https://doi.org/10.2352/ISSN.2470-1173.2021.6.IRIACV-334

Methods and Comparisons Between Computer Vision and Radar Based Vehicle Location

Deniz Aykac, Regina Ferrell, Nisha Srinivas, Thomas Karnowski
Electrification and Energy Infrastructure Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831

Abstract

Measuring vehicle locations relative to a driver's vehicle is a critical component in the analysis of driving data from both post-analysis (such as in naturalistic driving studies) or in autonomous vehicle navigation. In this work we describe a method to estimate vehicle positions from a forward-looking video camera using intrinsic camera calibration, estimates of extrinsic parameters, and a convolutional neural network trained to detect and locate vehicles in video data. We compare the measurements we achieve with this method with ground truth and with radar data available from a naturalistic driving study. We identify regions where video is preferred, where radar is preferred, and explore trade-offs between the two methods in regions where the preference is more ambiguous. We describe applications of these measurements for transportation analysis.

Keywords: *camera calibration, data fusion, autonomous driving, scene understanding*

** corresponding author's E-mail: aykacdb@ornl.*

Introduction

The Second Strategic Highway Research Project (SHRP2) [1], included an extensive naturalistic driving study (NDS) where approximately 3000 drivers were recorded with dedicated data acquisition systems (DAS) in their personal vehicles [2,3] as shown in **Figure 1**. The study was completed in 6 data collection sites in US between 2010 and 2013. The vehicles in the SHRP2 study used a collection of low-cost, integrated instrumentation including face, dash, front and rear cameras, radar, and other sensors. The DAS was developed by Virginia Tech Transportation Institute (VTTI) that installed the systems in the cars and now warehouses the data [2]. The focus of SHRP2 was to identify conditions that led to unsafe driving, and as such is a valuable resource for researchers to use in their analyses.

Radar units were used in this study as a tool to represent the actual driving scenarios for researchers. Radar provides information about the position and velocity of the target vehicle relative to the participant vehicle and up to 8 target vehicles in a frame [4-6], with a goal of increasing the knowledge of the situation surrounding the crash and near crash cases to gain insight into driving behavior. The central concept of radar is that a pulse is sent from the transmitter in the radar unit and the reflected signal from objects (vehicles) are detected by an antenna. The SHRP2 radar data goes through a series of steps to make the raw data easier to interpret such as timestamp adjustments, removing targets that were placed in the wrong tracks. There are various issues with radar correctly detecting the target location. For example, based on the shape of the target the radar signature might not be clean leading to error in localizing the targets. Another phenomenon happens when the irregularities on the road causing the radar's aim to change, possibly leading to an erroneous decrease in range of the target in travel direction.

SHRP2 radar units were built with wider horizontal angle, compared to standard and more narrow horizontal view, to be able to track the adjacent lanes better. This adjustment has a trade-off by causing the signals to get weaker and thus transforming the raw data into correct tracks becomes more difficult. Another problem with radar is that signal reflected from non-uniform surfaces may be noisy. The process of translating raw radar data through multiple post processing steps into correct tracks is not easy. Because of the complexity of tracking targets, we investigate how accurate these low-cost radar data is and how it compares with modern video processing methods. In addition, we make comparisons between radar and video data in terms of consistency and reliability.

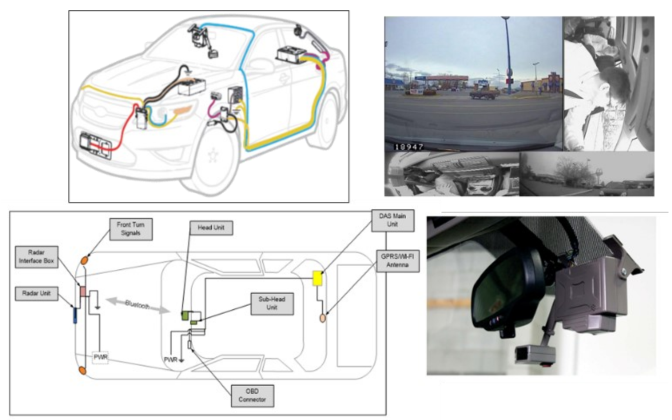


Figure 1 *SHRP2 – Data acquisition system; Left top and bottom show the placement of sensors and DAS in the vehicle from two different views, Right top shows the views from four cameras in the vehicle, Right bottom shows the placement of front-face-dash-view camera unit.*

