Journal of the Society of Automotive Engineers Malaysia
Volume 1, Issue 3, pp 198-207, September 2017
e-ISSN 2550-2239

Single Camera Object Detection for Self-Driving Vehicle: A Review

S. Herman[1] and K. Ismail[*1]

[1]Faculty of Mechanical Engineering, Universiti Teknologi Malaysia (UTM) Johor Bahru, 81310 Johor, Malaysia

[*]*Corresponding author: kamarulafizam@utm.my*

REVIEW Open Access

Article History:

Received
22 Apr 2017

Received in revised form
30 June 2017

Accepted
28 Aug 2017

Available online
1 Sep 2017

Abstract – *The development of technologies for autonomous vehicle (AV) have seen rapid achievement in the recent years. Commercial carmakers are actively embedding this system in their production and are undergoing tremendous testing in the real world traffic environment. It is one of today's most challenging topics in the intelligent transportation system (ITS) field in term of reliability as well as accelerating the world's transition to a sustainable future. The utilization of current sensor technology however indicates some drawbacks where the complexity is high and the cost is extremely huge. This paper reviews the recent sensor technologies and their contributions in becoming part of the autonomous self-driving vehicle system. The ultimate focus is toward reducing the sensor count to just a single camera based on the single modality model. The capability of the sensor to detect and recognize on-the-road obstacles such as overtaking vehicle, pedestrians, signboards, bicycle, road lane marker and road curvature will be discussed. Different feature extraction approach will be reviewed further with the selection of the recent Artificial Intelligent (AI) methods that are being implemented. At the end of this review, the optimal techniques of processing information from single camera system will be discussed and summarized.*

Keywords: Self-driving vehicle, autonomous vehicle

Journal homepage: www.journal.saemalaysia.org.my

1.0 INTRODUCTION

International Society of Automotive Engineering (SAE) define automated self-driving vehicle in 5 different levels. A vehicle is considered fully autonomous or driverless when it can control the operation of steering and motion (acceleration and deceleration), fully rely on the system in monitoring driving environment and fall-back performance of the driving task (Peng, 2016). A brief flurry about self-driving vehicles makes some of the carmakers to actively compete in their production towards fully autonomous vehicle. According to Ford's media on 16th Aug 2016, the intention on having a fully autonomous SAE level 4 capable vehicle was announced to be commercialized in 2021. On the other hand, Audi USA's press release on 5th January

2017 had also announced on the expansion into AI for the long-time partners in bringing a fully automated driving to the roads starting in 2020.

What makes autonomous technology significantly different from conventional automotive technology is the ability to make judgements about the external environments of the vehicle on behalf of the driver (Sanchez, 2015). For technologies at the lower end of the automation spectrum or known as Advanced Driver-Assistance System (ADAS), driver still retains some control of the vehicle at all times where he is ultimately responsible for interpreting the environment and determining whether the autonomous functions in the vehicle such as driver-warning systems and adaptive cruise control should be used (Sanchez, 2015). However, for vehicle with a high degree of autonomy, standards and testing are necessary to cover all aspects of the situation in which it will operate safely (Sanchez, 2015).

Jiang et al. (2015) in a study stated that one of the main hitch of a self-driving vehicle is the cost in which Google had taken about $200,000 in building its 2014 self-driving vehicle. There are various sensors technologies in Google's driverless vehicle including sonar device, stereo camera, laser, radar, and also Velodyne 64-beam laser (LiDAR – light detection and ranging) where the usage of LiDAR itself is extremely expensive. Different from most of automobile manufacturing companies, Tesla's business model of its self-driving car owns the entire supply chain from manufacturing to distribution. This strategy is driven by the ultimate goal of lowering manufacturing costs and costs of goods sold, thereby assuring business' sustainability (Bilbeisi & Kesse, 2017). Without embedding an expensive LiDAR sensor, Tesla autopilot combines a forward looking camera, radar, and 360 degree sonar sensors with real time traffic updates in its model S which has also being recognized as a 5-star rating in all categories of the National Highway Traffic Safety Administration (NHTSA) crash test.

Continuous improvement is being made in supporting the use of cheaper sensors in dealing with the reception of an expensive autonomous vehicles production. Carmakers are now actively collaborate with scientists and researchers and trying to figure out the most optimal sensors technologies that will be used in AI expansion. However, all critical components are required to meet high manufacturing, installation, repair, testing, and maintenance standards, because the failure of the system could be fatal to both vehicle occupants and other road users, which probably make it relatively expensive. Unlike the automobile industry, seldom consumer want to make a purchase on certain vehicle just to obtain a new technology updates (Litman, 2017).

