Multimodal sensing prototype for robust autonomous driving under adverse weather conditions

Gerard deMas-Giménez[a], Adrià Subirana[a], Pablo García-Gómez[b], Eduardo Bernal[a], Josep R. Casas[c], and Santiago Royo[a,b]

[a]Center for Sensors, Instrumentation, and Systems Development (CD6-UPC), Universitat Politècnica de Catalunya, Rambla de Sant Nebridi 10, 08222, Terrassa, Spain
[b]Beamagine, S.L., Carrer de Bellesguard, 16, 08755, Castellbisbal, Spain
[c]Image Processing Group, TSC Department, Universitat Politècnica de Catalunya (UPC), 08034 Barcelona, Spain

ABSTRACT

Current autonomous driving datasets face significant limitations in adverse weather conditions diversity and sensor generalization. Additionally, commonly used sensors like visible cameras or rotating LiDARs, struggle to perform under harsh weather conditions. To address these challenges, this work introduces a multimodal data acquisition system that integrates high-resolution solid-state LiDAR, automotive RADARs, a combination of visible, thermal, SWIR, and polarimetric cameras, and a GNSS/INS system for odometry and localization. This diverse sensor suite ensures robust performance in low-visibility environments, such as fog or heavy rain, by providing complementary and redundant information. Controlled by an autonomous-safe Nvidia DRIVE AGX with a ROS-based architecture, the system enables precise spatial calibration, temporal synchronization, and real-time data fusion and perception algorithms across an overlapped field of view of 60°x 20°. With this system, this work aims to publish in the coming months an open-source multimodal labeled dataset for autonomous driving complemented by synthetic data generation through a digital twin in Nvidia's Omniverse platform. The data set will have a similar structure as the well-known nuScenes, and will have a similar developer kit to navigate through it with ease. The data set will support a wide range of autonomous driving applications like 3D multimodal object detection and tracking, SLAM, depth completion, and perception enhancement through challenging scenarios.

Keywords: Multimodal sensing, autonomous driving, perception, data set, LiDAR, adverse weather conditions

1. INTRODUCTION

Autonomous driving represents a pivotal shift in contemporary transportation, holding the potential to fundamentally alter the landscape of mobility and significantly improve road safety standards. The advancement and deployment of autonomous vehicles (AVs) rely upon sophisticated sensing and perception technologies, enabling these vehicles to interpret and navigate their surroundings in real-time. As the global need for safer and more efficient transportation alternatives intensifies, the autonomous driving industry has strengthened its position as a fiercely competitive market. Companies and researchers around the globe are engaged in developing unique and varied strategies to address the challenges of autonomous navigation to get an edge over their competitors.

A crucial element in these advancements is the data acquisition systems employed by AVs, as they are in charge of feeding the algorithms that power autonomous driving. While some strategies focus on integrating

Further author information: (Send correspondence to G.D.M.G.)
G.D.M.G.: E-mail: gerard.de.mas@upc.edu
A.S.: E-mail: adria.subirana@upc.edu
E.B.: eduardo.bernal@upc.edu
P.G.G: pablo.garcia@beamagine.com
J.R.C.: josep.ramon.casas@upc.edu
S.R.: santiago.royo@upc.edu

multiple sensor modalities, others have opted for more specialized approaches focusing on specific data types to optimize performance.

One clear example of a specialized focus is Tesla's vision-centered approach.[1] Their sensor suit consists of multiple visible cameras to get a 360° field of view (FOV) around the vehicle. Tesla's " Autopilot" system detects objects and estimates depth with only images and sophisticated neural networks based on stereoscopic vision. While this system maintains reduced costs, as visible cameras are generally cheaper than other options like light detection and ranging (LiDAR), it comprises safety and robustness in certain environmental conditions where visible cameras might fail, like dense fog scenarios or low-light conditions.

