A Low-Cost Single Board Computer-Based Automated System for Measuring Lateral Passing Distance of Vehicles Overtaking Cyclists

Wuihee Yap[a,*] , Milan Paudel[b] , Fook Fah Yap[c], Nader Vahdati[d], Oleg Shiryayev[e]

[a] *College of Arts & Sciences, University of Washington, Seattle, WA 98195, USA*

[b] *Transport Research Center @ NTU, Singapore*

[c] *School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore*

[d] *Department of Mechanical Engineering, Khalifa University of Science and Technology, Healthcare Engineering Innovation Center, SAN Campus, Abu Dhabi P.O. Box 127788, United Arab Emirates*

[e] *Department of Mechanical Engineering, University of Alaska Anchorage, 3211 Providence Dr., ECB 301, Anchorage, AK 99508, USA*

Abstract

The safety of cyclists in urban environments remains a critical concern in sustainable urban planning. A primary factor affecting this safety is the lateral passing distance of motor vehicles overtaking cyclists. This study aims to use current technology to provide an automated and robust framework, integrating a low cost, small single-board computer with a distance measuring sensor and a camera, to collect and analyze vehicle-cyclist passing distance[1] and speed. Preliminary deployments in Singapore have demonstrated the system's efficacy in capturing high-resolution data under varied traffic conditions[2]. Our setup, using a Raspberry Pi 4, LIDAR distance sensor, OAK-1 AI camera, and the DBSCAN clustering technique, had a 100% success rate for correctly identifying the number of close vehicle passes for distances between 1 to 1.5m. The insights garnered from this integrated setup promise not only a deeper understanding of vehicle-cyclist interactions but also a roadmap for data-driven urban safety interventions.

1. Introduction

The shift towards cycling is a global phenomenon, not just restricted to Singapore. Driven by environmental concerns, health advantages, and economic factors, urban centers worldwide have seen an uptick in the use of bicycles. The advantages of cycling have been widely recognized and documented, encompassing not only health benefits which lead to a longer life span, but also environmental benefits that reduce pollution and traffic congestion (de Hartog et al., 2010). Bicycles, being zero-emission vehicles, offer a tangible solution to mitigate the detrimental environmental impacts traditionally associated with vehicular transportation (Lindsay et al., 2011). Furthermore, in urban hubs like Singapore, where car

---

[1] The code used for the sensors and Raspberry Pi can be found at: https://github.com/wuihee/cycling-safety-code.

[2] The data collected and code used for data visualization and analysis can be found at: https://github.com/wuihee/cycling-safety-analysis

ownership costs are soaring and Certificate of Entitlement (COE) prices exceed SGD$100,000 (Land Transport Authority, 2023), cycling offers both an eco-friendly and financially sound alternative.

Modern urban planning champions sustainability, and a significant part of this entails prioritizing cycling. Singapore's Land Transport Authority acknowledges the increasing prevalence of cycling and the need to encourage sustainable transport. As a result, the Land Transport Master Plan 2040 features new cycling infrastructure that will lead to a "45-Minute City with 20-Minute Towns" (Land Transport Authority, 2019), all of which would result in an increased number of cyclists on the road. Other cities are also developing better infrastructure to support cycling. Paris plans to be completely cyclable by 2026 under Plan Velo, which plans to add 130 kilometers of bike paths, as well as more bike parking and booth maintenance (La Ville de Paris, 2021).

With the rise in cycling, there has been an unfortunate increase in cycling-related accidents. The World Health Organization reports that 41,000 cyclists face fatal accidents annually during commutes (World Health Organization, 2020). In Singapore, the growth of food delivery services necessitates more cyclists on the road, further escalating the risk. Data reflects an 11% increase in food consumed/ordered via food delivery services during the COVID-19 pandemic (Grab & NielsenIQ, 2021) which corresponds to a 25% rise in cycling-related accidents during the same period (Singapore Police Force, 2020). Nonetheless, many European countries and the USA, Canada, and Australia, where cycling is prevalent, have resisted effective techniques like traffic calming to improve the safety of cyclists and road safety in general. (Pucher & Buehler, 2017). In fact, there seems to be very little cyclists can do to prevent vehicles from overtaking them dangerously (Walker et al., 2014), which suggests that structural or legal measures must be taken to reduce cycling-related accidents.

To address the safety concerns, some cities, including Singapore, have set a 1.5m passing distance guideline (Ministry of Transport, 2019). However, the effectiveness and enforcement of these guidelines remain contentious. Moreover, a study from Germany showed variable adherence levels, where cars meet the legal passing distance of 1.5m only 30% of the time (Stülpnagel et al., 2022). Table 1 shows different countries and their different Mandatory Passing Distance (MPD) rules.

Table 1
Countries With a Minimum Passing Distance Guideline or Law.

Country	MPD Advised	MPD Mandated
Austria	Yes - 1.5m	No
Belgium	No	Yes - 1m
Chile	Yes - 1.5m	No
France	No	Yes – 1m on roads with

		≤50km/h speed limit, and 1.5m on roads with >50km/h speed limit.
Germany	No	Yes - 1.5m in urban areas, and 2m out of town.
New Zealand	Yes - 1.5m	No
Singapore	Yes - 1.5m	No
United States	Yes -Varies by state	Yes - Varies by state

Note. Table containing the mandatory passing distance rules for various countries (Road Safety Authority, 2019).

Despite guidelines, adherence remains an issue. Uncertainty about distance measurements from motorists and a lack of quantifiable evidence from cyclists create complications. Current methods to measure passing distances are labor-intensive, error-prone, and often costly. A study from Australia which concluded that every 1-in-17 event was a close pass of less than 100cm, relied on GoPro devices paired with ultrasonic sensors, necessitating manual data extraction post-collection (Beck et al., 2019). Another study from Spain which concluded that a 1.5m MPD does not guarantee safety, featured an exhaustive setup including three video cameras, 2 speed sensors, 2 distance sensors, a VBOX data logger, PC and batteries (Llorca et al., 2017). While the results of such studies are comprehensive, their results may be difficult to replicate due to the manual labor required to collect the data or the complexity of their setups.

This study aims to investigate the potential of an automated data collection method that reduces manual effort, offering a solution for policy-makers, researchers, and consumers. We leverage modern technology for cost-effectiveness. Our system is mounted on a bicycle and identifies vehicles passing within 1.5m or closer without requiring manual logging or video cross-referencing. The total cost of our setup is around USD$200. Through the integration of a distance sensor, microcontroller, and an AI camera with computer vision capabilities, we ensure automatic and precise data recording. Post-collection, clustering analysis identifies vehicle passes, providing a clear count of close encounters.

2. Methodology
We start by reviewing various distance measuring sensors: laser, LIDAR, and ultrasound. These were assessed for accuracy, range, and reliability through tests against a stationary wall at varying distances. From this evaluation, we selected the most suitable candidates for outdoor experiments. In our outdoor test, the sensors were attached to a stationary bike on the roadside to gauge their proficiency at detecting passing vehicles. Our preferred choice was then integrated with the OAK-1 AI camera and mounted on a bicycle. Real-world data collection was simulated as the bike navigated actual roads with traffic. The system was programmed with Python using the Raspberry Pi, employing modules like Pyserial and Smbus to facilitate communication between the sensors and the Pi. The AI camera was

programmed through the DepthAI API by Luxonis. Post-data collection, noise was filtered out to emphasize clusters indicative of vehicle passes. Finally, the machine learning clustering method DBSCAN was employed to automatically identify vehicle passes in the data collected, the accuracy of which was validated using GoPro footage.

3. Sensor Selection

To ensure a robust and accurate data collection system, our study employed an iterative process of selection, testing, and deployment of three types of measuring sensors. The methodology consisted of preliminary indoor and outdoor tests, followed by on-road stationary and mobile tests.

Table 2 shows a general comparison between the different types of sensors available. From this comparison alone, it is difficult to determine which type or model of sensor is the most appropriate. Hence, the tests aim to guide our selection.

Table 2

General Comparison Between Different Types of Sensors

	LIDAR	Laser	Ultrasonic
Working Principle	Uses laser beams to measure distances	Uses a laser beam to measure distance based on reflection	Uses sound waves to measure distances based on echo timing
Range	Typically 100m - 300m, but can go up to 1km for some models	Shorter, often less than 100m	Typically 2 cm - 5m, though specialized models can go further
Accuracy	±2cm to ±10cm depending on model and conditions	±1mm to ±5mm	±1cm for close range but can vary based on conditions
Resolution	Fine; can be sub-cm in some models	Fine, often sub-mm	Coarser, often in the cm range
Pros	High resolution, long-range, works in various lighting conditions	High precision, works in many lighting conditions	Simple, cheap, works in the dark and through many materials
Cons	Expensive, can be affected by atmospheric conditions and reflective surfaces	Range limited, can be affected by reflective surfaces or ambient light	Affected by sound-absorbent materials, less accurate at longer distances

3.1 Indoor and Outdoor Fixed Distance Tests

To determine the compatibility of the sensors with our project's specific needs, we executed a comprehensive set of tests. Our benchmarks necessitated the precise measurement of distances up to 3 meters in outdoor daylight and the capability to detect high speed vehicular movements.

