2017 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2017)
November 16 - 18, 2017, Daegu, Korea

Implementation of Semantic Segmentation for Road and Lane Detection on an Autonomous Ground Vehicle with LIDAR

Kai Li Lim, Thomas Drage and Thomas Bräunl

Abstract— **While current implementations of LIDAR-based autonomous driving systems are capable of road following and obstacle avoidance, they are still unable to detect road lane markings, which is required for lane keeping during autonomous driving sequences. In this paper, we present an implementation of semantic image segmentation to enhance a LIDAR-based autonomous ground vehicle for road and lane marking detection, in addition to object perception and classification. To achieve this, we installed and calibrated a low-cost monocular camera onto a LIDAR-fitted Formula-SAE Electric car as our test bench. Tests were performed first on video recordings of local roads to verify the feasibility of semantic segmentation, and then on the Formula-SAE car with LIDAR readings. Results from semantic segmentation confirmed that the road areas in each video frame were properly segmented, and that road edges and lane markers can be classified. By combining this information with LIDAR measurements for road edges and obstacles, distance measurements for each segmented object can be obtained, thereby allowing the vehicle to be programmed to drive autonomously within the road lanes and away from road edges.**

I. Introduction

The Renewable Energy Vehicle (REV) Project at the University of Western Australia conducts research into electric vehicles, vehicle automation and autonomous driving systems. Recent projects include the development of an Autonomous Formula-SAE Electric car [1]. This vehicle is an open-wheeled, electric drive race car, with electronic drive-by-wire and electromechanical brake/steering actuation. The vehicle serves as a compact, flexible test-bed for sensor testing and the development of autonomous driving algorithms.

Prior research has been conducted on road and road edge detection through optical systems [2], radar [3] as well as using Light Density and Ranging (LIDAR) sensors such as in the winning entry in the 2007 DARPA Urban Challenge [4]. The methodology described in [5] utilises a feature-extraction algorithm while other algorithms such as [6] rely on the presence of curbs and seek to identify and track curbs as features in the LIDAR data. More recently, there has been an increase in the use of cameras to achieve this [7], giving rise to visual road detection. Methodologies to achieve this include feature extraction and classification [7], horizon and vanishing point detections [8], and artificial neural networks (ANNs) [9].

The authors are with The REV Project at the School of Electrical, Electronic and Computer Engineering, The University of Western Australia, Perth WA 6009 Australia. {kaili.lim, thomas.braunl}@uwa.edu.au, thomas.drage@research.uwa.edu.au

The problem of path-finding can be described as: "Given a start state, a goal state, a representation of the robot and a representation of the world, find a collision-free path that connects the start with the goal satisfying the system constraints" [10]. In mobile robotics, a proven method to obtain the requisite representation of the world is via the use of LIDAR data to generate a virtual map in real-time both as the sole sensor [11] and in conjunction with data from additional sensors [12]. Similar LIDAR based map building approaches have been shown to be suitable for outdoor terrain [13]. These generated maps vary from simple two-dimensional maps suitable for basic path planning consisting of traversable regions, obstacles and unexplored regions [14] to more complex three-dimensional maps from which sophisticated cost maps are generated [15]. A more detailed map can be built by supplementing the camera in addition to LIDAR. These additional details can include a combination of vehicle detection and classification [16], road sign recognition [17], and scene recognition [18].

Visual cameras and LIDAR are often incorporated in autonomous driving systems. Works that combines LIDAR and camera sensors for autonomous driving include the approach from Zhang, Clarke and Knoll [16], where they have proposed the fusion of LIDAR and the camera as a compromise for each sensor's drawbacks, with LIDAR providing range information, and the camera identifies objects and scenes. The authors achieved low false alarm rates and a high detection rate for vehicles in urban environments. A similar fusion of multiple LIDAR, radar, and camera sensors to achieve object detection and tracking was proposed by Cho et al. [19]. By tracking pedestrians and vehicles, the system could detect and track vehicles from 150m away, and pedestrians and cyclists within a 20 m radius. To the best of our knowledge, works that incorporate semantic segmentation onto a LIDAR-based autonomous ground vehicle has not been established at the time of writing.