In contrast, companies like Waymo, LLC, have embraced a multimodal approach,[2] integrating LiDAR, radar, and visible cameras to enhance perception accuracy and robustness. For instance, low-light conditions will not present as many problems as with this configuration LiDAR and radars, being active sensors, do not rely on external illumination. However, low-visibility scenarios such as fog, heavy rain, and snow will still compromise the system of the ride, as the water suspended particles in the air greatly absorb the operational wavelength of LiDARs, typically located in the infrared region of the electromagnetic spectrum.

Adverse weather conditions such as fog, heavy rain, and snow pose significant challenges to perception systems of autonomous vehicles, often degrading the performance of individual sensors due to physical limitations. These limitations underscore the critical need for a robust multimodal approach, where complementary sensing modalities are fused to ensure perception continuity. By leveraging the strengths and offsetting the weaknesses of each modality, multimodal architectures significantly enhance system reliability, enabling safer decision-making and navigation in edge-case scenarios where monomodal systems may fail.

This paper explores the development of a multimodal autonomous system prototype designed to produce a rich dataset for research and development of novel autonomous driving perception algorithms under adverse weather conditions. By integrating unconventional sensors, such as thermal and polarimetric cameras or solid-state LiDAR,[3] alongside more traditional modalities like radar and visible cameras, the prototype aims to create a diverse and accurate dataset for testing and enhancing autonomous driving systems. Through the careful collection, temporal synchronization, and spatial registration of such multimodal data, this approach has the potential to offer new insights into the performance and reliability of autonomous systems under any weather condition.

2. STATE OF THE ART

The advancement of autonomous driving technologies has been made possible thanks to the availability of large-scale, richly annotated datasets that support the development and benchmarking of perception, localization, and decision-making algorithms. Over the past decade, several benchmark datasets have emerged, each contributing uniquely to the research landscape through different combinations of sensor modalities and environmental contexts. However, while these datasets have established foundational tools for progress, they remain constrained in terms of adverse weather diversity and sensor modality coverage. This section provides a critical review of the most influential publicly available autonomous driving datasets—including KITTI,[4] Waymo Open,[2] nuScenes,[5] and SeeingThroughFog[6](STF)—highlighting their sensor configurations, environmental variety, and their respective limitations.

The KITTI[4] dataset is one of the most widely known benchmarks in autonomous vehicle research. It was introduced in 2012 by the Karlsruhe Institute of Technology and the Toyota Technological Institute at Chicago. It is considered the pioneer of modern autonomous driving datasets as it was the first to include 3D laser scanner and 3D annotations in real-life scenarios. They equipped a Volkswagen Passat (B6) with: two color and two grayscale visible cameras, a 64-beam spinning LiDAR with a maximum range of 100m, and a Global Positioning System (GPS)/Inertial Measurement Unit (IMU) localization system. The setup has been calibrated to enable data fusion following their calibration pipeline.[7] Despite its foundational impact, the KITTI[4] dataset exhibits several limitations that restrict its applicability to more generalized and safety-critical autonomous driving scenarios. Most notably, KITTI[4] was collected under predominantly clear-weather and daylight conditions, which limits its utility for developing and benchmarking perception systems under adverse weather or low-light environments. The dataset primarily captures urban and semi-urban settings around Karlsruhe, Germany, lacking diversity

in geographical, meteorological, and traffic conditions. Furthermore, while KITTI[4] was groundbreaking for introducing 3D LiDAR and stereo camera data, it does not incorporate additional sensor modalities to deal with the aforementioned challenging scenarios.

Released in 2019, the Waymo Open Dataset[2] represents Google's large-scale initiative to advance autonomous driving research. It is widely regarded as one of the most comprehensive datasets in terms of spatial coverage and sensor redundancy. Waymo's[2] data acquisition prototype features five rotating LiDAR sensors and five high-resolution pinhole cameras. All sensors are precisely calibrated and temporally synchronized to enable accurate multimodal data fusion. The dataset contains 1150 driving segments, with 1000 reserved for training and 150 for validation and testing, and each segment lasts 20 seconds. It includes dense 2D and 3D annotations across a broad spectrum of driving environments, including urban, suburban, and highway settings. Moreover, data was recorded at different times of the day and under varying traffic conditions.