Table 3 delineates a comparative analysis of various sensors tested, all of which exhibit notable accuracy in detecting vehicles within a maximum distance of 3m. Despite their evident precision and sufficiently high frequency to execute the task, it's crucial to note that the performance of laser and LIDAR sensors may be compromised under daylight conditions.

Table 3
Comparative Specifications of Select Distance Sensors Tested.

	WaveShare TOF	JRT BB2X Laser	LIDAR Lite V4	A02YYUW
Sensor Type	Laser	Laser	LIDAR	Ultrasonic
Max. Range	5m	100m	10m	4.5m
Frequency	10Hz	10-20Hz	200Hz	10Hz
Accuracy	±1 cm short/medium distance ± 1.5 cm long distance	±3mm	±1 cm to 2m ± 2 cm to 4m ± 5 cm to 10m	±1cm
Affected by Light	Yes	Yes	Yes	No

We tested each sensor at fixed intervals of 0.5 meters, spanning a range from 0 to 5 meters. These tests took place in both internal and external environments to assess the sensor performance under varying conditions. At every interval, around 500 distance data points were gathered. The sensors were either operated using their default software or integrated with a Raspberry Pi for control.
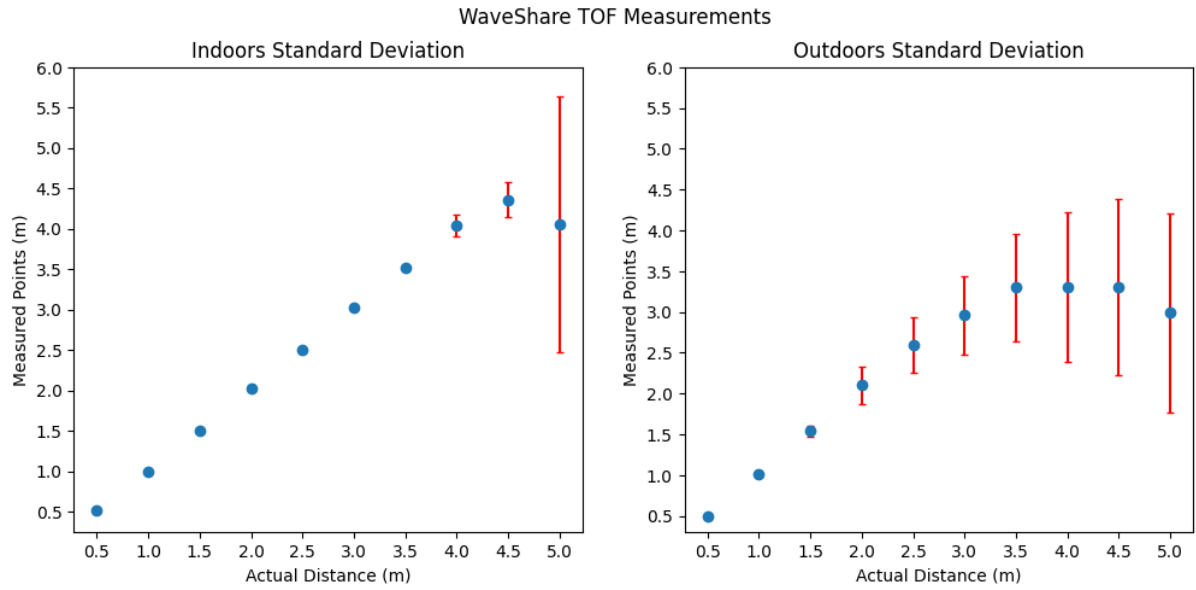
Fig. 1. Depicts the mean distance measured by the WaveShare TOF sensor across distances of 0.5m to 5.0m. The red bars represent the standard deviation of the measurements at each interval (Yap).

Table 4

Standard Deviation of Measurements at Different Intervals for TOF Laser Range Sensor

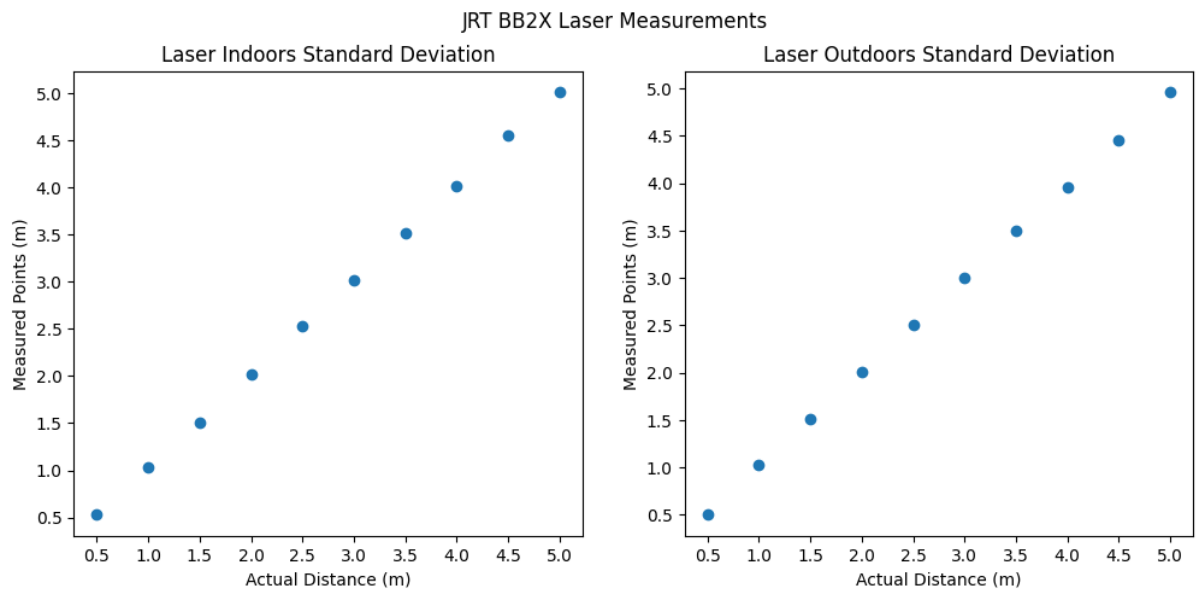|	0.5m	1.0m	1.5m	2.0m	2.5m	3.0m	3.5m	4.0m	4.5m	5.0m
Indoors	0.01	0.01	0.01	0.01	0.01	0.02	0.03	0.14	0.22	1.58
Outdoors	0.00	0.01	0.07	0.23	0.34	0.48	0.66	0.92	1.08	1.22

Fig. 2. Illustrates the mean distance measured by the JRT BB2X Laser across distances of 0.5m to 5.0m. The red bars represent the standard deviation of the measurements at each interval.

Table 5

Standard Deviation of Measurements at Different Intervals for BB2X JRT Laser

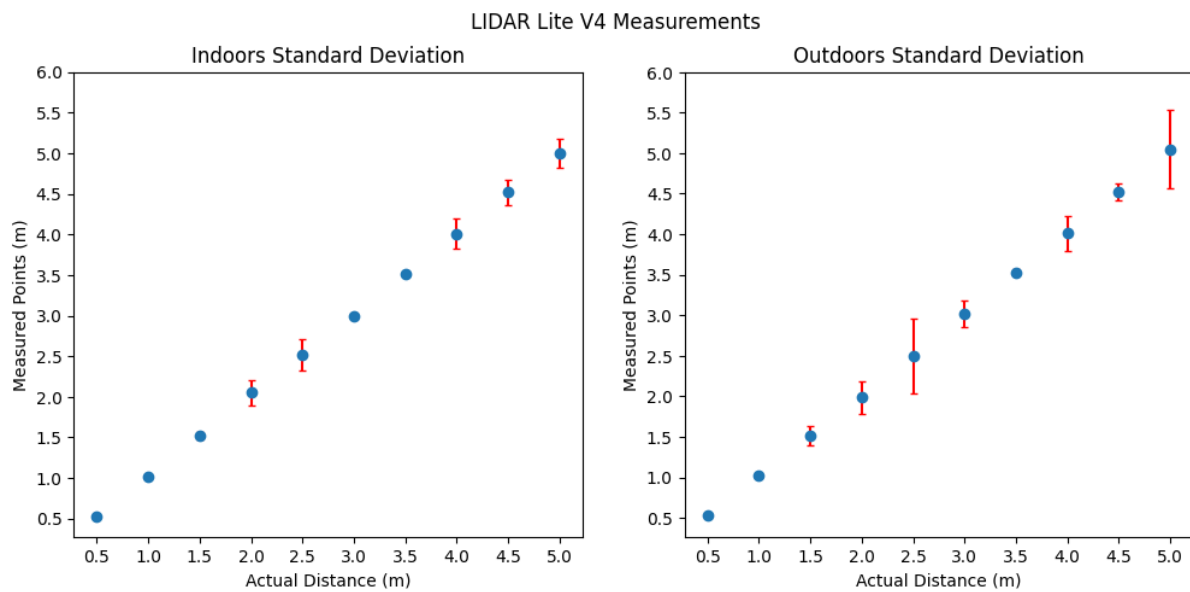|	0.5m	1.0m	1.5m	2.0m	2.5m	3.0m	3.5m	4.0m	4.5m	5.0m
Indoors	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Outdoors	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Fig. 3. Depicts the mean distance measured by the LIDAR Lite V4 from the 0.5m to 5.0m range. The red bars represent the standard deviation of the measurements at each interval.