Localizing the target using computer vision

Motivation

Measuring vehicle locations relative to a participant's vehicle is a critical component in the analysis of driving data in naturalistic driving studies (NDS). Relative vehicle locations provide valuable information about driving behaviors and interaction with the other vehicles [7], especially the vehicles traveling on the adjacent lanes and near the participant vehicle are of particular interest since they are most likely to get into accident with the participant vehicle. We seek to understand how the low-cost SHRP2 radars compare with modern video processing methods. Our experiments focus on localizing the targets using visual methods and comparing the results with radar readings.

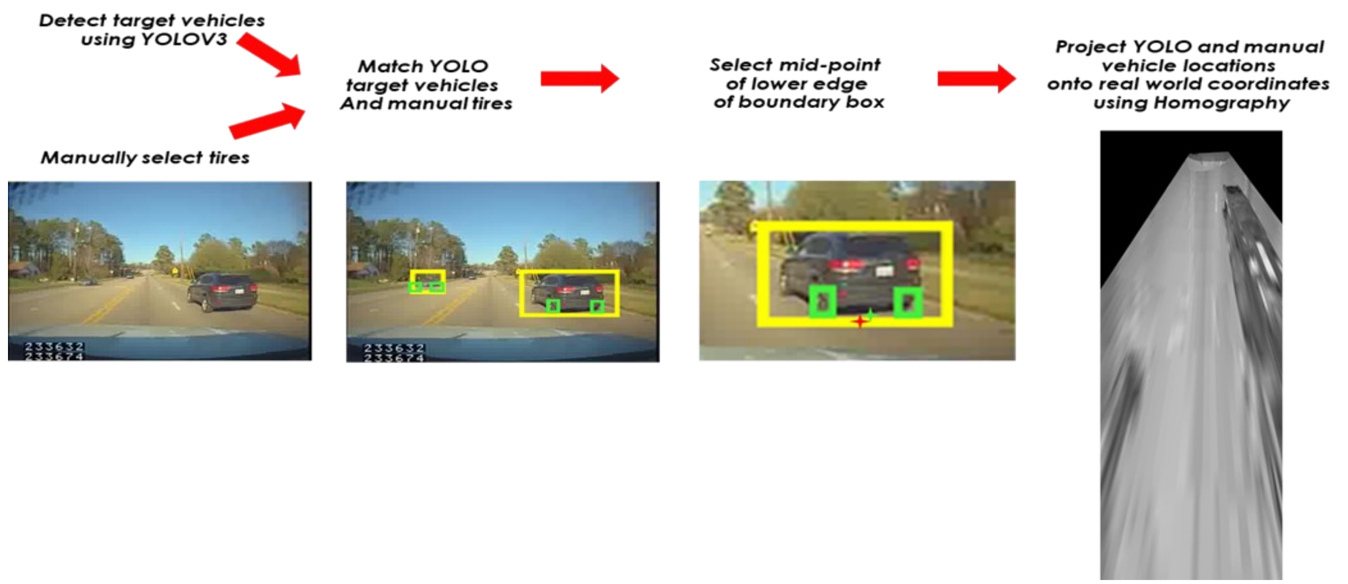


Figure 2 Flow of visual method

Procedure

The goal of the visual method described here is to detect and localize vehicle locations relative to the participant's vehicle using only the image data. The flow of the entire visual method is given in **Figure 2**. We start with detecting the vehicles using YOLOV3 pre-trained model [8]. For ground truth vehicle locations two tires were manually selected for each vehicle. The YOLOV3 predicted boundaries often require post-processing as it may find multiple boundaries for the same vehicle as shown in **Figure 3**. After matching two tires with each YOLO predicted boundary, we select the boundary which covers the tires more and thus maximizes the intersection over union (IOU) with the tires. Once designation of each vehicle boundary with two tires is complete the YOLO representative vehicle position (image point) was chosen to be the middle of the lower edge of the YOLO boundary (marked in red on left in **Figure 4**). The manual representative vehicle position is chosen as the middle of the line connecting two tires (marked in green on left in **Figure 4**). After these positions were found the next step was to estimate the extrinsic parameters (tilt and pan) for the cameras in vehicles used in the experiment. Then the frames were corrected with estimated tilt and pan parameters [9], which are assumed to be constant for a particular vehicle in the experimental data. YOLO and manual vehicle locations (image points) were then projected onto real world coordinates (planar points) using homography. During the process planar images (birds-eye-view images) were also generated as shown on the right side of **Figure 4**. The distances from participant to target vehicle