Released in 2019 by Motional, the nuScenes dataset[5] marked a significant evolution in autonomous driving benchmarks. nuScenes features a comprehensive sensor suite mounted on a vehicle: six cameras, five RADAR sensors, and a spinning 32-beam LiDAR. This configuration allows for full 360º perception coverage, synchronized and calibrated for precise data fusion. The dataset contains 1000 driving scenes, each 20 seconds long, annotated with 3D bounding boxes, object tracking IDs, and semantic labels for 23 object classes. nuScenes was also notable for introducing a developer kit and metadata-rich JSON structure to streamline research and benchmarking across detection, tracking, and prediction tasks.

Previous data sets only focused on good environmental conditions. Even nuScenes'[5] and Waymo's[2] rainy scenes have fairly good visibility. As a result, existing autonomous systems perform well under normal imaging conditions but fail under adverse weather due to the bias toward clear scenes. Mercedes-Benz, alongside Ulm and Princeton universities, tackled the issue by presenting the first multimodal data set under adverse weather conditions. STF has over 10000km of driving in northern Europe during winter. Even under these circumstances, adverse weather conditions occur rarely. For instance, dense fog conditions only appear 0.01% of the time. Nevertheless, autonomous systems must still be able to drive safely. To overcome the challenges of adverse weather conditions, they presented a sensor suit sensitive to the visible, mm-wave, Near Infra-Red (NIR), and Far Infra-Red (FIR) bands. They used: two visible stereo cameras, a gated NIR camera, a radar, a 64-beam spinning LiDAR, a thermal camera, and an IMU.

Given the current publicly available data sets for autonomous vehicles, there is a need to continue the work of STF[6] by expanding and complementing autonomous driving data in adverse weather conditions with new sensors that explore different regions of the electromagnetic spectrum.

3. METHODOLOGY

To enable the collection of a comprehensive and resilient dataset for autonomous driving under diverse and challenging environmental conditions, we designed and implemented a novel multimodal data acquisition prototype. The system is engineered to support high-resolution, synchronized data capture from a wide range of complementary sensors, enabling both robust real-world perception and controlled synthetic scenario replication. This section outlines the methodology used in developing the data acquisition platform. The first subsection details the hardware architecture, sensor selection criteria, and physical integration strategy, including spatial layout and field-of-view considerations. The second subsection describes the procedures for multi-sensor temporal synchronization, spatial calibration, and data fusion, which ensure accurate alignment and interoperability across different sensing modalities.

3.1 Design and Sensor Suite

As seen in section 2, there is no publicly available dataset -except for STF[6]- that has a wide range of sensor and is tested under harsh environmental conditions. Therefore, this work intends to fill the gap in the literature by developing a design for robust multimode autonomous driving under adverse weather conditions that complements and updates the work done in STF.[6]

Our proposed sensor suite can be found in Figure 1. We equipped a Dacia Duster ECO-G with the following sensors: three RGB high-resolution visible cameras, one short-wave infrared (SWIR) camera, one thermal (LWIR)

camera, one polarimetric camera, two automotive frequency modulated continuous wave (FMCW) radars, one solid-state LiDAR, and a Global Navigation Satellite System (GNSS) / Inertial Navigation System (INS) system.

