

Evaluating Roadside Perception for Autonomous Vehicles: Insights from Field Testing

Rusheng Zhang, Depu Meng, Shengyin Shen, Tinghan Wang, Tai Karir,
Michael Maile, Henry X. Liu

Abstract—Roadside perception systems are increasingly crucial in enhancing traffic safety and facilitating cooperative driving for autonomous vehicles. Despite rapid technological advancements, a major challenge persists for this newly arising field: the absence of standardized evaluation methods and benchmarks for these systems. This limitation hampers the ability to effectively assess and compare the performance of different systems, thus constraining progress in this vital field. This paper introduces a comprehensive evaluation methodology specifically designed to assess the performance of roadside perception systems. Our methodology encompasses measurement techniques, metric selection, and experimental trial design, all grounded in real-world field testing to ensure the practical applicability of our approach.

We applied our methodology in Mcity¹, a controlled testing environment, to evaluate various off-the-shelf perception systems. This approach allowed for an in-depth comparative analysis of their performance in realistic scenarios, offering key insights into their respective strengths and limitations. The findings of this study are poised to inform the development of industry-standard benchmarks and evaluation methods, thereby enhancing the effectiveness of roadside perception system development and deployment for autonomous vehicles. We anticipate that this paper will stimulate essential discourse on standardizing evaluation methods for roadside perception systems, thus pushing the frontiers of this technology. Furthermore, our results offer both academia and industry a comprehensive understanding of the capabilities of contemporary infrastructure-based perception systems.

Index Terms—Roadside Perception Systems, Automated Vehicles, Standardized Evaluation Method, Roadside Perception System Benchmarks

I. INTRODUCTION

A key component of autonomous driving is the perception system, which enables the vehicle to sense and understand its surroundings. However, the onboard perception system, which

This research was partially funded by the U.S. Department of Transportation (USDOT) Advanced Transportation and Congestion Management Technologies Deployment Award (693JJ32150006) and Mcity of the University of Michigan. Corresponding authors: Rusheng Zhang and Henry X. Liu.

R. Zhang, D. Meng, and H.X. Liu are with the Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI, 48109, USA (email: {rushengz, depum, henryliu}@umich.edu)

S. Shen is with the University of Michigan Transportation Research Institute, 2901 Baxter Rd, Ann Arbor, MI, 48109, USA. (email: shengyin@umich.edu)

T. Wang is with the Department of Mechanical Engineering, University of Michigan, 2350 Hayward St, Ann Arbor, MI 48109, USA. (email: tinghanw@umich.edu)

T. Karir is with Huron High School, Ann Arbor, Michigan, USA. (email: tai.karir@gmail.com)

M. Maile is with Ivie Communications, (Email: michael.a.maile@gmail.com)

H. X. Liu is also with Mcity of the University of Michigan.

¹<https://mcity.umich.edu/>

relies on sensors mounted on the vehicle, such as cameras, LiDARs and radars, may face limitations in complex scenarios, harsh weather, and lighting conditions, due to occlusions, blind spots, sensor noise and environmental diversity. Therefore, a complementary roadside perception system is needed to enhance the onboard perception and provide more systematic and reliable scene information. Roadside perception leverages sensors installed on the roadside infrastructure to detect and track vehicles and other objects in the region of interest, and communicate the perception results to the onboard system via vehicle-to-infrastructure (V2I) communication technologies [1], [2].

Roadside perception is a rising field that has attracted increasing attention from both academia and industry in recent years. Several off-the-shelf products have been introduced to the market, which provide roadside perception solutions based on camera, radar and LiDAR sensors. For example, Derq [3] offers AI-driven roadside perception software that can detect and predict road users' behavior and prevent collisions. Ouster [4] produces high-resolution digital LiDAR sensors that can be used for roadside perception and V2X communication. These examples represent only a small fraction of the burgeoning market of off-the-shelf products aimed at enhancing roadside perception for autonomous driving. There are numerous other companies, as detailed in [1], [2], contributing innovative solutions in this sector. Together, they illustrate both the immense potential this field holds for the future of autonomous driving.

Despite the burgeoning market of roadside perception systems, there currently exists a notable deficit in this field: the absence of a standardized, fair comparison between these diverse systems. Various off-the-shelf products, each with their unique features and performance claims, are currently operating without a clear means of direct, fair comparison. This could result in an unregulated landscape, where the absence of fair competition hampers innovation and quality assurance in roadside perception systems.

In response to this pressing need, this paper presents an evaluation methodology specifically devised for a systematic assessment of roadside perception systems. The proposed methodology encompasses aspects such as measurement techniques, metrics selection, and experimental trial design, all derived from field testing experience. Importantly, these components are deeply rooted in real-world conditions, making this evaluation methodology not only practical but also readily implementable. The focus on real-world application ensures that our proposed system can serve as a useful tool in

advancing fair comparison in this field.

To illustrate the viability of our proposed methodology, we implemented it within Mcity, a controlled yet realistic testing environment [5]. We evaluated three off-the-shelf perception systems, providing a practical demonstration of how our methodology enables detailed analysis of system performance. By evaluating these three distinct off-the-shelf perception systems within Mcity, our goal is more than just sharing the outcomes of these assessments. In fact, these evaluations serve as a practical illustration of a standardized methodology for assessing roadside perception systems. Our aim is to make a significant stride towards establishing standardized benchmarks for roadside perception systems. We believe that such an advancement will bring substantial benefits to both industry and academia, providing a solid foundation for further research and development in this sector.

II. RELATED WORKS

Roadside perception systems have experienced remarkable progress in recent years, driven by the dynamic advancements in autonomous driving technologies [1], [2]. These systems enhance the onboard vehicle perception, offering more reliable and precise scene comprehension, especially in challenging scenarios involving complex environments or severe weather and lighting conditions [6], [7]. Notably, the academic literature presents a variety of roadside perception systems, utilizing different sensor technologies. This includes systems based on LiDAR [8], and camera-based systems [7], [9], [10], each offering distinct advantages in capturing environmental data.

One of the applications of roadside perception systems is cooperative driving, which refers to the coordination and collaboration among vehicles and infrastructure to achieve safer and more efficient traffic flow. Cooperative perception aims to share locally perceived data with other vehicles and roadside infrastructure, and is one of the necessary components of cooperative driving [11]. This technology can enhance the situational awareness of the vehicles, overcome the limitations of onboard sensors, such as long-range, occlusion, and blind-spot issues [6], [7], and increase the accuracy and robustness of detection results [12], [13], [14].

Even as cooperative perception technologies advance and an array of off-the-shelf products emerge, a critical hurdle remains in the field of roadside perception systems: the absence of standardized evaluation methods and benchmarks. Related fields, including autonomous driving and object tracking, have established substantial benchmarks and evaluation metrics that have greatly propelled research and application. For instance, in the domain of autonomous driving, several resources such as the KITTI dataset [15], nuScenes [16], and the Waymo Open Dataset [17], have emerged as pivotal benchmarks for academic and industrial progression. These resources have also played a crucial role in the development of standardized metrics to accurately assess the performance of autonomous driving systems. A parallel trend is observed in the sphere of video-based multiple object tracking (MOT), where challenges like MOT16/17/20 [18], [19], and TAO [20] have defined rigorous, objective and fair metrics to measure performance.

Despite these advancements in related fields, there is a pressing need to formulate or adapt such comprehensive evaluation methodologies specifically for roadside perception systems.

Standardized evaluation of roadside perception systems presents notable complexities, especially compared to traditional benchmarking methodologies used for MOT or onboard detection algorithms. These complexities originate from the variety of sensors employed, such as cameras, LiDAR, and radar, as well as their assorted combinations. Additionally, practical limitations in accessing the internal algorithms of these systems further compound the challenges faced.