Table 6

Standard Deviation of Measurements at Different Intervals for LIDAR Lite V4

	0.5m	1.0m	1.5m	2.0m	2.5m	3.0m	3.5m	4.0m	4.5m	5.0m
Indoors	0.01	0.02	0.05	0.15	0.19	0.03	0.03	0.18	0.16	0.18
Outdoors	0.02	0.02	0.12	0.20	0.46	0.16	0.04	0.21	0.10	0.49

Note. We noticed an anomaly where the standard deviation for the LIDAR Lite V4 sensor around 2.0m to 2.5m spikes. This could be due to an internal algorithm in the sensor.
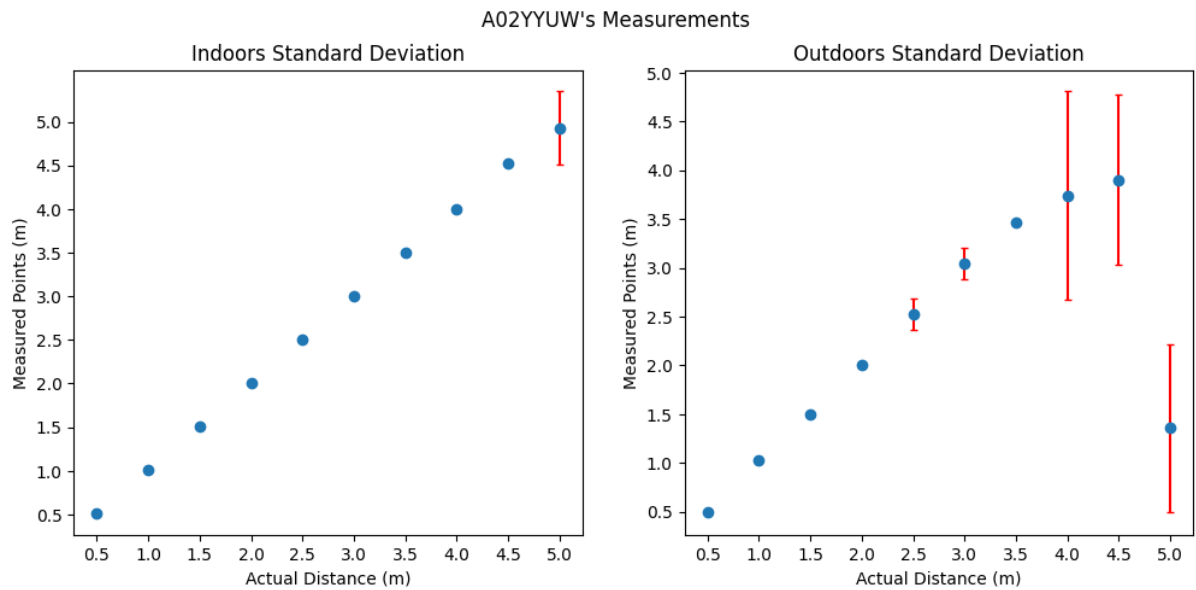
Fig. 4. Depicts the mean distance measured by the A02YYUW from the 0.5m to 5.0m range. The red bars represent the standard deviation of the measurements at each interval.

Table 7

Standard Deviation of Measurements at Different Intervals for A02YYUW

	0.5m	1.0m	1.5m	2.0m	2.5m	3.0m	3.5m	4.0m	4.5m	5.0m
Indoors	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.42
Outdoors	0.02	0.02	0.00	0.02	0.16	0.16	0.00	1.07	0.87	0.86

The graphs and tables presented in the findings utilize various visual elements to depict sensor performance. Solid dots on each graph represent the average distance recorded by the sensor at specified intervals, while red bars indicate standard deviations.

Focusing initially on the WaveShare TOF sensor (Figure 1 and Table 4), it demonstrated reliable performance up to 4.5m in indoor settings. However, a noticeable decline in accuracy and consistency was observed outdoors, where the standard deviation surged from 0.34m at a distance of 2.5m to 1.22m at 5.0m.

In contrast, both the JRT BB2X Laser and LIDAR Lite V4 sensors maintained accuracy across all tested environments. Notably, the JRT BB2X Laser showcased high precision, with an almost negligible standard deviation, as depicted in Figure 2 and Table 5. Although the LIDAR Lite V4 sensor (Figure 4 and Table 5) manifested accurate readings, it did not match the laser sensor's precision, with its highest standard deviation being under 0.5m—still notably superior to the WaveShare TOF sensor.

Figure 4 and Table 7 shows that the A02YYUW ultrasonic sensor's reliability is effective 4.5, and 3.5m for indoors and outdoors respectively. Unlike other sensors, its measurement inconsistencies are not due to environmental sunlight conditions but rather its high sensitivity. While dependable at short ranges with a 10 Hz frequency, the sensor's effectiveness diminishes beyond 2.5m, demanding precise positioning to accurately capture data. Even slight deviations can lead to entirely incorrect readings. Given these limitations, the A02YYUW's sensitivity and requirement for exact alignment make it an impractical choice for experiments necessitating robust performance, especially in dynamic settings like motion-based cycling.

The findings from our tests indicate that the manufacturers' specifications for the sensors hold true only under optimal conditions. In real-world scenarios, the sensors exhibited increased unreliability, underscoring the necessity of these tests to ascertain whether the sensors meet our criteria.

3.2 Stationary Bike Test

We mounted the sensors onto a stationary bicycle positioned beside a road with moderate traffic. Using a GoPro camera in tandem with each sensor, we assessed their capability to identify passing vehicles. This setup provided insights into the sensors' performance under controlled yet realistic conditions.

3.2.1 Laser Sensor

Figure 5 presents the results from the stationary bike test using the JRT BB2X Laser sensor. Despite its high reliability in distance measurement, the sensor's frequency proved inadequate for accurately recording the distances of passing vehicles, even in moderate traffic. Although the product documentation advertised a frequency of 10-20Hz, real-world tests showed it to be inconsistent and often much lower.
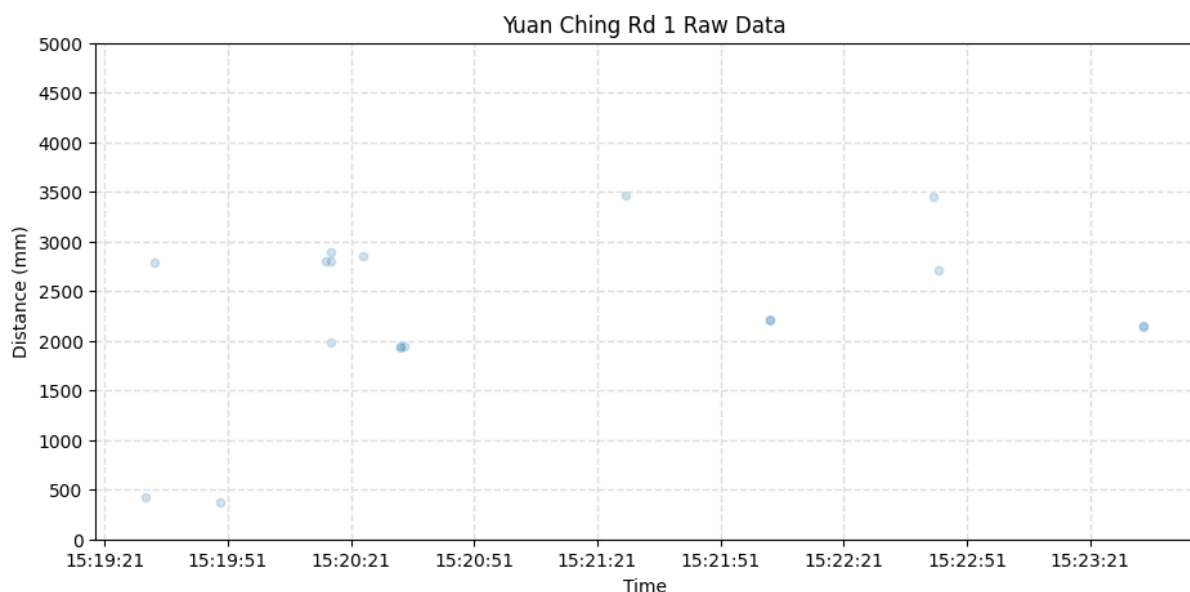

Fig. 5. Results from the JRT BB2X Laser sensor's stationary test.