2.0 OBJECT DETECTION TECHNOLOGY

2.1 Camera as the Object Detection Sensor Technology

What makes the object detection a crucial task in autonomous driving nowadays is finding the solutions to the combination of the perception sensors, where image based object detection is still consider irreplaceable (Wu et al., 2016). According to Woodside Capital Partners (2016), the most intuitive sensors that are similar to the function of human vision are the camera-based system where it is believed to play an important part in either AV or ADAS. Unlike LiDAR or Radar based systems, the highest resolution with spatial information and minute details can only be captured by image sensors in camera systems (Woodside Capital Partners, 2016). Besides its cheaper price, it is also known as much evolving technology where most of its data are usable as compared to radar and LiDAR (Woodside Capital Partners, 2016).

There are three types of camera that are mostly used in the development of the AV which are known as single camera (monocular vision), dual camera (stereo vision) and specialized camera (built in camera). All of these cameras are essential in providing some of the ADAS application such as, forward collision warning, pedestrian detection, traffic signal detection, lane departure warning, headway monitoring, blind spot detection, parking assist system and intelligent headlight control. According to Woodside Capital Partners (2016), for blind spot detection, cameras are best mounted near the side view mirror which its can provide the extended view on both sides. Moreover, at least six cameras are needed to provide 360 views on the AV where two of it will be placed at each side and one for front and back views (Woodside Capital Partners, 2016).

Different camera specification gives different results in the lists of AV's application. Jeon et al. (2016) used a single stereo camera of 640x360 pixel resolutions named VSTC-V260 with 24fps in the speed average of 40kmph in testing his pedestrian detection, traffic light and traffic sign recognition. According to McBride et al. (2006), the object detection can also perform well in the ease of a parking lot scene just by using a low-cost single-stereo camera. Haloi and Jayagopi (2015) used a wide angle camera sensor mounted in the vehicle's roof in capturing broad road environments which can give up to 440x680 images size at the speed of 45km/h. On the other hand, Miao et al. (2012) used a camera with 320x240 resolutions of 100 fps from Lumenera Corporation for his lane detection system. For a best vehicle behaviour prediction, Wang et al. (2016) used a Canon camera with high definition quality that gives up to 1920x 1080 pixels with 30fps. All of these specifications are highly important in determining the results of the detection and recognition. Without relying on the expensive sensors, scientist and researchers are already told us on what a single camera is capable to do in the development of AV.

2.2 Camera Application in AV

In order to sense and monitor the behaviour of its external environment and to take action where required, an autonomous vehicle requires some range of technologies (Sanchez, 2015). Key technologies that include functions of LiDAR and camera are summarized in Table 1.

Table 1: Components of autonomous vehicle technology data (Forrest & Konca 2007)

Sensors	Data Processing	Mechanical Control Systems	Communication	Infrastructure
3D camera	Decision making	Driving wheel control	Vehicle to vehicle communication	Physical infrastructure
Radar (LiDAR)	User interface	Throttle control	GPS, digital maps	Optimisation

Referring to Table 1, combination of sensors are required to make sense of the external environment, gathering information and allowing vehicles to accurately localize its position (Sanchez, 2015). Data processing will extract relevant information as source of initial decision and manage the interaction between computer and driver, where 3D camera in this case is better in decision making than LiDAR. Mechanical control systems that exist in 3D camera can control the vehicle's driving wheel in order to perform the desired action such as braking, accelerating and turning. On the other hand, communication and networking in 3D camera is likely to have vehicle to vehicle interaction than pinpointing a location (Sanchez, 2015).

Nowadays, audio visual has become an example on how fast the ICT complement existing vehicle technologies in order to offer a new better function in personal mobility transport (Sanchez, 2015). In conjunction to a better enhancement on video camera sensor, the study by Bojarski et al. (2016) suggested the end to end learning of a convolutional neural network (CNN) in mapping raw pixels from a single front-facing camera directly to steering commands. They found that the recorded steering wheel angle applied by human driver with the single images sampled from the video can generate to the desired steering command after it was trained with CNN embedded system called, DAVE-2 as shown in Figure 1.

