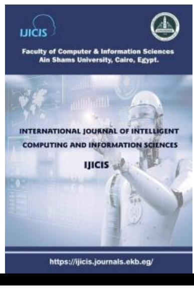
IJICIS, Vol.22, No.3, 54-69
DOI: 10.21608/ijicis.2022.117646.1158

International Journal of Intelligent Computing and Information Sciences

https://ijicis.journals.ekb.eg/



VEHICLES DETECTION AND TRACKING IN ADVANCED AND AUTOMATED DRIVING SYSTEMS: LIMITATIONS AND CHALLENGES

Mona A. Sadik
Department of Computer Science,
Faculty of Computer and Information Sciences,
Ain Shams University,
Cairo, Egypt
mona.sadik@cis.asu.edu.eg

Ahmed S. ElSayed
Department of Computer Science,
Faculty of Computer and Information Sciences,
Ain Shams University,
Cairo, Egypt
ahmed_salah@cis.asu.edu.eg

Sherin M. Moussa*
Department of Information Systems,
Faculty of Computer and Information Sciences,
Ain Shams University,
Cairo, Egypt
sherinmoussa@cis.asu.edu.eg

Z. T. Fayed
Department of Computer Science,
Faculty of Computer and Information Sciences,
Ain Shams University,
Cairo, Egypt
ztfayed@cis.asu.edu.eg

Received 2022-01-22; Revised 2022-05-27; Accepted 2022-05-30

Abstract: *Automated Driving Systems (ADS) and Advanced Driving Assistance Systems (ADAS) are widely investigated for developing safe and intelligent transportation systems. A common module in both systems is road objects monitoring, in which the semantic segmentation for road scene understanding has encountered lots of challenges. Due to the rapid evolution in technologies applied in vision-based systems in many fields, diverse techniques and algorithms have emerged to tackle such limitations, as invariant-illumination conditions, shadows, false positives, misdetections, weather conditions, real time processing and occlusions. A comparative study is conducted in this paper for vehicle detection and tracking methods applied on images and streams produced from monocular cameras and sensors in ADAS and ADS in terms of the aforementioned problems, the used dataset, along with the extracted features and the associated evaluation criteria. The study deduces the limitations of the current state-of-art techniques in such particular systems and highlights the main directions that can be adoted for future research and investigations.*

Keywords: *ADAS, ADS, Vehicle Detection, Vehicle Tracking, Autonomous Driving, Road Objects Monitoring, Road Objects Classification*

* Corresponding author: Sherin M. Moussa
Department of Computer Science, Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt
E-mail address: sherinmoussa@cis.asu.edu.eg

1. Introduction

Safe and Intelligent transportation systems are becoming a need nowadays for providing more safe and luxurious driving environment. New Autonomous Driving Systems (ADS) or Advanced Driving Assistance Systems (ADAS) are taking a lead to be developed by most automobile industries [1]. In the last few years, car industries equipped modern cars with ADAS functionalities [2] such as lane departure warnings, electronic stability controls, lane keep assists, anti-lock brake systems, …, etc. Incremental steps have then been taken towards more advanced assistance systems, like collision avoidance, automated parking and systems for on-road vehicles detection and tracking [3].

ADS have also been investigated widely, taking part in car industries to manufacture a car that can drive and take decisions or apply maneuvers on its own with the same efficiency as humans [4] or may be better in avoiding human possible mistake that could happen due to the driver's distraction or fatigue [5]. ADS and ADAS have many common modules that must be fulfilled in order for their functionalities to operate properly, such as road monitoring [6]. Monitoring roads requires accurate capturing and analysis of road images (captured by cameras) or readings (captured by sensors) or both [7].
To apply road analysis, road objects such as the road itself, lane lines, traffic signs, pedestrians, vehicles with different types or any obstacle that may appear in the middle of the road accidentally, must be specified at first. After that each road object can be either detected [8] or detected and tracked [9] to analyze its position and\or its motion with respect to the road scene captured by either monocular camera(s) or multiple sensors, as illustrated in Figure 1. For detecting and tracking road objects, both ADS and ADAS still face many challenges and problems that require lots of research and experiments before they can be applied in real life, such as weather conditions [10], [11], lighting conditions [12], occlusions [13], shadows [14] and real time processing [15]. These problems resulted either when using cameras or other sensors, even after applying different methods to handle images with low qualities [16].

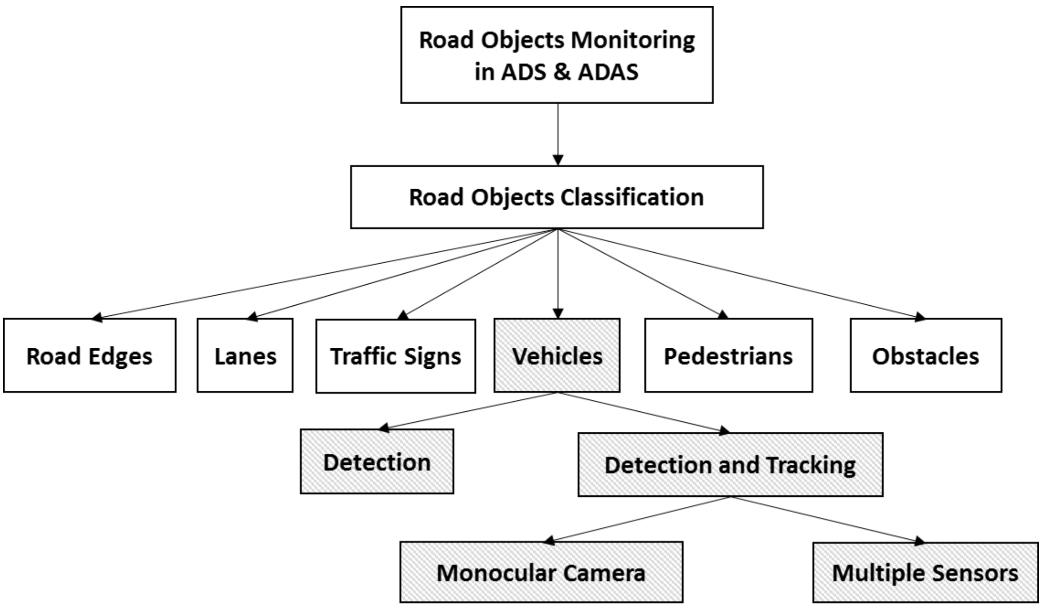


Figure. 1: Road Objects Classification for Road Objects Monitoring in ADS and ADAS

The main scope of this study focuses on vehicles as a single kind of road objects. It investigates the methods used for vehicles detection only, and vehicles detection and tracking in terms of the previously mentioned problems, while differentiating between the methods that applied tracking using monocular cameras and fusion between cameras and other sensors, such as Radars and Lidars [17].