3.2.2 LIDAR Sensor

The LIDAR sensor, with its impressive frequency reaching up to 200 Hz, was adept at detecting passing vehicles, as evidenced by the clusters of dark points in Figure 6. Having validated the LIDAR sensor's proficiency, we progressed to the mobile bike assessment.
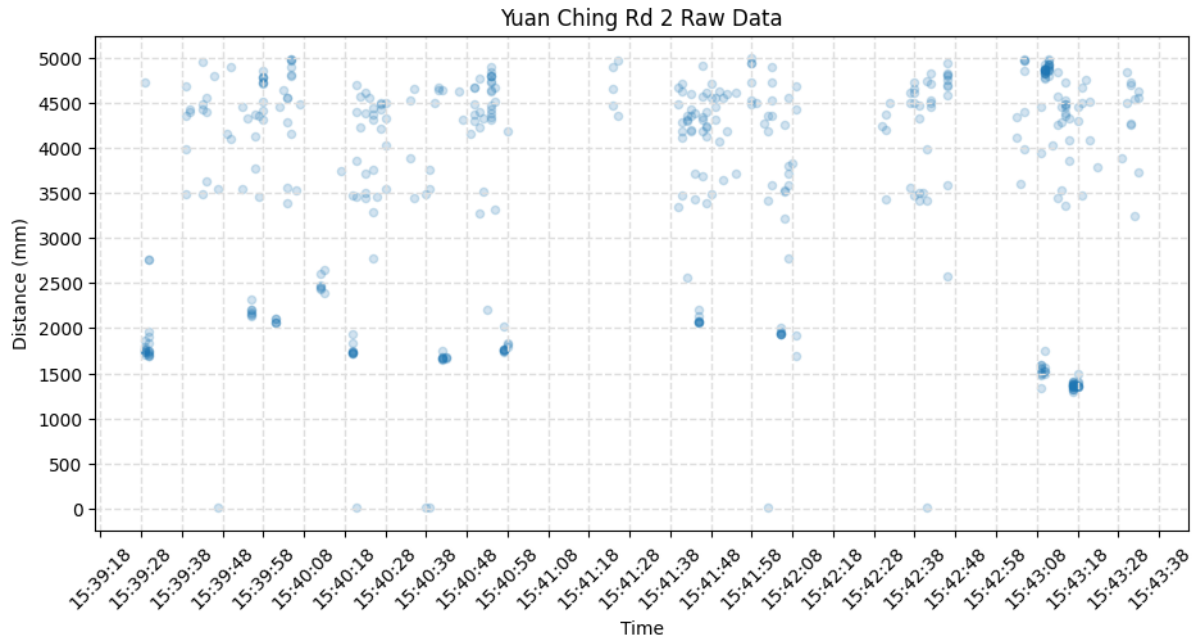

Fig. 6. Results from the LIDAR Lite V4' s stationary test.

Next, we cross-referenced the video footage from the GoPro with the data collected. Each red circle in Figure 7 highlights the points representing a vehicle pass. Out of 12 vehicles, the sensor only failed to detect a single car at 15:42:08, resulting in an approximate success rate of 92% based on this limited dataset. This confirms that the dark point clusters generated by the sensor indeed represent the passing of vehicles. With this verification, we advance to the final evaluation of the complete setup.
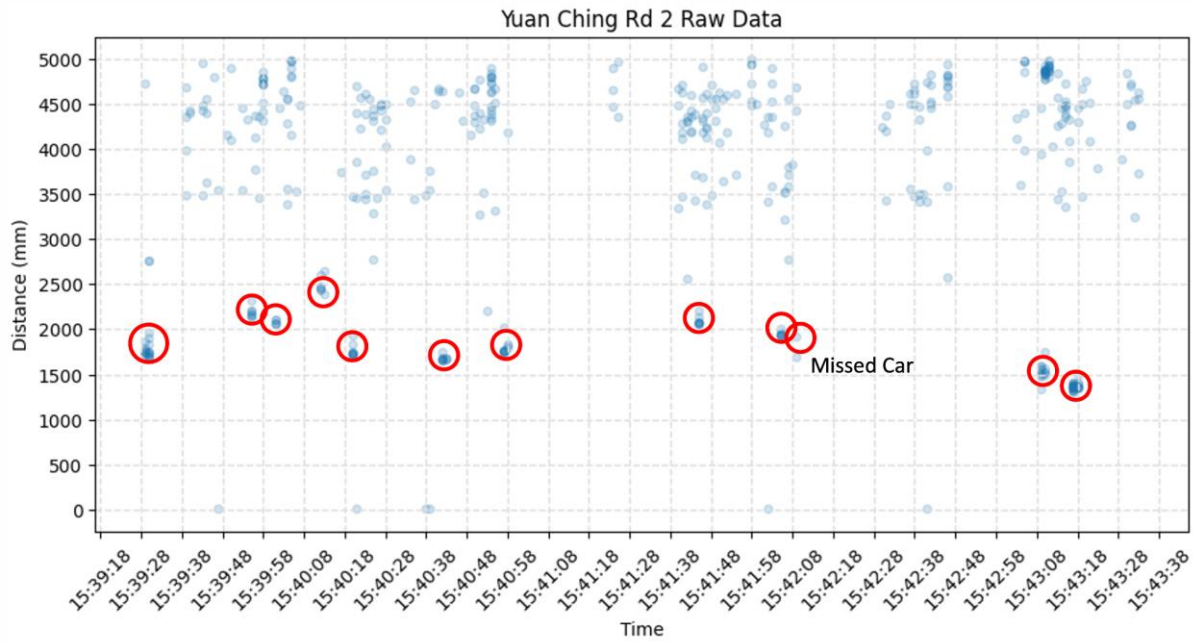
Fig. 7. Results from the LIDAR Lite V4's stationary test annotated.

4. Mobile Bike Test

To mimic the real-world scenario of a cyclist on the move, our sensor was mounted on a bicycle and cycled along a 5km route alongside moderate traffic in Singapore's Jurong West area as shown in Figure 8. This phase aimed to assess the challenges and practicality of each sensor during typical cycling activities.
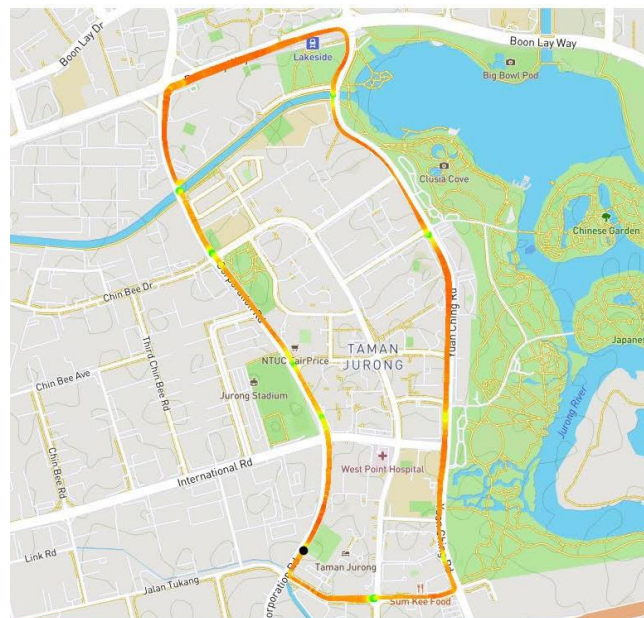

Fig. 8. The route cycled for the data collection.

4.1 Setup

Our configuration integrated three primary components: the LIDAR sensor, a Raspberry Pi with its power supply, and the OAK-1 Lite camera by Luxonis. The schematic in Figure 9 breaks down the components of the setup, which together work to detect passing vehicles,

