A Sensor Fusion and Localization System for Improving Vehicle Safety in Challenging Weather Conditions

December 2021 | Final Report

SAFE-D
SAFETY THROUGH DISRUPTION

Texas A&M Transportation Institute

San Diego State University
Leadership Starts Here

VT | Virginia Tech Transportation Institute
Virginia Tech

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. 04-117	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle A Sensor Fusion and Localization System for improving Vehicle Safety in Challenging Weather Conditions		5. Report Date December 2021 6. Performing Organization Code:
7. Author(s) Abhay Singh (TAMU) Vamsi Krishna Vegamoor (TAMU) Sivakumar Rathinam (TAMU)		8. Performing Organization Report No. Report 04-117
9. Performing Organization Name and Address: Safe-D National UTC Texas A&M University Texas A&M Transportation Institute 3135 TAMU College Station, Texas 77843-3135 USA		10. Work Unit No. 11. Contract or Grant No. 69A3551747115/ Project 04-117
12. Sponsoring Agency Name and Address Office of the Secretary of Transportation (OST) U.S. Department of Transportation (US DOT)		13. Type of Report and Period Final Research Report 14. Sponsoring Agency Code

15. Supplementary Notes
This project was funded by the Safety through Disruption (Safe-D) National University Transportation Center, a grant from the U.S. Department of Transportation – Office of the Assistant Secretary for Research and Technology, University Transportation Centers Program, and, in part, with general revenue funds from the State of Texas.

16. Abstract
SAE Level 5 autonomy requires the autonomous vehicle to be able to accurately sense the environment and detect obstacles in all weather and visibility conditions. This sensing problem becomes significantly challenging in weather conditions that include such events as sudden change in lighting, smoke, fog, snow, and rain. There is no standalone sensor currently on the market that can provide reliable perception data in all conditions. We demonstrate that a combination of Long Wave Infrared (LWIR) cameras with radar provide a viable sensing system that is robust to adverse visibility conditions. We have validated this prototype system both in simulation as well as in real-world traffic using a 2017 Lincoln MKZ operating in College Station, TX.

17. Key Words Sensor Fusion, Thermal imaging, Radar, autonomous vehicles.	18. Distribution Statement No restrictions. This document is available to the public through the Safe-D National UTC website, as well as the following repositories: VtechWorks, The National Transportation Library, The Transportation Library, Volpe National Transportation Systems Center, Federal Highway Administration Research Library, and the National Technical Reports Library.

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 17	22. Price $0

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Abstract

SAE Level 5 autonomy requires the autonomous vehicle to be able to accurately sense the environment and detect obstacles in all weather and visibility conditions. This sensing problem becomes significantly challenging in weather conditions that include such events as sudden change in lighting, smoke, fog, snow, and rain. There is no standalone sensor currently on the market that can provide reliable perception data in all conditions. We demonstrate that a combination of Long Wave Infrared (LWIR) cameras with radar provide a viable sensing system that is robust to adverse visibility conditions. We have validated this prototype system both in simulation as well as in real-world traffic using a 2017 Lincoln MKZ operating in College Station, TX.

Acknowledgements

This project was funded by the Safety through Disruption (Safe-D) National University Transportation Center, a grant from the U.S. Department of Transportation – Office of the Assistant Secretary for Research and Technology, University Transportation Centers Program.

We are also grateful to Dr. Dezhen Song from the Computer Science and Engineering department at Texas A&M University for serving as the Subject Matter Expert for this report and for his valuable feedback.

Table of Contents

List of Figures

List of Tables

Introduction

It is well known that over 90% of traffic accidents on roads are caused by human errors [1] [2]. For human-driven vehicles, 22% of all crashes in the United States occur during adverse weather conditions and during poor visibility. The primary goal of fully autonomous driving is to eliminate such accidents by removing the driver from the equation. For such vehicles to operate safely, they need to be able to perceive other road users (cars, bikes, pedestrians, etc.) just as well as, if not better than, a human driver. Currently, no single sensor exists on the market that can perform this task reliably. Instead, autonomous driving systems utilize a combination of sensors for environmental perception.

A fully autonomous vehicle, as defined by SAE, must be able to perform all driving functions (including perception) under all conditions [3]. In order to achieve Level 5 autonomy, we need to design systems that work reliably even in adverse weather and visibility conditions, without the need for human intervention. Radars, lidars and RGB cameras have different operating characteristics in different situations. A summary of the relevant performance aspects obtained from a study by [4] is provided in Table 1.

Table 1: Performance of Typical Sensors Used

Aspect of Performance	Humans	Radar	Lidar	RGB Camera
Object Detection	Good	Good	Good	Fair
Object Classification	Good	Poor	Fair	Good
Visibility Range	Good	Good	Fair	Fair
Dark/Low Light	Poor	Good	Good	Poor

It has been demonstrated that lidar sensors perform poorly in heavy rain or fog due to reflections from the water droplets [4, 5]. RGB cameras yield limited information at night due to low lighting. They can only sense the area covered by the headlights and are further hindered by headlight bloom from oncoming vehicles. Moreover, they are severely disrupted by direct glare while driving directly in the direction of sunrise or sunset, as shown in Figure 1.