were estimated on the plane of the road. The assumption is that middle of the dashboard to both YOLO and manual representative vehicle positions are considered relative participant to target distance. Converting image points into planar points and estimating the difference in homography plane resulted in this relative distance measurement. And hance, the target location relative to participant's vehicle was estimated in real world coordinates by using the visual method.

SHRP2 data sets have radar information for each trip depicting target location and velocity relative to participant's vehicle for each frame. Radar also tracks the targets through the frames. The targets are put into tracks based on the reflected signals coming from the vehicles in the direction of travel or the oncoming traffic. The relative distances are in meters and the location of the radar is right above the license plate. Based on this information about radar we were able to match the estimated target location from visual method to radar readings. There were cases in which radar failed to register target in a track. and there were other cases in which radar and visual method matched. All the analyses were performed on the "matched" radar and visual method's target locations. In the following section more specific cases, where radar fails and where visual method fails, will be presented in addition to inconsistencies in radar detection and in-depth target location analysis based on the target's travel lane.

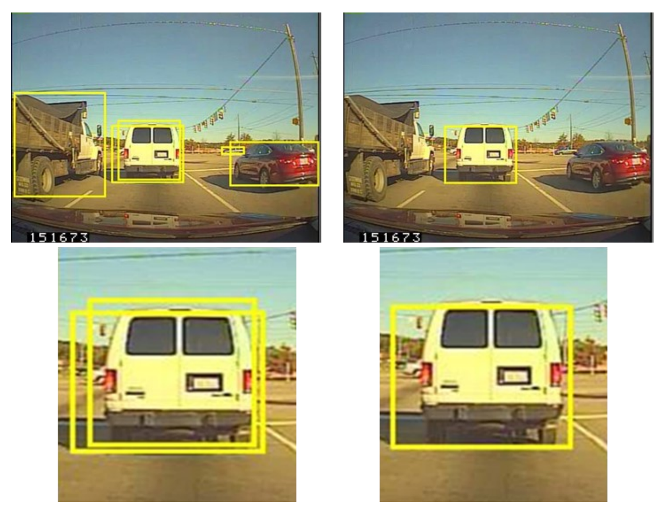


Figure 3 *Top; left: Example of multiple YOLO predicted boundaries for the same vehicle before processing, right: Single YOLO predicted boundary after post processing and matching with radar entry. Bottom; zoomed boundaries for the same vehicle before and after post processing*

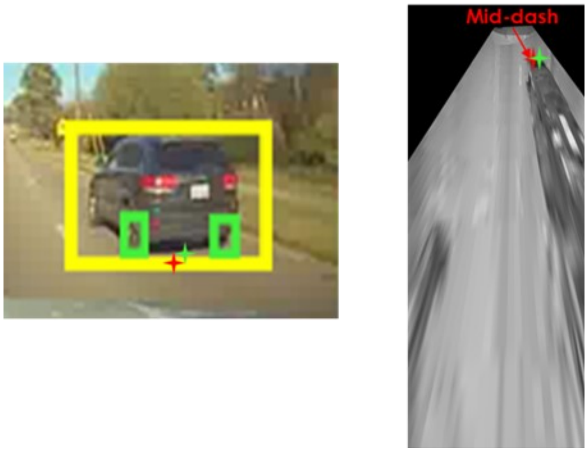


Figure 4 *Left: Selection of target locations for YOLO boundary box and manual tires. Right: Representative target locations on the planar image. Red is for YOLO vehicle location and green is manual vehicle location.*

Results

We used a total of 158 frames from 3 trips for 3 different vehicles from the SHRP2 training data set [5] in the analysis of target location. The frames were chosen to represent a homogenous distribution of vehicles traveling in same, left, and right lane with respect to the participant. We note that good weather and daylight conditions were chosen to minimize any other parameter affecting the performance of the visual method.