Thus, the summarization of the main contributions of this study can be listed as follows:

- It provides a comprehensive study for the methods considered for vehicles detection, as well as both vehicles detection and tracking in ADAS and ADS, revealing their main strengths and limitations
- It deduces the common limitations encountered by both ADAS and ADS with respect to vehicles detection and tracking in the road objects monitoring phase.
- It highlights the main challenges that should be considered in future research directions for further investigation.

The next sections are organized as follows. Section two discusses the methods considered for vehicle detection only. Section three presents the methods investigated for both vehicle detection and tracking. Finally, section four provides an analytical discussion for the current state-of-the-art in regard to these systems and concludes the main directions that can be adopted in the future for further investigation and research.

2. Techniques for Vehicles Detection Only

Detecting objects is the initial step in recognizing or tracking that object and sometimes the detection methods can be used without modifying in tracking that object. Hence, this section discusses the methods that detect and recognize vehicles in road images or video streams in terms of the five considered problems: weather conditions, lighting conditions, occlusions, shadows and real time processing, as well as the used dataset for evaluation along with the evaluation criteria, the extracted features, and their limitations.

In [18], authors proposed a system that detects vehicles based on monocular and stereo visions. In the monocular based vision, they used the symmetry map for detection and tested their system on ARGO [19], which is a modified vehicle used for experiments only, that is used for vision algorithms testing applied with autonomous systems. However, they still need to make a better integration between software and hardware for better performance and speed. Authors in [20] proposed vehicle detection system that uses Gabor filter and SVM for feature extraction and classification, respectively. They applied their system on their own datasets, and they still need to conduct further comparisons using other types of features, Gabor filter parameters more optimized and perform a feature selection step or may be features fusion. In [21], authors have designed a system that detect vehicles composing of two steps. Multi-scale hypothesis generation was applied as the first step at which they tried to detect the rear view of the vehicle by applying an edge detector. While appearance-based hypothesis verification was applied at the second step, at which they used Haar Wavelet Transform and SVM. The system was applied on their dataset and succeeded to detect vehicles. However, when they tried their system on abnormal weather and lighting conditions the system's performance degraded and the hypothesis step was taking so much time.

In [22], a system for detecting multi-vehicles from images captured by a traveling vehicle was proposed. This system introduced an approach called evolutionary Gabor filter optimization (EGFO) which optimizes the parameter set of Gabor filter and they integrated it with genetic algorithms with a clustering approach applied incrementally for feature extraction. As for the classification stage they used the SVM algorithm. They tested their system on their own dataset, and they found that they still needed to reevaluate their system using other data sets and different filters types and schemes for selection by explicitly encoding selection in the chromosome. A vehicle detection system was proposed in [23] where they used gradient maxima localization and AdaBoost for feature extraction and classifications, respectively. They tested their system on their own dataset, and they found that very distant vehicles were not detected by their system and that they needed to increase their training set.

In [24], the Faster R-CNN performance was improved to detect vehicles more efficiently. Their method was applied on KITTI's dataset [25], but when it was compared to some state-of-the-art methods, it showed a more enhanced performance but only on the easy set of videos. Whereas when it was applied on the moderate and hard sets, which contained more challenging lighting conditions and occlusions, it was not good. Authors in [26] proposed a system that can detect vehicles in complex environment with diverse vehicle types at real-time scenarios. Two different HOG descriptors were generated to extract the features of vehicles and then SVM and AdaBoost classifier were used at the classification step. They used the GTI-UPM vehicle dataset [27] for training and real traffic video scenes while testing. They have achieved a good detection rate however occlusion was the main challenge for them that they could not overcome.

In [28], authors detected the position of the front vehicle and estimated the inter-vehicle distance for forward collision warning systems in urban/suburb scenes using canny edge detection, Hough transform and density-based spatial clustering of applications with noise techniques [29] for detecting vanishing points. The front vehicles were detected using the shadow feature and validated the hypothesis generation (HG) by applying HOG and SVM as feature descriptor and classifier, respectively over five test videos. However, some limitations were encountered, including light reflection that can be very strong, which may lower detection rate and increase FAR. In addition, real-world applicability was not examined, and thus, several factors such as weather conditions (as a sunny/rainy day), road environments (either urban/suburb roads, structured roads, or highways), lighting conditions (whether day/night times) and traffic flow were not considered in their test. Another two staged cascade detectors have been combined in [30] in a HybridNet system. They used CNN-based networks in their systems with adding extra transitional stage to map proposals on high resolution feature maps between the two stages. They tested their system on both KITTI and PASCAL VOC2007 [31] datasets. But yet, they had to enhance their performance in terms of accuracy and response time to meet up the requirements of a real time system.

In [32], authors proposed a method for detecting vehicles using color intensity segregation that was applied on video sequences with different resolutions from KITTI vision dataset. This method composed of two phases. The first phase can be considered as the feature extraction stage at which they extracted the ROI by applying multiple convolution operations and Hough Transform. At the second stage, they applied Gaussian filter followed by Intensity gradient generation and finally Hysteresis threshold operation to detect the vehicle. The performance of this method was proven to be robust to different illumination conditions, including rainy/sunny/cloudy circumstances, as well as to disorderly/messy backgrounds and shadows when applied on various video qualities. However, it was not used for vehicle tracking, and sufficient details must exist in the video for proper ROI extraction.

An adaptive system robust to lighting conditions was presented in [33] for vehicle detection that can work efficiently in real time scenarios. Three different lighting categories were considered: day, dusk and dark and depending on which category was found by the reconfigurable part a choice is done to use which path. Deep belief networks (DBN) and SVM were used for detection and classification, respectively in dark light scenarios, whereas HOG and SVM were used for detection and classification, respectively, in day and dusk scenarios. They trained and tested their system on GTI-UPM vehicle dataset for dusk and day scenarios and used the vehicle dataset for night times (SYSU) [34] for dark light scenarios. Authors in [35] have conducted a comparative study between five deep learning techniques for detecting objects on KITTI dataset for vehicle detection, which are RetinaNet, R-FCN, faster R-CNN, YOLOv3, and SSD. This study claimed that R-FCN and SSN achieved high accuracy levels. However, the best choices that can meet up with real time scenarios were SSD, YOLOv3 and RetinaNet. Nonetheless, they also faced a challenge balancing the accuracy and real-time performance.

