

## References


1. Singh, S., "Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey," A Brief Statistical Summary DOT HS 812 506, USDOT, Washington, D.C., 2018.
2. Milakis, D., van Arem, B., and van Wee, B., "Policy and Society Related Implications of Automated Driving: A Review of Literature and Directions for Future Research," *J. Intell. Transp. Syst.*, 21(4):324–348, 2017, doi:10.1080/15472450.2017.1291351.
3. Fraade-Blanar, L., Blumenthal, M.S., Anderson, J.M., and Kalra, N., "Measuring Automated Vehicle Safety: Forging a Framework," RR-2662, *RAND Corporation, Santa Monica, California*, 90, 2018.
4. Wishart, J., Como, S., Elli, M., Russo, B. et al., "Driving Safety Performance Assessment Metrics for ADS-Equipped Vehicles," *SAE Int. J. Adv. Curr. Pract. Mobil.* 2(5):2881–2899, 2020, doi:10.4271/2020-01-1206.
5. Elli, M., Como, S., Wishart, J., Dhakshinamoorthy, S., and Weast, J., "Evaluation, Validation and Refinement of Operational Safety Metrics for Automated Vehicles in Simulation," 2020.
6. Saunier, N. and Sayed, T., "Automated Analysis of Road Safety with Video Data," *Transp. Res. Rec.* 2019(08):57–64, 2007, doi:10.3141/2019-08.
7. Atev, S., Arumugam, H., Masoud, O., Janardan, R., and Papanikolopoulos, N.P., "A Vision-Based Approach to Collision Prediction at Traffic Intersections," *IEEE Trans. Intell. Transp. Syst.* 6(4):416–423, 2005, doi:10.1109/TITS.2005.858786.
8. Chen, P., Zeng, W., Yu, G., and Wang, Y., "Surrogate Safety Analysis of Pedestrian-Vehicle Conflict at Intersections Using Unmanned Aerial Vehicle Videos," *Research Article, Hindawi*e5202150, 2017, doi:10.1155/2017/5202150. https://www.hindawi.com/journals/jat/2017/5202150/.
9. Jackson, S., Miranda-Moreno, L.F., St-Aubin, P., and Saunier, N., "Flexible, Mobile Video Camera System and Open Source Video Analysis Software for Road Safety and Behavioral Analysis," *Transp. Res. Rec. J. Transp. Res. Board* 2365(1):90–98, 2013, doi:10.3141/2365-12.
10. Zhu, Q., Chen, L., Li, Q., Li, M., Nuchter, A., and Wang, J., "3D LIDAR Point Cloud Based Intersection Recognition for Autonomous Driving," in *2012 IEEE Intelligent Vehicles Symposium*, IEEE, Alcal de Henares, Madrid, Spain, ISBN 978-1-4673-2118-1, 456–461, 2012, doi:10.1109/IVS.2012.6232219.
11. Aycard, O., Baig, Q., Bota, S., Nashashibi, F. et al., "Intersection Safety Using Lidar and Stereo Vision Sensors," 869, 2011, doi:10.1109/IVS.2011.5940518.
12. Khattak, A. and Gopalakrishna, M., "Remote Sensing (LIDAR) for Management of Highway Assets for Safety," 41.

13. Wu, J., Xu, H., Zheng, Y., and Tian, Z., "A Novel Method of Vehicle-Pedestrian Near-Crash Identification with Roadside LiDAR Data," *Accid. Anal. Prev.* 121:238–249, 2018, doi:10.1016/j.aap.2018.09.001.

14. Zyner, A., Worrall, S., and Nebot, E.M., "ACFR Five Roundabouts Dataset: Naturalistic Driving at Unsignalized Intersections," *IEEE Intell. Transp. Syst. Mag.* 11(4):8–18, 2019, doi:10.1109/MITS.2019.2907676.

15. Xie, K., Yang, D., Ozbay, K., and Yang, H., "Use of Real-World Connected Vehicle Data in Identifying High-Risk Locations Based on a New Surrogate Safety Measure," *Accid. Anal. Prev.* 125:311–319, 2019, doi:10.1016/j.aap.2018.07.002.

16. Liu, J. and Khattak, A.J., "Delivering Improved Alerts, Warnings, and Control Assistance Using Basic Safety Messages Transmitted Between Connected Vehicles," *Transp. Res. Part C Emerg. Technol.* 68:83–100, 2016, doi:10.1016/j.trc.2016.03.009.

17. Zhang, M. and Khattak, A.J., "Identifying and Analyzing Extreme Lane Change Events Using Basic Safety Messages in a Connected Vehicle Environment," 2018.

18. Zhao, S. and Zhang, K., "Observing Space-Time Queueing Dynamics at a Signalized Intersection Using Connected Vehicles as Mobile Sensors," 2017.

19. Joint Committee for Guides in Metrology (JCGM), "Evaluation of Measurement Data - Guide to the Expression Of Uncertainty in Measurement," 2008.

20. Shalev-Shwartz, S., Shammah, S., and Shashua, A., "On a Formal Model of Safe and Scalable Self-driving Cars," *ArXiv170806374 Cs Stat*, 2018.

21. Choi, E.-H., "Crash Factors in Intersection-Related Crashes: An On-Scene Perspective: (621942011-001)," 2010, doi:10.1037/e621942011-001.

22. J2735D: Dedicated Short Range Communications (DSRC) Message Set Dictionary™ - SAE International, 2016.

23. Connected Vehicle ISD Message Creator, 2016, https://webapp.connectedvcs.com/isd/.

24. Zhou, K., "Multi-Modal Intelligent Traffic Signal Systems (MMITSS) Phase III: Task 3.2 Specifications for MAP Engine Library," 2018.

## Contact Information

**Niraj Altekar**
The University of Arizona
nvaltekar@email.arizona.edu

## Acknowledgments

Authors are grateful to the Institute of Automated Mobility (IAM) for generously funding this work and Maricopa County Department of Transportation (MCDOT) for enabling infrastructural resources required for this work.

## Definitions/Abbreviations

**ADS** - Automated Driving System
**AV** - Automated Driving System equipped Vehicle
**BSM** - Basic Safety Message
**FOV** - Field of View
**GPS** - Global Positioning System
**IAM** - Institute of Automated Mobility
**JSON** - Java Script Object Notation
**LIDAR** - Light Detection and Ranging
**MCDOT** - Maricopa County Department of Transportation
**MSDV** - Minimum Safe Distance Violation
**MU** - Measurement Uncertainty
**OBU** - Onboard Unit
**OSA** - Operational Safety Metrics
**RADAR** - Radio Detection and Ranging
**RSU** - Roadside Unit
**RTK** - Real Time Kinematics
**RTSP** - Real-time Streaming Protocol
**SPaT** - Signal Phase and Timing
**TRI** - University of Arizona Transportation Research Institute
**V2I** - Vehicle to Infrastructure



ISSN 0148-7191