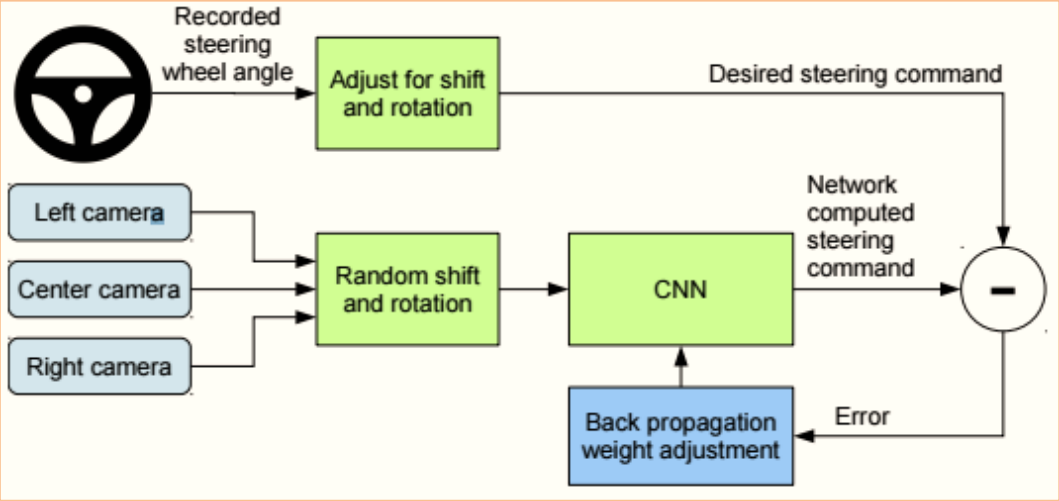

Figure 1: DAVE-2 training system

Likewise what Bojarski et al. (2016) mentioned in their study, Wu et al. (2016) give a fully CNN object detection that simultaneously fulfil all the AV's safety definition with additional real time inference speed control on a sudden vehicle (SqueezeDet). As similar to Wu et al. (2016), LeCun et al. (2005) developed off-road obstacle through a single trained function by mapping raw color images from two forward-pointing cameras mounted on the robot to a set of possible steering angles. The revolutionized CNN used in the development of sensor technology are moving closely towards the performance in a real time. Romera et al. (2016) proposed full image segmentation in unifying and simplifying most of the detection tasks as one of the approach required in AV as shown in Figure 2.

Different from the traditional detector approach that separates different detector in every possible obstacles, unification and simplification of this new approach uses the segmented image output in detecting all of the obstacles close to the real time. According to Tang (2013), extensive technology advances in camera-based vision systems have released possible results in terms of acceptable accuracy, with the power of computing (both software and hardware), and image processing algorithms. Due to limited view of a single camera, there might be some problem that are not adequately resolved although the traffic performance measurements is quiet encouraging (on average 70-90%) but when combining multiple camera views to obtain a joint tracking, it is typically much better than single-camera tracking (Tang, 2013).

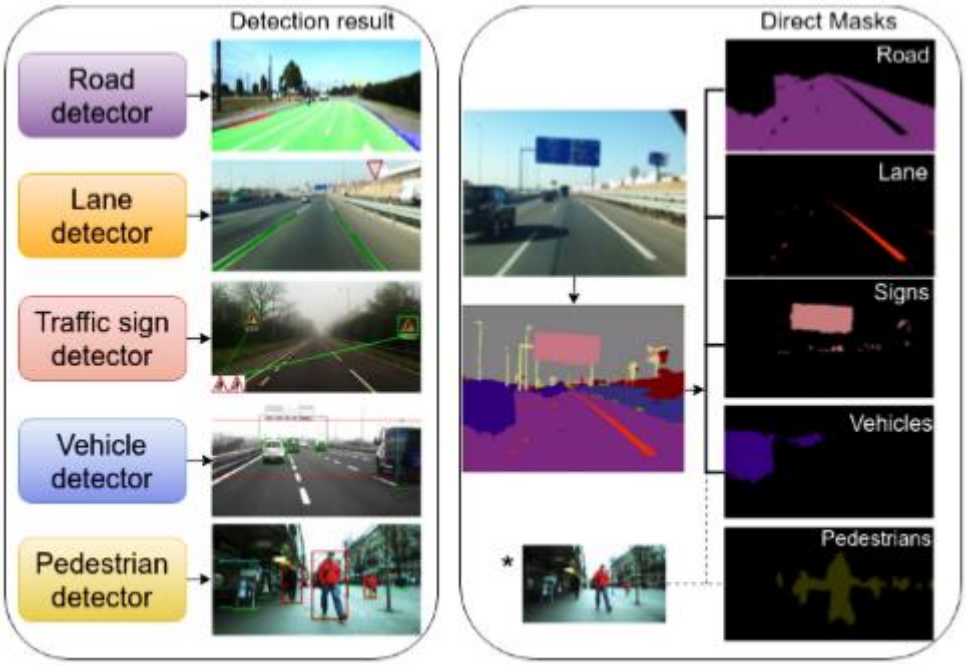

Figure 2: Traditional approach versus proposed approach
*A different example image was added for pedestrians (due to lack in the first input image) but these are also segmented by the same system