Multi-vehicle detection and tracking in complex urban environment were investigated in [36], in which a combination of features was used for Harr and HOG feature extraction approaches. The cascaded structured AdaBoost classifier was considered to select some HOG features for classification to apply SVM classifier over INRIA Cars dataset. However, the proposed algorithm was not evaluated on a real ADAS and was not tried in extended scenarios. The fast target detection, occlusions and slow feature training were not handled. Besides, false positive was sometimes produced because of the existence of interfering targets. A vehicle detection method was presented in [37] based on multi-sensor fusion. A Lidar and a camera were used together for extracting and detecting vehicles in an image. A max-min elevation map was constructed, and some morphological and clustering operations were applied on it. After that the classification were done using YOLOv3 algorithm. They tested their system on KITTI's dataset. However, some drawbacks were found in their system, for instance if the vehicle object was too far the lidar could not scan it. Also, if a large object was too close a part of it could be detected as w new object which fails in the classification stage. The vehicle detection methods are comprehended in Table 1.

3. Techniques for Both Vehicles Detection and Tracking

In this section, we discuss the methods that detected and tracked vehicles in road images or video streams taken by monocular camera or multiple sensors in terms of the five problems in concern, the used dataset for evaluation, along with the evaluation criteria, extracted features and their limitations.

3.1 Using Monocular Cameras

In [38], authors introduced a system that detected and tracked vehicles and can operate in real time. For detection, they used Harris features and an edge following algorithm while for tracking they used Hidden Markov Model (HMM) and they tested their system on their own video dataset sequences. However, they found some limitations in their system such as more complex probability density functions for unsteady motion during several driving conditions, distant vehicles can't be detected all times and finally, the importance of assembling indication for a motion during time periods when the spectator vehicle travels at fairly high rapidity. Authors in [39] proposed a system which is able to work at real-time scenarios to detect and track vehicles using WaldBoost detector (WB) and modified the Tracking Learning Detection (TLD) in [40] for better tracking results. They also introduced the Toyota

Motor Europe (TME) dataset and tested their system on it, achieving acceptable tracking accuracy results in various lighting conditions and partial occlusions. However, the system had some limitations such as: target whose width was beyond 60 pixels caused accuracy problems, close targets take a lot of validation time and low performance on trucks.

In [41], authors studied algorithms to detect and track vehicles from their own video streams. In the detection stage they used HOG and an improved version of Adaboost for feature extraction and classifying, respectively. While for tracking, they used Harris detector, and normalized cross correlation classifier. Their proposed method can be used for real time systems however their algorithms are a bit slow they always assume that the ground is flat which does not coincide with reality and finally if they studied the vehicles behavior, they could improve their accuracies. Authors in [42], introduced a new dataset that is built using four high resolution monocular cameras which presents different viewpoints for a full surround analysis. They used their dataset to test vehicle detection, tracking and to understand the vehicle's behavior which can help in ADAS. They applied deformable part models pre-trained on KITTI's dataset with HOG descriptor for vehicle detection, and modified Markov decision process tracker for tracking.

In [43], a detection and tracking system was presented using a symmetric computation based on HOG along with Adaboost classifier for detection and a motion-based model using the adaptive Kalman filter for tracking. They tested their system on TME motor vehicle dataset [39] which was built using a camera that captured the front vehicle on urban roads. They still needed further investigation as false detection rates can be produced from the shadows casted on the road by signs on roadside, buildings, and trees. A vehicle tracking model was introduced in [44] based on Gaussian Mixture Probability Hypothesis Density filter (GMPHD). They used Haar-like features and Adaboost classifier for feature extraction and detection, respectively. They tested their system on TME motor vehicle dataset and the results proved that it can be applicable to systems that operate at real time. However, their system can work well only on optimal light conditions as non-optimal lighting conditions limits the number of features available to track.

A simple tracker was introduced in [45] using the IoU (Intersection over Union) technique. Their main idea was to track vehicles by detection them. The detection was done using Evolving Boxes detector (EB) and after that the IoU was applied to simply track vehicles. They applied their method on DETRAC vehicle tracking dataset [46], which consisted of traffic surveillance videos. Their method was so fast but still lower frame rates and more complicated occlusions can reduce the system accuracy.

In [47], a system for vehicle detection and tracking that can operate at real time environments was proposed that was called DeepTrackNet. They evaluated their system using Mobilenet Single Shot Multi Box Detector, Faster-RCNN and R-FCN for detection and for tacking they used three feature-based online trackers and deep regression network for testing. They tested their system on the Visual tracker benchmark TB-100 video sequences [48]. Their system had proven to be reliable to be used in real time systems, but they may need to use quantization, pruning and network optimization for reducing the model size.

Authors in [49] proposed a vehicle detection and tracking system that can work in real time scenarios. They combined HOG, color histogram features and color spatial features together to represent a vehicle and for classification they used the SVM classifier. After that they designed a pipeline for tracking the detected objects by means of active heat-maps. Their system can operate fast enough such that it can be used for tracking. They tested their system on GTI-UPM and KITTI datasets. However, shadow patterns in complex scenes may cause problems. The detection and tracking methods using monocular cameras are summarized in Table 2.

3.2. Using Multi-Sensor Fusion

A multi-modal system that detects, classifies, and tracks pedestrians and vehicles was proposed in [50]. The detection phase was done in the laser space using the Lidar whereas the classification was done in both laser and vision spaces. In the vision space the Haar-like features extracted in the detection phase were passed to the AdaBoost classifier and after that the classification decisions in both spaces were fused using a Bayesian sum decision rule. Finally, tracking was done using Kalman Filter. They tested their system on CALTECH dataset [51]. However, they still needed to increase the field of view by integrating more cameras in the system and investigate new methodologies and classifiers and test different classifier combinations for better performance

Authors in [52] presented a system that detects and tracks moving objects using the fusion of different sensors which are cameras, Lidar, and radar. They extracted the rear-view of the front vehicle using stream captured from the cameras installed at the front of the vehicle. They used the extended Kalman filter for tracking and applied it on their own video streams. For improving their system, they still need to make further investigation to recognize the lane marks and sidewalks. Moving objects such as trucks, bikes, pedestrians and vehicles were detected and tracked using a multi-sensor fusion system in [53]. Authors used three sensors to test and evaluate their system such as: camera, lidar and radar. Our main concern is the vision-based module in the system at which they proposed a sparse version of HOG descriptor for object detection and discrete AdaBoost for object classification. They also adapted the Markov Chain Monte Carlo approach for tracking all of the moving objects. They tested their system on four datasets of their own: two datasets from highways areas and other two datasets from urban areas.