and Figure 9 depicts the final setup mounted on a bike. The principle behind this setup was to leverage the object detection capability of the OAK camera to identify an approaching car in proximity to the cyclist. Upon detection, the LIDAR sensor would activate to record distances. This approach not only streamlined the data collection process but also significantly reduced the occurrence of false positives and irrelevant data. By focusing measurements exclusively during confirmed vehicle presence, we can enhance data accuracy. Unlike other systems, our refined data processing eliminates the need for tedious manual cross-referencing with video footage or reliance on manual interventions like button-presses to ascertain data quality. In our initial tests, we also attached a GoPro to the bike to manually cross-check and validate the collected data as shown in Figure 10. However, the GoPro was not part of the final configuration.
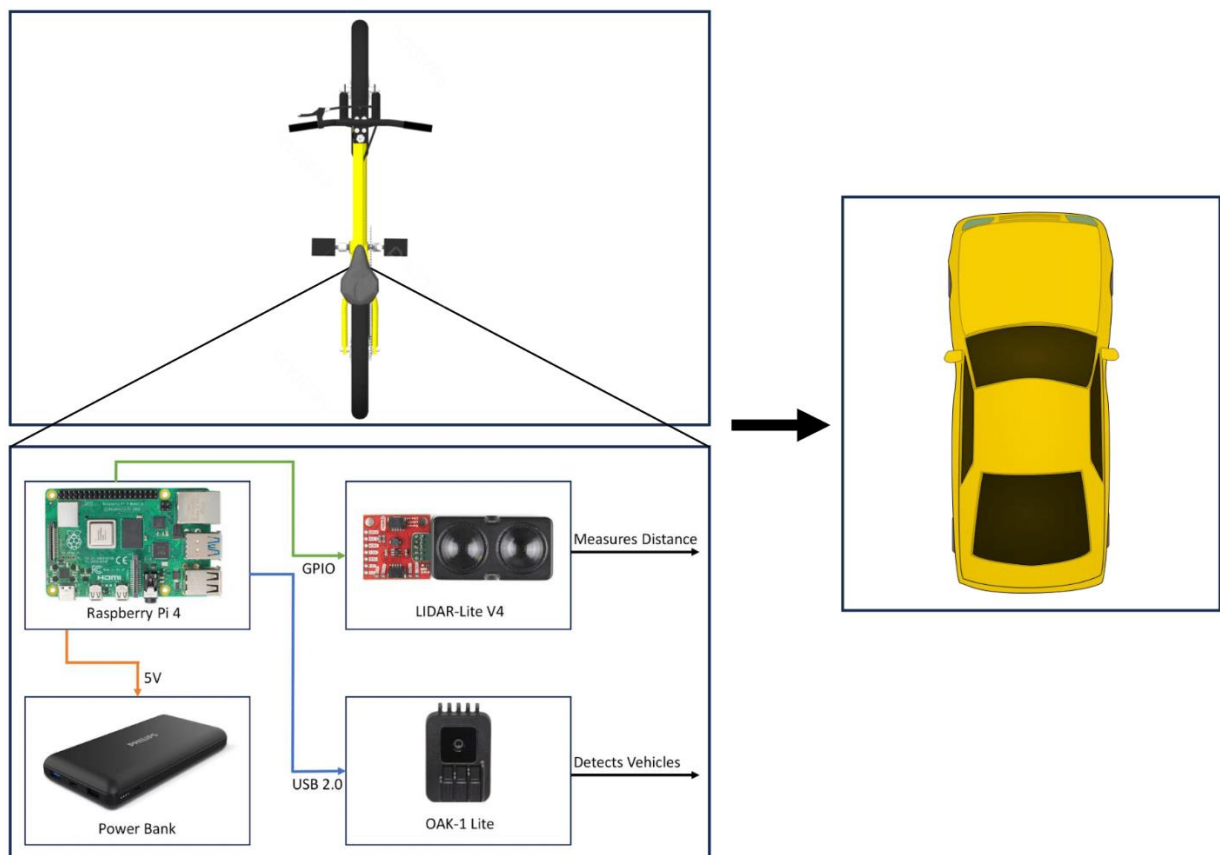
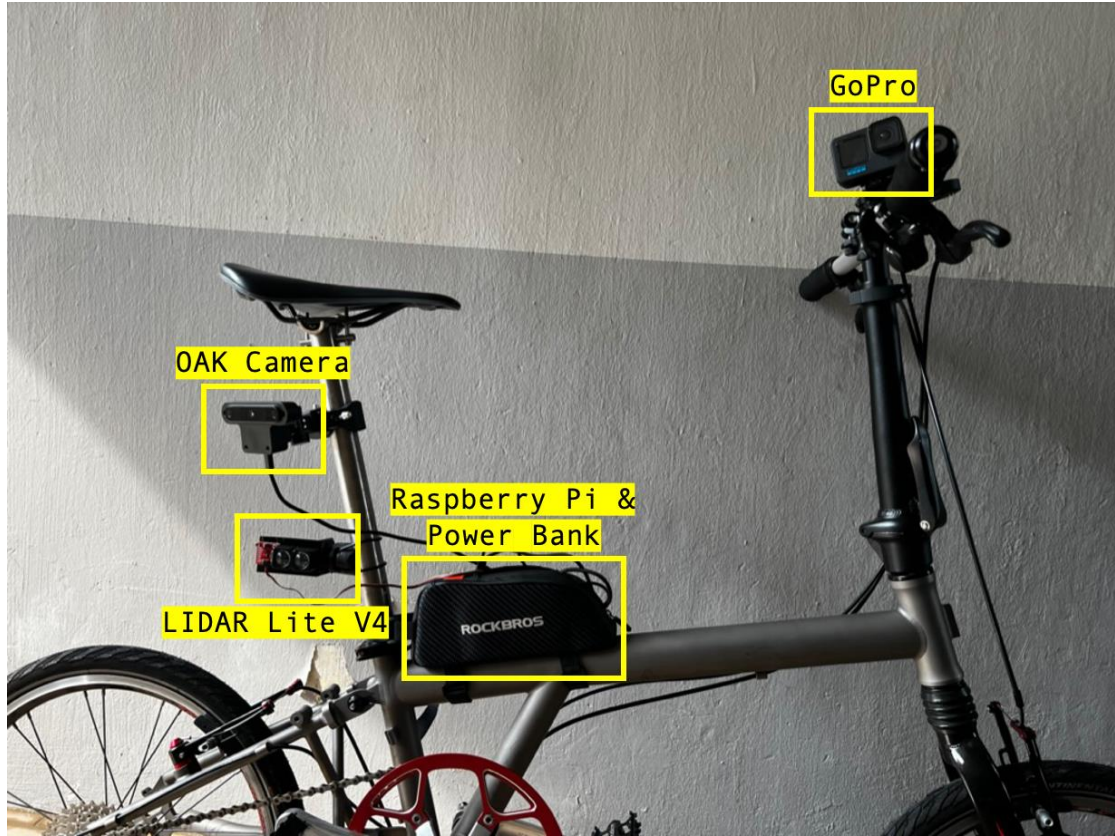

Fig. 9. Schematic of the setup.

Fig. 10. Raspberry Pi, LIDAR Lite V4, and OAK camera mounted on the bike with a GoPro.

4.2. Preliminary Results

At first glance at Figure 11, the data appeared extremely noisy. However, a detailed look reveals that most of the noise originates from readings above 3 meters, which are not important to the experiment as we are mainly concerned with close vehicle passes. Below this threshold, the data is notably cleaner, with pronounced dark clusters signaling vehicle passes. Interestingly, the broad gaps seen around 16:07:41 signify periods when the AI camera detected no vehicles, which contributed to reducing the overall noise in the dataset.
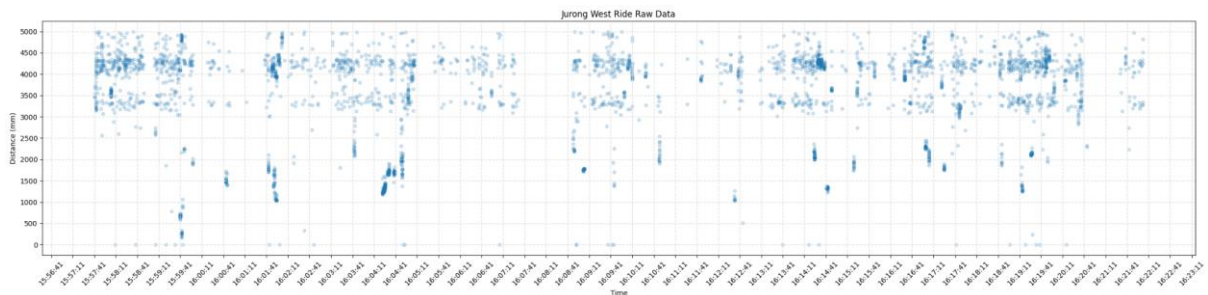

Fig. 11. Raw data collected by the LIDAR Lite V4 in conjunction with the OAK Camera.

To further clarify our findings, we cleaned our data in Figure 12 by first removing all data points above the 3m mark. We then accentuate clusters, which are simply a continuous set of non-null data points. We calculated the average of these points and replaced the entire cluster with this average.

Fig. 12. Cleaned data from the LIDAR Lite V4 and OAK Camera.

To verify that these dark clusters truly represented vehicles, we cross-referenced our findings with GoPro footage, circling all vehicle passings in red in Figure 12. Overall, the sensor adeptly identified vehicles passing the cyclist. Of the 21 vehicles that overtook the cyclist in the same lane, our setup correctly identified 7 passed at a proximity of 1.5m or closer. Points not highlighted in red either indicate false positives or noise. The subsequent clustering algorithm will identify false positives, while isolated faint points represent noise.

Fig. 13. Data collected with LIDAR and OAK Camera annotated using GoPro footage.

4.3. Automated Data Analysis

To enhance efficiency in the data collection process, we replace the time-consuming manual identification of clusters through cross-referencing with the GoPro with an algorithmic approach that automatically recognizes clusters representing vehicle passes.

4.3.1 Clustering Algorithm Selection

The sensor emits measurement signals at a frequency of approximately 20 Hz. Instead of producing one or two random data points, it should record a cluster of points that represent the distance of the overtaking vehicle during its pass. Hence, we require a cluster analysis to automatically identify vehicle passes.