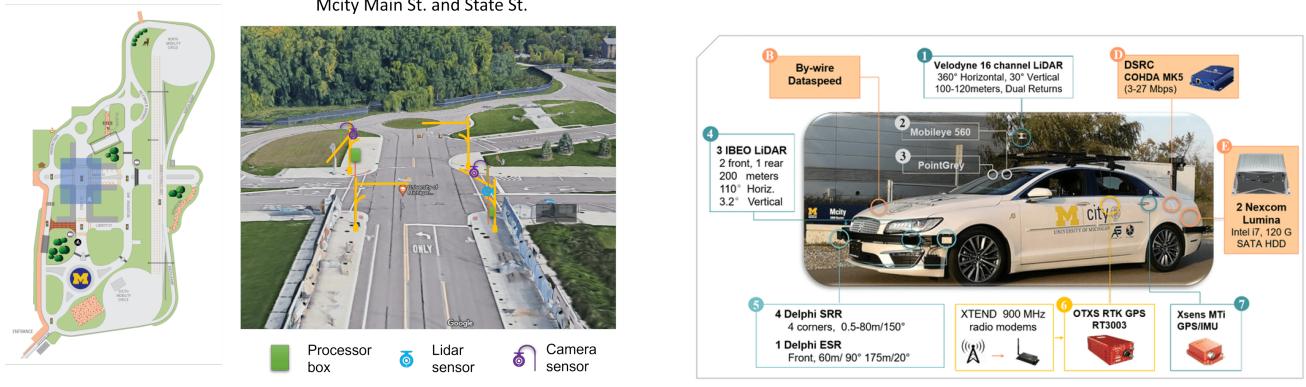
III. PROBLEM FORMULATION

The assessment process is carried out in Mcity, a controlled testing environment that replicates realistic urban road conditions. The perception systems are strategically deployed at a chosen intersection within Mcity. The positioning of the sensors is meticulously optimized to suit the unique detection capabilities of each sensor type, ensuring the best possible view of road user movements (Refer to Figure 1a for the sensor placement and an aerial view of Mcity). The evaluation of vehicle detection is facilitated through the use of two autonomous vehicles, each equipped with Real-Time Kinematic (RTK) Global Positioning System (GPS) technology. The RTK GPS provides highly precise trajectory data of the vehicle, with accuracy down to a few centimeters. This level of precision considerably surpasses the accuracy typically attainable by roadside perception systems and more than adequately meets the requirements for autonomous driving. Therefore, we use these recordings from RTK GPS as the ground truth in our evaluation process. Regarding pedestrian detection, a portable RTK GPS device will be carried by the pedestrian to capture their precise trajectory, thereby providing the ground truth for pedestrian movement. An experimental vehicle and its setup used in our experiment can be seen in Figure 1b.

Given the absence of direct access to the systems under evaluation and the 'black box' nature of their perception algorithms, the evaluation focuses solely on the final output of the systems. We operate under the assumption that the perception system will periodically output a list of detected entities. Each entry in this list should contain a minimum of four attributes: latitude, longitude, category (distinguishing between vehicle or pedestrian), and a unique identifier (id). During each experimental trial, we collect two sets of data: the detected trajectories from the perception system and the corresponding ground truth trajectories obtained from the RTK GPS data. The primary objective and problem being addressed in this paper is the establishment of a standardized, full-stack evaluation solution that hinges on the use of detection data and ground truth data collected as described above. The goal is to systematically evaluate, fairly compare, and establish benchmarks for roadside perception systems.

IV. METHODOLOGY

In this section, we will explore three essential parts of our proposed evaluation methodology. The first part centers around measurement techniques. Here, we introduce a method for



(a) Sensor placement and a view of Mcity

(b) Depiction of the experimental vehicle setup

Fig. 1: Illustrations of the experimental setup for roadside perception systems testing in Mcity.

estimating two key properties: sensor latency and positioning error. These values are convoluted within the raw experimental data and can't be separately measured, necessitating the development of an estimation method. The second element focuses on our selection of metrics. During this portion, we introduce the specific metrics that have been chosen to evaluate the performance of roadside perception systems. The final part pertains to the design of our experimental trials, specifying the planned movement patterns of vehicles and pedestrians during each trial.

A. Measurement Techniques

Our evaluation method is solely executed based on the detection data and ground truth data gathered as described in the Section III. The latency and positioning error values are intrinsically interwoven within the data, making their direct measurement unfeasible. Consequently, in this subsection, we focus on our developed estimation method specifically designed to assess these two critical values.

1) Assumptions and Mathematical Modeling: Considering a scenario where a single vehicle is traversing along a one-dimensional straight line, the vehicle's location at a specific time t is denoted as $G(t)$. Concurrently, we denote the detected location at time t as $D(t)$. Given a specific time, t_1 , the vehicle's location is $G(t_1)$. Simultaneously, we can identify a detection point generated by the detection system at the same location, timestamped as t_2 . This association establishes the relationship $G(t_1) = D(t_2)$. We denote the latency at this point as \mathbf{l} . The following equation emerges:

$$D(t_2) = G(t_2 - \mathbf{l}) + e_1 + \mathbf{e}_2 = G(t_1) \quad (1)$$

In this equation, e_1 represents the detection system's constant offset, and \mathbf{e}_2 represents the random error, which is a zero-mean random variable. It's crucial to note that Equation (1) always holds true, as any error can be deconstructed into a static offset plus a zero-mean random error.

2) Latency Measurement: Now, let's consider a scenario where the vehicle is traveling at a constant speed, v_0 . This results in:

$$G(t) = v_0 t \quad (2)$$

Substituting Equation (2) into Equation (1) and denoting $\tau = t_2 - t_1$, we derive:

$$\tau = t_2 - t_1 = 1 - \frac{e_1}{v_0} - \frac{\mathbf{e}_2}{v_0} \quad (3)$$

It's essential to observe that in Equation (3), τ, l, e_2 are random variables, while the rest are static terms. If we apply the expectation operation to both sides of Equation (3), we get:

$$E[\tau] = E[l] - \frac{e_1}{v_0} \quad (4)$$

Similarly, when the vehicle is traveling at speed $-v_0$:

$$E[\tau'] = E[l] + \frac{e_1}{v_0} \quad (5)$$

Adding Equations (4) and (5) yields:

$$E[\tau] + E[\tau'] = 2E[l] \quad (6)$$

In Equation (6), τ and τ' are the two random variables that can be directly observed and sampled. Consequently, the expectation of these two random variables can be estimated, thereby allowing us to estimate the expected value of the latency.

Following this mathematical model, we designed an experiment to sample the variables τ and τ' , and hence estimate the average latency. As depicted in Figure 2, one vehicle drives back and forth in a straight line during the trial. Each trip consists of three areas: an acceleration area, an area of constant speed, and a deceleration area. We have designed the experiment in such a manner that the area of constant speed is ideally located within the optimal detection range of the sensor under evaluation. In this area, the driver exerts utmost effort to maintain a constant speed of either v_0 or $-v_0$. A series of test points are established within this area. At each point, we record the time difference between the detection timestamp and the ground truth timestamp ($t_2 - t_1$ as in equation 3). Subsequently, by calculating the average

of all the sampled time differences, we use this value as the estimated average latency. It's important to note that this experiment must be conducted in pairs of trips (back and forth) in order to neutralize the static error term $\frac{e_1}{v_0}$ in equation (4). In practical implementation, a small speed variation is inevitable, Appendix A gives a brief analysis on the effect of speed variation, and Appendix B gives the analysis on the variance of this method.

3) *Estimating Positioning Error*: With the estimated average latency \tilde{l} from the method described above, we can compute the estimated positional error for a given detection point using the following estimator: $\tilde{e}_d = D(t) - G(t - \tilde{l})$. To demonstrate the viability of this estimator, we present the following theorem, a brief analysis of its variance is also presented in Appendix C.

Theorem 1. *The estimator \tilde{e}_d is an unbiased estimator of the positional error.*

Proof. We start with the expectation of the estimator:

$$E[\tilde{e}_d] = E[D(t) - G(t - E[l])]$$

Rearranging and splitting the terms gives us:

$$= E[D(t) - G(t - l) + G(t - l) - G(t - E[l])]$$

Substituting the expression from equation (1) into the expectation, we get:

$$E[\tilde{e}_d] = E[e_1 + e_2] + v_0 E[E[l] - l] = e_1$$

Thus, we conclude that the estimator is unbiased. \square

B. Metrics Selection

1) *Data Representation and Preprocessing*: The first step in our methodology requires an understanding of the fundamental concepts of data representation. To begin with, we define **Data Point** as point representation of a single measured object. At its core, a data point encapsulates at least three attributes: a timestamp, a location (latitude and longitude), and an ID. Data points serve as the most basic and atomic data structure in our evaluation data abstraction.

Next, we introduce two essential data structures: data frame and trajectory. A Data Frame consists of multiple data points, all sharing an identical timestamp. It encapsulates a singular time frame, encompassing all detected or ground truth objects at that specific time instance. A Trajectory is a collection of data points bearing different timestamps (arranged in ascending order) but sharing the same ID. This structure illustrates the movement path or trajectory of a particular detected or ground truth object over time. The objective of preprocessing is to parse the detection and ground truth data into these structures. We group each set of data by time to create multiple time frames and by ID to form various trajectories. All these data structures are then collectively stored as a '**trajectory set**'.

2) *Matching Mechanisms*: A significant portion of our chosen metrics rely on the correct alignment of data points, frames, and trajectories between the detection results and the ground truth data. For this purpose, we define two types of matching mechanisms: **Point Matching** and **Association Matching**.

a) **Point Matching**: In this process, we first establish a match between a data frame from the detected data and the ground truth data frame that possesses the closest timestamp (after subtracting the estimated latency as described in Section IV-A2). Subsequently, we execute a point-to-point matching between the data points in the detected data frame and the corresponding data points in the ground truth data frame. For this operation, we employ the Hungarian method [21], which minimizes the matching distance and subsequently achieves an optimal one-to-one correspondence between the points in both frames.

b) **Association Matching**:: This matching mechanism amplifies the concept of point matching and applies it to trajectories. Here, we first adopt the Hungarian method to pair detected trajectories with the ground truth trajectories, by maximizing the number of true positives. Following this, we perform **point matching** within each pair of the matched trajectories.