In [54], a vehicle detection system was presented which was based on the integration between a camera and a laser scanner output. The laser scanner was used to detect hindrances and the camera was used to identify these extracted regions and track them. Tracking was done using Joint Probabilistic Data Association and Unscented Kalman Filter algorithms. They used a set of 28 sequences of their own to test their system; however further testing with new Multiple Object Tracking (MOT) techniques should be tested. A summary of the detection and tracking methods using the fusion of multiple sensors is presented in Table 3.

4. Discussion and Conclusion

Automated Driving Systems (ADS) and Advanced Driving Assistance Systems (ADAS) are the most recently emerging autonomous systems. For these systems to work, both share the road objects monitoring module. In order to monitor a road, all the road's objects must be identified and monitored separately using the adequate methods for detecting and\or tracking. In this paper, the vehicle object is adopted to be the main scope of the study. A comparative study is conducted to investigate the different vehicles detection and\or tracking methods that have been considered in scenes understanding, acquired by various sensors.
Major challenges have been deduced, affecting both detection and tracking mechanisms, such as weather conditions, lighting conditions, occlusions, shadows, and real time processing. Accordingly, a comprehensive analysis has been considered for the diverse methods of vehicles detection and tracking in ADAS and ADS with respect to the adopted challenges. The comprehensive study has considered three different perspectives:

- Comparison between the methods detecting vehicles only.
- Comparison between the methods detecting and tracking vehicles acquired by monocular camera(s).
- Comparison between the methods detecting and tracking vehicles acquired by multiple sensors.

In each perspective, the methods were compared in term of the used algorithm for detection and\or tracking, whether for a single or multiple vehicle objects, streaming type, the view of the capturing device, the extracted features, dataset used for evaluation, evaluation metrics, limitations, and most importantly, how each method handled the main challenges found. It was found that:

- Some methods focused on tackling one or more challenges. Yet, they failed to manipulate some of them, while ignored the other challenges.
- Other methods handled some of the challenges but ignored the rest.

- A single method claimed to succeed in handling all challenges on its own built database. However, it was not validated for different datasets, rather, it was a very customized testing environment.

From this conducted study, it can be concluded that:

- The challenges faced by ADAS and ADS differ in their difficulty levels and they can be arranged as follows: Level 1: Occlusions (The most difficult problem to tackle), Level 2: Shadows, Level 3: Weather conditions, Level 4: Real time processing, and finally Level 5: Lighting conditions (The simplest problem).
- Most research studies have been directed to tackle or enhance the results of solving the mentioned challenges from the simplest to the most difficult one, while many studies did not try considering the difficult challenges.

- Challenges analysis was conducted and illustrated in Figure 2. For each challenge, the number of succeeded tackling attempts, failed tackling attempts, and studies that did not consider the challenge were tracked. It was found that Lighting conditions challenge was tackled by 65.5% of the studies, while 11.5% failed to tackle it, and 23% did not consider it. Real time processing challenge was tackled by 42.3% of the studies, while 19.2% failed to tackle it, and 38.5% did not consider it. Weather conditions challenge was tackled by 30.8% of the studies, while 19.2% of the studies failed to tackle it, and 50% did not consider it. Shadows challenge was tackled by 23% of the studies, while 7.7% failed to tackle it, and 69.3% of studies did not consider it. Finally, Occlusions challenge was tackled by 19.2% of the studies, while 19.2% of the studies failed to tackle it, and 61.6% of studies did not consider it.
- Future studies should focus on handling the most difficult problems, such as occlusions, shadows, and weather conditions to build more accurate ADAS and ADS. This would elevate autonomous systems to the next level of automation; promoting safe and reliable self-driving cars.

Table 1. A summarized comparison between the vehicle detection techniques in ADAS and ADS.