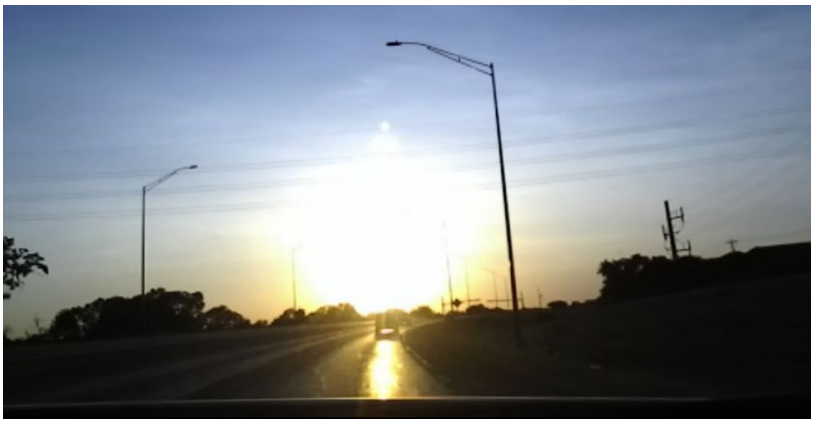

Figure 1: Glare from sun severely affecting visibility as seen from an RGB camera while driving.

Loss of visibility under intense sunlight can be fatal, as demonstrated by the failure of Tesla's AutoPilot system to detect the side of a trailer in 2016 [6]. Recently, [7] presented an approach for improving visibility in direct sunlight by using a High Dynamic Range (HDR) camera that captures multiple frames at different exposure settings to create a composite image. The same approach is not guaranteed to work for night-time autonomous driving, especially with a moving camera.

Thermal cameras, especially those operating over the Long Wavelength Infrared (LWIR) spectrum (8,000 to 14,000 nanometers) are inherently resilient to direct sunlight, as demonstrated in Figure 2, which presents the same scene as in Figure 1, but as seen from a thermal camera. Moreover, a study by [8] demonstrated the superior performance of LWIR cameras over lidars and RGB cameras in foggy conditions for identifying pedestrians. Thermal perception systems for autonomous driving have not received the same amount of attention in published research as RBG cameras; one of the goals of this study is to bridge this gap.

Figure 2: Thermal cameras are resilient to direct sunlight.

Since a radar sensor is also resilient to poor visibility conditions, in this project, we combine information from thermal cameras with that from radar to create an accurate map of the vehicle's

surroundings. We explore common data association and tracking approaches to determine the best approach for this application. We also validate our approach using simulations and experimental data.

Background

A variety of papers have been published in the field of perception and sensor fusion for autonomous driving. Many of these focus on some combination of RGB camera, lidar and radar, as this combination is commonly used. Recently, the self-driving industry has been making progress towards higher levels of autonomy (SAE Level 4/5), resulting in a rise in studies that explore the utilization of thermal camera information. Azam et al [9] utilized a thermal camera in combination with a lidar for object detection and tracking. As mentioned earlier, lidar sensors are not always reliable in foggy conditions. Moreover, the prohibitive cost of high-resolution lidar poses a challenge to large-scale adoption. Consequently, we do not use lidar information for sensor fusion in this project and instead use it only for validation.

Approaches for data association for multiple target tracking have been published in literature dating back to the late 1970s [10-12]. Many of the conventional approaches can be broadly separated into two classes: Single-Frame vs Multi-Frame data association, depending on whether the measurement-to-track association is made on a frame-by-frame basis, or if a history of "likely assignments" are stored for a complete decision to be made later [13].

The simplest of the single-frame methods is the Nearest Neighbor association, where the sensor measurement closest to the track is associated with the track and the rest are discarded as clutter. An optimal version of this algorithm, referred to as the Global Nearest Neighbor association, uses the Kuhn-Munkres/Hungarian algorithm [14]. Alternatively, Joint Probabilistic Data Association (JPDA) first proposed by [15], uses a Bayesian approach to data association, effectively utilizing a weighted sum of all measurements in the neighborhood of a track. Multiple Hypothesis Tracking (MHT) is an example of the multi-frame association [16] and newer approaches also exist based on Random Set theory and Particle Hypothesis Density [17]. A review of data fusion strategies is also available in [18].

In this project, we have primarily focused on JPDA and MHT approaches with the goal of evaluating each to determine the best fit for this application.

Methodology

Sensors

Five Automotive Development Kit (ADK) thermal cameras were acquired from FLIR, with a field of view (FOV) of 50° each. For the radar, a 77 GHz Delphi ESR module was used, mounted to the front bumper of the car. For data collection, a 2017 Lincoln MKZ owned by the Mechanical Engineering Department at Texas A&M University was utilized as the driving

platform. An Xsens IMU mounted in the trunk of the vehicle was used for collecting acceleration and orientation information. An RGB camera from PointGrey was also mounted to the inside windshield of the car to serve as a reference. The data from the RGB camera was not used in the sensor fusion algorithm.

An intel NUC 8th Gen (NUC8i7HVK) with 4 cores and 32 GB RAM was used to interface all the sensors, including the vehicle's drive-by-wire system. All data was recorded on board and sensor fusion was done offline for convenience using an Nvidia RTX 2070 graphics card. An Ubuntu 18.04 install of Robot Operating System (ROS) was used as the primary middleware.