The two matching mechanisms will yield a set of matched data point pairs from the detected data points to the ground truth data points. It's noteworthy that these matching mechanisms slightly diverge from those utilized in common image-based tracking benchmarks [18], [19], [20]. In our case, we need to incorporate an extra nearest-time matching step. This distinction arises because, unlike standard scenarios where the ground truth is annotated on the identical frames, our ground truth is independently captured by a standalone GPS device, which operates asynchronously with the the detection system.

3) *True Positives, False Positives, False Negatives, and ID Switches*: We employ standard terminology used in binary classification tasks to identify and analyze the outcomes of our matching process. An illustrative example is provided below (Figure 3) for further clarification.

- **True Positives (TP)**: A True Positive instance arises when a detected data point is successfully matched with a ground truth data point, with the distance between the two not exceeding 1.5 meters. This indicates that the detection system has accurately identified an object and correctly estimated its location within an acceptable margin of error.
- **False Positives (FP)**: False Positives denote cases where a data point is detected, but no corresponding ground truth data point exists, or the distance between the detected point and the matched ground truth data point exceeds 1.5 meters. These scenarios suggest that the detection system has erroneously detected an object or incorrectly estimated an object's location beyond the acceptable tolerance.
- **False Negatives (FN)**: False Negatives occur when a ground truth data point exists, but the detection system fails to identify a corresponding detected data point. This represents a detection failure by the system.
- **Identity Switches (ID Switches)**: An ID Switch happens when the identity assigned to a specific trajectory by the detection system gets wrongly associated with another trajectory. In other words, the tracking system loses track of the original object and erroneously associates the track of a different object to it.

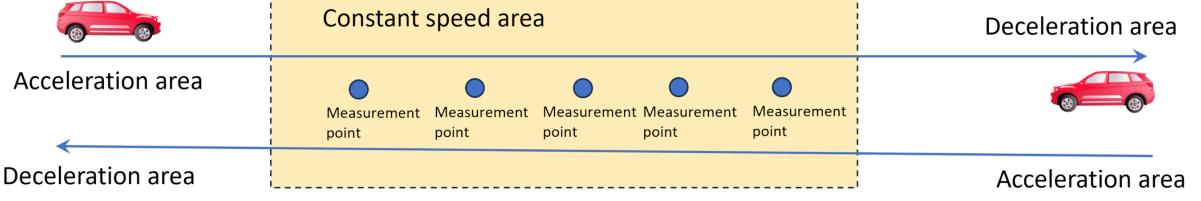


Fig. 2: Depiction of the experimental setup for latency measurement, showcasing the vehicle's movement path involving acceleration, constant speed, and deceleration areas.

The matched pairs of points discussed above can be the outcome of either **Point Matching** or **Association Matching**. For simplicity, we denote the true positives, false positives, false negatives obtained from **Point Matching** as TP, FP, FN, and those from **Association Matching** as TPA, FPA, and FNA. The 1.5m threshold in accordance with the SAE2945 standard [22].

4) *Multiple Object Tracking Precision (MOTP)*: Multiple Object Tracking Precision (MOTP) [23] is a measure of the tracking algorithm's precision in localizing the detected objects. It is computed as the total distance between each True Positive (TP) and its corresponding ground truth object, divided by the total number of True Positives. Mathematically, it can be defined as:

$$MOTP = \frac{\sum_{t,i} d_{t,i}}{\sum_t c_t} \quad (7)$$

where $d_{t,i}$ is the distance of the i th match in frame t , and c_t is the number of matches in frame t . Lower MOTP values correspond to better localization precision of the detection algorithm.

5) *Multiple Object Tracking Accuracy (MOTA)*: Multiple Object Tracking Accuracy (MOTA) [23] evaluates the overall tracking performance, taking into account False Positives (FP), False Negatives (FN), and Identity Switches (IDS_w). It is calculated as follows:

$$MOTA = 1 - \frac{\sum_t (FP_t + FN_t + IDS_w)_t}{\sum_t g_t} \quad (8)$$

where FP_t is the number of false positives in frame t , FN_t is the number of false negatives in frame t , IDS_w_t is the number of identity switches in frame t , and g_t is the number of ground truth objects in frame t . Higher MOTA scores indicate better tracking accuracy. Notice the FP, FN are calculated with **point matching**.

6) *IDF1 Score*: The IDF1 score [23] is a measure that encapsulates both the precision and recall of the identification process in multi-object tracking. It is particularly designed for tasks where maintaining consistent identities of the tracked objects is of significant importance.

To elaborate further, the IDF1 score is the harmonic mean of Identification Precision (IDP) and Identification Recall (IDR):

- **Identification Precision (IDP)**: IDP is the proportion of correctly identified detections out of all detections. Mathematically, it's defined as $IDP = \frac{TPA}{TPA+FPA}$,

where TPA is the number of true positive associations and FPA is the number of false positive associations.

- **Identification Recall (IDR)**: IDR, on the other hand, is the ratio of correctly identified detections to the total number of ground truth objects. It's defined as $IDR = \frac{TPA}{TPA+FNA}$, with FNA representing the number of false negative associations.

Combining these two measures, the IDF1 score is computed as:

$$IDF1 = \frac{2 \times IDP \times IDR}{IDP + IDR} = \frac{2 \times TPA}{2 \times TPA + FPA + FNA} \quad (9)$$

In essence, a high IDF1 score indicates that not only the detection system correctly identified and located the objects, but also consistently maintained their identities throughout the tracking process.

7) *Higher Order Tracking Accuracy (HOTA)*: Higher Order Tracking Accuracy (HOTA) [24] is a comprehensive metric that accounts for both detection and association accuracy in object tracking scenarios. HOTA is based on two core components: Detection Accuracy (DetA) and Association Accuracy (AssA).

Detection Accuracy (DetA) assesses the ability of the system to correctly detect objects, without considering their identities. The formula to calculate DetA is given by:

$$DetA = \frac{TP}{TP + FP + FN} \quad (10)$$

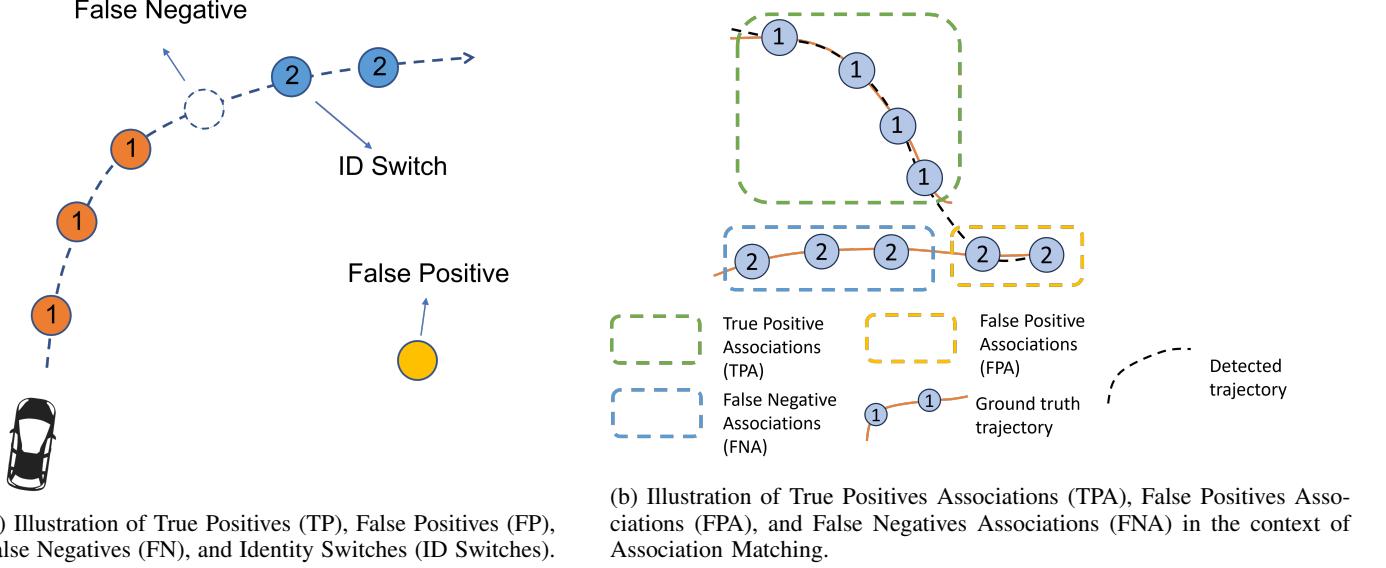
Association Accuracy (AssA), on the other hand, evaluates the capacity of the system to correctly associate detections to the same object over time, in essence, evaluating the tracking aspect. AssA is computed using:

$$AssA = \frac{TPA}{TPA + FPA + FNA} \quad (11)$$

The overall HOTA score is then calculated by taking the geometric mean of DetA and AssA, which balances both detection and association aspects.