2.3 Feature Extraction

A well-organized object recognition technique is very helpful in the ways of possessing a good algorithm (Kaur & Marwaha, 2017). In the application of image processing, multi-object detection is considered very important. According to Aly (2014), the first successful step in the system was generated the Inverse Perspective Mapping (IPM) of the roads image before it was filtered by two-dimensional Gaussian kernel. Using line detection and a new Random Sample Consensus (RANSAC) spline fitting technique, the refinement of spline fitting can be achieved efficiently. On the other hand, Kaur and Marwaha (2017) stated that thresholding based approach is one of the vital approach in the image segmentation.

In the end to end learning, Bojarski (2016) used 9 layers of network that consist of normalize layer, 3 fully connected layers, and 5 convolutional layers. The image normalization in normalize layer is hard-coded and cannot be adjusted which allows alteration of the network architecture. Feature extraction was chosen empirically in the designed convolutional layers through series of experiments and various layers configuration. Fully connected layers were designed to control the steering that leads to the output control value. Similar to Bojarski (2016), LeCun (2005) used 6 feature maps where the input was a single left/right paired of unprocessed low-resolution images. Massive amounts of data need to be trained to emulate the behavior of a human driver in avoiding the upcoming obstacles which at the same time allowed the network to learn on the low-level and high-level features.

2.4 Underlying Artificial Intelligence (AI) Algorithm

In a goal of creating safer self-driving vehicle, carmakers are pouring billions of dollars into AI research where the end to end approach have successfully learn what driver did in various situation (Baik and Greenblatt, 2016). Chen et al. (2015) had built a state of art model from

deep convolutional Neural Network (ConvNet) framework. Trained data sets which focus on 3 lanes configurations were gathered from the open racing car simulator (TORCS) video game, in which 13 affordance indicators were collected as shown in Figure 3. In order to drive a host, a controller need to computes the driving commands that will be sent back to TORCS based on the current speed and indicators present.

always:
1) angle: angle between the car's heading and the tangent of the road
"in lane system", when driving in the lane:
2) toMarking_LL: distance to the left lane marking of the left lane
3) toMarking_ML: distance to the left lane marking of the current lane
4) toMarking_MR: distance to the right lane marking of the current lane
5) toMarking_RR: distance to the right lane marking of the right lane
6) dist_LL: distance to the preceding car in the left lane
7) dist_MM: distance to the preceding car in the current lane
8) dist_RR: distance to the preceding car in the right lane
"on marking system", when driving on the lane marking:
9) toMarking_L: distance to the left lane marking
10) toMarking_M: distance to the central lane marking
11) toMarking_R: distance to the right lane marking
12) dist_L: distance to the preceding car in the left lane
13) dist_R: distance to the preceding car in the right lane

Figure 3: Lists of affordance indicators

According to Fan et al. (2016), performance of Faster R-CNN on vehicle detection can be improves through some appropriate parameter tuning and algorithmic modification. Comprehensive experiments has been done on both training-test scale size, number of proposals, localization versus recognition and iterative training in order to tune the most suitable approach of Faster R-CNN by using a KITTI benchmark dataset. Different from both Chen et al. (2015) and Fan et al. (2016), Miao et al. (2012) has developed a real time monocular vision system in which the design was taken from Open Source Computer Vision Library (OpenCV) using K-means cluster algorithm. It is said to have the ability to locate the actual position of the road in most inexpensive computational cost.

3.0 DISCUSSION

On-the road obstacle detection is the primary condition that has to be achieved optimally before autonomous driving could occur. Lane detection and lane marking are very crucial because this is the only system that keeps the vehicle on the road. Image processing is the key technology for this as the vision system is the only sensor that can look and find this marker. Somehow image processing is subjected to many external factors as simple as ambient light intensity. The method has to be robust and adaptive. For a single stereo camera VSTC-V260 that had been used by Jeon et al. (2016), 95.4% of traffic sign had successfully been recognized in 30m recognition distance, where the changes of traffic light colors had been detected fairly in 30ms processing time for 35m of maximum distance. Because pedestrians is a complex object/subject to determined, Jeon et al. (2016) stated that his pedestrians detection took longer than traffic light and traffic sign recognition where its took up to 180ms processing time for a maximum 40m in distance.