The frames were processed using YOLOv3 pre-trained model to detect vehicles. The predicted YOLO boundaries were then post-processed to eliminate multiple detections since it seems to be the biggest issue with YOLO vehicle detection. Manual tires were matched with these post-processed YOLO vehicle boundary boxes for each vehicle detection making sure that each vehicle was matched with a pair of tires. Then the estimated distance from participant to target location using visual method was matched with radar entry. Some anecdotal cases are worthy of mention to illustrate issues with the sensors and their processing. In **Figure 5** the radar was not able to detect the target on the left adjacent lane for 11 frames whereas the visual method was able to detect as close as 4.5 meters away from the participant vehicle. Inconsistency in radar detection is again apparent in **Figure 6**, in which the vehicle closer to the participant vehicle are detected but not the vehicle farther away.

The target vehicles were grouped into three different sections: those traveling in same, left, and right lane with respect to the participant vehicle. We computed the root-mean-square differences between the radar estimate and visual estimate for these three lanes to see if any of the lanes represent any significant error when compared. The results of this calculation is given in Table 1 for each vehicle as well as averaged over all trips/frames for all the vehicles. Note that the x-direction is the direction across lanes and y-direction is the travel direction. The difference between radar and visual method both for the travel direction and across lanes in the same lane is the lowest among the three lanes, with the same lane difference in x-direction was 0.72m and y-direction was 2.84m. The difference for left and right lane in both directions were comparable to each other and highest error was 4.44m in y-direction in right lane as shown in Table 1. The estimated target location based on the manual tires and both radar and visual method averaged over all lanes in both x and y directions are shown in Table 2.

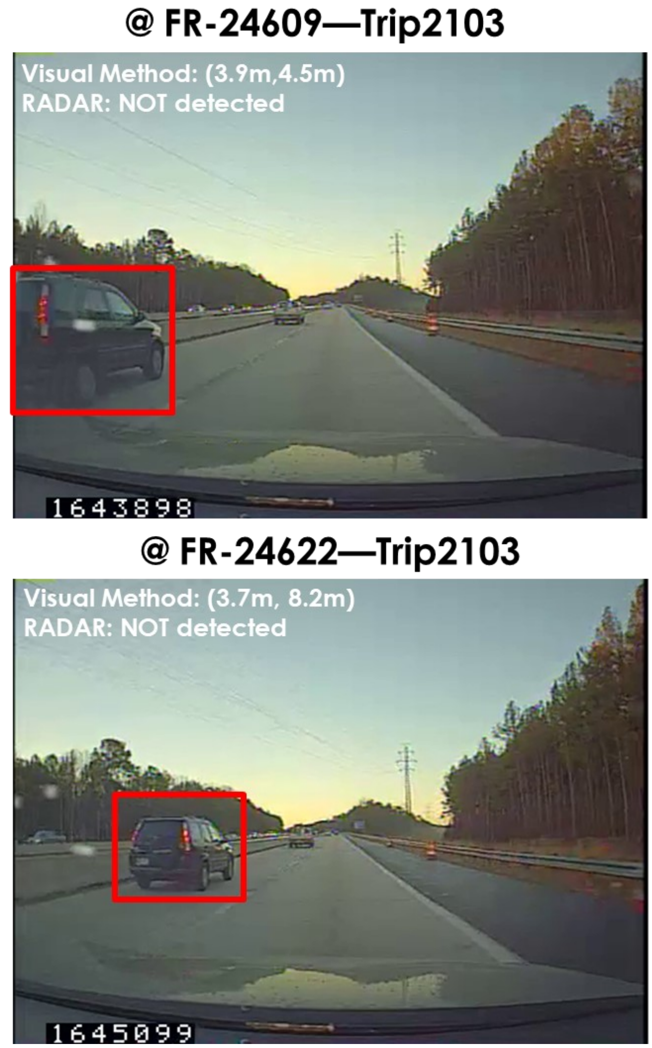


Figure 5 *Top: Radar doesn't detect the vehicle on the adjacent left lane @ FR-24609. Bottom: The same target vehicle at a later frame (@ FR-24622) is still not detected by radar as indicated by red boundary box.*