Ref.	Main Problem	Algorithms used	Stream-ing Type (Images/ Videos)	Dataset	Field of view	Extracted Features	Evalua-tion Criteria	Main Problems Tackling						Limitations
								RT.	WC.	LC.	Sh.	Oc.	S\M	
[18]	Vehicle detection	Symmetry map	Video (12.5 fps)	Videos captures by ARGO	Front view	Bottom two corners of the vehicle	Distance in meter	-	-	-	-	-	S	A better integration between software and hardware
[20]	Feature extraction and classification for rear-view vehicle detection	• Gabor filters • SVM	Images	Property	Front view	Sub-divided parts of whole image	• False negatives • False positives • Average Accuracy	-	-	√	-	-	S	• Perform comparisons using other types of features. • Gabor filters parameters optimization. • Apply feature selection and/or fusion.
[21]	Pre-crash vehicle detection system	• Edge detector • Haar wavelet transform • SVM	Images	Property	Front view	Rear view vehicle	-	-	X	√	√	-	M	• The hypothesis generation step is slowing the system. • Abnormal weather and lighting conditions caused the system's performance to degrade.
[22]	Vehicle detection from images acquired by a moving vehicle	• Evolutionary Gabor Filter Optimization (EGFO) • Genetic algorithm • SVM	Images	Property	Front view	Edges	Detection error	-	-	√	-	-	M	• Reevaluate their system using different types of filters and different data sets. • Test diverse filter selection structures by selection encoding in the chromosome explicitly.
[23]	Real-time vision-based vehicle detection system	• Gradient maxima localization • AdaBoost	Video (10 fps)	Property	Front view	Shadows under-neath vehicles	• False detection • Non-detection	√	-	√	√	-	M	• Very distant vehicles were not detected. • Increasing the training set.
[24]	Vehicle detection	Faster R-CNN	Video	KITTI	Front view	Whole image	Average Precision (AP)	-	X	X	-	X	-	The performance is not good at challenging lighting conditions and occlusions
[26]	Vehicle Detection in complex environment and diverse types.	• HOG • AdaBoost classifier • SVM	Video (30 fps)	• GTI-UPM • Real traffic scene	Front view	Shadows under vehicles	Detection accuracy	√	-	√	-	X	M	Occlusions
[28]	Detection of front vehicles and estimation of the inter-vehicle distance for forward collision warning systems in urban/suburb scenes.	• Canny Edge Detection, • Hough Transform (HT) • DBSCAN • HOG • SVM	Video (20–25 fps)	Propriety	Front view	Shadow regions at the bottoms of front vehicles	• Detection Rate (DR) • False Alarm Rate	√	X	X	-	-	M	• Light reflection. • Weather and lighting conditions, traffic flow, and road environments were not considered.
[30]	Vehicle detection.	Two-stage cascade detector (CNN based) + IoU	Video	• KITTI • PASCALVO C2007	Front view	Whole image	AP	X	X	-	-	√	M	Enhance the performance in terms of speed and accuracy
[32]	Detection of on-road vehicles.	• Convolution operations • HT • Gaussian filter • Intensity gradient generation • Hysteresis threshold operation	Video > 38 fps for input resolutions less than [375*1242]	KITTI	Front view	Color intensity segregation	DR	√	√	√	√	-	S	• Tracking of detected vehicles. • Sufficient details must exist in the video for extracting the ROI properly.
[33]	Real-time vehicle and Pedestrian detection	• Deep belief networks (DBN) + SVM • HOG + SVM	Video (50 fps)	• GTI-UPM • SYSU night-time	Front view	Whole image	Accuracy	√	-	√	-	-	-	-
[36]	Detection and tracking of multi-vehicle targets in complex urban environment.	• Features of Harr • HOG • AdaBoost classifier features • SVM	Video Nearly 8 fps (137 ms per frame)	INRIA database	Front view	Complete rear view of front vehicles	• True Positive (TP) • False Positive (FP) • Non- DR	X	√	√	√	X	M	fast target detection, occlusions, slow feature training and false positive can be produced. No real-world applicability is examined. Camera and background vibration causing multifaceted interfering, the appearance/disappearance/shelter of vehicles were not considered.
[37]	Vehicle detection in real time using camera and Lidar.	• Max-Min • Elevation map + Morphological operations and Clustering • YOLO	Images	KITTI	Front view	Whole image	• Mean AP	√	√	√	-	√	M	Far away vehicles were not detected by Lidar Large near objects can be falsely detected as a separate object.

RT: Real-Time speed, WC: Weather Condition, LC: Lighting Conditions, Sh.: Shadows, Oc.: Occlusions, S\M: Single\Multiple objects, FPS: Frame Per Second.

Table 2. A summarized comparison between the vehicle detection and tracking techniques using Monocular Cameras in ADAS and ADS.

Ref.	Main Problem	Algorithms used		Streaming Type (Images/ Videos)	Dataset	Field of view	Extracted Features	Evaluation Criteria	Main Problems Tackling						Limitations
		Detection	Tracking						RT.	WC.	LC.	Sh.	Oc.	S\M	
[38]	Real-time in-car video analysis to detect and track vehicles	• Harris Feature • Edge-following algorithm	Hidden Markov model (HMM)	Video (50 fps)	Property	Front view	Corners and line segments	Confusion matrix	√	√	√	√	√	M	• More complex probability density functions for instable motion are required under severe driving conditions. • Hard to detect far vehicles. • It is inevitable to accumulate evidences of motion for a duration of time, especially at the time the observer vehicle is at a high rapidity.
[39]	Detection and Tracking Vehicles	WaldBoost detector	TLD tracker	Video	TME motoway	Front view	Rear view vehicle	• Precision • Recall	√	-	√	√	√	M	• Cause problems when the target is having a width exceeding 60 pixels • Closer targets take a lot of time to be validated. • On trucks systems face low performance
[41]	Vehicle detection and vehicle tracking in FCW systems.	• HOG • Improved Adaboost	• Harris detector • Normalized cross correlation	Video 24 fps	Property	Front view	Detection: shadows under the vehicles Tracking: Bottom line of the bounding box for recently tracked vehicles.	• TP • FP	X	X	√	-	-	M	• They always assume that the ground is flat which does not coincide with reality. • The used algorithms are a bit slow. • Studying vehicle behavior.
[42]	Detection, tracking, and vehicle behavior understand - 3D tracking via 4 cameras.	Deform-able part models with HOG	Modified Markov decision process (MDP) tracker	Video 12 fps	Property	Front Rear Right Left	Whole image	• Recall • AP • MOTA, • MOTP, • Frag, IDS • MT, ML	-	-	-	-	√	M	-
[43]	Front vehicle detection and tracking	• Symmetric computation based on HOG • Adaboost classifier	Adaptive Kalman filter	Video 1/25 second per frame	TME motorway	Front view	Whole image	• DR • FR • MR • FPS	X	-	-	X	-	M	Increased false detection rate resulting from the shadows caused on the road cast via roadside signs, buildings and trees.
[44]	Vehicle Tracking	Adaboost and Haar-like features	• Gaussian Mixture Probability • Hypothesis Density filter (GMPHD)	Video	TME motorway	Front view	Whole image	Mean Error	X	-	X	-	-	M	Non-optimal lighting conditions
[45]	Tracking Vehicles by detection	Evolving Boxes detector (EB)	IoU Tracker	Video 25 fps	DETRAC vehicle tracking dataset.	Traffic surveillance camera	Whole image	• Precision • Recall	-	-	-	-	X	M	The success rate is decreased due to the lower frame rates and Heavier occlusions.
[47]	Real-time detection, localization & tracking for autonomous vehicles	• Single Shot Multi Box Detector • Faster-RCNN • R-FCN	Feature based online trackers and deep regression network	Video 100 fps	• VTB TB-100 dataset • Imagenet dataset	Front view	Whole image	• Overall accuracy. • Failure rate	√	-	-	-	-	M	Large model size can be reduced by network optimization, pruning and quantization
[49]	Vehicle detection-and-tracking	• HOG • Color spatial features & olor histogram features • SVM	A pipeline includes building active heat-maps	Video 25 fps	• GTI-UPM • KITTI	Front view	Whole image	• Testing Accuracy	√	-	√	X	-	M	Shadow patterns in complex scenes

RT: Real-Time speed, WC: Weather Condition, LC: Lighting Conditions, Sh.: Shadows, Oc.: Occlusions, S\M: Single\Multiple objects, TME: Toyota Motor Europe Dataset MOTA: Multi-Object Tracking Accuracy, MOTP: Multi-Object Tracking Precision, MT: Mostly Tracked, ML: Mostly Lost, FR: False Detection Rate, MR: Missing Rate, FPS: Frame Per Second

Table 3. A summarized comparison between multi-sensor fusion techniques for vehicle detection and tracking in ADAS and ADS.