According to McBride et al. (2006), the weakness from a stereo image of a low-cost camera can be settled down by matching its geometrical models. In his detection over a parking lot scene, the detection rate recorded was 81.5% where it successfully detected 106 vehicles out of 130. On the other hand, Haloi and Jayagopi (2015) in their research recorded 94.25% of the right boundaries detected over Indian road, and their algorithm had also hit over 95% accuracy in both KITTI and Caltech datasets. By using the Lm085 camera, Miao et al. (2012) in his research had successfully located all the road lanes and boundaries where it was applied in various road scenes includes both marked and unmarked roads as shown in Figure 4.

Figure 4: Lane and boundary detection results on straight road, curve road, and unstructured road in various illumination variations

For a vehicle behaviour prediction using HD Canon camera, Wang et al. (2016) stated the maximum pixels error achieved was 0.585 which clearly showed that his system was reliable, efficient and the most important was much cheaper. Move to the feature extraction of camera based application, a review by Kaur and Marwaha (2017), stated that when no remarkable changes on the grey levels between foreground and background, threshold determination image cannot produce an efficient results. On the other hand, the IPM approach used by Aly (2014) showed 96.34% correct detection for 2-lane mode and 90.89% correct detection for all-lanes mode. This impressive results show the effectiveness of IPM approach used in detecting lanes roads in a vary condition. However, Chen et al. (2015) recorded a strong response on the detection of nearby car and lane marking but its false positive was much higher than the testing sample on the DPM baseline.

The use of artificial intelligent and machine learning have shown and proved that it is relevant in achieving level 5 autonomous driving. Many algorithms were tested and the result seems to be very promising especially when single camera is used for obstacle detection.

Powerful intelligence is very essential because of the data dimension is limited only to the captured image. On a simulation conducted by Bojarski et al. (2016), the percentage of the time the network could drive was 90% for 10 interventions recorded in 600 seconds, while on-road test reached approximately 98% of the autonomous behavior in 10 miles. LeCun et al. (2005) reported a several reasons on the high error recorded in both of his training and testing results (25.1%, 35.8%) such for a given image pair, there may be numbers of legitimate steering angles where the commands may be valid on the obstacles. From the underlying of AI used in the perception of detection, 1500 and 1800 test scale models used in Faster R-CNN by (Fan et al., 2016) has been designated as their benchmark.

4.0 CONCLUSION

The establishments of sensory system for AV have seen various approaches using knowledge from various backgrounds. The initial motivations have always been set to finding new and novel sensors which can accurately measure and detect elements that are useful for safe navigation towards level 5 of autonomous driving. Camera sensor has shown some promising extension in the intelligent transportation system. Specification plays an important role in determining the results of the camera application, despite the lighting condition (sunny day, gloomy day, road shadowing and so on). The use of radar, LiDAR and multi-dimensional imaging techniques have shown promising outcome. As a matter of fact, automotive industry already incorporated these sensors into their autonomous driving product. However, this sophisticated sensory system comes with high cost and will affect the maintenance and warrantee in the after sales framework. Furthermore, such cost and complexity will only fit certain market segment. The use of simplified and single modality sensory system such as vision and imaging has drawn attention of researchers who work in this field. The need of cost efficient and low maintenance system has moved the focus from designing complex and sophisticated sensors to pushing the ability of vision system to be able to do more things. This is achieved by empowering the processing algorithm combined with proper artificial intelligent and machine learning engine. This article has shown the tremendous efforts and remarkable achievements by scientists so far towards achieving level 5 autonomous driving using only with single camera system.

ACKNOWLEDGEMENTS

The author would like to thank the Universiti Teknologi Malaysia for funding this research through Research University Grant Scheme (RJ130000.7724.4J236) and the Ministry of Higher Education Malaysia.

REFERENCES

Aly, M. (2014). *Real time detection of lane markers in urban streets*. California: California Institute of Technology (Caltech).

Audi USA (6 January 2017). *Press release: Audi and NVIDIA team up to bring fully automated driving to the roads starting in 2020 accelerated with artificial intelligence*. Retrieved from https://www.audiusa.com

Baik, S., & Greenblatt, N. (2016). *A big question for self-driving car developers*. Sidley Publication.