Thermal System Mount

After multiple iterations through rapid prototyping (3D printing), a final design was created such that the thermal camera system provided an output with a 190° total FOV. This was to demonstrate the ability of such systems to perceive a large portion of the scene, similar to a lidar. It is possible to develop a 360° FOV system with more cameras or different models of the ADK.

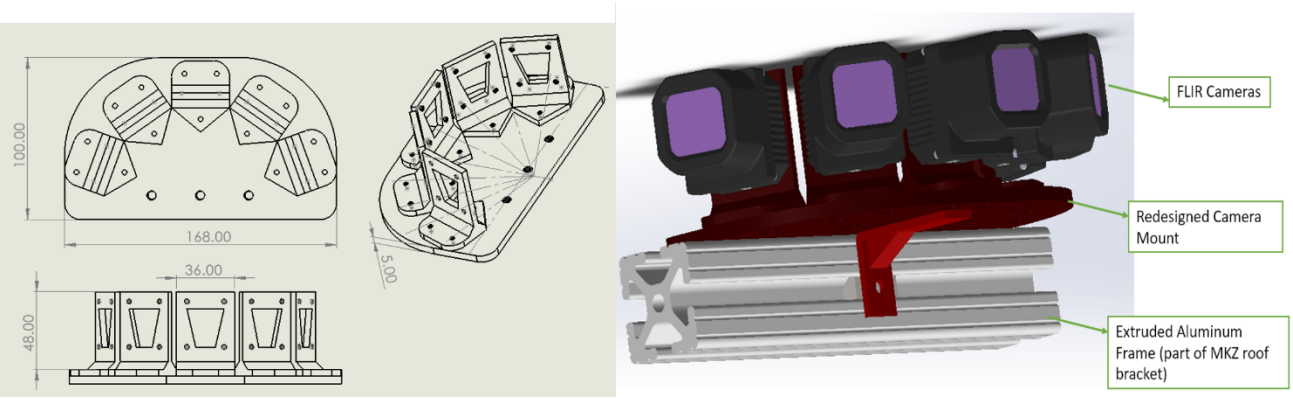

Figure 3: Final design of thermal camera system.

Figure 3 shows the key dimensions (in millimeters) of the designed mount, as well as a SolidWorks render. This design was machined in Aluminum and mounted to the vehicle rooftop. Figure 4 shows the thermal system after machining and assembly as well as the mounting location on the car.

Figure 4: Machined product and mounting location.

Thermal Panorama

Since the relative angle between the optical axes of the cameras are fixed (and known), a script was written to stitch the video feeds from all the cameras to create a panoramic video.

For this purpose, OpenCV tools were used to transform all cameras—1 through 5—to the image frame of the center camera (camera #3 in Figure 4). Then, overlapping portions of the images were cropped out to create a complete 190° FOV panorama. A screenshot of the stitched video is presented in Figure 5, showing a portion of the dataset collected on University Drive in College Station, TX. ROS device drivers for the FLIR ADK were supplied by the manufacturer but were designed for use with only one camera at a time. Consequently, the driver package was modified to accommodate simultaneously collecting video streams from multiple cameras. The video was collected at 20 frames per second. The ROS package that was developed will be made available on GitHub at the completion of this project.

Figure 5: Stitched thermal video panorama.

Each of the five thermal cameras performed their own Automatic Gain Control as set by the manufacturer. This caused the "banding" effect as seen in Figure 5, in spite of an attempt to mitigate this issue by using a simple gradient-based smoothening at the edges. More advanced image blending methods exist in published literature (for example, in [19]). Nevertheless, this issue did not adversely affect object detection performance, so no further action was taken.

Object Detection Through Machine Learning

From the generated panoramic video, objects on the road need to be identified—specifically, vehicles, pedestrians, and bicycles. For this project, the YOLOv4 (You Only Look Once) algorithm [20] was used. While pre-trained weights are available for images from regular cameras (RGB), they do not work well for thermal (grayscale) images. Moreover, the YOLO algorithm works well on images with an aspect ratio ≈ 1:1. Given that the stitched image from the five FLIR cameras has an aspect ratio ≈ 1:5, there was a need to retrain an image detector for this project.

FLIR offers an annotated thermal dataset[1] for image recognition training. It contains over 8,000 images taken at 512x640 resolution (aspect ratio ≈1:1). Since the training input images should have a similar aspect ratio to that of the panorama, each image in the dataset was edited by padding zeros (equivalent to black pixels in grayscale) on either side of the image to achieve the required aspect ratio. The annotations for each image were also transformed accordingly. Figure 6 and Figure 7 show an example of the original image from the dataset and the padded image. The size of the padding on either side was randomized to improve robustness of the image detector.

Figure 6: A sample image from the FLIR dataset with a 1:1 aspect ratio.

Figure 7: Random padding added to images from dataset for training, to impart 1:5 aspect ratio.