Ref.	Main Problem	Algorithms used	Streaming Type (Images/ Videos)	Dataset	Field of view	Extracted Features	Evaluation Criteria	Main Problems Tackling						Limitations
								RT.	WC.	LC.	Sh.	Oc.	S\M	
[50]	Detect, Track, and classify entities in semi-structured outdoor scenarios for intelligent vehicles using Camera and Lidar	• Haar Like Features • AdaBoost Classifier • Bayesian sum decision rule	Video (13 fps)	CALTECH dataset	Front view	Whole image	• Hit Rate (HR) • False Positives (FP)	-	√	√	-	-	M	• More cameras should be added in order to increase the view field. • Investigate new methodologies and classifiers and test different classifier combinations for better performance
[52]	Moving Object Detection and Tracking using Camera, Radar and Lidar	Extended Kalman Filter	Video	Property	Front view	Vehicle rear-view	• DR • FP per minute	-	√	√	-	-	M	Additional contextual information should be studied concerning the urban traffic environments, like the existence of sidewalks and lane-markings.
[53]	Detection and Tracking moving object using Camera, Radar and Lidar	• S- HOG • Discrete AdaBoost • MCMC	Video	Property	Front view	Vehicle rear-view	DR	√	√	√	-	-	M	-
[54]	Vehicle detection in a single lane using camera, Laser and GPS	• Unscented Kalman filter. • Joint Probabilistic Data Association	Video	Property	Front view	Whole image	• Positive detection. • Misdetection	-	√	√	-	X	S	Testing new tracking techniques

RT: Real-Time speed, WC: Weather Condition, LC: Lighting Conditions, Sh.: Shadows, Oc.: Occlusions

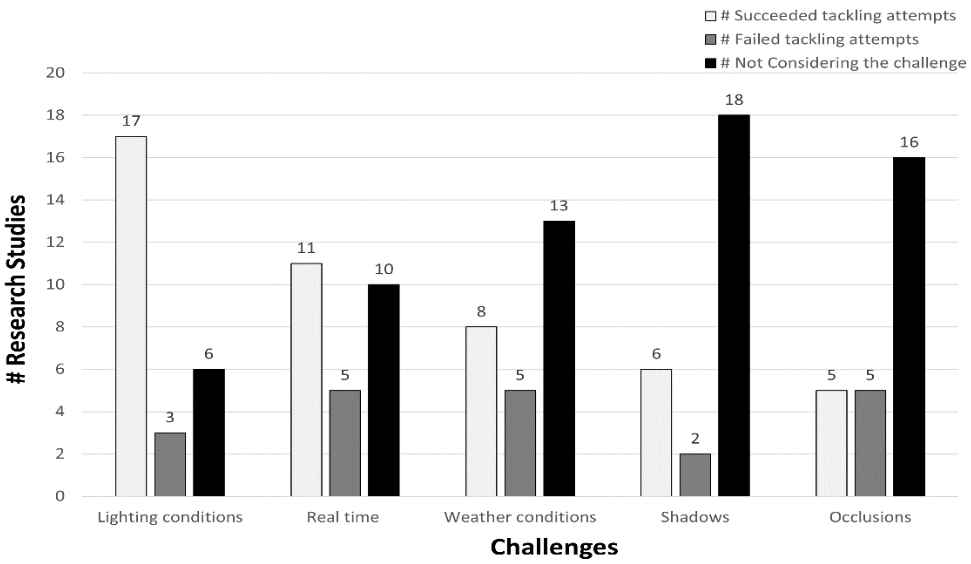


Figure. 2: Challenges Analysis

References

1. K. Bayoudh, F. Hamdaoui, and A. Mtibaa, "Transfer learning based hybrid 2D-3D CNN for traffic sign recognition and semantic road detection applied in advanced driver assistance systems," *Appl. Intell.*, vol. 51, no. 1, pp. 124–142, Jan. 2021, doi: 10.1007/s10489-020-01801-5.
2. S. Hamsini and MKathiresh, "Automotive Safety Systems," in *Automotive Embedded Systems: Key Technologies, Innovations, and Applications,* M. Kathiresh and R. Neelaveni, Eds. Cham: Springer International Publishing, 2021, pp. 1–18.
3. A. S. Tigadi, N. Changappa, S. Singhal, and S. Kulkarni, "Autonomous Vehicles: Present Technological Traits and Scope for Future Innovation," in *Automotive Embedded Systems: Key Technologies, Innovations, and Applications*, M. Kathiresh and R. Neelaveni, Eds. Cham: Springer International Publishing, 2021, pp. 115–143.
4. S. Solmaz, F. Holzinger, M.Mischinger, and J.Reckenzaun, "Novel Hybrid-Testing Paradigms for Automated Vehicle and ADAS Function Development," in *Towards Connected and Autonomous Vehicle Highways*, Springer, 2021, pp. 193–228.
5. J. Ma, H. Xie, Song, and H. Liu, "A Bayesian Driver Agent Model for Autonomous Vehicles System Based on Knowledge-Aware and Real-Time Data," *Sensors*, vol. 21, no. 2, p. 331, Jan. 2021, doi: 10.3390/s21020331.
6. K. Li, D. Yu, JLiu, J.Wang, "An End-to-End Multi-Task Learning Model for Drivable Road Detection via Edge Refinement and Geometric Deformation," *IEEE Trans. Intell. Transp. Syst.*, pp. 1–11, 2021, doi: 10.1109/TITS.2021.3084058.
7. D. R. Niranjan, VinyKarthik, and Mohana, "Deep Learning based Object Detection Model for Autonomous Driving Research using CARLA Simulator," in *2021 2nd International Conference on*

Smart Electronics and Communication (ICOSEC), Oct. 2021, pp. 1251–1258, doi: 10.1109/ICOSEC51865.2021.9591747.