The exploration of clustering algorithms reveals distinct methodologies and applications across various domains, each with unique working principles, cluster determination methods, and optimal use cases (Jain et al., 1999). For instance, K-Means, renowned for its application to large datasets with approximately spherical cluster shapes, necessitates a specified number of clusters and exhibits high sensitivity to outliers (MacQueen, 1967). Contrastingly, DBSCAN, a density-based algorithm, autonomously determines the number of clusters, showcasing adeptness in identifying clusters of varied shapes and sizes while proficiently managing noise and outliers (Ester et al., 1996). Hierarchical clustering, particularly

Agglomerative clustering, and Gaussian Mixture Model (GMM) require a specified number of clusters and demonstrate moderate sensitivity to outliers, with the former being suitable for small datasets requiring a tree-like structure and the latter being apt for scenarios demanding probabilistic cluster assignments (Johnson, 1967), (Reynolds, 2009). Mean Shift, a mode-seeking algorithm, determines clusters automatically and is notably utilized in computer vision and image segmentation (Cheng, 1995). Table 8 provides a brief comparison between the different clustering algorithms available.

Table 8

General Comparison Between Different Types of Clustering Algorithms

	Working Principle	Number of Clusters	Sensitivity to Outliers	Use Cases
K-Means	Partitioning	Needs to be specified	High	Large datasets; when the shape of clusters is approximately spherical.
DBSCAN	Density-Based	Determined automatically	Low	When cluster shape is irregular and noise/outliers are present.
Hierarchical	Agglomerative or divisive	Visualized using dendrogram, cut at desired level	Moderate	Small datasets; when tree-like structure or hierarchy is required.
Agglomerative	Groups data into objects into a tree of clusters.	Needs to be specified	Moderate	When a hierarchical approach is desired.
Gaussian Mixture Model (GMM)	Uses probability distributions	Needs to be specified	Moderate	When clusters are elliptical or when probabilistic cluster assignments are desired.
Mean Shift	Mode seeking	Determined automatically	Moderate	Image segmentation, computer vision.

For the specific task of identifying close vehicle passes, DBSCAN emerges as a preferred choice among clustering algorithms. Unlike K-Means or Gaussian Mixture Models, DBSCAN does not require a pre-defined number of clusters, making it adept at discovering clusters of varied shapes and sizes. Moreover, DBSCAN's inherent ability to segregate noise or outliers ensures that sporadic, unrelated data points do not form erroneous clusters, a challenge that can affect some other algorithms. This noise-handling capability makes DBSCAN particularly suited for real-world data where outliers are common. For a dynamic setting like traffic with unpredictable passing patterns and potential anomalies, DBSCAN offers the robustness and adaptability necessary to discern meaningful clusters effectively.

4.3.2 DBSCAN

Given a set of points in some space, DBSCAN groups together points that are closely packed together (points with many nearby neighbors), marking as outliers points that lie alone in low-density regions.

Let D be the data set of all points, Q be the query point, and $N(Q)$ represent the neighborhood of Q, which includes all points within distance ε of Q. As illustrated in Figure 14 by the orange point, a point Q is a *core point* if at least $minPts$ points are within distance ε of it.

$$Q \text{ is a } \textit{core point} \text{ if } |N(Q)| \geq minPts \tag{1}$$

A point P is *directly reachable* from Q if point P is within distance ε from core point Q, shown by the blue points to the orange point.

$$P \text{ is } \textit{directly reachable} \text{ from } Q \text{ if } P \in N(Q) \text{ and } |N(Q)| \geq \tag{2}$$
$$minPts$$

As illustrated by the green points, point P is *reachable* from Q if there exists a sequence of points:

$$p_1, p_1, \ldots, p_n \text{ with } p_1 = Q \text{ and } p_n = P \text{ where } p_{i+1} \text{ is directly reachable from} \tag{3}$$
$$p_i$$

Formally, a cluster in DBSCAN is a non-empty subset $C \subseteq D$ satisfying:
1. For any two points P and Q in C, if Q is reachable from P, then P and Q are part of the same cluster.
2. For any point Q in C if there is a point P in C such that Q is reachable from P, and P is a core point.

All other not reachable points are considered *outliers* or noise as indicated by the gray points and can be represented as:

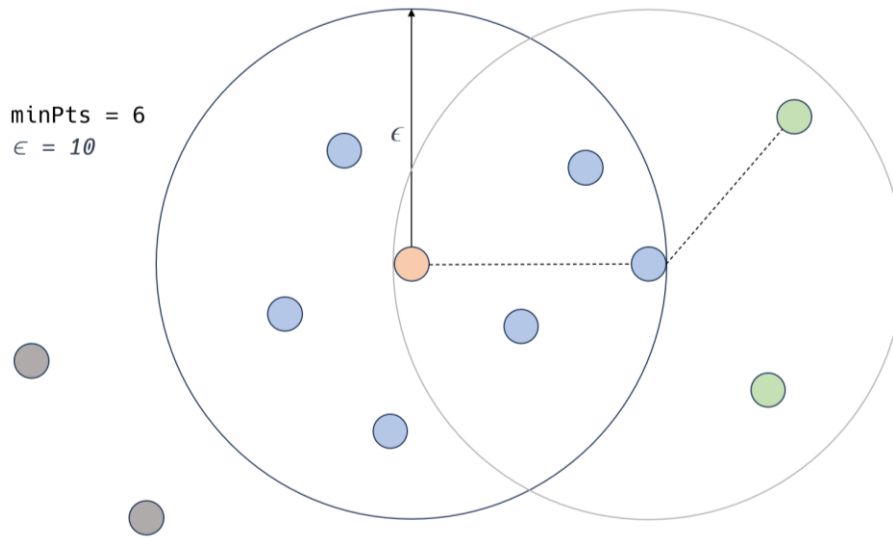$$D \setminus \bigcup_{all\ clusters\ C} C \tag{4}$$

Fig. 14. DBSCAN illustration.

Therefore, the DBSCAN algorithm can be understood as a region query, where $N_\epsilon(p)$ is the ε-neighborhood of the point Q in dataset D:

$$N_\epsilon(Q) = \{P \in D \mid dist(Q,P) \leq \epsilon\} \tag{5}$$

In our experiment, we calibrated ε to 0.02 and $minPts$ to 6. The results, as shown in Figure 15, depict the clusters identified by DBSCAN circled in blue which represent vehicle passes. There were four false positives detected by the algorithm, as marked with green squares, and the remaining points being noise. However, when distinguishing close vehicle passes ranging from 1m to 1.5m, DBSCAN demonstrated 100% accuracy. Clusters indicating passes under 1m are assumed to be anomalies, given the low likelihood of such close overtakes. Similarly, passes over 2m likely belong to vehicles in different lanes, and these can also be filtered. Therefore, with our preliminary results, we show that it is possible to create an automated framework that easily collects and analyses close vehicle passes with no reliance on manual techniques.
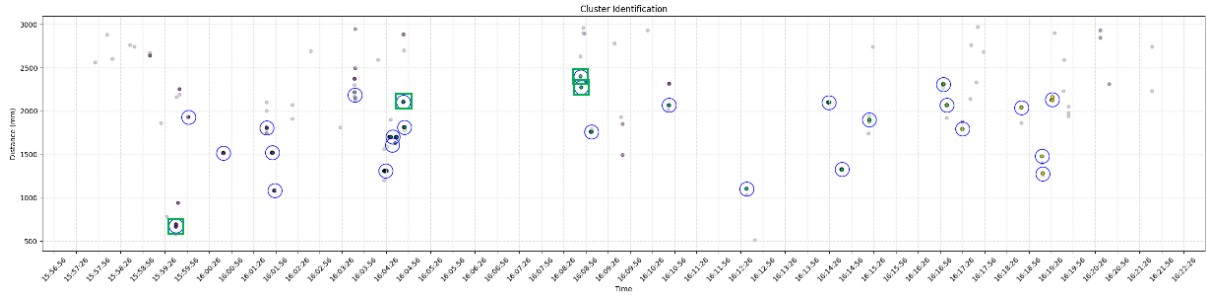
Fig. 15. Clusters identified using DBSCAN.