Bilbeisi, K.M., & Kesse, M. (2017). *Tesla: A successful entrepreneurship strategy*. Morrow, GA: Clayton State University.

Bojarski, M., Testa, D.D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., Jackel, L.D., Monfort, M., Muller, U., Zhang, J., Zhang, X., & Zhao, J. (2016). *End to end learning for self-driving cars*. Holmdel, NJ: NVIDIA Corporation.

Chen, C., Seff, A., Kornhauser, A., & Xiao, J. (2015). *DeepDriving: Learning affordance for direct perception in autonomous driving*. Computer Vision Foundation (CVF).

Fan, Q., Brown, L., & Smith, J. (2016). *A closer look at Faster R-CNN for vehicle detection*. Paper presented at IEEE Intelligent Vehicles Symposium (IV), Gothenburg, Sweden.

Haloi, M., & Jayagopi, D.B. (2015). *Vehicle local position estimation system*. Paper presented at 2014 IEEE International Conference on Vehicular Electronics and Safety (ICVES'14), Hyderabad, India.

Jeon, J., Hwang, S-H, & Moon, H. (2016). *Monocular vision-based object recognition for autonomous driving in a real driving environment*. Paper presented at 13th International Conference on Ubiquitous Robots and Ambient Intelligent (URAI), Xian, China.

Jiang, T., Petrovic, S., Ayyer, U., Tolani, A., & Husain, S. (2015). Self-driving cars: Disruptive or incremental. *Applied Innovation Review, June 2015*(1), 3-22.

Kaur, R., & Marwaha, C. (2017). A review on the performance of object detection algorithm. *International Journal of Engineering and Computer Science, 6*, 20572-20576.

LeCun, Y., Urs Muller, U., Ben, J., Cosatto, E., & Flepp, B. (2005). Off-road obstacle avoidance through end-to-end learning. *Advances in Neural Information Processing Systems* 18 – *Proceedings of the 2005 Conference*, 739-746.

Litman, T. (2017). *Autonomous vehicle implementation predictions, implications for transport planning.* Victoria, Canada: Victoria Transport Policy Institute.

McBride, J., Snorrason, M., Eaton, R., Checka, N., Reiter, A., Foil, G., & Stevens, M.R. (2006). Object detection with single-camera stereo. *Proceedings* Volume 6230, *Unmanned Systems Technology VIII*, 623002. doi: 10.1117/12.669024

Miao, X., Li, S., & Shen, H. (2012). On-board lane detection system for intelligent vehicle based on monocular vision. *International Journal on Smart Sensing and Intelligent Systems*, *5*(4), 957-972.

Peng, H. (2016). Connected automated vehicle, the roles of dynamics and control. *The Magazine of ASME: Mechanical Engineering, Technology that Moves the World 12*,*138*, 4-11.

Romera, E., Bergasa, L.M., & Arroy, R. (2016). *Can we unify monocular detectors for autonomous driving by using the pixel-wise semantic segmentation of CNNs?* Paper presented at Workshop at the IEEE Symposium on Intelligent Vehicles 2016 (IV16-WS), Gothenburg, Sweden.

Sanchez, D. (2015). *Collective technologies: Autonomous vehicles (Working paper)*. Securing Australia's Future (SAF), Project 05. Melbourne, VIC: Australian Council of Learn Academies (ACOLA).

Tang, H. (2013). *Development of a multiple-camera tracking system for accurate traffic performance measurements at intersections (CTS 13-10)*. Duluth, MN: Center for Transportation Studies, University of Minnesota.

The Ford Motor Company (16 Aug 2016). *Ford targets fully autonomous vehicle ride sharing in 2021; invests in new tech companies, Doubles Silicon Valley team.* Retrieved from https://media.ford.com

Wang, W., Wang, F., Liu, P., He, Y., & Zhang, X. (2016). A monovision-based 3D pose estimation system for vehicle behaviour prediction. *Proceedings of the 2016 IEEE International Conference on Vehicular Electronics and Safety (ICVES)*, 1-6. doi: 10.1109/ICVES.2016.7548177

Woodside Capital Partners (2016). *Beyond the headlights: ADAS and automation sensing*. California: WoodsideCap.

Wu, B., Iandola, F., Jin, P.H., & Keutzer, K. (2016). *SqueezeDet: Unified, small, low power fully convolutional neural networks for real time object detection for autonomous driving.* Computer Vision and Pattern Recognition.