Once the object detector was trained, it was run offline on the previously collected data. The algorithm provides a list of the positions and types (cars/people/bicycles) of objects detected for each frame. Figure 8 shows a sample frame of the panorama with the detections overlaid at the

[1] https://www.flir.com/oem/adas/adas-dataset-form

intersection of College Ave. and University Dr. in College Station, TX. The dataset will be made publicly available through the SAFE-D Dataverse[2].

Figure 8: A sample output of the trained image detection algorithm.

Sensor Fusion Algorithm

Data Association and Tracking Using MHT

MHT is one of the widely used algorithms for multiple target tracking. The key idea behind MHT is to delay the data association process until more information is obtained. To achieve this, separate tracks are maintained for each possible data association. At every time step, predicted track position from a Kalman Filter is used to establish the gating area for each track. For any new observations that lie inside the gating area of a track, a new track is generated. A track is also kept to represent a missed detection case. A hypothesis is the collection of tracks which do not share observations at any time step.

In an earlier approach of MHT known as Hypothesis-Oriented MHT, all the hypotheses were propagated to the next time step. In this case, a single track can produce multiple tracks based on the number of observations inside the gating area, which in turn will produce even more possible hypotheses. Soon, the total number of possible hypotheses can blow up in size and it becomes very computationally expensive to handle them. To solve this problem, we used Track-Oriented MHT, where only the tracks to future time steps are passed on. Now we need a track scoring system which can be used to obtain the best possible hypotheses and remove unlikely tracks. Following the work by Sitlller [21], we used the log likelihood ratio between the target hypothesis and the null hypothesis as the track score. The target hypothesis assumes that the sequence of observations comes from the same target, and the null hypothesis assumes that the sequence of observations comes from the background.

After getting the score for each track, we wanted to determine the most likely hypothesis and then eliminate unlikely tracks by using N-scan pruning. The score of a hypothesis is the sum of the scores of all the tracks in the hypothesis. In order to find the most likely hypothesis, we generated a graph where each node represents a track with weight equal to the track score. An edge was

[2] Can be downloaded from dataverse.vtti.vt.edu/dataverse/safed. Please see the 'Additional Products' section of this report for the project URL.

added between two nodes if they shared an observation at any time step. Next, we calculated the Maximum Weighted Independent Set (MWIS) to find the most likely hypothesis.

N-scan pruning is an important step in MHT due to the exponential increase in the number of track hypotheses over time. First, we identified all the tracks which are the most likely hypothesis ("best tracks"). Let *k* be our current frame, then we keep all the tracks which are branching out from these best tracks after *k*-*N* frames and delete all others. In other words, we consolidate the data association decisions for old observations up to frame k−(N −1). Now we pass on the remaining tracks to the next frame.

Other practical issues have also been addressed, such as track initiation and keeping the number of tracks in check. As the MWIS is a computationally expensive problem to solve, keeping the number of tracks under control is essential for live tracking. Our implementation of MHT is divided into two stages. First, we will show the results based on radar data alone and next we will use information from the thermal image as well. There are two fundamental ways in which we can combine the information from radar sensor and thermal cameras:

1. Sensor fusion before tracking where we initiate and update tracks based on combined information from both sensors.
2. Sensor fusion after tracking where we initiate and update tracks for both sensors independently and then combine tracks using suitable weights for thermal and radar tracks.

We opted to start with implementing sensor fusion before tracking. As a prerequisite for this algorithm, we obtained bounding boxes for vehicles in the thermal image using a YOLO object detection algorithm. We then used this information along with radar information to improve tracks. One of the major contributions of bounding box data is that we can get one measurement for each vehicle. This is not feasible with radar data alone, as radar can give multiple measurements for the same vehicle.

Another practical issue we encountered is that the radar only provides horizontal distance and angle to the object in front of it. It is a two-dimensional data where we lose the detected object's height information. This becomes especially important when there is some elevation or depression in the road ahead. In such cases, projected track position on the thermal image may not accurately represent the observed car. We used the bounding box data to largely eliminate this shortcoming; however, the problem still can show up in certain situations when we were not able to detect the vehicle on thermal image.

Data Association and Tracking using JPDA

While the nearest neighbor approach (simply selecting the nearest sensor return at the next time step) works well for single object tracking, it performs poorly when clusters of objects are present in the scene. Moreover, it was observed that the Delphi ESR radar installed on the car often returns multiple measurements for the same object. For example, a single vehicle in front could return two or three "blips" on the radar. This violates the one-to-one correspondence

assumption of the nearest neighbor tracker. The Multiple Detection Joint Probability Data Association (MD-JPDA) tracker presents a Bayesian approach to tracking, which takes into account the possibility of many-to-one measurement to target correspondence [22]. For the JPDA approach, we assumed that the cluster size was small (less than four measurements per cluster), which allowed us to use the fast version of the algorithm proposed by Zhou and Bose [23]. Indeed, from the dataset we collected for tracking vehicles, this assumption held well, though it may need to be revisited for tracking pedestrians. An overview of the JPDA approach is as follows:

First, tracks need to be initiated. For this, we compared data from two consecutive frames for each sensor. If any measurement in the second frame fell within a small neighborhood (5 pixels for the camera and 30 cm for the radar) of the measurement in the first frame, then we initiated a track. These bounds on the neighborhoods are reasonable since the sensors collect data at 20 Hz, with only 50 milliseconds between two consecutive measurement frames.