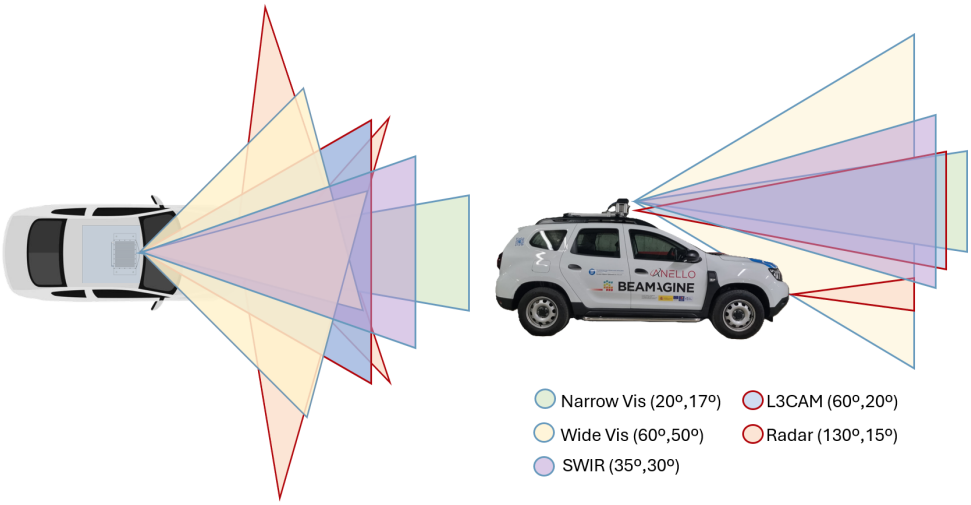
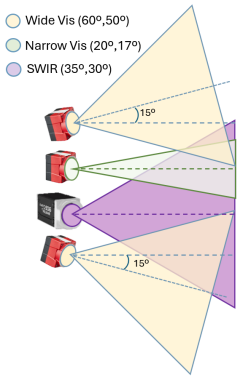


Figure 1: Design of the developed data acquisition prototype.

The RGB visible cameras are all Allied Vision Alvium G1-319c. Two of them have a wide FOV of 60° by 50°. They are 15° and -15° oriented with respect to the forward direction of the car. They also have part 30° of their overlapped FOV to enable stereo imaging algorithms. That leads to a combined horizontal FOV of 90°. The other RGB camera has a narrow FOV of 20° by 17° that focuses on small, far objects in the forward direction of the car. The SWIR camera is a Basler Ace 2X a2A640-240gmSWIR sensitive in the range from 400nm to 1700nm. This region has a remarkable penetration through turbid media such as fog or smoke, furthermore, it is also very sensitive to sources of light. The camera is able to see the lights of cars in dense fog environments even before visible cameras. The SWIR camera has a FOV of 35° by 30°. The distribution of the mentioned cameras can be seen in Figure 2.

(a) Structure supporting the 2D sensors on top of the ego vehicle.



(b) Cameras FOV representation.

Figure 2: Physical and graphic representation of the 2D sensors of the multimodal prototype.

Both radars are Smartmicro sensor UMRR-96 Type 153 Automotive, they are based on FMCW technology, and they operate in the range of 76-81GHz and have a nominal range of up to 120m with 130° of FOV. This model aims at short to medium range and very wide horizontal angular coverage, providing short, medium- and long-range modes. It is almost unaffected by bad weather, turning it into a very reliable sensor against adverse weather conditions. The radars are placed no more than 0.8m above ground behind the car's bodywork.

The thermal camera is a Seek Thermal C304SP that operates in the range from 7.8μm to 1400μm with a FOV of 56° x 42°. Thermal cameras are particularly useful in low-light and night-time driving, as well as early detection of warm bodies such as pedestrians, cyclists, and animals. It is also useful in smoke, yet struggles in fog as infrared radiation is significantly affected in foggy environments.[8]

The polarimetric camera is a LUCID Vision Labs Phoenix™ PHX050S1-QC. This camera gives RGB information as well as linear polarization intensity for 0°, 45°, 90°, and 135°. Polarization can be used effectively to distinguish different objects and filter undesired reflections in bliding light scenarios.

