4th year Project Proposal
(Team proposal)

Project title	Autonomous Driving Sensor Fusion
Student names	Maksym Bolotov, Rebecca Chen, Maxim Chu, Catherine Wu
Supervisor	Prof. Matt Ritchie
Date	18/10/2024

Abstract

The continuous emergence of automated vehicles (AVs) makes crucial the development of safe and robust sensing systems, which are capable of working accurately through all types of weather conditions and faults, as safety is the most important challenge in automated driving. This project proposes the combination of four low-cost sensors: a stereoscopic camera, Light Detection and Ranging (LiDAR), Radio Detection and Ranging (RADAR) and an Infrared (IR) camera, via sensor fusion to achieve road detection with precise measurements along with reliable human, vehicle and animal detection. The sensor fusion will follow a deep learning approach using online datasets, consequently being deployed to physical experiments. The project aims to house the sensor array on a moving platform connected to a single board computer (SBC) to simulate real world driving. Moreover, additional sensors are proposed including ultrasonics, GPS (for Simultaneous Localization and Mapping (SLAM)) and other camera types.

Project description and literature review

The current level of development in vehicle automation includes vehicles that are able to drive using some automated technology, such as advanced driver assistance systems (ADAS) [1]. The next level of development is partial driving automation, which involves using two or more driver assistance features. Some work has been completed; however, these vehicles are not yet approved for customer use [1]. This area of research is important for enhancing the safety of autonomous vehicles for everyone. The development of fully autonomous vehicles will hopefully eliminate human error, which is a leading cause of road traffic incidents [2]. Furthermore, this development will provide opportunities for people with mobility issues to become more independent due to AVs.

The aim of the project is to perform sensor fusion using RADAR (AWR1642), LiDAR (RPLIDAR A2), Camera (Intel RealSense D435), and an IR Camera to improve road safety. Furthermore, multiple algorithm will be developed to detect and classify objects, SLAM, path planning and tracking which will predict the movement of objects and avoid obstacles and hazards.

Depth cameras are essential for providing both depth and contextual information about the environment, which is a crucial aspect of AVs. The stereoscopic camera captures two images simultaneously from different angles, the disparity between corresponding pixels is then used to calculate depth. Additionally, the camera incorporates an IR projector to enhance depth perception in low-texture environments where standard visual cues are insufficient [3].

Stereo vision resembles human vision and hence enables the detection of obstacles such as other vehicles [4][5] and pedestrians [6], which is crucial for safe driving. From stereo data, regions of interest can be identified, objects classified, and 3D point clouds generated. In addition to object detection, cameras can capture visual cues such as lane markings [7][8], which assist in estimating the road layout and understanding the traffic scene. In the cases where visual data is ambiguous, probabilistic models have been employed to improve decision-making, ensuring a smoother, more reliable experience [9][10]. Cameras provide autonomous vehicles with the ability to make decisions like human drivers as they can detect, recognise, and extract valuable information. Furthermore, cameras have also been adapted to solve SLAM problems [11].

LiDAR has widely been used in autonomous driving experiments [12]. LiDAR is able to perceive an environment as a point cloud, accurately being able to calculate the distance, but is very limited in object classification ability. In robotics, LiDAR has been used to solve SLAM problems [13] as well as object detection and classification using machine learning [14]. In [13] a 2D LiDAR sensor was used to differentiate between objects and the background. Due to the nature of the LiDAR sensor, data is received sequentially through the angle at which the sensor is pointing and the distance to the nearest object. In [13] clusters of points were setup to detect when a large discrepancy between sequential data points was detected, implying an object was closer than the background. In this project, a similar approach can be used to analyse the LiDAR data and detect and verify the objects distance.