C. Experiment Trial Design

The experimental trials are designed to evaluate different metrics across a diverse range of scenarios. For this purpose, we choose the intersection of State Street and Main Street in Mcity as our trial location. This intersection, with two lanes on each street, represents a medium-sized intersection scenario



(a) Illustration of True Positives (TP), False Positives (FP), False Negatives (FN), and Identity Switches (ID Switches).

Fig. 3: Illustration of True Positives, False Positives, False Negatives, and Identity Switches in the context of our matching process.

that is commonplace in real-world urban transportation networks. The trials are organized into four distinct categories: latency trials, one-vehicle trials, one-vehicle-with-pedestrian trials, and two-vehicle-with-pedestrian trials.

The first two trials are designed specifically for latency evaluation. Each of these trials involves a vehicle driving back and forth in a straight line, maintaining a consistent speed within predefined constant speed zones. The first trial is conducted along the east-west direction, while the second trial follows the north-south direction. These trials are designed according to section IV-A2 and shown in Fig. 4a. The latency for each trial is estimated independently, and the final latency value is calculated as the average of the two trials. This estimated latency is then utilized in the evaluation of subsequent trials.

The second set of trials comprises eight individual trials, each involving a single vehicle. These trials cover all possible maneuvers at the intersection, providing a comprehensive examination of the detection system's performance in diverse vehicle movement scenarios as shown in Fig. 4b. This collection of trials aims to provide direct observation on the system's geographical coverage and its performance in handling diverse movement scenarios. Moreover, these trials offer detailed, granular insights into the system's performance by gauging its specific responsiveness to individual vehicle movements. By examining the system's performance under these conditions, we can gain a nuanced understanding of its capabilities and potential areas for improvement.

The subsequent pair of trials incorporates both a vehicle and a pedestrian. One trial directs the vehicle along an east-west axis, while the alternate trial positions the vehicle on a north-south trajectory. These trials, as depicted in Fig. 4c, have been crafted to examine the detection system's performance under a mixed scenario, where both vehicles and pedestrians are in play. The system's performance under these scenarios directly informs us of its capabilities in ensuring the safety

of vulnerable road users (VRUs) when implemented in a production setting.

The final set of trials involves three tests, each involving two vehicles and a pedestrian. Two trials simulate car-following situations (one in the east-west direction, as shown in Fig. 4d and the other in the north-south direction, as shown in Fig. 4e), and the third involves two vehicles approaching the intersection from perpendicular directions, as illustrated in Fig. 4f. These trials are designed to provide an evaluation of the system under complex scenarios. Factors such as occlusion of the following vehicle by the leading vehicle in car-following situations and the need to distinguish multiple detected objects pose additional challenges. Particularly, the trial involving perpendicularly approaching vehicles tests the system's capacity to warn of intersection crashes and red-light violations.

V. EXPERIMENTS AND RESULTS

A. Experiments Setup

The experiments were conducted on October 13th, 2023, in Mcity. Two autonomous vehicles and one pedestrian, each equipped with RTK GPS, were utilized for the trials to provide high-precision ground truth data. The same experiments were performed to evaluate the Three different roadside perception systems. To respect the confidentiality agreements in place, we will not be revealing the names of these roadside perception systems under evaluation in our study. Instead, we will refer to them as **System A**, **System B** and **System C** throughout the rest of the paper. Please note that this does not in any way affect the validity or integrity of our evaluation process and results. The **System A** employs Lidar technology, while **System B** and **System C** uses fisheye image sensors. These three systems represent the most popular choices in current roadside perception sensor technology. We also offer the source code used in this study for evaluating these trial data for

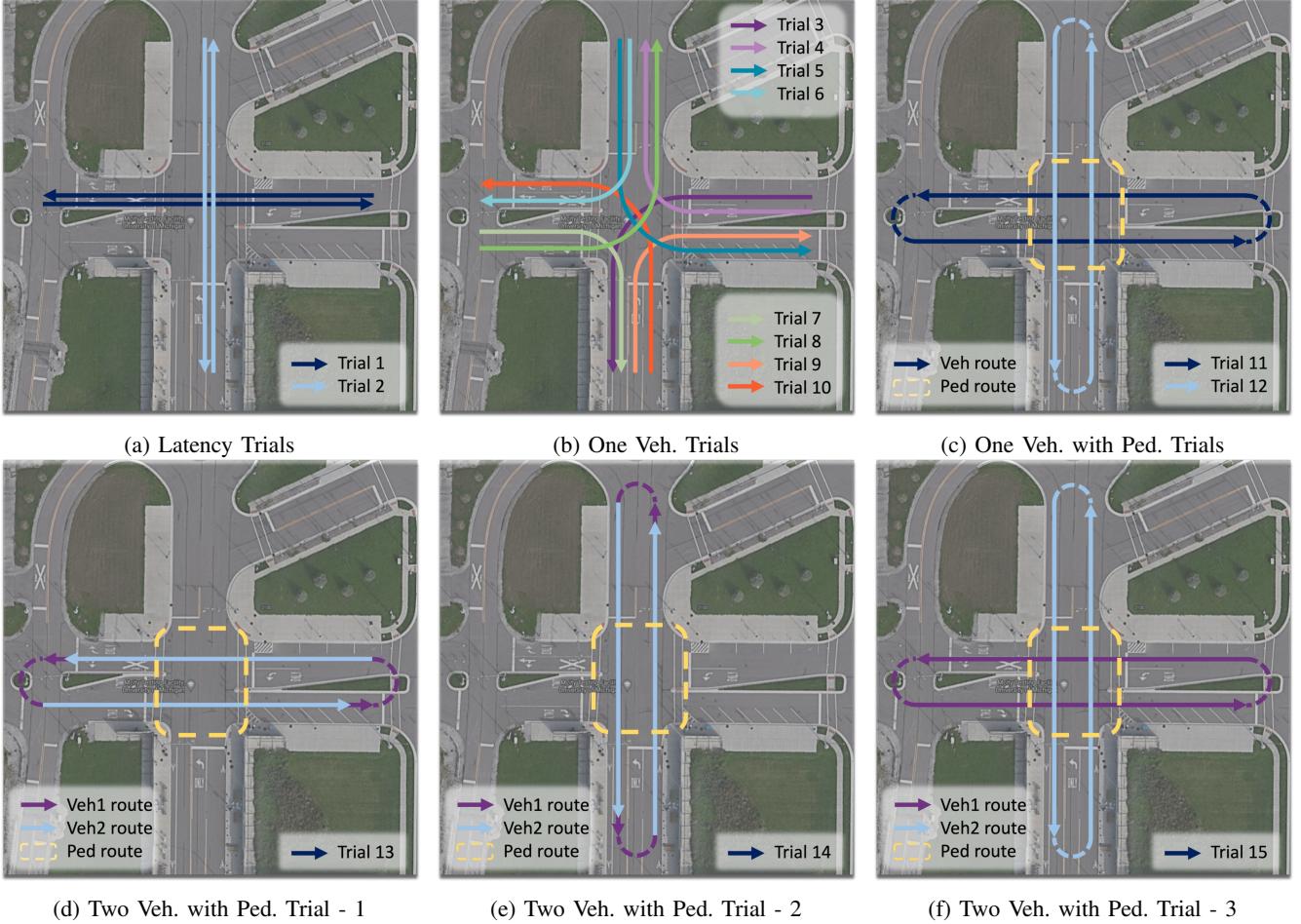


Fig. 4: Illustrations of all trials designed for evaluation. (a) Latency trials, (b) One Vehicle Trials, (c) One Vehicle with Pedestrian Trials, and (d), (e), (f) Two Vehicles with Pedestrian Trials.

further research facilitation. The code is accessible on GitHub at [25].

We conducted the trials at two different times: during the daytime at 13 PM and in evening at 6 PM. This strategic choice allowed us to evaluate the performance of the perception systems under varying light conditions. We executed the experimental trials, as outlined in the previous section, for each system. The detection results and the ground truth data were diligently recorded for subsequent analysis. The details and results of the experiments are presented in the following section.

B. Results

TABLE I: Latency Measurements of Perception Systems

System	North-South	East-West	Mean	Std.
System A	41 ms	54 ms	48 ms	0.025
System B	137 ms	153 ms	145 ms	0.106
System C	1740 ms	1690 ms	1715 ms	0.061

1) *Latency Measurement Results:* We conducted two trials for each perception system to measure the latency. This was achieved with the method described in Section IV-A2 by driving vehicles back and forth along predetermined routes with

constant speed. The first trial involved driving in the north-south direction, and the second in the east-west direction. Each trial consisted of five rounds of driving. The results of these measurements are presented in Table I. System A demonstrates a latency of 48 milliseconds, which is well within the acceptable range for cooperative driving tasks, showcasing its efficiency in real-time responsiveness. System B, with a latency of 145 ms, also fits within a threshold suitable for a broad spectrum of applications. In contrast, System C exhibits a latency of 1715 ms, this level of latency falls short of the requirements for many real-time applications, highlighting a critical area for improvement.