The last two cameras are integrated into a multimodal embedded system, an L3CAM produced by Beamagine, S.L. The L3CAM counts as well with a solid-state microelectromechanical system (MEMS) LiDAR with 60°by 20°of FOV. Solid-state LiDARs offer higher resolution and range compared to their spinning counterparts. In addition, they are more resilient to decalibrations due to mechanical misalignment. However, they cannot have a 360º horizontal FOV. Thanks to this sensor, we can also study how a higher resolution point cloud affects feature extraction and detection in state-of-the-art 3D perception neural networks. Additionally, the prototype also includes an Anello Photonics, Inc., EVK. A GNSS/INS system used for localization and odometry.

Table 1: Failure mode comparison for each sensor in adverse weather conditions. There is always at least one 3D sensor and one 2D sensor operating properly.

	LiDAR	RADAR	RGB	POL	LWIR	SWIR
Daylight	✓	✓	✓	✓	∼	✓
Night	✓	✓	✗	✗	✓	∼
Rain	∼	✓	∼	∼	∼	✓
Snow	∼	✓	∼	∼	✓	✓
Fog	∼	✓	✗	∼	✗	✓

While our system has considerably more imaging modes than other data acquisition prototypes, our FOV is restricted to just the frontal part of the vehicle. Nevertheless, there is a FOV of 60° by 20° where all modalities overlap. This gives the system a lot of redundancy and robustness against adverse weather conditions. For instance, if the prototype is driving in low-light conditions, the visible, polarimetric, and SWIR cameras will not produce good data, yet we can still use the thermal camera combined with radars and the solid-state LiDAR to identify the environment well enough. Table 1 shows a qualitative comparison of the performance of each sensor in our sensor suite. It can be seen that there is always at least a 2D sensor and a 3D sensor in a good performance regime independently of the weather condition.

3.2 Data fusion calibration

Accurate multimodal perception in autonomous systems critically depends on the precise alignment of sensor data both in time and space. Given the diversity of our sensor suite, robust data fusion requires meticulous calibration workflows. This subsection presents the methodology used for achieving temporal synchronization and spatial registration across all sensors in the system.

Temporal synchronization ensures that data from all modalities are temporally aligned to a common time base, enabling frame-level consistency and minimizing temporal drift. Due to the specifications of some of the sensors used, we were unable to set triggers for the sensors. Alternatively, we used a best-effort synchronization approach. Every sensor captures at its maximum frame rate, and data from each sensor is paired with the principal sensor. In our setup, our principal sensor was the solid-state LiDAR. The LiDAR works at a frame rate of 7 Hz, whereas the other cameras go from 11 to 15 Hz, depending on the external light conditions. Figure 3 shows the temporal delay from each sensor with respect to the primary sensor, the LiDAR, in more than 1200 frames. It can be seen that very rarely there is a delay of more than 50ms with respect to the LiDAR. This is an acceptable delay in the driving regime, as typical velocities on roads without separation from the opposite lane