Using DBSCAN, we isolated points between 1m to 2m and averaged each cluster to represent a vehicle pass with a single point on a scatter plot. We then extracted relevant screenshots from the GoPro video and integrated them into an interactive plot. Users can click on a point to view the corresponding overtaking vehicle screenshot and its passing distance. In Figure 16, the clicked point is highlighted with a black circle, showcasing an overtaking car at that specific distance.
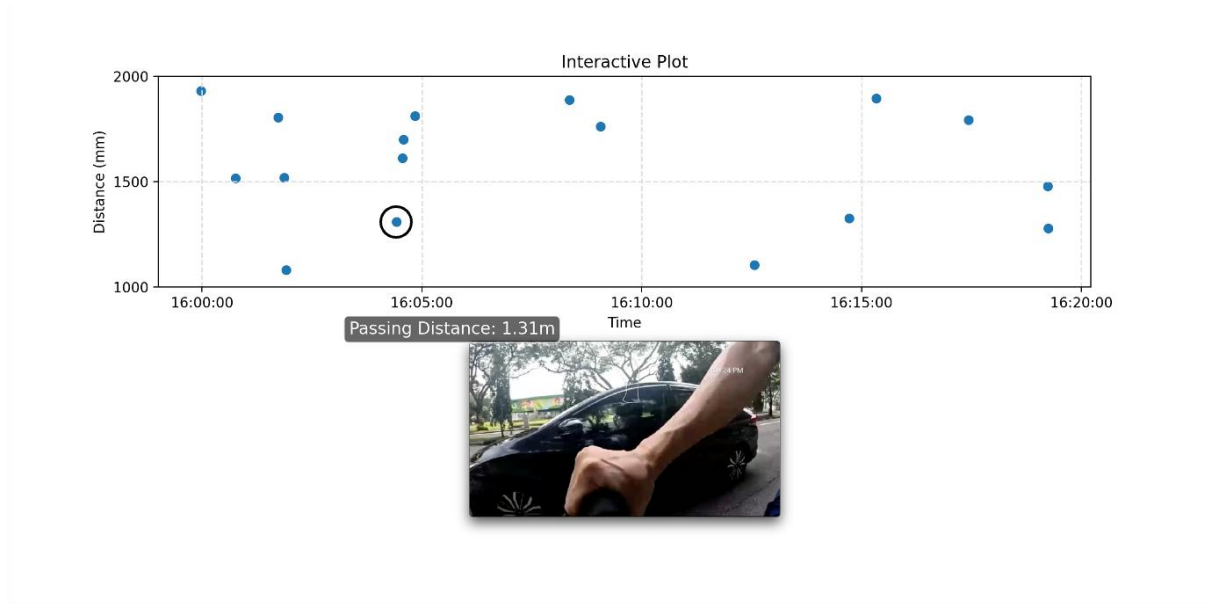
Fig. 16. Interactive plot depicting a vehicle overtaking at 1.31m.

5. Speed Considerations

Our current system lacks the capability to monitor the speed of passing vehicles. Considering that some countries adjust passing distance rules based on vehicle speed, this feature is significant and would offer deeper insights into cyclist and vehicle safety interactions.

To estimate the speed of a passing car from video footage, we first need to determine the pixel-to-real-world scale for the different passing distances between the bike (on which the camera is mounted) and the car. This can be done by measuring a known object in the image and then dividing the pixel length of the object by its real-world length. The pixel-to-real-world scale is inversely proportional to the distance of the camera from the object. From an experiment using an OAK 1 camera and equally spaced checkerboard images on a vertical wall as shown in Figure 17, we show that the number of pixels on a camera image that corresponds to a horizontal line of length 1 meter is almost inversely proportional to the distance of the camera from the wall, d in Figure 18.
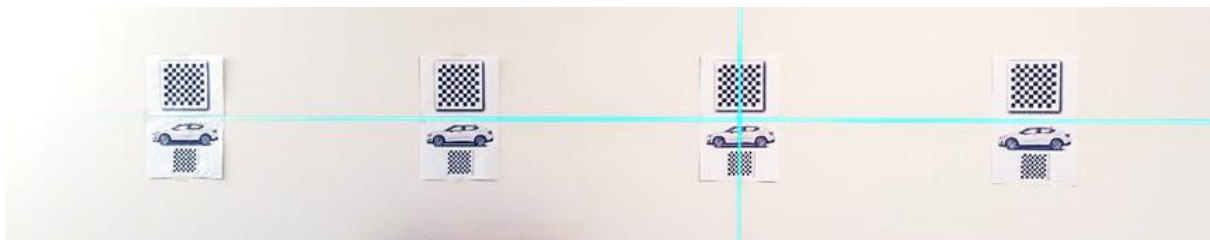

Fig. 17. Checkerboard images at 1m horizontal spacing on a vertical wall. The green laser lines are used to guide the placement of images.
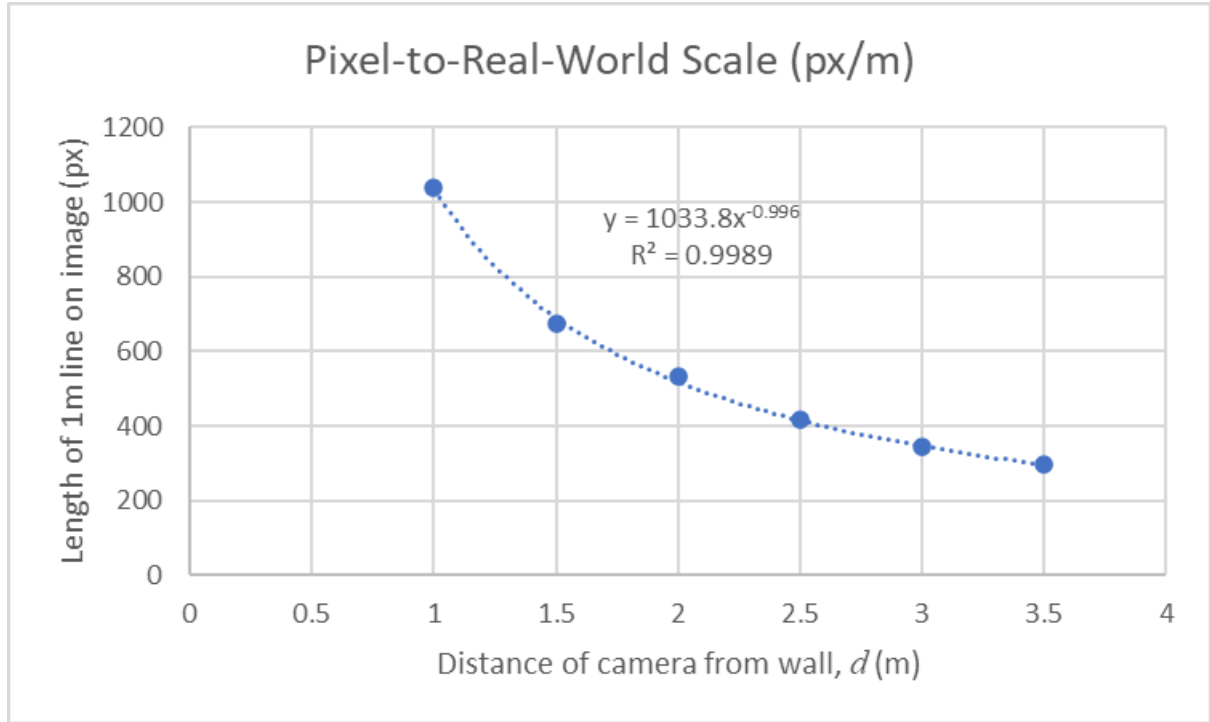
Fig. 18. A trend line through the data points shows a near-perfect inverse relationship between pixel-to-real-world scale and working distance.

Thus, for any horizontal displacement of length Δp pixels on an Oak-1 camera image, the actual displacement, s, as illustrated in Figure 19, is given by:

$$s = \Delta p\left(\frac{1}{f}d\right) = \Delta p\left(\frac{1}{1033.8}d\right) \tag{6}$$

where d is the distance between the camera and the car, and f is the focal length of the camera which can be determined from a regression analysis of the experimental data.
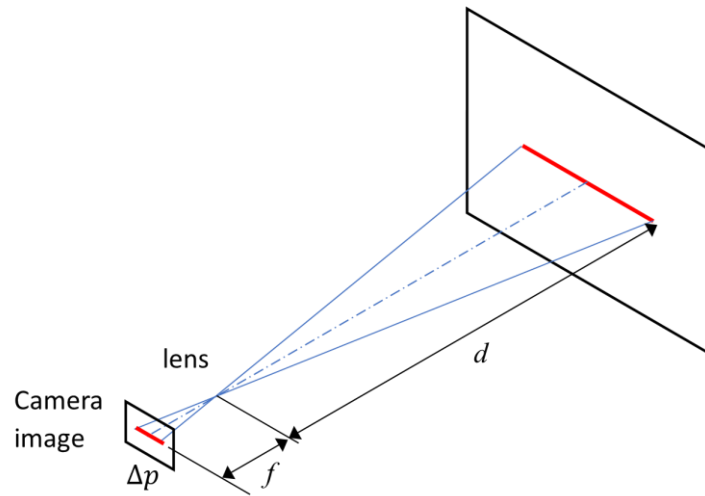

Fig. 19. Camera projection model.

From a video footage of a passing car, we can extract the image frames and the time of each frame. We can then calculate the displacement of the car, s, between any two frames using equation 6. The speed of the car relative to the bike, v, can be estimated by using $v = s\frac{s}{\Delta t'}$, where Δt is the time difference between the two frames. The speed estimation can also be done automatically using image recognition by tracking the displacement of the bounding box of the passing car (or part of the car such as the front wheel). Figure 20 highlights the minor discrepancies between manually measured distances in car images and those computed through image recognition. From experiments, the errors in the displacement estimates are less than 5%. This works out to be 2 km/h for a relative speed of 40 km/h.
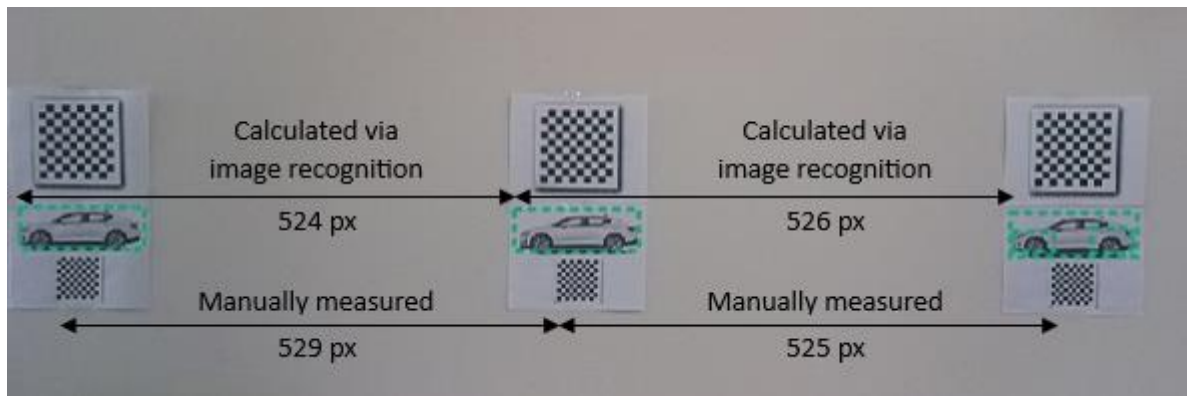

Fig. 20. Comparing manually measured and image recognition-calculated displacements.