As observed in Table I, the latency measured during the North-South trials is close to the latency observed in the East-West trials for both systems, with a small proportional variance. This consistency provides validation for our measurements, reinforcing our confidence in the reliability of the latency estimates. The latency measurements obtained here are utilized in the subsequent assessments described in the following sections.

2) *Qualitative Results Review:* The qualitative analysis is predicated on a visual comparison between the detected trajectories and the ground truth data, as illustrated in Figure

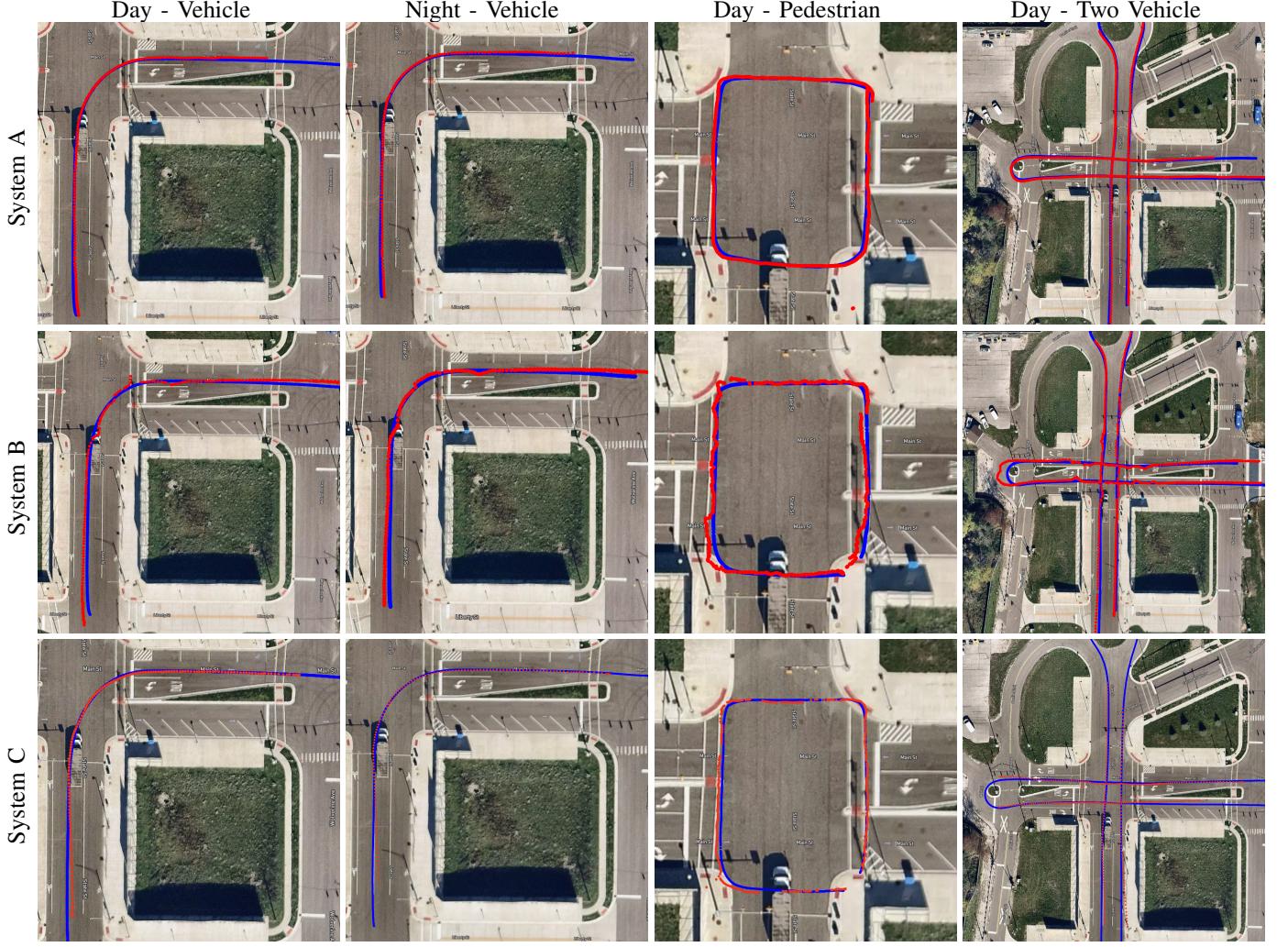


Fig. 5: The detection results of the System A (first row), the System B (second row), and the System C (third row) in different trials. The red dots are detection results and the blue dots are ground truth.

5. In this figure, each row corresponds to results from the same system, while each column represents results from the same trial across different systems. The detection results are indicated by red dots, and the ground truth data is represented by blue dots.

The first column of the figure displays the detection results from Trial 3, and the second column presents the results from the same trial conducted at night. Notably, all systems perform relatively well under varying light conditions. This performance is particularly surprising for System B and C, which are image-based detection systems, implying that its underlying algorithm effectively handles the challenges posed by low light conditions. The third column of the figure presents the pedestrian detection results. In this scenario, System A and B provide better localization results. The fourth column in the figure shows the results from Trial 15, which involves two vehicles approaching the intersection from perpendicular directions. All systems demonstrate capability in tracking multiple vehicles, with results aligning closely with the ground truth data.

In summary, the visualized results indicate that all Systems perform well in general, accurately detecting and localizing road users in various conditions.

3) Quantitative Results: Detailed quantitative results for each trial are outlined in six separate tables. Table II presents the results for daytime trials conducted using System A, Table IV displays same results for System B and VI for System C. The performance of all systems under night conditions are presented separately in Table III for System A, Table V for System B and Table VII for System C. In each table, each row corresponds to a specific trial. It is important to note that for Trials 11 to 15, the metrics were calculated separately for vehicles and pedestrians, despite the simultaneous detection of both during the trials.

Insights drawn from Tables II and III reveal that System A demonstrates notable efficacy, particularly in the domains of vehicle detection, tracking, and localization. This system records impressive scores in MOTA, IDF1, and HOTA across most vehicle trials conducted during both daytime and nighttime conditions. However, its performance in pedestrian

TABLE II: Experimental Results for System A in Daytime

Category	Trial	FP Rate (%) ↓	FN Rate (%) ↓	IDS ↓	MOTA (%) ↑	MOTP (m)↓	IDF1 (%) ↑	HOTA (%) ↑
Vehicle	Trial 3	0.0	7.2	0	92.0	0.402	96.3	96.3
	Trial 4	0.0	7.6	0	91.8	0.478	96.1	96.1
	Trial 5	0.0	0.5	0	99.3	0.386	99.7	99.7
	Trial 6	0.0	0.7	0	99.3	0.373	99.7	99.7
	Trial 7	0.0	0.5	0	99.4	0.507	99.8	99.8
	Trial 8	0.0	0.6	0	99.4	0.424	99.7	99.7
	Trial 9	0.0	0.4	0	99.3	0.415	99.8	99.8
	Trial 10	0.0	1.9	0	97.8	0.376	99.0	99.0
	Trial 11	0.0	2.2	0	97.9	0.408	98.9	98.9
	Trial 12	0.0	1.0	0	98.8	0.555	99.5	99.5
	Trial 13	0.0	2.7	0	97.3	0.470	98.6	97.3
	Trial 14	0.0	0.7	0	99.2	0.568	99.6	96.6
	Trial 15	0.0	1.5	0	98.5	0.523	99.2	96.7
Pedestrian	Trial 11	6.5	0.0	0	93.5	0.383	96.8	93.6
	Trial 12	7.8	0.0	0	92.2	0.370	96.0	92.2
	Trial 13	4.8	0.0	0	95.2	0.378	97.8	95.5
	Trial 14	11.1	0.0	0	88.9	0.336	94.5	89.3
	Trial 15	8.6	0.0	0	91.4	0.368	95.6	91.4

TABLE III: Experimental Results for System A in Night

Category	Trial	FP Rate (%) ↓	FN Rate (%) ↓	IDS ↓	MOTA (%) ↑	MOTP (m)↓	IDF1 (%) ↑	HOTA (%) ↑
Vehicle	Trial 3	0.0	7.2	0	92.1	0.342	96.3	96.3
	Trial 4	0.0	9.5	0	89.5	0.335	95.0	95.1
	Trial 5	0.0	0.7	0	99.3	0.413	99.6	99.6
	Trial 6	0.0	0.5	0	99.1	0.367	99.7	99.7
	Trial 7	0.0	0.7	0	99.3	0.343	99.6	99.6
	Trial 8	0.0	0.8	0	99.2	0.247	99.6	99.6
	Trial 9	0.0	0.2	0	100.0	0.409	99.9	99.9
	Trial 10	0.0	0.0	0	100.0	0.339	100.0	100.0
	Trial 11	0.0	2.3	0	97.7	0.413	98.8	98.8
	Trial 12	0.0	0.2	0	99.7	0.354	99.9	99.9
	Trial 13	0.0	2.1	0	97.9	0.390	98.9	97.8
	Trial 14	0.0	0.4	0	99.5	0.390	99.8	97.4
	Trial 15	0.0	1.6	0	98.3	0.400	99.2	97.2
Pedestrian	Trial 11	1.9	0.0	0	98.1	0.382	99.0	98.1
	Trial 12	2.9	0.0	0	97.1	0.345	98.7	97.2
	Trial 13	5.6	0.0	0	94.4	0.341	97.2	94.5
	Trial 14	2.1	0.0	0	97.9	0.324	99.0	98.0

detection is marginally less effective, as evidenced by a higher incidence of false positives. This observation aligns with expectations, considering that pedestrians, being smaller and consequently less prominent in the LiDAR point cloud, pose a greater challenge for accurate detection.