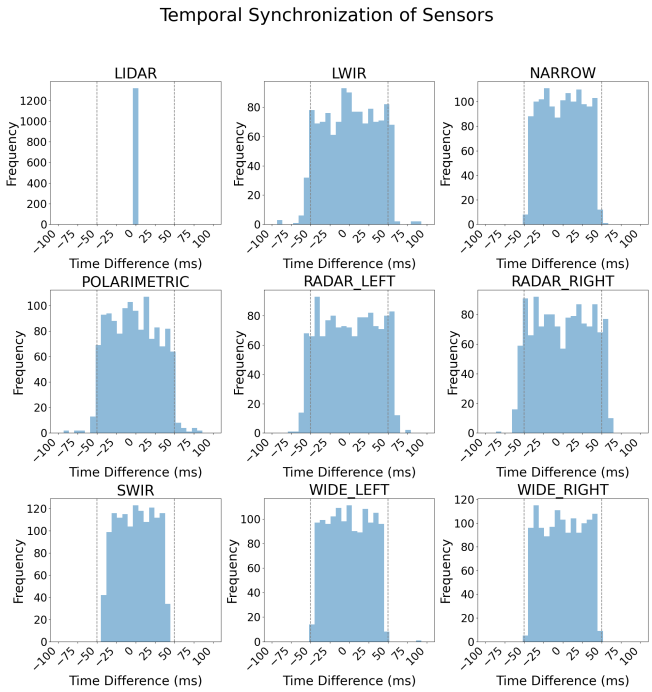


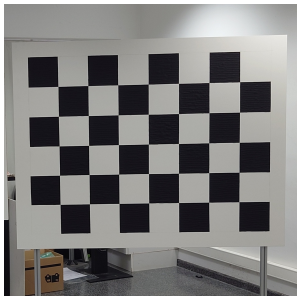
Figure 3: Time synchronization

are around 90km/h. If we consider the velocity of a vehicle with respect to another driving through the opposite lane, 180km/h, i.e., 50 m/s. In that scenario, a 50 ms delay represents 2.5m of discrepancy.

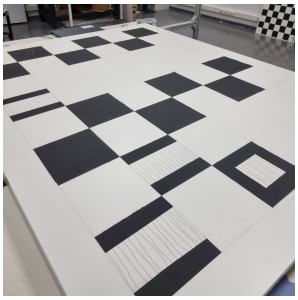
Spatial registration, on the other hand, focuses on determining the extrinsic parameters between sensors, allowing point clouds, images, and localization data to be fused into a unified coordinate frame. Common reference features have to be found in every mode with the solid-state LiDAR.[9] This is not trivial, as each sensor has its own sensitivity range, and they may not share it with other sensors in the system. Therefore, we designed two different calibration boards. One for visible, thermal, and LiDAR calibration and another for visible, radar, and LiDAR calibration.

The visible-thermal-LiDAR calibration board can be seen in Figure 4. It consists of a 5 by 7 square checkerboard. Each square measures 20 by 20 cm. Each square is made of a tape that greatly absorbs 1064nm light, the LiDAR's operational wavelength, so no returns from the LiDAR are obtained from them. Behind the tape, nichrome wire is carefully placed. Due to the Joule effect, the nichrome wire greatly heats up when current passes through it, making the squares visible for the thermal camera. The intersection between the squares can be obtained through classical image processing algorithms, making the process semi-automatic.

The previous board cannot be used to register the radars, as the metallic wire used to detect the pattern with the thermal camera interferes with the radar waves, resulting in a poorly localized return. To avoid the issue, a second calibration board was designed. The second board can be seen in Figure 5. This board was designed using the previous ideas of J. Domhof, et. al.[10] Similar to our previous board, we generate a checkerboard panel with the same absorbing tape, allowing us to identify it using both LiDAR and visible cameras. The board is made of styrofoam, which is invisible to radar waves. Behind the board, a carefully designed, metallic retroreflector is placed to force a high-intensity return from the radar. With the visible camera, we identify the board in the image. Taking advantage of the fact that the camera-LiDAR pair is already calibrated, we use the LiDAR to locate the board in the 3D space. We transform the radar's point cloud using a rough geometrical estimation of the position of the radar with respect to the LiDAR in the prototype. To fine-tune the initial transformation, we search for the radar high-intensity return from the retroreflector near the lidar's board location. With several captures, we used Kabsch's method[11] to obtain the final transformation that best fitted our set of points. The process is repeated for both radars individually.

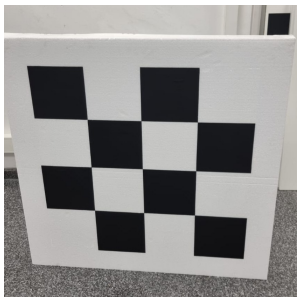


(a) Complete view of the visible-thermal-LiDAR calibration board.

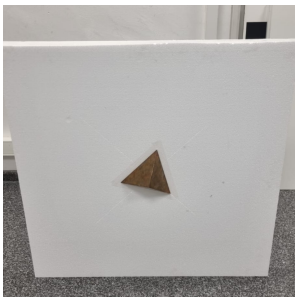


(b) Nichrome wire under the 1064nm absorbing tape of the visible-thermal-LiDAR calibration board.