Table 1 Comparison of target locations in three lanes between radar and the visual method

	Left Lane	Same Lane	Right Lane
VehicleID_152851	(1.67m, 4.50m)	(0.26m, 3.68m)	(0.87m, 1.88m)
VehicleID_318769	(1.16m, 3.84m)	(0.80m, 3.84m)	(1.19m, 5.35m)
VehicleID_296344	(0.95m, 4.60m)	(0.72m, 2.48m)	(0.98m, 4.70m)
Averaged over all trips/frames for 3 vehicles	**(1.35m, 4.38m)**	**(0.72m, 2.84m)**	**(1.02m, 4.44m)**

Table 2 Comparison of target location in both directions between methods

	X-direction (across lanes)	Y-direction (travel direction)
Visual Method-Manual	0.62m	5.13m
Radar-Manual	1.10m	5.72m

Figure 6 *Inconsistency in radar detection: Top frame shows radar detected a vehicle 5.2m away but bottom frame shows radar was unable to detect a vehicle further away*

The distribution of difference in both directions between radar and visual method were investigated in more detail and presented in histogram plots in **Figure 7**. The plots show that radar and visual method agreed more in x-direction than y-direction which is expected because of the limitation in homography in y direction. The most important thing is to see how much of that limitation is a concern. To analyze it further the differences in y-direction were grouped into three categories, 0-15m, 16-30m and 31-45m as seen in **Figure 8**. These groups represent the distances to targets. The difference is within less than 5m for majority of cases in all groups. And maximum difference in distance to target appears to be the case when the target is farther away from the participant vehicle especially greater than 30 meters.

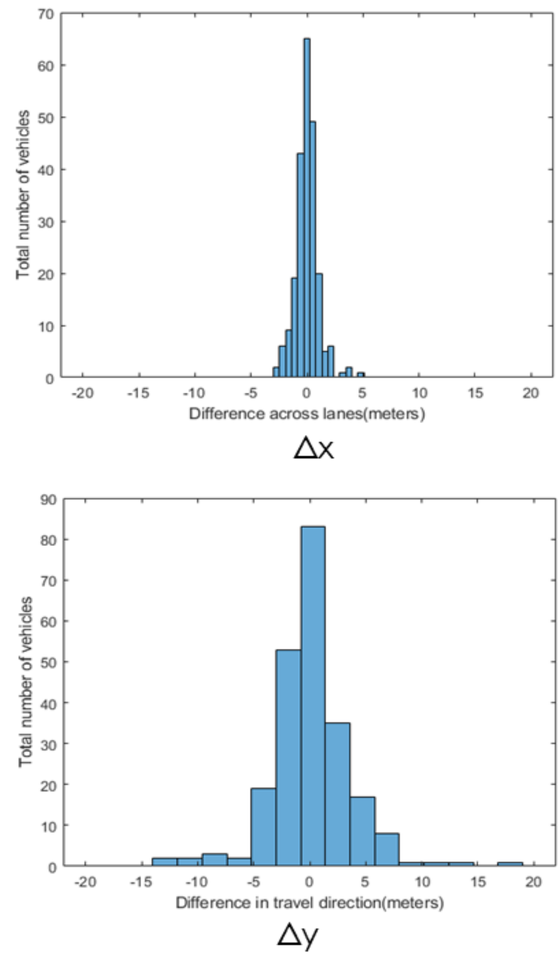


Figure 7 *Distribution of differences between radar and visual method in both x and y-directions*

The effects of YOLO boundary on localizing the target was apparent in cases as shown in **Figure 9**. Radar and visual method agreed on detecting the target vehicles but the difference in localizing is more pronounced in Target 1. There are two reasons for this error: the first is that the YOLO boundary doesn't cover the tires giving the false impression that the target is farther away than it is, whereas in Target 3 case the YOLO boundary is big enough to cover the tires but also tight enough to not cover more of the road features resulting in the agreement of radar and visual method. Another reason for Target 1 difference is that resolution of homography limits the localization if the target is farther away from the participant. In **Figure 10** we focused on YOLO boundary more closely and found out that variation in YOLO boundary boxes made the measurements inconsistent.