Our work is an enhancement to the work done by Drage, Churack and Bräunl [20], where we have proposed a LIDAR-based road edge detection approach on the same vehicle. Our algorithm could detect road curbs and edges by measuring the differences in surface smoothness, which in turn allows the positioning of road edges and curbs.

II. Implementation

This section describes the addition of visual perception to the LIDAR-based autonomous SAE car as described in [20], which includes sections that describe our testing environment, and its applicable procedures to achieve visual

978-1-5090-6064-1/17/$31.00 ©2017 IEEE

autonomous driving. By mounting a monocular camera onto the chassis of the vehicle, above the LIDAR (see Fig. 1), road recognition and obstacle detection are achieved using semantic segmentation. This camera supplements the LIDAR, where the LIDAR is responsible for providing distance measurements for objects and road edges detected by the camera. Semantic segmentation was achieved using SegNet [21], a convolutional neural network (CNN) architecture for semantic segmentation that is often used for road scenes. Its architecture uses an encoder-decoder network that is followed by a pixelwise classification layer, where the encoder and decoder networks consist 13 convolutional layers each. The Caffe [22] implementation of SegNet is used for this project. To interface the sensors for autonomous driving, SegNet is installed onto an Nvidia Jetson TX1 [23], and the LIDAR interfaces directly to a Raspberry Pi 3 [24], which drives a control system. A GPS module and an inertial measurement unit (IMU) module also connects to the Raspberry Pi 3 for positioning and localisation.

The LIDAR system used in this project consists of an IBEO Lux automotive LIDAR with specifications as shown in Table I. This sensor utilises reflected infra-red light to measure distance (via time-of-flight) and can build a 3D point cloud by scanning horizontally in four vertical layers. The IBEO sensor has sophisticated internal data processing functionality including object detection and classification. Data is delivered using TCP/IP over an Ethernet connection and includes scan data in polar coordinates and object data in x-y coordinates referenced to the sensor.

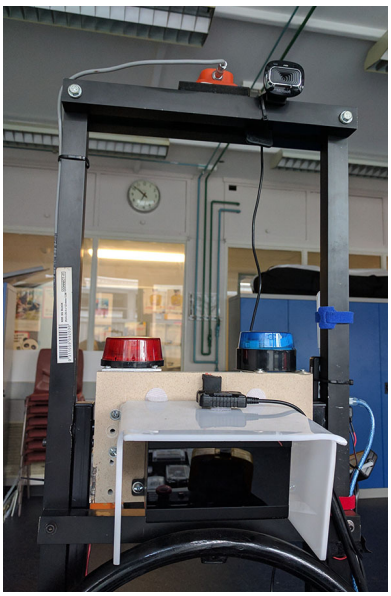

Fig. 1: The camera is mounted above the LIDAR system from [20], beside the IMU

The following subsections describe the process of achieving visual autonomous driving for our project using semantic segmentation with respect to its application environment and its driving sequences.

TABLE I: LIDAR Characteristics

Specification	Value
Technology	Time of flight (output of distance and echo pulse width)
Range	200m
Field of View (Horiz / Vert)	85 / 3.2
Layers	4
Echo Detection	3 measurements per pulse
Update Rate	Up to 50Hz
Accuracy	10cm

A. Application Environment

SegNet was tested within the grounds of the University of Western Australia (UWA), which is the same location that the LIDAR system was tested in [20]. The roads within UWA offers a similar drive environment to standard suburban roads. These single carriageway roads are of low traffic density, with views of pedestrians, faculty buildings, and vegetation for SegNet to recognise and segment. As a feasibility test, we also tested SegNet off a car-mounted dashcam recording while driving on local roads.

This application environment was selected to test the suitability of using SegNet for autonomous driving locally and to gauge the visual autonomous navigation performance of the vehicle. To achieve a successful autonomous drive using SegNet on the vehicle, road edges and lane markings must be properly recognised, before the application can be subsequently expanded onto a road-licensed vehicle.