The AWR1642 is a frequency modulated continuous wave (FMCW) millimetre wave (mmWave) radar that operates within the 76-81 GHz frequency range [15] [16]. The antennas system consists of 4 receivers and 2 transmitters [16], making it a MIMO (Multiple Input Multiple Output) radar. Radio waves are emitted from the transmitters and reflect off the surfaces of objects. These reflected waves are received by the radar to calculate the objects' distance, angle and velocity [17]. The distance to an object is influenced by the time it takes for the radio waves to return [18]. Distance calculated by radar will be useful for object detection, as it operates reliably in all lighting and weather conditions, unlike optical and infrared sensors [16].

The angle of the object provides it direction relative to the radar and is determined by the difference in angle received by multiple receivers, this is known at the angle-of-arrival (AoA) [19]. The use of angles can be implemented in object localization to find the direction of an object relative to the AV [20]. The velocity of the object is determined by the shift in frequency of the received signal in relation to the transmitted signal, this is known as the Doppler effect [19]. This information will be useful for tracking the speed of objects in the surrounding environment.

An IR camera is sensitive to thermal radiation, creating a thermal image; a pixel grid of temperatures in a frame. IR cameras have advantages, such as being able to work in low-light conditions, through glare, shadows, fog and are able to distinguish objects differently (via temperature rather than colour). This is especially important as identifying people, cars and animals which are all objects of importance in the context of safe driving.

By incorporating an IR camera into our AV sensor array, identification of people, cars and animals will be more accurate, reliable and possible at larger ranges. Combining the strengths of an RGB camera, the relatively low resolution and frame rate are remedied, and object classification can be cross compared improving reliability and therefore safety.

Some work incorporating IR sensors in driving has been done such as by FLIR [21], a thermal sensor manufacturer. They provide additionally provide a dataset of fully annotated thermal images which may be used to train a convolutional neural network (CNN) to identify objects and classify them.

The core of the project is to effectively combine the strengths of the four sensors, Stereo Camera, LiDAR, RADAR and IR Camera into one cohesive system via sensor fusion.

The literature covers various combinations of these sensors, predominantly the pair of stereo camera and LiDAR, and the pair of stereo camera and RADAR.

LiDAR-Camera fusion is motivated by the powerful imaging of the camera being combined with the more accurate range measurements of the LiDAR. It is typically tackled via a deep learning approach [22, 23] such as the use of CNNs. These approaches train a machine learning model for road detection, usually using data from open datasets such as KITTI [24] or Waymo [25]. With some intrinsic and extrinsic calibration, the images captured by RGB cameras can be aligned with the LiDAR point clouds. The results from this fusion can be used to tackle different tasks such as accurate classification of objects especially vehicles, semantic segmentation, an ablation study on the robustness of the sensor fusion [26] or to simply improve depth estimation or build 3D mappings of objects [27]. Due to the highly informative nature of both of these sensors, LiDAR-Camera fusion [28] is the most common sensor fusion approach and is continually improving in performance and efficiency.

RADAR-Camera [29][30] fusion takes advantage of the accuracy and robustness of the RADAR combining it with the camera image to provide precise depth estimations at points in the camera image. Using a multiple input multiple output (MIMO) radar, and the time of flight (ToF) principle the RADAR data can be processed and converted to a point cloud that can be represented on a 2D image plane. Similarly to LiDAR-Camera fusion the more sparse depth estimations of the LiDAR/RADAR are combined with the image to complete tasks like object detection, classification and semantic segmentation. Data is commonly sourced from nuScenes [31] as the data is captured with a full sensor suite, including 5x RADARs.

Even less common, is literature regarding the combination of all three LiDAR, RADAR and Camera sensors. Through benchmarks such as the nuScenes leaderboard and the like, this sensor fusion has shown some SOTA performance. One newer approach uses a Bayesian Neural Network (BNN) [32] as an extension to use of a Deep Neural Network (DNN) in Multiple Object Detection (MOD). The motivation behind this is to measure uncertainty and has also improved the precision of MOD compared to its predecessor. Another paper [33] finds that in fusing multiple sensors together, the order / structure of fusion is of importance. They conclude that most effective way to combine the sensors is to first project the LiDAR onto the camera image, then consequently fuse the LiDAR points with the RADAR points to improve accuracy achieving SOTA performance.