Figure 4: Different stages of the visible-thermal-LiDAR calibration board.



(a) Front face of the board



(b) Rear part of the board, where the metallic retroreflector is placed.

Figure 5: Visible-LiDAR-radar calibration board.

4. RESULTS

Once the prototype was finally assembled and all sensors were calibrated, we began the on-field tests. The prototype was driven around Catalonia under diverse, highly dynamic driving scenarios like roads, highways and crowded urban scenes in Barcelona.

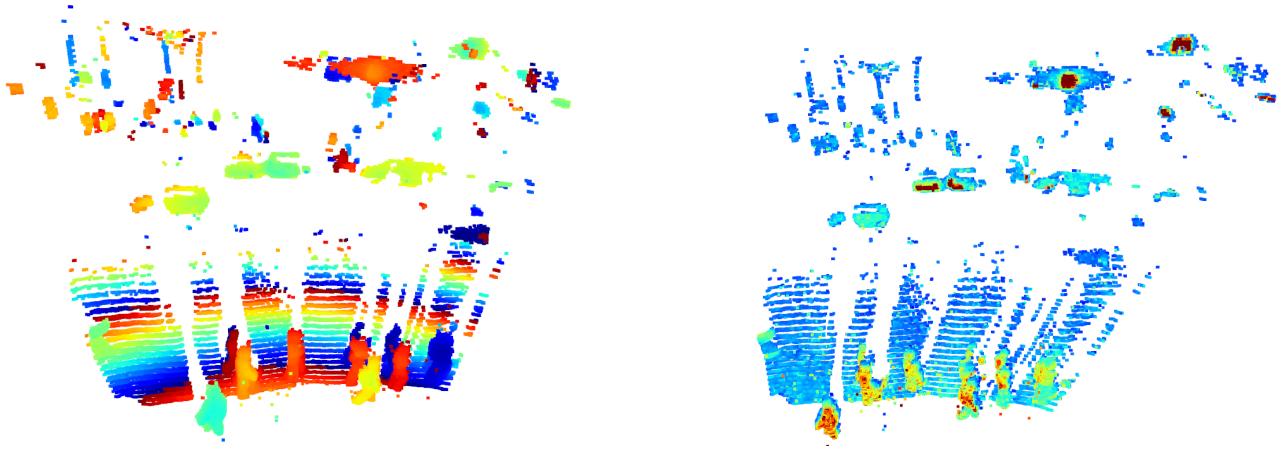
For instance, Figure 6 and Figure 7, show the 2D and 3D outputs from our sensors, respectively. It can be seen that under good weather conditions, every sensor performs in good regime. Thanks to the high definition, visible cameras, we obtain a clear semantic understanding of the scene. The narrow FOV camera helps us focus on further objects directly aligned with the vehicle. The thermal camera clearly distinguishes warm bodies from the background. Pedestrians and vehicles' wheels are clearly visible. In Figure 6, the polarimetric image is not decoded, so it does not give much qualitative information. Lastly, as the SWIR camera is also sensitive to visible light, the scene is clear. It can be seen that the SWIR camera is very sensitive to sources of light, see the several cars' lights in Figure 6, this effect is very useful to detect cars under low-visibility scenes such as fog or smoke.

Data fusion consistency even with dynamic scenes was also tested. In Figure 8, a point cloud projected onto two images from different sensors of the same frame, as in Figure 6 can be seen. We can see how the points precisely coincide with the shapes of the objects in the image. The point's colors represent the intensity of the return pulses. Also, the density of the solid-state LiDAR can be appreciated. Data fusion is consistent for a wide range of distances. In Figure 8, pedestrians crossing the crosswalk are around 8m from the ego vehicle, whereas the bus is further than 60m from the prototype and its retroreflectors are properly projected on both wide left and wide right cameras.