Tables IV and V suggest that System B's performance is notably inferior to that of System A. Despite the visually satisfactory outputs depicted in Figure 5, System B's quantitative results fall short of expectations. This divergence is largely due to the system's latency variations, characterized by a high standard deviation, as detailed in Table I. Such fluctuations play a critical role in contributing to localization errors. The reason for this is that when calculating localization errors, we match the detection point with the ground truth data from the same time frame. However, visually, all detection

points are plotted on the map irrespective of their detection time, masking the impact of latency variation. These factors predominantly affect the system's localization accuracy, resulting in comparable performance levels during both daytime and nighttime trials. However, System B excels in pedestrian detection, demonstrating significantly improved results. This enhancement is largely because the slower movement of pedestrians lessens the impact of latency variation on their localization.

Tables VI and VII present the performance metrics of System C. These tables reveal that System C outperforms System B in vehicle detection, primarily due to its reduced latency variation. However, there still exists a noticeable performance gap when compared to System A. Regarding pedestrian detection, System C shows less optimal results,

TABLE IV: Experimental Results for System B in Daytime

Category	Trial	FP Rate (%) ↓	FN Rate (%) ↓	IDS ↓	MOTA (%) ↑	MOTP (m) ↓	IDF1 (%) ↑	HOTA (%) ↑
Vehicle	Trial 3	37.2	33.6	1	30.1	0.758	47.4	39.7
	Trial 4	13.7	22.1	0	61.8	0.847	81.9	72.5
	Trial 5	33.1	39.8	0	22.6	0.919	63.4	48.3
	Trial 6	44.7	54.1	0	-9.8	0.885	50.2	35.8
	Trial 7	69.1	75.4	0	-64.0	0.897	27.4	17.0
	Trial 8	35.1	37.9	0	25.3	0.754	63.5	47.3
	Trial 9	15.5	25.1	0	55.8	0.723	79.4	69.4
	Trial 10	29.3	43.2	0	16.7	0.881	63.0	50.1
	Trial 11	26.3	30.9	1	40.5	0.781	71.3	58.3
	Trial 12	34.3	39.4	0	22.9	0.770	63.1	47.5
	Trial 13	35.4	46.5	0	8.5	0.786	58.5	42.2
	Trial 14	39.1	45.1	0	10.8	0.745	57.7	41.0
	Trial 15	39.4	44.6	0	11.9	0.825	57.9	40.9
Pedestrian	Trial 11	5.9	5.5	0	88.6	0.640	94.3	91.4
	Trial 12	5.9	4.1	0	90.2	0.519	95.0	92.3
	Trial 13	6.9	4.5	0	88.7	0.670	94.3	89.9
	Trial 14	7.8	6.3	0	86.0	0.650	92.9	89.0
	Trial 15	6.3	6.7	0	87.0	0.666	93.5	90.0

TABLE V: Experimental Results for System B in Night

Category	Trial	FP Rate (%) ↓	FN Rate (%) ↓	IDS ↓	MOTA (%) ↑	MOTP (m) ↓	IDF1 (%) ↑	HOTA (%) ↑
Vehicle	Trial 3	41.7	35.2	1	25.8	0.817	43.9	35.7
	Trial 4	34.8	24.1	1	43.7	0.846	46.2	41.7
	Trial 5	41.7	45.4	0	9.8	0.794	56.4	40.3
	Trial 6	28.7	38.9	0	25.5	0.883	65.8	52.1
	Trial 7	65.2	65.2	0	-30.4	1.090	34.8	21.1
	Trial 8	31.1	32.2	0	36.1	0.790	68.3	52.4
	Trial 9	22.1	32.8	0	40.0	0.798	72.1	60.0
	Trial 10	10.2	20.8	0	66.1	0.793	84.2	77.0
	Trial 11	31.8	43.3	0	16.2	0.808	61.9	48.2
	Trial 12	31.3	37.5	0	27.3	0.786	65.4	50.4
	Trial 13	22.3	31.4	0	42.1	0.803	72.9	58.6
	Trial 14	33.7	39.0	0	23.9	0.770	63.5	46.9
	Trial 15	27.7	34.5	0	34.2	0.839	68.7	53.1
Pedestrian	Trial 11	47.2	0.0	0	52.8	0.591	81.8	60.2
	Trial 12	45.9	0.0	0	54.1	0.603	84.2	62.7
	Trial 13	47.5	0.0	0	52.5	0.588	82.1	60.1
	Trial 14	45.7	0.0	0	54.3	0.551	84.5	62.8

with this shortcoming being particularly pronounced during daytime trials.

Table VIII provides a comprehensive summary of the mean results from all trials conducted with Systems A, B, and C, encompassing various conditions and targeting both vehicles and pedestrians. The data consolidated in this table corroborate our earlier discussions, clearly indicating that the overall performance of Systems B and C falls short of System A's efficiency. This disparity is particularly evident in image-based systems, which appear to struggle with latency issues more than lidar-based systems. This is likely due to more intensive processing tasks required by image sensors. Addressing this latency challenge is evidently a critical area for improvement in these systems.

Furthermore, Table VIII sheds light on the influence of light-

ing conditions on the three systems. It becomes evident that lighting conditions do not affect System A, a LiDAR-based system, demonstrating its robustness in varied environments. In contrast, Systems B and C exhibit noticeable differences in pedestrian detection performance under varying lighting conditions. Between System B and System C, System C demonstrates superior vehicle detection, and System B shows better performance in pedestrian detection.

It is also significant to note that the Mean Object Tracking Precision (MOTP) metric for System A is comparatively lower. A lower MOTP value indicates superior localization accuracy, an area where LiDAR-based detection systems, such as System A, are widely recognized to excel. This advantage is primarily due to the inherent capabilities of LiDAR technology in precisely mapping object locations, a feat that often challenges

TABLE VI: Experimental Results for System C in Daytime

Category	Trial	FP Rate (%) ↓	FN Rate (%) ↓	IDS ↓	MOTA (%) ↑	MOTP (m) ↓	IDF1 (%) ↑	HOTA (%) ↑
Vehicle	Trial 3	1.4	4.5	1	93.0	0.458	55.1	60.8
	Trial 4	6.6	22.1	0	66.9	0.496	85.0	80.4
	Trial 5	19.4	46.5	1	9.2	0.705	34.2	37.5
	Trial 6	33.7	56.1	0	-18.4	0.514	52.8	42.2
	Trial 7	19.4	41.5	0	23.4	0.645	67.8	58.9
	Trial 8	24.5	48.4	1	3.6	0.596	36.2	36.7
	Trial 9	10.9	27.2	0	55.5	0.875	80.1	73.3
	Trial 10	16.1	35.7	1	36.4	0.584	38.2	41.5
	Trial 11	22.4	38.6	0	28.9	0.583	68.5	57.6
	Trial 12	12.6	26.4	0	55.8	0.678	79.9	71.9
	Trial 13	21.4	45.9	1	11.6	0.627	57.4	48.9
	Trial 14	18.3	32.9	2	41.0	0.786	57.4	50.3
	Trial 15	21.6	38.0	2	29.9	0.657	61.6	47.6
Pedestrian	Trial 11	6.8	16.9	4	73.3	0.422	40.7	48.8
	Trial 12	7.2	17.1	4	72.8	0.429	60.4	63.3
	Trial 13	7.3	28.3	7	54.5	0.386	26.8	37.9
	Trial 14	6.9	21.3	2	67.3	0.435	41.7	49.5
	Trial 15	7.1	18.1	3	71.6	0.466	51.9	57.0