It should be noted that our initial implementation uses the trained dataset from the University of Cambridge, CamVid [21]. This dataset was recorded in the City of Cambridge, England. Like most British cities, Cambridge's roads are often narrow, with dense buildings by the side. There is also a large pedestrian population due to it being an academic city. Comparatively, roads in Perth are generally wider, with a sparser build-up density than Cambridge. Its low population density means that there are fewer road pedestrians as compared to Cambridge. The ground terrain around Perth is mostly flat, with long sunshine hours. This means that one can generally expect excellent road visibility on Western Australian roads on most days. However, during poor visibility and night time drives, suburban roads around Perth are generally poorly lit, which may affect road segmentation accuracy. By testing this dataset, we will subsequently contemplate on the need to record and use a dataset from Perth for more accurate segmentation results.

B. Autonomous Driving Procedures

To use SegNet's output for autonomous driving, we assume that the car begins with a position in the middle of the road lane and that the road incline is flat. By mounting the camera at a fixed position on the car, the central driving position of the car can be obtained, along with distance measurements from the road edges to the left and right sides of the car. With the camera, this is done by identifying and segregating a fixed trapezoidal image region that

encapsulates the road segment, which is then transformed into a birds-eye view (BEV) perspective to obtain vehicle's position with respect to the road's centre. By scaling the road width according to the Australian standards of 3.3–3.5 m, along with the detected road edges and/or lane markings on SegNet's output, the distance from the vehicle's centre to the left and right road edges/lanes can is obtained as Fig 2, with its confidence value determined by the successful detection of road edges or lane markings. From these three sections distance thresholds for the left and right edges or lane markers distances for the car to autonomously centre itself on the road while driving. We call this road centring. In the event where lane markers are not found, road edges will be used instead.

To perform road centring, the car must steer itself in the opposite direction when it crosses the distance threshold to either the left or right road edge/lane markers. The distance from the car to the road edges or lane markers are constantly analysed. If the car is too close to the edge or lane marker, the road centring algorithm will then send commands to the drive system to steer away from the edges with fine adjustments, until the car is cleared from the distance threshold.

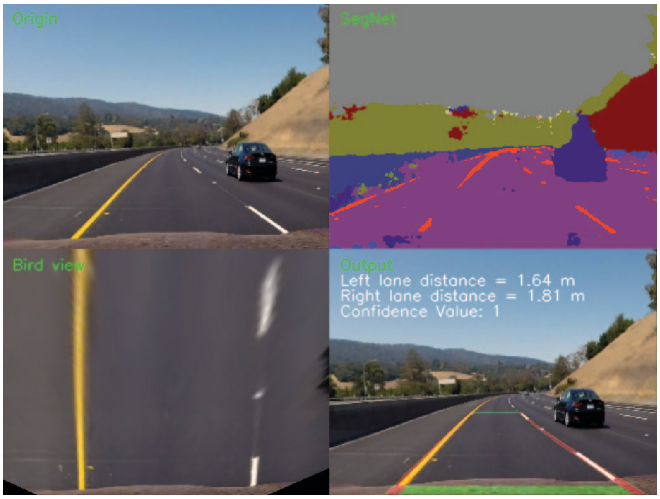

Fig. 2: Lane distance measurements with SegNet's output on Udacity's Self-Driving Car Nanodegree recording [25] as the input in *Origin*.

III. Testing and Evaluations

A. Methodology

Testing begins with the calibration of the camera, where distance measurements in the real world will be represented in pixel ratios on SegNet's output. Here, we calibrated a Microsoft LifeCam HD-3000 camera. This was done by measuring the distances between road bollards in front of the parked vehicle on the road as illustrated in Fig. 3.

The bollards are placed a three-row formation to allow distance measurements from two-point distances on the frame. The first row of bollards on the car establishes the starting distance, with the centre bollard measuring the distance from the camera to the front of the car, while ensuring that the

Fig. 3: Photo illustrating the bollards' position with reference to the car in the centre.

camera is pointed to the centre of the car. All distances are measured from the base at the centre of each bollard. Subsequently, a topological representation of the measurements can be illustrated in Fig. 4, which is then represented again in the camera frame in Fig. 5. These measurements are verified with the LIDAR plot at that position as illustrated in Fig. 6, where the bollards (represented as dot plots) are clearly present around the 2 m, 8 m and 10 m mark on the y-axis. From Fig. 5, each image pixel was calculated to represent 17 mm and 23 mm when measured from 8.5 m and 11.6 m respectively from the camera, and that a level road will converge at around 41°on the camera frame.