Once a set of tracks was initiated, the camera and radar trackers grouped the incoming sensor measurements into clusters. This was done using ellipsoidal validation regions, which were chosen to maximize the probability that the true measurement from the track was within this region while minimizing the volume of the region. Such clusters were created independently for the thermal imaging system and the radar. These tracks were maintained using a Kalman filter. For the radar, the state vector consist of the relative position of the target with respect to the ego vehicle, the target velocity, and the heading angle. For the thermal camera system, the state vector consists of the coordinates of the center point bounding boxes along with dimensions of the box (in pixels), as well as their rate of change (pixels/second).

Finally, tracks had to be destroyed as the object being tracked moved out of the frame. This was achieved by maintaining a time-to-death counter for each tracked object. The counter was dropped by one for every time step when none of the new measurements could be associated with an existing track and is reset if a measurement was successfully assigned to that track in the future. If the counter dropped below a threshold, the track was deleted.

Results and Discussion

Over 200 GB of data was collected in and around College Station, TX in a variety of conditions, such as evening glare, nighttime, and rain. A video link showing the tracking performance for MHT and JPDA algorithms over a portion of the data set is available for viewing online.

- MHT (Night Time): Video Link
- JPDA (Sunset): Video Link

Validation

While we have a preliminary implementation of both the data association approaches, it is difficult to compare their performance based solely on visual representation of the sensor fusion

algorithms. Definitive object tracking metrics are available in the literature [24], but these require the ground truth trajectories of all objects in the environment. We pursued a two-pronged approach to solve this issue. First, we validated the algorithm using simulated data where the ground truth is known. Then, we used a high resolution lidar (Ouster OS1-128) and manual annotations to experimentally validate the algorithm.

Validation Through Simulation

By creating a detailed urban environment with objects of interest (cars, bicycles, pedestrians), we can obtain the exact location and pose information of all objects in the scene. Then by running each of the data association algorithms in this virtual environment, we can compare the predicted tracks with the known trajectories to evaluate performance. MATLAB's automated driving toolbox was selected, as it allowed simulation of cameras and radars as well as integration with Unreal Engine for better visualization. A summary of the simulation setup is provided below.

Sensor Simulation

Similar to the panoramic camera system mounted to the real vehicle, a set of five cameras was implemented using the Simulation 3D Camera block in MATLAB. Output from each of these cameras was published as an ROS topic with the same characteristics as that of the real thermal cameras (resolution, frame rate, field of view, etc.). Consequently, we were able to generate a complete panorama in a similar manner as the experimental data. An example of the stitched panorama from the simulated camera sensors in a virtual environment is shown below in Figure 9. The images were converted to gray scale to maintain a similarity with the output of the thermal cameras.

Figure 9: A screenshot showing stitched video, as seen through simulated cameras, from the ego vehicle.

Since the characteristics of the Delphi ESR radar have been experimentally measured by other researchers [25], we were able to readily use this information for tuning the parameters of the "Simulation 3D Probabilistic Radar" block in MATLAB to simulate the forward-facing radar. Again, the simulation was set up to publish ROS topics in the same format as that of the real radar on the car.

Driving Scenario Designer

A custom traffic scenario was designed using the Driving Scenario Designer tool in MATLAB. The scenario involved two vehicles in front of the ego vehicle as well as two vehicles traveling

in the opposite direction. Non-ego vehicles were set to perform lane change maneuvers so that we could test the performance of the sensor fusion algorithms in tracking vehicles that cross each other as well as those that are occluded. A video of the raw driving scenario designed is available here: (https://drive.google.com/file/d/1KdB9S-PPvEmavxYHTVpVgC2dPJF-vaPF/view?usp=sharing).

The final step for validation through simulation was to calculate the object tracking score. Since we knew the exact (ground truth) positions of the simulated vehicles in the scene, we were then able to compare this with the output from the fusion algorithms. The output is given in the camera coordinates. The output of the sensor fusion algorithm and the ground truth can be written into file where each line is in the following format:

<frame ID> <Track ID> <X> <Y> <Width> <Height>

This encapsulates the X-Y coordinates of each bounding box in the camera frame (in pixels) as well their width and height. In this way, the validation was done frame-by-frame. In each frame, both the number of objects detected and their position accuracy were validated. In order to standardize this validation, we used an open-source library[3] (py-motmetrics).

A comparison of the JPDA and MHT using the simulated sensor data is presented below.

Table 2: Comparison of MOT metrics for JPDA and MHT for Simulated Data

	Identification Precision (IDP)	Identification Recall (IDR)
JPDA	44.7%	38.2%
MHT	68%	49%

In this phase, for virtual validation, it appears that the MHT algorithm outperforms the JPDA algorithm by a notable margin. As a point of reference, open-source benchmark results[3] show an Identification Precision of 73% and an Identification Recall of 45.1%, so our results are comparable with those in published literature

Validation Through Experiments

For the validation of our developed tracking systems, we collected datasets with both a lidar and a thermal camera system. The Ouster OS1-128 lidar was also mounted on the roof of the vehicle, 15 cm vertically below the thermal camera assembly. This distance was small enough to assume that the camera and lidar had the same origin. Ground plane was segmented out from the lidar data and removed. Using the remaining lidar points, vehicles of interest were manually annotated

[3] https://github.com/cheind/py-motmetrics

using an open-source package[4]. A screenshot showing the annotations after ground plane removal is shown in Figure 10.