The IR camera has been fused with RGB camera [34], LiDAR [35] and RADAR [36] in the past. For obvious reasons, the conclusions of the studies of IR with LiDAR and RADAR are almost identical to that of RGB with those same sensors, albeit with different justifications and use cases. It is more important to focus on the benefits of the IR camera versus the RGB camera which is covered above in the discussion of IR, in our case, the identification of people being the most key. There is literature regarding the fusion of RGB and thermal images [37] but this is not currently an area of interest for this project as the combined image cannot be properly utilised due to lack of appropriate training data.

Goals and objectives

1. Implement and calibrate individual sensors (Camera, LiDAR, Radar, IR)
    a. Collect and process the raw data from each sensor.
    b. Develop the algorithms to interpret the sensor data and extract essential information.
    c. Conduct sensor calibration and performance testing to understand the limitations of the sensor and to estimate errors.
2. Achieve sensor fusion between the sensors
    a. Preprocess the raw data to ensure compatibility for fusion.
    b. Implement and test sensor fusion techniques, starting with one sensor at a time (e.g. Camera-LiDAR) and then progressively add additional sensors until all sensors are incorporated (Camera-LiDAR-Radar-IR).
    c. Correctly time synchronise all sensors to an accuracy of 100ms.
3. Evaluate obstacle detection, object tracking and SLAM algorithms
    a. Develop and apply algorithms to detect objects and to navigate an unknown environment.
    b. Test and compare the performance of various algorithms, measuring the accuracy, robustness, and computational efficiency.
    c. Obtain a mean average precision (mAP) of 0.5 or above on the nuScenes detection task [31].

Project schedule

Referring to the Gantt chart on page six, the tasks that need to be performed are listed under five headings: Assignments, Sensors, Sensor Fusion, Sensor Perception, and Other Sensors. The tasks listed under “Sensors” indicate what is needed to prepare the sensors for sensor fusion. Interpreting the data involves understanding what the graphs, charts, and data represent. Validating the data includes checking for consistency and determining how far the data deviates from the true value. The tasks listed under 'Sensor Gantt Chart Fusion' specify all the necessary steps to fuse sensor data. Offline fusion will initially be attempted with two sensors, gradually building up to three and four sensors. Additional sensors may be added later and have been accounted for under the heading “Other Sensors”. The main focus of the project is “Sensor Perception”. This section is dedicated to programming the sensor perception and ensuring it functions correctly with simulated data before testing the algorithms on an RC car or robot.

Allocation of team resources

Catherine

- Radar set up
- Radar data interpretation/preprocessing (calibration and error investigation)
- Physical vehicle set up (incl. 3D printing)

Maksym

- LiDAR set up
- LiDAR data interpretation/preprocessing (calibration and error investigation)
- SLAM and localisation algorithm development

Maxim

- IR sensor set up
- IR data interpretation/preprocessing (calibration and error investigation)
- Sensor fusion algorithm development

Rebecca

- Camera set up
- Camera data interpretation/preprocessing (calibration and error investigation)
- Object detection/classification