5. CONCLUSIONS AND FUTURE WORK

A multimodal data acquisition prototype specially designed to be able to explore perception under adverse weather conditions has been developed and described. Furthermore, two novel multimodal extrinsic calibration

Figure 6: Images from the 2D imaging sensors on the prototype in a crowded urban scene in Barcelona. From left to right, top row: wide left, narrow, and wide right visible RGB cameras. Bottom row: thermal, polarimetric, and SWIR cameras.



(a) Wide left (b) Wide right

Figure 7: Distance (left) and intensity (right) point clouds from the sample shown in Figure 6 in Barcelona.

boards have been developed to enable data fusion across all sensors. Results show that even though each sensor presents failure modes and weaknesses, relevant features from the environment can be obtained regardless of the weather conditions. As future work, a multimodal annotated data set for autonomous driving under adverse weather conditions with our prototype is currently under development and is expected to be publicly open-sourced and published later this year.

ACKNOWLEDGMENTS

This work has been sponsored by the Spanish Ministry of Science through projects TED2021-132338B-I00 and PID2023-152427OB-I00 and by the European Commission through project GA-101139941. This work is part of a Ph.D. thesis with grant number 2023 FI-1 00229 co-funded by the European Union.

(a) Wide left visible camera

(b) Wide right visible camera

Figure 8: LiDAR point cloud projected onto wide left and wide right images of the same frame in a dynamic crowded scene in Barcelona.

REFERENCES

[1] Tesla, "Tesla vision update: Replacing ultrasonic sensors with tesla vision," (December 2024). Accessed: 2025-06-02.

[2] Sun, P., Kretzschmar, H., Dotiwalla, X., Chouard, A., Patnaik, V., Tsui, P., Guo, J., Zhou, Y., Chai, Y., Caine, B., Vasudevan, V., Han, W., Ngiam, J., Zhao, H., Timofeev, A., Ettinger, S., Krivokon, M., Gao, A., Joshi, A., Zhang, Y., Shlens, J., Chen, Z., and Anguelov, D., "Scalability in perception for autonomous driving: Waymo open dataset," in [*2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*], 2443–2451 (2020).

[3] Royo, S. and Ballesta-Garcia, M., "An overview of lidar imaging systems for autonomous vehicles," *Applied Sciences* **9**(19) (2019).

[4] Geiger, A., Lenz, P., and Urtasun, R., "Are we ready for autonomous driving? the kitti vision benchmark suite," in [*2012 IEEE Conference on Computer Vision and Pattern Recognition*], 3354–3361 (2012).

[5] Caesar, H., Bankiti, V., Lang, A. H., Vora, S., Liong, V. E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., and Beijbom, O., "nuscenes: A multimodal dataset for autonomous driving," in [*2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*], 11618–11628 (2020).

[6] Bijelic, M., Gruber, T., Mannan, F., Kraus, F., Ritter, W., Dietmayer, K., and Heide, F., "Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather," in [*2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*], 11679–11689 (2020).

[7] Geiger, A., Moosmann, F., Car, Ö., and Schuster, B., "Automatic camera and range sensor calibration using a single shot," in [*2012 IEEE International Conference on Robotics and Automation*], 3936–3943 (2012).

[8] Li, H., Wen, S., Li, S., Wang, H., Geng, X., Wang, S., Zhai, J., and Zhang, W., "The research on infrared radiation affected by smoke or fog in different environmental temperatures," *Scientific Reports* **14** (2024).

[9] García-Gómez, P., Royo, S., Rodrigo, N., and Casas, J. R., "Geometric model and calibration method for a solid-state lidar," *Sensors* **20**(10) (2020).

[10] Domhof, J., Kooij, J. F., and Gavrila, D. M., "An extrinsic calibration tool for radar, camera and lidar," in [*2019 International Conference on Robotics and Automation (ICRA)*], 8107–8113 (2019).

[11] Kabsch, W., "A solution for the best rotation to relate two sets of vectors," *Acta Crystallographica Section A* **32**, 922–923 (Sep 1976).