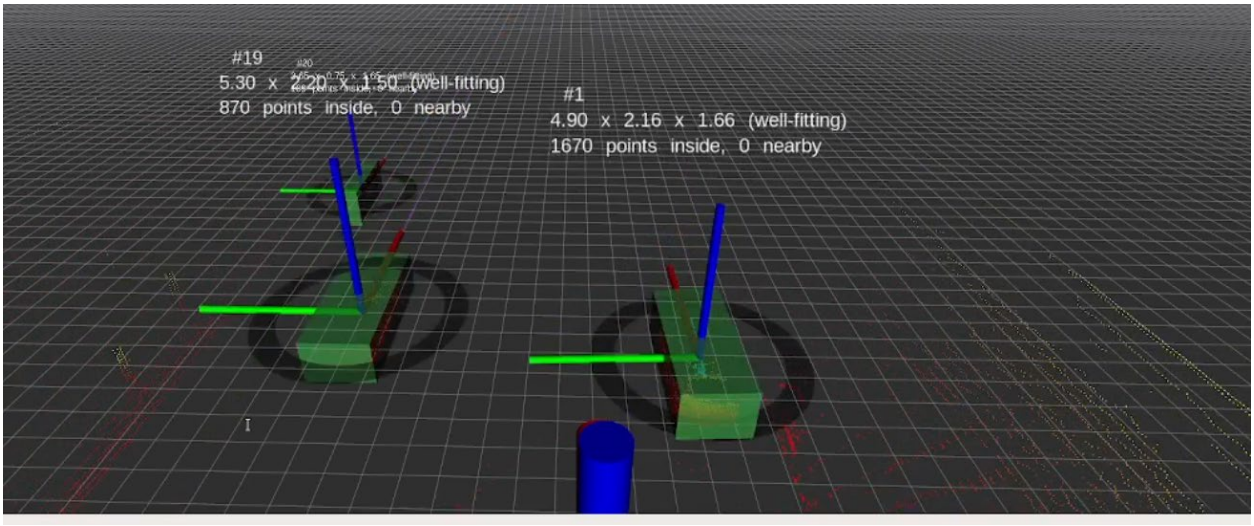

Figure 10: Manual lidar annotations after removal of ground plane.

As with the simulated validation, a text file was generated using the annotated data as the ground truth and was compared with the output of the sensor fusion algorithm.

Table 3: Comparison of MOT metrics for JPDA and MHT for Simulated Data

	Identification Precision (IDP)	Identification Recall (IDR)
JPDA	24.6%	22.2%
MHT	23.3%	27.4%

We observed that the scores were significantly lower for the experimental validation using lidar. There are two possible reasons for this.

First is the quality of the annotations themselves. Since the precision score is decided at a pixel level, the positioning of bounding boxes in the annotations significantly affects the result. In our case, the annotations were semi-automatically generated. The annotation process consisted of manually drawing bounding boxes for key frames and interpolating between these frames. We are currently working on fine tuning the annotations, which is a labor-intensive task.

Second, the lidar sensor's range is only up to 120 m, but the thermal camera system was able to detect vehicles at further distances for which no lidar points existed in the point cloud. An example of this is shown in Figure 11. This resulted in the lower Recall score. Nevertheless, this reinforces the superiority of the thermal perception system. In hindsight, we realize that a long-range lidar would have better suited the validation task, as it is difficult to selectively validate

[4] https://github.com/Earthwings/annotate

only the objects visible with the current lidar. This is something that can be pursued in a future extension of this work.