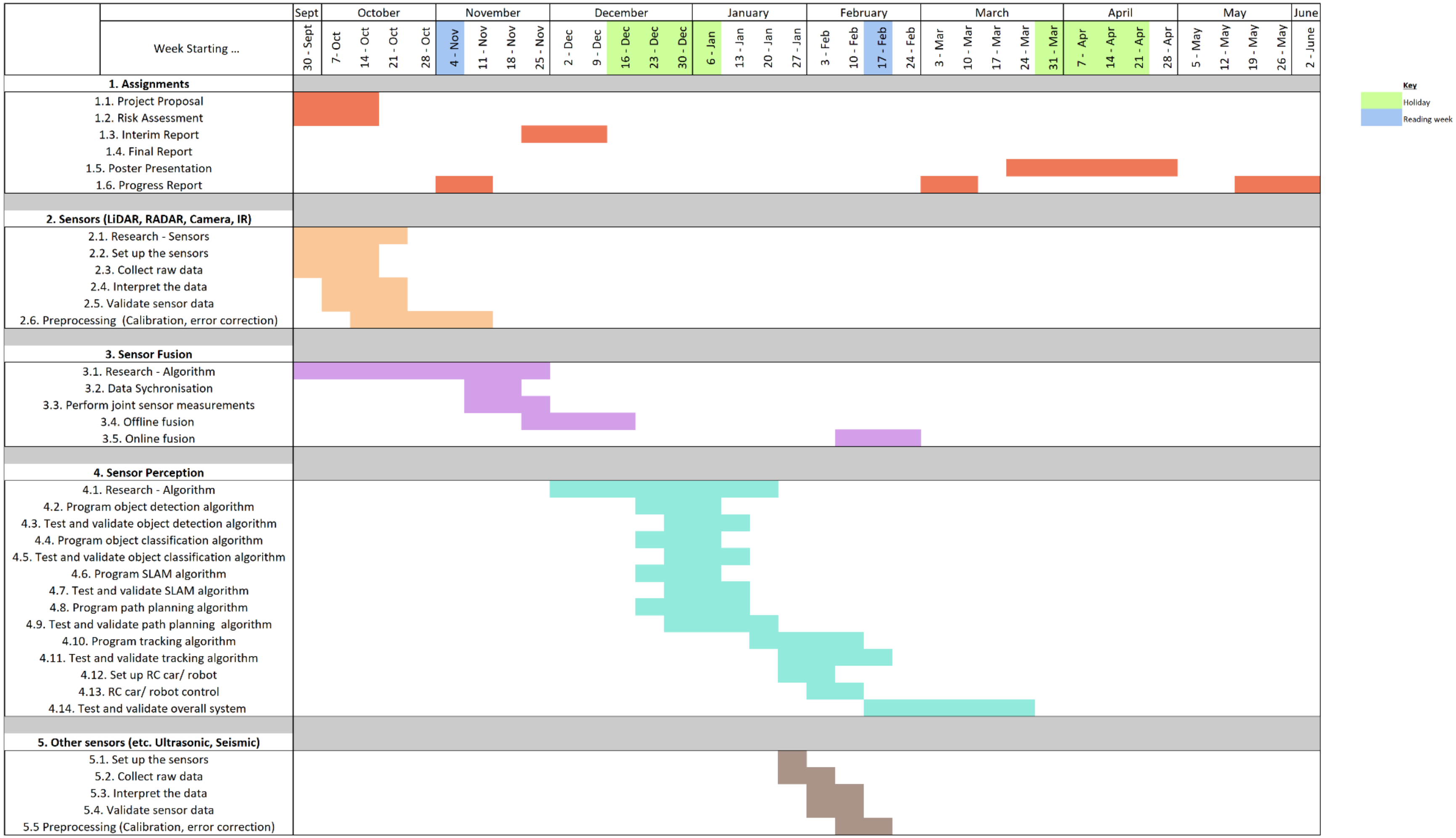
Failure risks and mitigation

Physical damage to the sensors while prototyping could cause substantial delay to the project due to the necessity of all the sensors for the fusion algorithm. To ensure this would not happen it will be necessary to ensure the sensors are mounted securely to any prototype built and are not damaged in transportation. This could happen throughout the duration of the project and could impact the project by up to 3 weeks to replace a sensor.

Throughout the project any of the team members could fall sick being unable to complete their project work on time. A sickness is an unpredictable event and could happen at any stage in the project. To mitigate this the group will have a shared Github space where every team member will have access and logs of work completed and can pick up another team members part if necessary.

Moreover, a potential source of error in the fusion algorithm could be data synchronisation. If the data that is captured from each perspective sensor is out of sync in a time domain then it could be detrimental to the fusion process and the testing of the algorithm. To mitigate against this it is necessary to consider running all the sensors from one laptop and hence capture data locally when all the sensors are synchronised. Alternatively, a synchronised server and algorithm that causes all sensors to start recording at once could be used.