TABLE VII: Experimental Results for System C in Night

Category	Trial	FP Rate (%) ↓	FN Rate (%) ↓	IDS ↓	MOTA (%) ↑	MOTP (m) ↓	IDF1 (%) ↑	HOTA (%) ↑
Vehicle	Trial 3	2.3	5.0	1	91.7	0.519	55.3	60.6
	Trial 4	0.0	18.1	0	77.9	0.506	90.0	90.5
	Trial 5	18.3	46.1	1	10.8	0.617	38.5	40.6
	Trial 6	20.6	50.7	0	-2.9	0.560	60.8	53.6
	Trial 7	19.8	37.0	0	33.0	0.460	70.5	60.4
	Trial 8	26.3	30.6	1	40.4	0.523	39.1	39.2
	Trial 9	7.8	24.7	0	62.2	0.767	82.9	77.8
	Trial 10	8.6	20.8	1	66.4	0.568	48.8	52.3
	Trial 11	27.3	42.7	0	18.6	0.596	64.1	53.5
	Trial 12	13.5	26.9	2	53.7	0.559	60.8	57.8
	Trial 13	21.3	37.9	1	30.6	0.566	66.2	54.3
	Trial 14	13.2	30.8	1	47.2	0.655	47.2	46.5
	Trial 15	12.8	24.8	1	53.4	0.522	78.8	58.0
Pedestrian	Trial 11	58.8	29.3	8	23.5	0.380	19.2	19.6
	Trial 12	30.9	29.3	5	39.7	0.400	21.4	27.8
	Trial 13	0.0	54.1	9	-20.7	0.376	20.4	33.7

image-based detection systems.

C. Discussion

From the results discussed above, we observed a notable performance gap between the LiDAR-based and image-based methods. This disparity is primarily attributed to the stringent 1.5-meter distance threshold applied during our evaluation, a standard derived from the SAE2945 [22], which stipulates 1.5 meters as the maximum acceptable localization error for autonomous vehicles using perception results. Figure 6 illustrates the false positive rate and false negative rate of all three systems when different thresholds are applied. The figure reveals that the FP and FN rates of the three systems asymptotically approach a lower value. Notably, System A, the LiDAR-based system, reaches this lower value with a

much smaller localization threshold. This finding indicates that while image-based detection systems do accurately detect road objects, their higher localization errors result in many correctly detected objects being categorized as false positives when evaluated against the 1.5-meter threshold. Consequently, these results suggest a clear area for improvement in image-based systems: enhancing the accuracy of localization in future iterations.

VI. CONCLUSION

In this paper, we have presented a systematic evaluation methodology for roadside perception systems, encompassing measurement techniques, metrics selection, and experimental trial design. We conducted comprehensive experiments on Three representative off-the-shelf roadside perception systems (referred to as System A and System B) within a controlled

TABLE VIII: The Mean Results of System A, System B, and System C over all trials

	FP Rate (%) ↓	FN Rate (%) ↓	IDS ↓	MOTA (%) ↑	MOTP (m)↓	IDF1 (%) ↑	HOTA (%) ↑
Vehicle detection in daytime							
System A	0.0	2.1	0.0	92.0	0.402	96.3	96.3
System B	34.8	41.4	0.2	17.9	0.813	60.4	46.9
System C	16.9	33.8	0.7	37.2	0.641	62.4	56.8
Vehicle detection in night							
System A	0.0	2.0	0.0	97.8	0.365	99.0	98.5
System B	32.5	37.0	0.6	27.7	0.832	61.9	49.0
System C	14.8	30.8	1.0	44.7	0.570	61.7	58.0
Pedestrian detection in daytime							
System A	7.8	0.0	0.0	92.2	0.367	96.1	92.4
System B	6.6	5.4	0.0	88.1	0.629	94.0	90.5
System C	8.1	20.8	3.8	66.5	0.440	44.0	50.9
Pedestrian detection in night							
System A	3.1	0.0	0.0	96.9	0.348	98.5	96.9
System B	46.6	0.0	0.0	53.4	0.583	83.1	61.4
System C	29.9	37.6	7.3	61.5	0.385	20.3	27.0

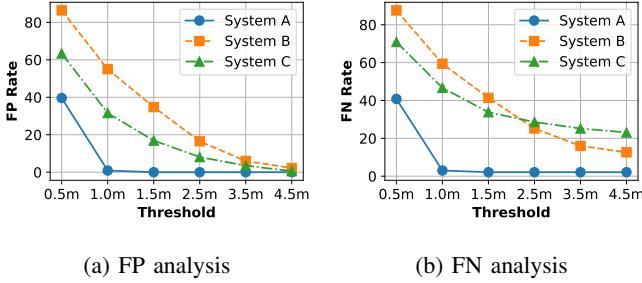


Fig. 6: Evaluation results under different distance threshold.

testing facility. Our experiments evaluated these systems under varying conditions and across diverse traffic scenarios involving vehicles and pedestrians.

The results showcase the viability of our proposed evaluation methodology in enabling standardized, comparative analysis of different roadside perception systems. We were able to quantitatively analyze numerous performance metrics for both systems and identify their respective strengths and weaknesses.

Our study and proposed methodology make significant headway towards establishing standardized benchmarks and evaluation practices for roadside perception systems. We believe widespread adoption of systematic evaluation approaches like ours will greatly benefit the research and development of infrastructure-based perception for autonomous driving. Standardized benchmarks will allow for more rigorous testing, direct comparison between products, identification of limitations, and targeted improvements.

APPENDIX

A. Analysis on Impact of Slight Speed Deviations

In the practical implementation of the latency measurement described in Section IV-A2, the driver endeavors to maintain a constant speed. However, it is inevitable that the speed

will exhibit minor fluctuations. This section offers a brief analysis of the effects of such speed variability. Adhering to the mathematical notation in Section IV-A2, when the speed is variable, we denote the speed $v(t)$ as a continuous random process. Define $G(t) = \int_0^t v(t)dt$. Incorporating this into formula (1), we obtain:

$$\int_0^{t_2-1} v(t)dt + e_1 + \mathbf{e}_2 = \int_0^{t_1} v(t)dt$$

$$e_1 + \mathbf{e}_2 = \int_0^{t_1-t_2+1} v(t)dt$$

Given that $t_1 - t_2 + l$ is a small quantity, the following approximation is viable:

$$\int_0^{t_1-t_2+1} v(t)dt \approx (v_0 + \mu)(t_1 - t_2 + l)$$

Here, μ represents a zero-mean random variable that captures the variability of speed $v(t)$ over a brief duration. Then, the variation in speed can be explicitly formulated in the following equation:

$$(v_0 + \mu)(t_1 - t_2 + l) = e_1 + \mathbf{e}_2 \quad (12)$$

Taking the expectation on both sides confirms that (4) remains valid:

$$E[\tau] = E[l] - \frac{e_1}{v_0}$$

B. Variance Analysis of Latency Measurement

This section presents a variance analysis for the method outlined in Section IV-A2. Employing equation (12), we obtain:

$$\tau = l - \frac{e_1 + \mathbf{e}_2}{v_0 + \mu} \quad (13)$$

When calculating the variance of both sides, we derive:

$$Var(\tau) = Var(l) + Var\left(\frac{e_1}{v_0 + \mu}\right) + Var\left(\frac{\mathbf{e}_2}{v_0 + \mu}\right)$$

This leads to the following approximation, assuming minor fluctuations in μ :

$$\text{Var}(\tau) \approx \text{Var}(l) + \frac{1}{v_0^2} \text{Var}(e_2) \quad (14)$$

Equation (14) indicates that the variability of the measured latency τ is influenced by both the latency and localization errors. It also suggests that choosing a higher traveling speed v_0 during the experiment can reduce the variance, bringing it closer to the true variance of the latency. Furthermore, as $\text{Var}(l)$ remains constant across different locations, the measured latency variance can serve as an indicator of positional accuracy in varying locations.

C. Variance Analysis of Position Measurement

This section provides a concise analysis of the variance associated with the estimator \tilde{e}_d , as utilized in Section IV-A3. Employing equation (12), we derive:

$$\begin{aligned} \text{Var}[\tilde{e}_d] &= \text{Var}[e_1 + e_2] + \text{Var}[(v_0 + \mu)(E[l] - l)] \\ &= \text{Var}[e_2] + (v_0^2 + \text{Var}[\mu])\text{Var}[l] \end{aligned}$$

Assuming a small variance of the term μ , the variance of the estimator $\text{Var}[\tilde{e}_d]$ can be approximated as:

$$\text{Var}[\tilde{e}_d] \approx \text{Var}[e_2] + v_0^2 \text{Var}[l] \quad (15)$$

This equation highlights the compounded effect of latency and positioning errors, echoing the findings in equation (14). It also reveals an intrinsic relationship between (14) and (15), characterized by a factor of v_0^2 .