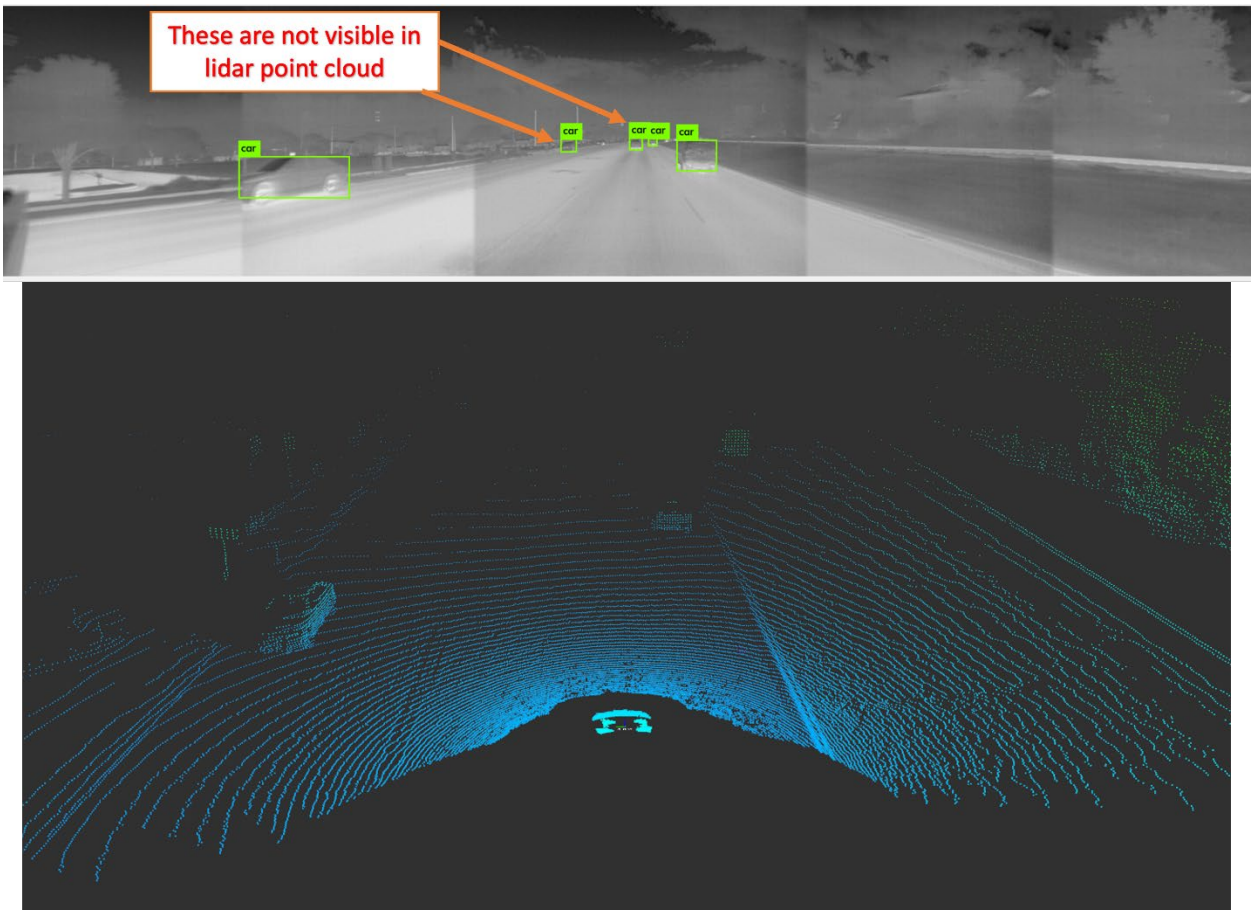

Figure 11: Cars that are far away are detectable by the thermal camera system but not by the lidar.

Conclusions and Recommendations

Conclusion

The broad objective of this project was to explore the feasibility of relying on a thermal camera to overcome adverse visibility conditions that pose challenges to other commonly used sensors for autonomous driving. A prototype hardware mount was designed and fabricated along with accompanying software that allows tracking of vehicles in a 190° FOV of the ego vehicle. The results were validated using simulated data as well as using a state of the art lidar sensor. It was observed that the combination of thermal and radar sensor can be used to supplement the existing perception stack. In fact, we observed that using thermal cameras extended the range of perception system since vehicles can be detected in the thermal image even when they are beyond the lidar's detectable range.

Future Work

It would be worthwhile to extend the dataset using the artificial rain tunnel at the RELLIS campus once construction is complete and to use artificial fog machines to simulate foggy conditions. While portions of our existing dataset contain these scenes, they were not long enough to draw strong conclusions. Furthermore, the current radar FOV is only 120°, so addition of side facing radars to create a panoramic radar sensor (similar to the thermal system) would be a natural extension of this work. Likewise, real-world validation from this study can be further strengthened by using long-range lidar sensors.

Additional Products

The Education and Workforce Development (EWD) and Technology Transfer (T2) products created as part of this project can be downloaded from the project page on the Safe-D website. The final project dataset is located on the Safe-D Dataverse.

Education and Workforce Development Products

Graduate Students (Abhay Singh and Vamsi Vegamoor) had their PhD dissertations partially supported by this Safe-D project.

Technology Transfer Products

1. The following conference paper was published at the SAE World Congress 2020. An additional journal paper summarizing the project is also planned for submission

   Bhadoriya, Abhay Singh, Vamsi Krishna Vegamoor, and Sivakumar Rathinam. (2021). *Object Detection and Tracking for Autonomous Vehicles in Adverse Weather Conditions*. No. 2021-01-0079. SAE Technical Paper.

2. Graduate students participated in the ENDEAVR (Envisioning the Neo-traditional Development by Embracing Autonomous Vehicles Realm) a non-profit event organized in Nolanville, TX. A prototype of the sensor fusion system was demonstrated and the public was educated about the capabilities of autonomous vehicles. URL: http://endeavr.city/

3. Abhay Singh served as a judge for the Texas Junior Academy of Science (TJAS) in 2021, which involved high school students from across the state.

Data Products

1. ROS Package for this project with all code and drivers.

   URL: https://github.com/VegaVK/flir_adk_multi

2. Datasets (ROS bag files) collected over the course of this project including lidar, radar, thermal camera and vehicle CAN bus data. To be uploaded to SAFE-D Dataverse.

URL: https://doi.org/10.15787/VTT1/B3VKEA

References

[1] J. R. Treat *et al.*, "Tri-level study of the causes of traffic accidents: final report. Executive summary," Indiana University, Bloomington, Institute for Research in Public Safety, 1979.