8. D. Feng, A. Harakeh, S. L. Waslander, and K. Dietmayer, "A Review and Comparative Study on Probabilistic Object Detection in Autonomous Driving," *IEEE Trans. Intell. Transp. Syst.*, pp. 1–20, 2021, doi: 10.1109/TITS.2021.3096854.
9. H. Chiu, A.Prioletti, "Probabilistic 3d multi-object tracking for autonomous driving," *arXiv Prepr.*, p. arXiv:2001.05673, 2020.
10. G. Volk, S. Muller, A. von Bernuth, D. HospachBringmann, "Towards Robust CNN-based Object Detection through Augmentation with Synthetic Rain Variations," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, Oct. 2019, pp. 285–292, doi: 10.1109/ITSC.2019.8917269.
11. M. Masmoudi, HGhazzai, and Y. Massoud, "Object Detection Learning Techniques for Autonomous Vehicle Applications," in *2019 IEEE International Conference on Vehicular Electronics and Safety (ICVES)*, Sep. 2019, pp. 1–5, doi: 10.1109/ICVES.2019.8906437.
12. M. Haris and A. Glowacz, "Road Object Detection: A Comparative Study of Deep Learning-Based Algorithms," *Electronics*, vol. 10, no. 16, p. 1932, Aug. 2021, doi: 10.3390/electronics10161932.
13. S. Gilroy, E. Jones, and M. Glavin, "Overcoming Occlusion in the Automotive Environment—A Review," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 1, pp. 23–35, Jan. 2021, doi: 10.1109/TITS.2019.2956813.
14. P. M. Shakeel, EAbdulhay, "Automated multimodal background detection and shadow removal process using robust principal fuzzy gradient partial equation methods in intelligent transportation systems," *Int. J. Heavy Veh. Syst.*, vol. 25, no. 3–4, p. 271, 2018, doi: 10.1504/IJHVS.2018.094824.
15. H. Rashed, M. Ramzy,andV. Vaquero, "FuseMODNet: Real-Time Camera and LiDAR Based Moving Object Detection for Robust Low-Light Autonomous Driving," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, Oct. 2019, pp. 2393–2402, doi: 10.1109/ICCVW.2019.00293.
16. E. El-Ashmony, MDosuky, andElmougy, "Classification Of Low Quality Images Using Convolutional Neural Network and Deep Belief Network," *Int. J. Intell. Comput. Inf. Sci.*, vol. 16, no. 4, pp. 19–28, Oct. 2016, doi: 10.21608/ijicis.2016.19822.
17. K. Jayan and BMuruganantham, "Advanced Driver Assistance System Technologies and Its Challenges Toward the Development of Autonomous Vehicle," in *Automotive Embedded Systems: Key Technologies, Innovations, and Applications*, Springer, 2021, pp. 55–72.
18. A. Bensrhair, M. Bertozzi, PMiche, G. Toulminet, "A cooperative approach to vision-based vehicle detection," in *ITSC 2001. 2001 IEEE Intelligent Transportation Systems. Proceedings (Cat. No.01TH8585)*, 2001, pp. 207–212, doi: 10.1109/ITSC.2001.948657.
19. A. Broggi, M. Bertozzi, A. Fascioli, and G. Conte, *Automatic Vehicle Guidance: The Experience of the ARGO Autonomous Vehicle*. WORLD SCIENTIFIC, 1999.
20. Zehang Sun, G. Bebis, and R. Miller, "On-road vehicle detection using Gabor filters and support vector machines," in *2002 14th International Conference on Digital Signal Processing Proceedings. DSP 2002 (Cat. No.02TH8628)*, vol. 2, pp. 1019–1022, doi: 10.1109/ICDSP.2002.1028263.

21. Zehang Sun, R. Miller, G. Bebis, and D. DiMeo, “A real-time precrash vehicle detection system,” in *Sixth IEEE Workshop on Applications of Computer Vision, 2002. (WACV 2002). Proceedings.*, 2002, vol. 2002-Janua, pp. 171–176, doi: 10.1109/ACV.2002.1182177.
22. Z. Sun, G. Bebis, and R. Miller, “On-road vehicle detection using evolutionary gabor filter optimization,” *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 125–137, 2005, doi: 10.1109/TITS.2005.848363.
23. A. Khammari, F Nashashibi, C Laurgeau, “Vehicle detection combining gradient analysis and adaboost classification,” in *Proceedings. 2005 IEEE Intelligent Transportation Systems, 2005.*, 2005, vol. 2005, pp. 1084–1089, doi: 10.1109/ITSC.2005.1520202.
24. Q. Fan, L. Brown, and J. Smith, “A closer look at Faster R-CNN for vehicle detection,” in *2016 IEEE Intelligent Vehicles Symposium (IV)*, Jun. 2016, vol. 2016-Augus, no. Iv, pp. 124–129, doi: 10.1109/IVS.2016.7535375.
25. A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for autonomous driving? The KITTI vision benchmark suite,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2012, pp. 3354–3361, doi: 10.1109/CVPR.2012.6248074.
26. G. Yan, M. Yu, Y. Yu, and L. Fan, “Real-time vehicle detection using histograms of oriented gradients and AdaBoost classification,” *Optik (Stuttg).*, vol. 127, no. 19, pp. 7941–7951, Oct. 2016, doi: 10.1016/j.ijleo.2016.05.092.
27. J. Arróspide, L. Salgado, and M. Nieto, “Video analysis-based vehicle detection and tracking using an MCMC sampling framework,” *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 1, p. 2, Dec. 2012, doi: 10.1186/1687-6180-2012-2.
28. D. Huang, C.-H. Chen, T. Chen, , “Vehicle detection and inter-vehicle distance estimation using single-lens video camera on urban/suburb roads,” *J. Vis. Commun. Image Represent.*, vol. 46, pp. 250–259, Jul. 2017, doi: 10.1016/j.jvcir.2017.04.006.
29. D. Birant and A. Kut, “ST-DBSCAN: An algorithm for clustering spatial–temporal data,” *Data Knowl. Eng.*, vol. 60, no. 1, pp. 208–221, Jan. 2007, doi: 10.1016/j.datak.2006.01.013.
30. X. Dai, “HybridNet: A fast vehicle detection system for autonomous driving,” *Signal Process. Image Commun.*, vol. 70, pp. 79–88, Feb. 2019, doi: 10.1016/j.image.2018.09.002.
31. M. Everingham, L. Van Gool, and A. Zisserman, “The Pascal Visual Object Classes (VOC) Challenge,” *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, Jun. 2010, doi: 10.1007/s11263-009-0275-4.
32. M. S. Sravan, and B. J. Kailath, “Fast and accurate on-road vehicle detection based on color intensity segregation,” *Procedia Comput. Sci.*, vol. 133, pp. 594–603, 2018, doi: 10.1016/j.procs.2018.07.090.
33. M. Hemmati, M. Biglari-Abhari, and S. Niar, “Adaptive Vehicle Detection for Real-time Autonomous Driving System,” in *2019 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, Mar. 2019, pp. 1034–1039, doi: 10.23919/DATE.2019.8714818.
34. L. Chen, X. Hu, T. Xu, H. Kuang, and Q. Li, “Turn Signal Detection During Nighttime by CNN Detector and Perceptual Hashing Tracking,” *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 12, pp. 3303–3314, Dec. 2017, doi: 10.1109/TITS.2017.2683641.
35. H. Wang, Y. Yu, , and Q. Liu, “A Comparative Study of State-of-the-Art Deep Learning Algorithms for Vehicle Detection,” *IEEE Intell. Transp. Syst. Mag.*, vol. 11, no. 2, pp. 82–95, 2019, doi: 10.1109/MITS.2019.2903518.