REFERENCES

- [1] “Smart road-roadside perception industry report, 2021.” <https://www.globenewswire.com/news-release/2021/06/25/2253283/0/en/Smart-Road-Roadside-Perception-Industry-Report-2021.html>, 2021. Accessed: 2023-02-03.
- [2] “China smart-road roadside perception industry report 2022.” <https://www.globenewswire.com/en/news-release/2022/10/03/2526871/28124/en/China-Smart-Road-Roadside-Perception-Industry-Report-2022-The-Four-Tech-Tycoons-Huawei-Baidu-Alibaba-and-Tencent-HBAT-have-All-Entered-the-Smart-Road-Roadside-Perception-Market.html>, 2022. Accessed: 2023-07-06.
- [3] Derq, “Derq: Ai for safer and smarter roads.” <https://www.derq.com/>, 2020. Accessed: 2023-02-03.
- [4] Ouster, “Ouster: The leading provider of digital lidar sensors.” <https://ouster.com/>, 2020. Accessed: 2023-02-03.
- [5] University of Michigan, “Mcity: University of Michigan.” <https://mcity.umich.edu/>, 2023. Accessed: July 30, 2023.
- [6] M. Tsukada, T. Oi, M. Kitazawa, and H. Esaki, “Networked roadside perception units for autonomous driving,” *Sensors*, vol. 20, no. 18, p. 5320, 2020.
- [7] R. Zhang, Z. Zou, S. Shen, and H. X. Liu, “Design, implementation, and evaluation of a roadside cooperative perception system,” *Transportation research record*, vol. 2676, no. 11, pp. 273–284, 2022.
- [8] Z. Gong, Z. Wang, B. Zhou, W. Liu, and P. Liu, “Pedestrian detection method based on roadside light detection and ranging,” *SAE International Journal of Connected and Automated Vehicles*, vol. 4, no. 12-04-0031, pp. 413–422, 2021.
- [9] R. Zhang, D. Meng, L. Bassett, S. Shen, Z. Zou, and H. X. Liu, “Robust roadside perception for autonomous driving: an annotation-free strategy with synthesized data,” *arXiv preprint arXiv:2306.17302*, 2020.
- [10] Z. Zou, R. Zhang, S. Shen, G. Pandey, P. Chakravarty, A. Parchami, and H. X. Liu, “Real-time full-stack traffic scene perception for autonomous driving with roadside cameras,” in *2022 International Conference on Robotics and Automation (ICRA)*, pp. 890–896, IEEE, 2022.
- [11] Y. Han, H. Zhang, H. Li, Y. Jin, C. Lang, and Y. Li, “Collaborative perception in autonomous driving: Methods, datasets and challenges,” *arXiv preprint arXiv:2301.06262*, 2023.
- [12] S. Su, Y. Li, S. He, S. Han, C. Feng, C. Ding, and F. Miao, “Uncertainty quantification of collaborative detection for self-driving,” in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5588–5594, IEEE, 2023.
- [13] Z. Lei, S. Ren, Y. Hu, W. Zhang, and S. Chen, “Latency-aware collaborative perception,” in *European Conference on Computer Vision*, pp. 316–332, Springer, 2022.
- [14] N. Vadivelu, M. Ren, J. Tu, J. Wang, and R. Urtasun, “Learning to communicate and correct pose errors,” in *Conference on Robot Learning*, pp. 1195–1210, PMLR, 2021.
- [15] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, “Are we ready for autonomous driving? the kitti vision benchmark suite,” in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.
- [16] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, and Y. Pan, “nuScenes: A multimodal dataset for autonomous driving,” *arXiv preprint arXiv:1903.11027*, 2019.
- [17] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine, et al., “Scalability in perception for autonomous driving: Waymo open dataset,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2446–2454, 2020.
- [18] A. Milan, L. Leal-Taixé, I. Reid, S. Roth, and K. Schindler, “Mot16: A benchmark for multi-object tracking,” *arXiv preprint arXiv:1603.00831*, 2016.
- [19] P. Dendorfer, H. Rezatofighi, A. Milan, J. Shi, D. Cremers, I. Reid, S. Roth, K. Schindler, and L. Leal-Taixé, “Mot20: A benchmark for multi object tracking in crowded scenes,” *arXiv preprint arXiv:2003.09003*, 2020.
- [20] A. Dave, T. Khurana, P. Tokmakov, C. Schmid, and D. Ramanan, “Tao: A large-scale benchmark for tracking any object,” in *Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*, pp. 436–454, Springer, 2020.
- [21] H. W. Kuhn, “The hungarian method for the assignment problem,” *Naval research logistics quarterly*, vol. 2, no. 1-2, pp. 83–97, 1955.
- [22] S. International, “On-board system requirements for v2v safety communications,” *SAE J2945/1*, 2016.
- [23] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi, “Performance measures and a data set for multi-target, multi-camera tracking,” in *European conference on computer vision*, pp. 17–35, Springer, 2016.
- [24] J. Luiten, A. Osep, P. Dendorfer, P. Torr, A. Geiger, L. Leal-Taixé, and B. Leibe, “Hota: A higher order metric for evaluating multi-object tracking,” *International journal of computer vision*, vol. 129, pp. 548–578, 2021.
- [25] M. T. Lab, “Roadside perception sysmtem evaluation.” https://github.com/michigan-traffic-lab/perception_evaluation, 2023.



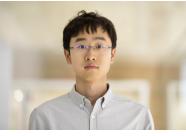
Rusheng Zhang received the B.E. degree in micro electrical mechanical system and second B.E. degree in Applied Mathematics from Tsinghua University, Beijing, in 2013. He received an M.S. and PhD degree in electrical and computer engineering from Carnegie Mellon University, in 2015, 2019 respectively. His research areas include artificial intelligence, cooperative driving, cloud computing and vehicular networks.



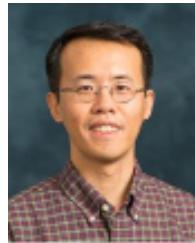
Depu Meng (Member, IEEE) is a Post Doctoral Research Fellow at the Department of Civil and Environmental Engineering, University of Michigan. He received his B. E. degree from the Department of Electrical Engineering and Information Science at the University of Science and Technology of China in 2018. He received his Ph. D. degree from the Department of Automation at the University of Science and Technology of China. His research interests include computer vision and autonomous driving systems.



Michael Maile received the Dipl. Phys. Degree from the University of Ulm in 1986. He is currently a Principal at Ivie Communications. His research areas include Sensor Fusion and Localization for Automated Vehicles and Infrastructure-Vehicle communication based safety applications.



Shengyin (Sean) Shen works as a Research Engineer in the Engineering Systems Group at the University of Michigan Transportation Research Institute (UMTRI). Sean holds an MS degree in Civil and Environmental Engineering from the University of Michigan, Ann Arbor, and an MS degree in Electrical Engineering from the University of Bristol, UK. He also earned a BS degree from Beijing University of Posts and Telecommunications, China. Sean's research interests are primarily focused on cooperative driving automation and related applications that use roadside perception, edge-cloud computing, and V2X communications to accelerate the deployment of automated vehicles. He has extensive experience in implementation of large-scale deployments, such as the Safety Pilot Model Deployment (SPMD), Ann Arbor Connected Vehicle Testing Environment (AACVTE), and Smart Intersection Project. Moreover, he has been involved in many research projects funded by public agencies such as USDOT, USDOE, and companies such as Crash Avoidance Metric Partnership (CAMP), Ford Motor Company, and GM Company, among others.



Tinghan Wang received the B.Tech. and Ph.D. degrees from Tsinghua University, Beijing, China, in 2016 and 2022. He is currently a research fellow with the Mechanical Engineering Department, the University of Michigan. His research interests include automated vehicle evaluation, cooperative driving, and end-to-end self-driving based on deep reinforcement learning.



Tai Karir is a senior at Huron High School. He is a member of the soccer team and also enjoys playing piano and reading. His research interests include software engineering, data analysis and machine learning.



Henry X. Liu (Member, IEEE) received the bachelor's degree in automotive engineering from Tsinghua University, China, in 1993, and the PhD. degree in civil and environment engineering from the University of Wisconsin-Madison in 2000. He is currently a professor in the Department of Civil and Environmental Engineering and the Director of Mcity at the University of Michigan, Ann Arbor. He is also a Research Professor at the University of Michigan Transportation Research Institute and the Director for the Center for Connected and Automated Transportation (USDOT Region 5 University Transportation Center). From August 2017 to August 2019, Prof. Liu served as DiDi Fellow and Chief Scientist on Smart Transportation for DiDi Global, Inc., one of the leading mobility service providers in the world. Prof. Liu conducts interdisciplinary research at the interface of transportation engineering, automotive engineering, and artificial intelligence. Specifically, his scholarly interests concern traffic flow monitoring, modeling, and control, as well as testing and evaluation of connected and automated vehicles. Prof. Liu is the managing editor of Journal of Intelligent Transportation Systems.