[2] N. H. T. S. Administration, "National motor vehicle crash causation survey: Report to congress," *National Highway Traffic Safety Administration Technical Report DOT HS,* vol. 811, p. 059, 2008.

[3] S. O.-R. A. V. S. Committee, "Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems," *SAE Standard J,* vol. 3016, pp. 1-16, 2014.

[4] B. Schoettle, "Sensor fusion: A comparison of sensing capabilities of human drivers and highly automated vehicles," *University of Michigan,* 2017.

[5] M. Kutila, P. Pyykönen, H. Holzhüter, M. Colomb, and P. Duthon, "Automotive LiDAR performance verification in fog and rain," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018: IEEE, pp. 1695-1701.

[6] K. Poland, M. P. McKay, D. Bruce, and E. Becic, "Fatal crash between a car operating with automated control systems and a tractor-semitrailer truck," *Traffic injury prevention,* vol. 19, no. sup2, pp. S153-S156, 2018.

[7] N. Paul and C. Chung, "Application of HDR algorithms to solve direct sunlight problems when autonomous vehicles using machine vision systems are driving into sun," *Computers in Industry,* vol. 98, pp. 192-196, 2018.

[8] K. M. Judd, M. P. Thornton, and A. A. Richards, "Automotive sensing: assessing the impact of fog on LWIR, MWIR, SWIR, visible, and lidar performance," in *Infrared Technology and Applications XLV*, 2019, vol. 11002: International Society for Optics and Photonics, p. 110021F.

[9] S. Azam, F. Munir, A. M. Sheri, Y. Ko, I. Hussain, and M. Jeon, "Data fusion of Lidar and Thermal Camera for Autonomous driving," in *Applied Industrial Optics: Spectroscopy, Imaging and Metrology*, 2019: Optical Society of America, p. T2A. 5.

[10] Y. Bar-Shalom and E. Tse, "Tracking in a cluttered environment with probabilistic data association," *Automatica,* vol. 11, no. 5, pp. 451-460, 1975.

[11] C. Morefield, "Application of 0-1 integer programming to multitarget tracking problems," *IEEE Transactions on Automatic Control,* vol. 22, no. 3, pp. 302-312, 1977.

[12] D. Reid, "An algorithm for tracking multiple targets," *IEEE transactions on Automatic Control,* vol. 24, no. 6, pp. 843-854, 1979.

[13] J. C. McMillan and S. S. Lim, "Data association algorithms for multiple target tracking," DEFENCE RESEARCH ESTABLISHMENT OTTAWA (ONTARIO), 1990.

[14] J. Munkres, "Algorithms for the assignment and transportation problems," *Journal of the society for industrial and applied mathematics,* vol. 5, no. 1, pp. 32-38, 1957.

[15] T. E. Fortmann, Y. Bar-Shalom, and M. Scheffe, "Multi-target tracking using joint probabilistic data association," in *1980 19th IEEE Conference on Decision and Control including the Symposium on Adaptive Processes*, 1980: IEEE, pp. 807-812.

[16] S. S. Blackman, "Multiple hypothesis tracking for multiple target tracking," *IEEE Aerospace and Electronic Systems Magazine,* vol. 19, no. 1, pp. 5-18, 2004.

[17] N. T. Pham, W. Huang, and S. Ong, "Probability hypothesis density approach for multi-camera multi-object tracking," in *Asian Conference on Computer Vision*, 2007: Springer, pp. 875-884.

[18] M. Liggins II, D. Hall, and J. Llinas, *Handbook of multisensor data fusion: theory and practice*. CRC press, 2017.

[19] Y. Xiong and K. Pulli, "Fast panorama stitching for high-quality panoramic images on mobile phones," *IEEE Transactions on Consumer Electronics,* vol. 56, no. 2, pp. 298-306, 2010.

[20] A. Farhadi and J. Redmon, "Yolov3: An incremental improvement," in *Computer Vision and Pattern Recognition*, 2018: Springer Berlin/Heidelberg, Germany, pp. 1804-02.

[21] R. W. Sittler, "An optimal data association problem in surveillance theory," *IEEE transactions on military electronics,* vol. 8, no. 2, pp. 125-139, 1964.

[22] B. Habtemariam, R. Tharmarasa, T. Thayaparan, M. Mallick, and T. Kirubarajan, "A multiple-detection joint probabilistic data association filter," *IEEE Journal of Selected Topics in Signal Processing,* vol. 7, no. 3, pp. 461-471, 2013.

[23] B. Zhou and N. Bose, "Multitarget tracking in clutter: Fast algorithms for data association," *IEEE Transactions on aerospace and electronic systems,* vol. 29, no. 2, pp. 352-363, 1993.

[24] K. Bernardin and R. Stiefelhagen, "Evaluating multiple object tracking performance: the clear mot metrics," *EURASIP Journal on Image and Video Processing,* vol. 2008, pp. 1-10, 2008.

[25] L. Stanislas and T. Peynot, "Characterisation of the Delphi Electronically Scanning Radar for robotics applications," in *Proceedings of the Australasian Conference on Robotics and Automation 2015*, 2015: Australian Robotics and Automation Association, pp. 1-10.