The LIDAR readings complement SegNet's output for road edge detection, whereby we use our Kalman Filtered Linear Regression Model as described in [20]. Our algorithm minimises the square residuals between the fit line y and the data (x_i, y_i), where the most suitable data line will be obtained for a given data set, and its success measured by the product-moment correlation coefficient r. The slope b and r values are tabulated as equation 1.

$$y = (\bar{y} - b\bar{x}) + bx, r = \frac{s_{xy}}{s_x s_y} \tag{1}$$

where:

$$b = \frac{s_{xy}}{s_x}^2, s_{xy} = \frac{\sum_{i=1}^{n} x_i y_i}{n} - \bar{x}\bar{y},$$
$$s_x{}^2 = \frac{\sum_{i=1}^{n} x_i{}^2}{n} - \bar{x}^2, s_y{}^2 = \frac{\sum_{i=1}^{n} y_i{}^2}{n} - \bar{y}^2$$

With a calibrated camera and LIDAR, the SAE car was driven around the application environment while the camera is recording. The recorded camera footage was used as an input for SegNet. For testing purposes, these footages were processed off-line with SegNet running off an Nvidia GTX Titan X GPU with a 480 by 360-pixel resolution, which resulted in a segmented image output at a consistent 10 frames per second (FPS) sampled at each video frame. This framerate is consistent with the visual autonomous driving results published by Nvidia in [26], making it adequate for autonomous driving. Further work is required for real-time on-line processing on the Jetson TX1. Likewise, LIDAR plots are also recorded during the duration of the drive, where timestamps are used to synchronise outputs.

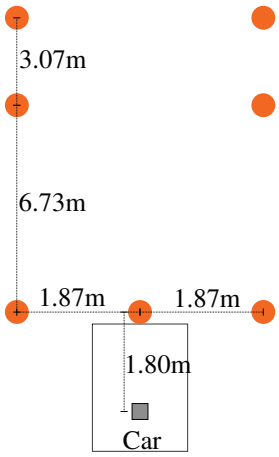

Fig. 4: The topological distance between bollards (dots) and the camera on the car (shaded square).

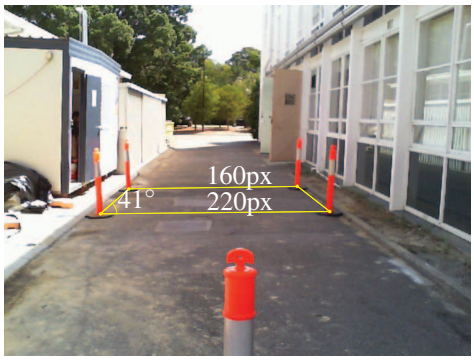

Fig. 5: Frame captured from the vehicle's camera for distance calibration

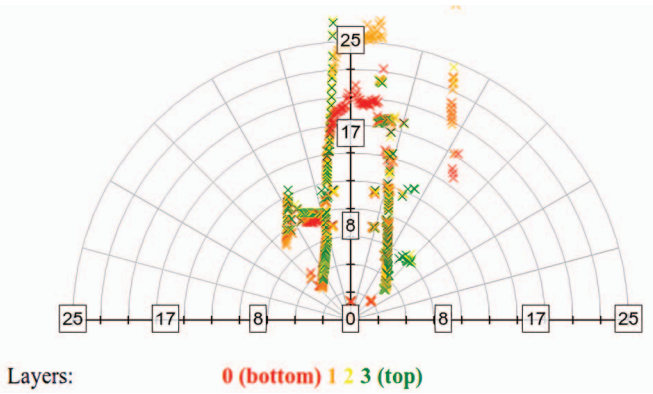

Fig. 6: LIDAR plot showing the detected road edges and the bollards' positions from the car where Fig. 5 was captured. The graph axes measure distances in metres.