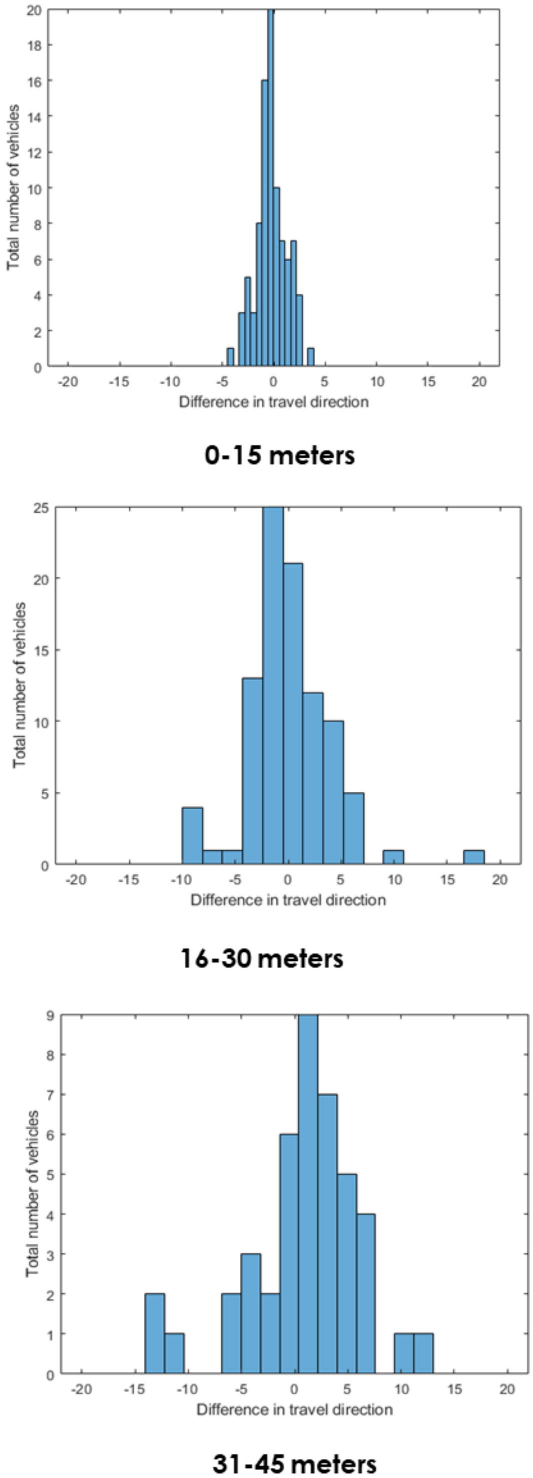
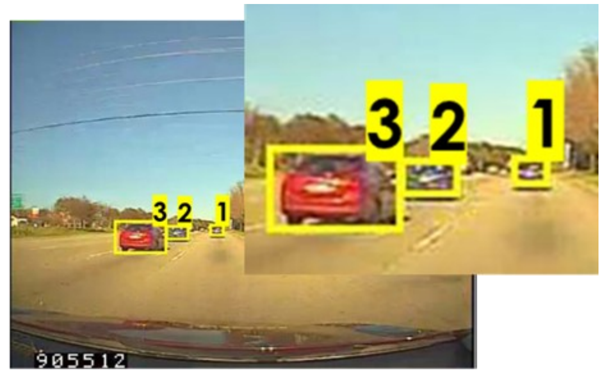


Figure 8 Distribution of difference in target location in y-direction grouped based on how far target away from the participant.