Furthermore, a lack of physical power supply when real world testing could cause incomplete or corrupted data. The radar particularly is powered from a main supply, hence if a main supply is unavailable this could mean the sensor is unable to contribute to the fusion process. To mitigate this it is necessary to look into safe portable outdoor mains power supplies and decide how the sensors will be powered outdoors.


Week Starting ...
Sept
October
November
December
January
February
March
April
May
June
30 - Sept
7- Oct
14 - Oct
21 - Oct
28 - Oct
4 - Nov
11 - Nov
18 - Nov
25 - Nov
2 - Dec
9 - Dec
16 - Dec
23 - Dec
30 - Dec
6 - Jan
13 - Jan
20 - Jan
27 - Jan
3 - Feb
10 - Feb
17 - Feb
24 - Feb
3 - Mar
10 - Mar
17 - Mar
24 - Mar
31 - Mar
7 - Apr
14 - Apr
21 - Apr
28 - Apr
5 - May
12 - May
19 - May
26 - May
2 - June
1. Assignments
1.1. Project Proposal
1.2. Risk Assessment
1.3. Interim Report
1.4. Final Report
1.5. Poster Presentation
1.6. Progress Report
2. Sensors (LiDAR, RADAR, Camera, IR)
2.1. Research - Sensors
2.2. Set up the sensors
2.3. Collect raw data
2.4. Interpret the data
2.5. Validate sensor data
2.6. Preprocessing (Calibration, error correction)
3. Sensor Fusion
3.1. Research - Algorithm
3.2. Data Sychronisation
3.3. Perform joint sensor measurements
3.4. Offline fusion
3.5. Online fusion
4. Sensor Perception
4.1. Research - Algorithm
4.2. Program object detection algorithm
4.3. Test and validate object detection algorithm
4.4. Program object classification algorithm
4.5. Test and validate object classification algorithm
4.6. Program SLAM algorithm
4.7. Test and validate SLAM algorithm
4.8. Program path planning algorithm
4.9. Test and validate path planning algorithm
4.10. Program tracking algorithm
4.11. Test and validate tracking algorithm
4.12. Set up RC car/ robot
4.13. RC car/ robot control
4.14. Test and validate overall system
5. Other sensors (etc. Ultrasonic, Seismic)
5.1. Set up the sensors
5.2. Collect raw data
5.3. Interpret the data
5.4. Validate sensor data
5.5 Preprocessing (Calibration, error correction)
Key
Holiday
Reading week

References

[1] A. Kumar, K. Stephen and A. S. Sabitha, "A Systematic Review on Sensor Fusion Technology in Autonomous Vehicles," 2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2023, pp. 42-48, doi: 10.1109/ICESC57686.2023.10193038

[2] World Health Organization, “Road traffic injuries, ” https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries.

[3] Intel-Realsense-D400-Series-Datasheet.pdf, https://www.intel.com/content/dam/support/us/en/documents/emergingtechnologies/intel-realsense-technology/Intel-RealSense-D400-Series-Datasheet.pdf

[4] K. Huh, J. Park, J. Hwang, and D. Hong, ‘A stereo vision-based obstacle detection system in vehicles’, Optics and Lasers in Engineering, vol. 46, no. 2, pp. 168–178, 2008, doi: 10.1016/j.optlaseng.2007.08.002.

[5] C. C. T. Mendes and D. F. Wolf, ‘Stereo-Based Autonomous Navigation and Obstacle Avoidance*’, IFAC Proceedings Volumes, vol. 46, no. 10, pp. 211–216, 2013.

[6] A. Howard, L. H. Matthies, A. Huertas, M. Bajracharya and A. Rankin, ‘Detecting Pedestrians with Stereo Vision: Safe Operation of Autonomous Ground Vehicles in Dynamic Environments’, Jet Propulsion Laboratory, California Institute of Technology.