Figure 21 shows an experiment to estimate the speed of a vehicle passing a stationary camera using the methods described above. The calculated values show close agreement with the speedometer readings (Table 9).
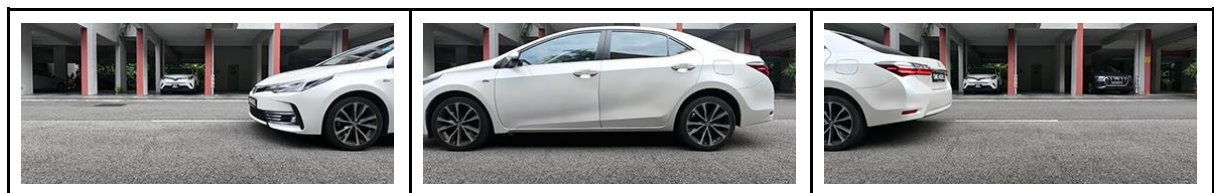

Fig. 21. Estimation of vehicle speed from image frames extracted from video.

Fig. 22. Speedometer readings at test run 1 and 2 respectively

Table 9
Comparison of measured and calculated speed values

Test run	Distance from car (m)	Calculated speed (km/h) - Manual	Calculated speed (km/h) - Image recognition	Speedometer reading (km.h)
1	2.28	29.5	28.3	30
2	1.92	30.9	30.0	32

6. Limitations and Future Work

The primary objective of this project was to explore the potential of new technologies like AI and machine learning to streamline and automate the process of collecting vehicle passing distances. While we have made progress, the current setup is not entirely complete. The combination of microcontrollers and sensors employed might not be the most efficient or cost-effective, suggesting that there could be alternatives better aligned with our goals and potentially more economical. Additionally, different hyperparameters or even a different clustering technique altogether could enhance our results by minimizing false positives. A comprehensive evaluation across diverse datasets is crucial to affirm the universal applicability and reliability of our chosen techniques.

Another small improvement would be to fully automate the process of creating the interactive graph. By utilizing the AI camera to take a picture during each vehicle pass, we can directly associate each cluster with its corresponding image. This automation will allow for a more convenient and instant identification of close pass incidents.

Looking forward, the project's evolution will encompass the assessment of a wider array of sensors, accompanied by extensive on-road data collection. This amassed data will help to test various clustering techniques, adjusting parameters to discern the optimal method. Upon establishing this foundation, our vision is to consolidate the hardware and software into a unified, market-ready product kit.

6. Conclusion

Our research hopes to use the latest technology to advance the boundaries of transportation safety. By integrating AI and machine learning with state-of-the-art sensors, we have unveiled a more streamlined, precise, and efficient methodology, hoping to offer a more efficient and easier alternative to traditional manual processes. With the foundation laid by our study, subsequent research can delve deeper into optimizing data collection techniques, thereby broadening the applicability and reliability of our findings. Moreover, the existence of this research paves the way for a host of applications. Planners and policymakers can now benefit from data that is more easily collected, aiding in crafting more effective transportation safety measures. Researchers can further build on this foundation by testing a broader range of sensors and algorithms, ensuring the continuous evolution and enhancement of our initial efforts. Ultimately, we hope that our research is able to work towards the goal of fostering safer roads and a more bicycle-friendly urban environment.

References

Beck, B., Chong, D., Olivier, J., Perkins, M., Tsay, A., Rushford, A., Li, L., Cameron, P., Fry, R., & Johnson, M. (2019). How much space do drivers provide when passing cyclists? Understanding the impact of motor vehicle and infrastructure characteristics on passing distance. *Accident Analysis & Prevention*, *128*, 253-260. https://www.sciencedirect.com/science/article/pii/S0001457518309990

Cheng, Y. (1995). Mean Shift, Mode Seeking, and Clustering. *IEEE transactions on pattern analysis and machine intelligence*, *17*(8), 790-799.

De Hartog, J. J., Boogaard, H., Nijland, H., & Hoek, G. (2010). Do the Health Benefits of Cycling Outweigh the Risks? *Environmental Health Perspectives*, *118*(8), 1109-1116. https://doi.org/10.1289/ehp.0901747

Ester, M., Kriegel, H.-P., Sander, J., & X, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *kdd*, *96*(34), 226-231.

Grab & NielsenIQ. (2021). *Foods Trends Report*. Grab. https://www.grab.com/sg/food/food-trends-report-2021/

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM computing surveys*, *31*(3), 264–323. https://doi.org/10.1145/331499.331504

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, *32*(3), 241-254.

Land Transport Authority. (2019). *Land Transport Master Plan 2040*. Land Transport Authority. https://www.lta.gov.sg/content/ltagov/en/who_we_are/our_work/land_transport_master_plan_2040.html

Land Transport Authority. (2023). *COE Open Bidding*. OneMotoring. https://onemotoring.lta.gov.sg/content/onemotoring/home/buying/coe-open-bidding.html

La Ville de Paris. (2021). *Le Plan vélo de Paris (2015-2020)*. https://www.paris.fr/pages/paris-a-velo-225

Lindsay, G., Macmillan, A., & Woodward, A. (2011). Moving urban trips from cars to bicycles: impact on health and emissions. *Australian and New Zealand Journal of Public Health*, *35*(1), 54-60. https://doi.org/10.1111/j.1753-6405.2010.00621.x.

Llorca, C., Angel-Domenech, A., Agustin-Gomez, F., & Garcia, A. (2017). Motor vehicles overtaking cyclists on two-lane rural roads: Analysis on speed and lateral clearance. *Safety Science*, *92*, 302-310. https://www.sciencedirect.com/science/article/pii/S0925753515002921

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, *1*(14), 281-297.

Ministry of Transport. (2019). *Government Accepts Recommendations From The Active Mobility Advisory Panel to Enhance Road Safety*. Ministry of Transport. https://www.mot.gov.sg/news/press-releases/Details/government-accepts-recommendations-from-the-active-mobility-advisory-panel-to-enhance-road-safety/

Pucher, J., & Buehler, R. (2017). Cycling towards a more sustainable transport future. *Transport Reviews*, *37*(6), 689-694. https://doi.org/10.1080/01441647.2017.1340234

Reynolds, D. (2009). Gaussian Mixture Models. *Encyclopedia of biometrics*, *741*, 659-663.

Road Safety Authority. (2019). *Examining the International Research Evidence in relation to Minimum Passing Distances for Cyclists*. Road Safety Authority. https://www.rsa.ie/docs/default-source/road-safety/r4.1-research-reports/safe-road-use/examining-the-international-research-evidence-in-relation-to-minimum-passing-distances-for-cyclists.pdf?sfvrsn=bf8ed37c_5

Singapore Police Force. (2020). *Annual Traffic Statistics 2020*. Singapore Police Force. https://www.police.gov.sg/-/media/170D31BB17EF441881138E1A556F210C.ashx

Stülpnagel, R. v., Hologa, R., & Riach, N. (2022). Cars overtaking cyclists on different urban road types – Expectations about passing safety are not aligned with observed passing distances. *Transportation Research Part F: Traffic Psychology and Behaviour*, *89*, 334-346. https://www.sciencedirect.com/science/article/pii/S1369847822001565

Walker, I., Garrard, I., & Jowitt, F. (2014). The influence of a bicycle commuter's appearance on drivers' overtaking proximities: An on-road test of bicyclist stereotypes, high-visibility clothing and safety aids in the United Kingdom. *Accident Analysis & Prevention*, *64*, 69-77. https://doi.org/10.1016/j.aap.2013.11.007

World Health Organization. (2020, November 5). *Cyclist safety: an information resource for decision-makers and practitioners*. World Health Organization (WHO). https://www.who.int/publications/i/item/cyclist-safety-an-information-resource-for-decision-makers-and-practitioners

Yap, W. Cycling Safety Analysis [Computer software]. https://github.com/wuihee/cycling-safety-analysis

Yap, W. Cycling Safety Code [Computer software]. https://github.com/wuihee/cycling-safety-code