B. Results and Discussions

Results from semantic segmentation are presented first on open roads with a dash cam recording, and then on the UWA campus ground with the SAE car. Image results in this section are presented with the left image showing segmentation input, and the right image showing its output through SegNet. In the case of false detections, they will be measured in their pixel accuracy (PA). This is done by finding the percentages of road regions, which is done by counting the number of falsely detected pixels against the total number of pixels in the road region for that image frame.

Segmentation results from the open road testing from the dashcam footage yielded consistently favourable results, as shown in Fig. 7. Recordings were captured on a clear day, driving on a low traffic dual carriageway. The road region is accurately segmented with negligible false detections. In addition to the road and its markings, SegNet can detect and segment the speed limit sign and the right turn sign ahead. All vehicles and vegetation were also properly detected and segmented. The minor false detections in the sky region are due to the CamVid dataset being trained on a cloudy day, but this should not affect road detection accuracy in any way. This result confirms the feasibility that a CamVid-trained SegNet can be implemented for autonomous driving in the Perth metro area.

Fig. 7: SegNet's input (left) and output (right) for a dual carriageway in the Perth metro area.

Tests were subsequently performed on the SAE car for a vision-LIDAR-based implementation. Here, runs on campus grounds were recorded using the camera while manually driven on the SAE car following a predetermined route on campus. The car traversed across roads and pavements and the segmentation results are as follows.

Fig. 8: Segmentation results from a parking area on campus grounds.

The test run began with a drive through the car park. SegNet's output in Fig. 8 shows that the image was segmented with good accuracy. With the exception of some minor (0.69%) false detections on the car's shadows on the left side, the road, parking lane, and pavements were properly segmented, along with the pedestrian and vehicles. With the LIDAR actively measuring the distance from the parked vehicles to the SAE car (see Fig. 9). Here, we adopt the

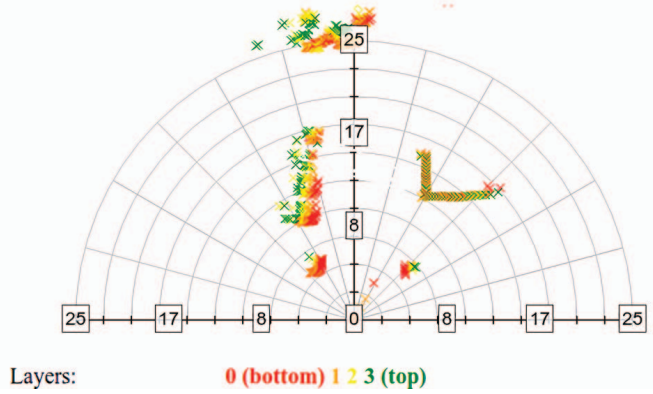

Fig. 9: LIDAR plot showing the detected parked vehicles at the position where Fig. 8 was captured. The graph axes measure distances in metres.

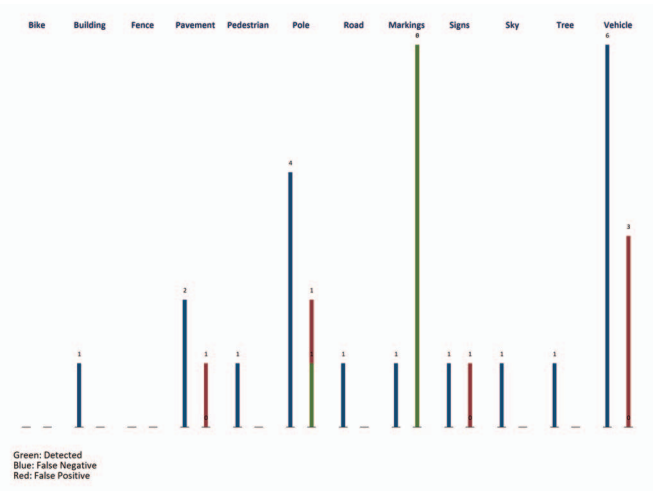

Fig. 10: The number of detected objects along their errors in detections according to the objects.

Linear Regression model that we described in Section III-A, which plots the road edge position from the parked vehicles so that a fixed distance can be kept between the autonomous cars and the parked vehicles. We have also counted the number of detected objects along with their positive and negative false detections, which are plotted according to their detection/error pairs in Fig. 10, whereby the labelled number on each bar indicated the number of correctly identified objects, if present.