[7] M. Aly, ‘Real time detection of lane markers in urban streets’, in 2008 IEEE Intelligent Vehicles Symposium, 2008, pp. 7–12, doi: 10.1109/IVS.2008.4621152.

[8] Z. Rahman and B. T. Morris, ‘LVLane: Deep Learning for Lane Detection and Classification in Challenging Conditions’, in 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC), 2023, pp. 3901–3907.

[9] H. Zhang, A. Geiger, and R. Urtasun, ‘Understanding High-Level Semantics by Modeling Traffic Patterns’, in 2013 IEEE International Conference on Computer Vision, 2013, pp. 3056–3063, doi: 10.1109/ICCV.2013.379.

[10] A. Geiger, M. Lauer, C. Wojek, C. Stiller, and R. Urtasun, ‘3D Traffic Scene Understanding From Movable Platforms’, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 5, pp. 1012–1025, 2014, doi: 10.1109/TPAMI.2013.185..

[11] C. Häne et al., ‘3D visual perception for self-driving cars using a multi-camera system: Calibration, mapping, localization, and obstacle detection’, Image and Vision Computing, vol. 68, pp. 14–27, 2017.

[12] J. Van Brummelen, M. O’Brien, D. Gruyer, and H. Najjaran, “Autonomous vehicle perception: The technology of today and tomorrow,” *Transportation Research Part C Emerging Technologies*, vol. 89, pp. 384–406, Mar. 2018, doi: 10.1016/j.trc.2018.02.012.

[13] M. Mihálik *et al.*, “A method for detecting dynamic objects using 2D LiDAR based on scan matching,” *Applied Sciences*, vol. 12, no. 11, p. 5641, Jun. 2022, doi: 10.3390/app12115641.

[14] Y. Wu, Y. Wang, S. Zhang, and H. Ogai, “Deep 3D object Detection Networks Using LiDAR data: A review,” *IEEE Sensors Journal*, vol. 21, no. 2, pp. 1152–1171, Aug. 2020, doi: 10.1109/jsen.2020.3020626.

[15] D. B. N, V. G. S, P. Srihari and B. Pardhasaradhi, "Real Time Vital Sign Monitoring System using AWR1642 Radar Module with Remote Access," 2022 IEEE International Symposium on Smart Electronic Systems (iSES), Warangal, India, 2022, pp. 191-195, doi: 10.1109/iSES54909.2022.00047.

[16] Texas Instruments, “AWR1642 Single-Chip 77- and 79-GHz FMCW Radar sensor,” AWR1642 datasheet, May 2017 [Revised Dec. 2021].

[17] D. Parekh, N. Poddar, A. Rajpurkar, M. Chahal, N. Kumar, G. P. Joshi, W. Cho, “A Review on Autonomous Vehicles: Progress, Methods and Challenges,” Electronics, 11, (2022), doi: 10.3390/electronics11142162.

[18] T. J. Daim and R. M. A. Lee, "IR-UWB Radar Sensor Based Object Distance Determination as a Preliminary Input for Human Motion Detection," 2020 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS), Shah Alam, Malaysia, 2020, pp. 201-205, doi: 10.1109/I2CACIS49202.2020.9140174.

[19] K. Messaoudi O. S. Oubbati, A. Rachedi, A. Lakas, T. Bendouma, N. Chaib, “A survey of UAV-based data collection: Challenges, solutions and future perspectives”, Journal of Network and Computer Applications, Vol 216, 2023, doi: 10.1016/j.jnca.2023.103670.

[20] E.Saad, M.Elhosseini, A. Y. Haikal, ”Recent achievements in sensor localization algorithms”, Alexandria Engineering Journal, Vol 57, Issue 4, pg. 4219-4228, 2018, doi: 10.1016/j.aej.2018.11.008

[21] – Teledyne FLIR. (2024). FLIR Thermal Dataset for Algorithm Training [Online]. Available: https://www.flir.co.uk/oem/adas/adas-dataset-form/

[22] L. Caltagirone, M. Bellone, L. Svensson, and M. Wahde, “LIDAR–camera fusion for road detection using fully convolutional neural networks,” *Robotics and Autonomous Systems*, vol. 111, pp. 125–131, Nov. 2018, doi: 10.1016/j.robot.2018.11.002.