36. Y. Wei, Q. Tian, J. Guo, W. Huang, and J. Cao, "Multi-vehicle detection algorithm through combining Harr and HOG features," *Math. Comput. Simul.*, vol. 155, pp. 130–145, Jan. 2019, doi: 10.1016/j.matcom.2017.12.011.
37. H. Wang, X. Lou, Y. Cai, Y. Li, and L. Chen, "Real-Time Vehicle Detection Algorithm Based on Vision and Lidar Point Cloud Fusion," *J. Sensors*, vol. 2019, pp. 1–9, Apr. 2019, doi: 10.1155/2019/8473980.
38. A. Jazayeri, H. Cai, J. Y. Zheng, and M. Tuceryan, "Vehicle Detection and Tracking in Car Video Based on Motion Model," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 583–595, Jun. 2011, doi: 10.1109/TITS.2011.2113340.
39. C. Caraffi, T. Vojir, J. Trefny, "A system for real-time detection and tracking of vehicles from a single car-mounted camera," in *2012 15th International IEEE Conference on Intelligent Transportation Systems*, Sep. 2012, pp. 975–982, doi: 10.1109/ITSC.2012.6338748.
40. Z. Kalal, J. Matas, and K. Mikolajczyk, "P-N learning: Bootstrapping binary classifiers by structural constraints," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2010, pp. 49–56, doi: 10.1109/CVPR.2010.5540231.
41. Zhipeng Di and DongzhiHe, "Forward Collision Warning system based on vehicle detection and tracking," in *2016 International Conference on Optoelectronics and Image Processing (ICOIP)*, Jun. 2016, pp. 10–14, doi: 10.1109/OPTIP.2016.7528490.
42. J. V Dueholm, M. S. Kristoffersen, R. K. Satzoda,"Multi-perspective vehicle detection and tracking: Challenges, dataset, and metrics," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, Nov. 2016, pp. 959–964, doi: 10.1109/ITSC.2016.7795671.
43. X. Zhang, H. Gao, C. Xue, J. Zhao, and Y. Liu, "Real-time vehicle detection and tracking using improved histogram of gradient features and Kalman filters," *Int. J. Adv. Robot. Syst.*, vol. 15, no. 1, p. 172988141774994, Jan. 2018, doi: 10.1177/1729881417749949.
44. F. García, and A. Broggi, "PHD filter for vehicle tracking based on a monocular camera," *Expert Syst. Appl.*, vol. 91, pp. 472–479, Jan. 2018, doi: 10.1016/j.eswa.2017.09.018.
45. E. Bochinski, V. Eiselein, and T. Sikora, "High-Speed tracking-by-detection without using image information," in *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Aug. 2017, no. August, pp. 1–6, doi: 10.1109/AVSS.2017.8078516.
46. L. Wen *et al.*, "UA-DETRAC: A new benchmark and protocol for multi-object detection and tracking," *Comput. Vis. Image Underst.*, vol. 193, p. 102907, Apr. 2020, doi: 10.1016/j.cviu.2020.102907.
47. D. K. Amara, Karthika, "DeepTrackNet: Camera Based End to End Deep Learning Framework for Real Time Detection, Localization and Tracking for Autonomous Vehicles," in *International Conference on Intelligent Computing, Information and Control Systems*, 2020, pp. 299–307.
48. "VTB TB-100 dataset." http://cvlab.hanyang.ac.kr/tracker_benchmark/datasets.html (accessed Dec. 20, 2021).
49. F. Wael, "A Comprehensive Vehicle-Detection-and-Tracking Technique for Autonomous Driving," *Int. J. Comput. Digit. Syst.*, vol. 9, no. 4, pp. 567–580, Jul. 2020, doi: 10.12785/ijcds/090405.
50. C. Premebida, and P. Peixoto, "A Lidar and Vision-based Approach for Pedestrian and Vehicle Detection and Tracking," in *2007 IEEE Intelligent Transportation Systems Conference*, Sep. 2007, pp. 1044–1049, doi: 10.1109/ITSC.2007.4357637.

51. Li Fei-Fei, R. Fergus, "One-shot learning of object categories," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 594–611, Apr. 2006, doi: 10.1109/TPAMI.2006.79.
52. H. Cho, YW. Seo, and RR. Rajkumar, "A multi-sensor fusion system for moving object detection and tracking in urban driving environments," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, May 2014, pp. 1836–1843, doi: 10.1109/ICRA.2014.6907100.
53. R.Chavez-Garcia and O. Acard, "Multiple Sensor Fusion and Classification for Moving Object Detection and Tracking," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 2, pp. 525–534, Feb. 2016, doi: 10.1109/TITS.2015.2479925.
54. F. Garcia, D. Martin, and J. M. Armingol, "Sensor Fusion Methodology for Vehicle Detection," *IEEE Intell. Transp. Syst. Mag.*, vol. 9, no. 1, pp. 123–133, 2017, doi: 10.1109/MITS.2016.2620398.