Fig. 11: Segmentation results on pavement between faculty buildings.

From the parking area, the car drives onto the pavement between faculty buildings. From Fig. 11, SegNet could discern pavements from roads as the grounds are now coloured blue. In addition, it was also able to detect the bicycles parked towards the right, and the pedestrians in the distance. False detections are present on the left side of the pavement, where it is incorrectly detected as buildings and pedestrians due to uneven lighting, accounting for 2.83% of the total pavement region.

Fig. 12: Segmentation results at a road junction.

Fig. 12 was captured when the car was stopping at a suburban road junction beside the campus. The segmentation output from this figure shows that while the objects in the distance were properly segmented, some parts at the bottom of the image was incorrectly classified as pavements and buildings, which makes up 16.19% of the road region. This was partly due to the clear weather resulting in a high brightness recording, while SegNet was expecting a darker surface to classify roads.

Fig. 13: Segmentation results on a road with pronounced shadows.

Fig. 13 was captured while the car was driving on a shadowed road on campus. The shadows on the road introduced a high contrast region, and that the bottom portion of the road was overexposed, resulting in a false classification of 13.93% and undetected lane markers. However, the overtaking vehicle, vegetation, road sign, and building were correctly segmented. When presented with a false detection on the road as with Figure 7, results from the LIDAR road edge detection system will compensate the regions of false detection, as the road and road edges were properly measured by the LIDAR system.

The false detection rates for each of the example figures is summarised as Table II, which also tabulates the number of road pixels in each detected road segments for each figure, along with the number of falsely classified pixels within that road segment. From these numbers, the detection error is

TABLE II: The number of road region pixels and its false detections pixels in that region for each figure along with their error percentages.

Fig.	Road Pixels	False Classifications	Error
7	41648	16	0.04%
8	86561	600	0.69%
11	76731	2168	2.83%
12	108684	17595	16.19%
13	83831	11679	13.93%

calculated as a percentage that corresponds to the area of each figure's road segments. The false detection rate is at its highest in Fig. 12, where the road region encompasses most of the frame, the unevenness in road surface and lighting is the main contributor to this error rate. Conversely, Fig. 7 records the lowest false detection rate its road segment, as the road area was well-defined, and the frame was properly exposed.

IV. Conclusion

We have presented a semantic segmentation-based visual navigation approach for autonomous ground vehicles. This approach improves on existing LIDAR-based vehicles to introduce object recognition and classification while driving. SegNet adequately performs semantic segmentation to recognise roads and lane markers, which in turn allows the vehicle to maintain a safe distance from the road and lane edges in addition to LIDAR measurements. We have also shown that the segmentation results from SegNet on the CamVid dataset is satisfactory on Perth metro roads. With a calibrated camera, visual autonomous driving can be achieved using real-time semantic segmentation. Future works will focus on the complete on-line implementation of SegNet on the SAE car for real-time visual autonomous driving.

ACKNOWLEDGMENT

The authors would like to thank The REV Project's Autonomous SAE team for their efforts in aiding the compilation of data for this paper, and Nvidia Corporation for sponsoring the GPU under its GPU Grant Program.

References

[1] T. Drage, J. Kalinowski, and T. Braunl, "Integration of drive-by-wire with navigation control for a driverless electric race car," *IEEE Intelligent Transportation Systems Magazine*, vol. 6, no. 4, pp. 23–33, 2014.

[2] P. Y. Shinzato, D. Gomes, and D. F. Wolf, "Road estimation with sparse 3d points from stereo data," in *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 2014, Conference Proceedings, pp. 1688–1693.

[3] K. Y. Guo, E. G. Hoare, D. Jasteh, X. Q. Sheng, and M. Gashinova, "Road edge recognition using the stripe hough transform from millimeter-wave radar images," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 825–833, 2015.

[4] W. Zhang, "Lidar-based road and road-edge detection," in *2010 IEEE Intelligent Vehicles Symposium*, 2010, Conference Proceedings, pp. 845–848.