[23] Y. Li *et al.*, “DeepFusion: LiDAr-Camera Deep Fusion for Multi-Modal 3D Object Detection,” *arXiv (Cornell University)*, Jan. 2022, doi: 10.48550/arxiv.2203.08195.

[24] KITTI 360 “The KITTI Vision Benchmark Suite.” (2024) https://www.cvlibs.net/datasets/kitti/

[25] Waymo “About – Waymo Open Dataset,” (2024) *Waymo*. https://waymo.com/open/

[26] K. Yu et al., “Benchmarking the Robustness of LiDAR-Camera Fusion for 3D Object Detection,” Arxiv, Jun. 2023, doi: 10.1109/cvprw59228.2023.00321.

[27] “2019 Index IEEE Robotics and Automation Letters Vol. 4,” *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 4611–4712, Oct. 2019, doi: 10.1109/lra.2019.2955867.

[28] H. Zhong, H. Wang, Z. Wu, C. Zhang, Y. Zheng, and T. Tang, “A survey of LiDAR and camera fusion enhancement,” *Procedia Computer Science*, vol. 183, pp. 579–588, Jan. 2021, doi: 10.1016/j.procs.2021.02.100.

[29] S. Yao *et al.*, “Radar-Camera Fusion for object detection and semantic Segmentation in Autonomous Driving: A Comprehensive Review,” *IEEE Transactions on Intelligent Vehicles*, vol. 9, no. 1, pp. 2094–2128, Jan. 2024, doi: 10.1109/tiv.2023.3307157.

[30] X. Dong, B. Zhuang, Y. Mao, and L. Liu, “Radar Camera Fusion via Representation Learning in Autonomous Driving,” Arxiv, Jun. 2021, doi: 10.1109/cvprw53098.2021.00183.

[31] Nuscenes “Nuscenes by motion” (2024) https://www.nuscenes.org/

[32] R. Ravindran, M. J. Santora, and M. M. Jamali, “Camera, LiDAR, and radar sensor fusion based on Bayesian Neural Network (CLR-BNN),” *IEEE Sensors Journal*, vol. 22, no. 7, pp. 6964–6974, Feb. 2022, doi: 10.1109/jsen.2022.3154980.

[33] Y. Li, J. Deng, Y. Zhang, J. Ji, H. Li, and Y. Zhang, “${\mathsf{EZFusion}}$: A Close Look at the Integration of LiDAR, Millimeter-Wave Radar, and Camera for Accurate 3D Object Detection and Tracking,” *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 11182–11189, Jul. 2022, doi: 10.1109/lra.2022.3193465.

[34] R. Schmidt, “Benefits of IR/visible fusion,” *Proceedings of SPIE, the International Society for Optical Engineering/Proceedings of SPIE*, vol. 6541, p. 654105, Apr. 2007, doi: 10.1117/12.720529.

[35] J. D. Choi and M. Y. Kim, “A sensor fusion system with thermal infrared camera and LiDAR for autonomous vehicles and deep learning based object detection,” *ICT Express*, vol. 9, no. 2, pp. 222–227, Jan. 2022, doi: 10.1016/j.icte.2021.12.016.

[36] R. Mobus and U. Kolbe, “Multi-target multi-object tracking, sensor fusion of radar and infrared,” *IEEE Intelligent Vehicles Symposium*, vol. 41, pp. 732–737, Nov. 2004, doi: 10.1109/ivs.2004.1336475.

[37] F. Farahnakian, J. Poikonen, M. Laurinen, and J. Heikkonen, “Deep Convolutional Neural Network-based Fusion of RGB and IR Images in Marine Environment,” *IEEE Intelligent Transportation Systems Conference (ITSC)*, Oct. 2019, doi: 10.1109/itsc.2019.8917332.