@FR-13540-Trip2981	Target 1		Target 2		Target 3	
Visual Method	2.1m	41.7m	5.5m	29.8m	4.8m	13.3m
Radar	1.3m	37.9m	4.2m	26.1m	4.4m	13m

***Figure 9** Problem with YOLO boundary box; difference between radar and visual method for Target 1 is related to the YOLO boundary box and how far the target is*

Conclusions

In this work we investigated the differences in computer vision-based target localization and radar, in particular trying to determine if the visual methods can be used in place of radar for cases where radar fails because of limitations of radar or in conjunction with radar for cases where radar output is noisy. In 158 frames used in our experiments, 189 target vehicles out of total of 423 were not detected by radar. Hence, driving behavior and interactions could be more accurately modeled by including the visual data. We again seek to emphasize that our studies were limited to daytime driving, as we know the visual methods are not trained on nighttime imagery.

To improve accuracy YOLOv3 pre-trained model parameters can be retrained or tuned to better predict the vehicle boundary so that it is large enough to maximize the IOU with two tires. Instead of using same vehicle calibration parameters for all the trips, individual trip calibration can be applied to the frames as the extrinsic parameters may vary from one trip to another as the camera's position might shift across trips. Overall, the visual method seems to be more likely to detect vehicles in close proximity to participant vehicle than radar and these vehicles are more likely to potentially impact driver behavior.

Figure 10 *Problem with YOLO boundary; top: @FR-37498 and bottom @FR-37505 with varying boundary boxes shown in the zoomed images.*

Acknowledgments
We would like to acknowledge the support and assistance of Virginia Tech Transportation Institute.
This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).
Work was funded by the Federal Highway Administration of the US Department of Transportation, Exploratory Advanced Research Fund.

References
[1] Walton, C. M., and B. L. Mallory. "Strategic Highway Research Program (Special Report 260). 2001." *Transportation Research Board* 17 (2006).

[2] Hankey, Jonathan M., Miguel A. Perez, and Julie A. McClafferty. *Description of the SHRP 2 naturalistic database and the crash, near-crash, and baseline data sets*. Virginia Tech Transportation Institute, 2016.

[3] SHRP2 https://intrans.iastate.edu/app/uploads/2018/03/1.-David-Plazak-SHRP-2-Data-2014-CUTC.pdf

[4] Gorman, T., et al, "SHRP2 Radar Post-Processing Task Report", https://insight.shrp2nds.us/projectBackground/download/1

[5] Li, Y. et al, "SHRP 2 Safety Data – Naturalistic Driving Study Training Dataset", https://insight.shrp2nds.us/projectBackground/zip_download/86

[6] Blatt Alan, et al, "Naturalistic Driving Study: Field Data Collection", SHRP 2 Report S2-S07-RW-1.

[7] Hammit,Britton, Rachel James, and Mohamed Ahmed. "Radar-vision algorithms to process the trajectory-level driving data in the SHRP2 Naturalistic Driving Study." *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018.

[8] https://github.com/AlexeyAB/darknet YOLO neural network for object detection

[9] Paone,Jeffrey, Karnowski Tom, Aykac Deniz, Ferrell Regina, Goddard James, Albright Austin, "Investigating Camera Calibration Methods for Naturalistic Driving Studies" *Electronic Imaging, Intelligent Robotics and Industrial Applications using Computer Vision* 2019, pp. 461-1-9.

Author Biography
Deniz Aykac received a BS in Physics from the Bogazici University, Turkey (1994) and an MS in Biomedical Engineering from The University of Iowa (2000). She has worked at Oak Ridge National Laboratory since 2002 extensively on 3D medical image processing, image, and video analysis.

Regina K. Ferrell received a BS in Electrical Engineering (1984) and a MS in Electrical Engineering (1994) from the University of Tennessee, Knoxville. She has worked for Oak Ridge National Laboratory since 1992 on a variety of projects with her primary focus on applications in image processing and big data.

Nisha Srinivas received her PhD and MSc. in Computer Science and Engineering from the University of Notre Dame in 2014. She has previously worked at Intel Labs, Oak Ridge National Labs and University of Wilmington North Carolina. She has been a Research Associate at ORNL since 2019. She has over 10 years of experience in the field of biometrics and machine learning.

Thomas P Karnowski received a BS in Electrical Engineering from the University of Tennessee (1988), a MS from North Carolina State University (1990), and a PhD from Tennessee (2010). He has worked at Oak Ridge National Laboratory since 1990 on a variety of research projects in applications of signal and image processing.