[5] M. Nikolova and A. Hero, "Segmentation of a road from a vehicle-mounted radar and accuracy of the estimation," in *Proceedings of the IEEE Intelligent Vehicles Symposium 2000 (Cat. No.00TH8511)*, 2000, Conference Proceedings, pp. 284–289.

[6] G. Zhao and J. Yuan, "Curb detection and tracking using 3d-lidar scanner," in *2012 19th IEEE International Conference on Image Processing*, 2012, Conference Proceedings, pp. 437–440.

[7] Y. Alkhorshid, K. Aryafar, S. Bauer, and G. Wanielik, "Road detection through supervised classification," in *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2016, Conference Proceedings, pp. 806–809.

[8] P. Y. Shinzato, V. Grassi, F. S. Osorio, and D. F. Wolf, "Fast visual road recognition and horizon detection using multiple artificial neural networks," in *Intelligent Vehicles Symposium (IV), 2012 IEEE*, 2012, Conference Proceedings, pp. 1090–1095.

[9] N. Abbas and V. Mahdi, *A Novel Neural Network based Voting Approach for Road Detection via Image Entropy and Color Filtering*, ser. Indian Journal of Science and Technology.

[10] S. M. LaValle, *Planning algorithms*. Cambridge university press, 2006.

[11] Y. Liu and Y. Sun, "Mobile robot instant indoor map building and localization using 2d laser scanning data," in *2012 International Conference on System Science and Engineering (ICSSE)*, 2012, Conference Proceedings, pp. 339–344.

[12] F. Mengyin, Z. Hao, Y. Yi, W. Meiling, and L. Yu, "Multiple map representations for vehicle localization and scene reconstruction," in *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 2014, Conference Proceedings, pp. 2241–2242.

[13] D. M. Cole and P. M. Newman, "Using laser range data for 3d slam in outdoor environments," in *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006.*, 2006, Conference Proceedings, pp. 1556–1563.

[14] Z. Hao, F. Mengyin, Y. Yi, W. Xinyu, and W. Meiling, "A path planning algorithm based on fusing lane and obstacle map," in *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 2014, Conference Proceedings, pp. 1442–1448.

[15] J. Gillula and J. Leibs, "How to teach a van to drive: an undergraduate perspective on the 2005 darpa grand challenge," *IEEE Control Systems*, vol. 26, no. 3, pp. 19–26, 2006.

[16] F. Zhang, D. Clarke, and A. Knoll, "Vehicle detection based on lidar and camera fusion," in *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 2014, Conference Proceedings, pp. 1620–1625.

[17] A. Gudigar, S. Chokkadi, and R. U, "A review on automatic detection and recognition of traffic sign," *Multimedia Tools and Applications*, vol. 75, no. 1, pp. 333–364, 2016. [Online]. Available: http://dx.doi.org/10.1007/s11042-014-2293-7

[18] J. Yang, S. Zhang, G. Wang, and M. Li, "Scene and place recognition using a hierarchical latent topic model," *Neurocomputing*, vol. 148, pp. 578–586, 2015.

[19] H. Cho, Y. W. Seo, B. V. K. V. Kumar, and R. R. 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)*, 2014, Conference Proceedings, pp. 1836–1843.

[20] T. Drage, T. Churack, and T. Braunl, "Lidar road edge detection by heuristic evaluation of many linear regressions," in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, 2015, Conference Proceedings, pp. 2465–2470.

[21] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *arXiv preprint arXiv:1511.00561*, 2015.

[22] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014, Conference Proceedings, pp. 675–678.

[23] NVIDIA Corporation, "Embedded systems," 2017. [Online]. Available: http://www.nvidia.com/object/embedded-systems-dev-kits-modules.html

[24] Raspberry Pi, "Raspberry pi 3 model b," 2017. [Online]. Available: https://www.raspberrypi.org/products/raspberry-pi-3-model-b/

[25] Udacity, "Carnd-lanelines-p1," 2017. [Online]. Available: https://github.com/udacity/CarND-LaneLines-P1

[26] M. Bojarski, D. Del Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, and J. Zhang, "End to end learning for self-driving cars," *arXiv preprint arXiv:1604.07316*, 2016.