Road obstacles positional and dynamic features extraction combining object detection, stereo disparity maps and optical flow data

Thiago Rateke$^{1,2}$ · Aldo von Wangenheim$^{1,2}$

1 Introduction

Obstacles detection in autonomous and drive-assisted vehicles concerns the detection of any other objects, static or in movement, on or near the road. In an intelligent autonomous vehicle navigation scenario it is, along with path detection, one of the most important tasks, because it involves not only the safety of the vehicle where the obstacles detection and recognition are performed, but also because it affects other participants in this scenario, such as: other vehicles, pedestrians, cyclists and animals. Based upon information continuously gathered by the obstacle detection, the behavior of an autonomous vehicle must adjust itself or, in the case of an Advanced Driver Assistance Systems (ADAS), it must generate alerts that allows drivers to adapt their driving to potential threats.

The state-of-the-art for obstacle detection is already quite robust, and with the recent advancements in convolutional neural network (CNN)-based deep learning approaches, has been obtaining excellent results. Prior to this work, we performed a Systematic Literature Review (Rateke and von Wangenheim (2018) and Rateke and von Wangenheim (2020)), based on the procedures described in Kitchenham and Charters (2007). And based on this literature review we were able to determine that the state-of-the-art in the road obstacle detection area with focus on vehicular navigation has many examples and different approaches.

The approaches we were able to identify vary widely. Some examples are: using only Stereo Vision (e.g.: Hne et al. (2017)), only Optical Flow (e.g.: Bouchafa and Zavidovique (2011)), Image Segmentation (e.g.: Poddar et al. (2015)), and recently works using Convolutional Neural Networks (e.g.: Prabhakar et al. (2017)). There are also several other approaches that use combinations...
between different methods, such as: Gupta et al. (2017), which uses methods based on Neural Networks, Stereo Vision and Image Segmentation, and Giosan and Nede-schi (2014) which uses methods based upon Stereo Vi-

sion, Optical Flow and Image Segmentation.

In moving obstacles detection, however, it is not only important to identify the kind of obstacle: other vehicle, a cyclist, a horse-rider, a pedestrian crossing the road or a stray animal. It is also important to be able to determine the potential path of these obstacles and be able to estimate if there exists the possibility of a collision, i.e., if each detected object has the potential to become a threat to the vehicle. For this purpose, it is necessary to be able to estimate the obstacles distance, velocity and direction of movement. In this context, we can understand Autonomous Vehicle Threat Assessment (AVTA) as the continuous active inspection of its sensorial data by a vehicle in order to identify road objects and traffic participants that could pose a threat to the vehicle’s navigation.

In the work we present here, we move on from obstacle detection to the next step, which is to identify, in the detected obstacles, features that are relevant to threat assessment in a vehicular navigation system context, such as: distance, velocity and also direction of movement from detected objects.

1.1 Objectives

The objective of this work is to investigate the feasibility of the development of a passive vision (PV)-based integrated moving obstacles detection and description approach that fulfills the following requirements:

- detects and classifies obstacles pertaining to a set of predefined classes;
- provides depth information about each obstacle, relative to the vehicle;
- provides information about the trajectory and speed of each obstacle, relative to the vehicle;
- is capable of determining this information employing only data gained from passive vision, without relying on additional data from LIDAR (Light Detection and Ranging) or other active sensors.

Furthermore, our work concentrated not in developing new image processing algorithms, but investigated if there exist already developed and mature technologies which could be combined in order to achieve the objective above.

1.2 Approach Outline

In our approach, in the obstacle detection step, we employ stereo images and a state-of-the-art CNN structure, the Mask R-CNN [He et al. (2017)], which in addition to the detection and recognition of objects, also determines the position and shape of these objects, providing, as a second layer of results, a semantic segmentation (SS) of the recognized objects. From the original images, obtained from data-sets that provide stereo data with two-camera captures, we also generate the Disparity Maps (DM) of the scene (depth map) for each pair of stereo frames. This DM we apply to the objects recognized and segmented by the Mask R-CNN, extracting the average depth information for each segmented object, allowing spatial localization of these objects. In addition, we also apply the Optical Flow calculation on these objects, being able to filter the average movement flows (motion direction and intensity) separately for each detected object.

1.3 Research Rationale

Different sensors can be used for the obstacle detection task. Some vehicles employ an ensemble of diverse sensors, not only cameras Urmson et al. (2008), and Fernandes et al. (2014). One of these sensor, present in many autonomous vehicle navigation projects, are active sensors named as LIDARs, which are laser sources used for active sensing of reflected light, in order to measure distances between the sensor and the target object LiDAR-UK.com (2015). In vehicular projects, the LIDAR employed is normally a laser of Class1, which is the category considered to present less danger. It employs light in the infra-red (IR) spectrum, in wavelengths in the order of 905nm.

Based on the studies of Commission (2001) and STAN-

DARD (2005), a single Class1 laser source poses no danger to the retina when it does not remain in direct contact with the human eye for a longer time. A categorization of the lasers and the possible damages caused by excessive exposure in different levels is presented in Commission (2001). Lasers that emit in a wavelength between 780nm and 1400nm can cause cataracts and burn the retina. Considering a scenario where autonomous vehicles are used on a large scale, situations of dense traffic could be responsible for a many-LIDAR-originated “lasersmog” and become a risk to the nearest humans, which would simultaneously be targeted by the signals of many laser sources.

Even if there exists no conclusive study of the impact of many-car generated lasersmog on pedestrians yet, we understand that stereo camera-based PV may
be a better alternative in a future autonomous vehicle scenario. For this purpose our work focuses on data achieved through passive stereo vision only, without any information supplementation through LIDAR data.

The remainder of this paper is organized as follows: In Section 2 we present the related works and their respective approaches. In Section 3 we present the datasets used in our experiments and also present the methods we apply in our approach. Our approach is presented in Section 4. Followed by the results obtained in Section 5. Finally, in Section 6 we conclude this paper with a discussion about the results and the next steps in future work.

2 Related work

Other authors have already tackled the PV-based extraction of relevant features from objects in the scene. Mitzel et al. (2011) performs a pedestrian detection with focus on multiple pedestrians tracking and uses Stereo Vision techniques for the detection step and the RANSAC framework for the pedestrian motion estimation and tracking.

In Chanawangsa and Chen (2013) the authors present an approach to do the tracking of detected vehicles in the scene with focus on identifying overtaking situations. The markings on the road, the lanes, are also detected to know when a vehicle may be entering in front, allowing to generate an alert. For the vehicles detection step are used Histogram of Oriented Gradients (HOG) and Support Vector Machine classifier (SVM). The Kalman filter is used for vehicle tracking step.

An obstacle detection that also employs Stereo Vision techniques is presented in Huang and Lin (2016). Based on the image generated by the DM, the object contours are found. Based on these contours the authors use the objects’ geometric information, such as area and height to classify objects (e.g.: people, vehicles and others).

Other works also use geometric information from the detected obstacles in order to classify the obstacles by types. In Li and Chen (2014), the authors present an approach that besides the geometric information of the detected obstacles (height and width) uses fuzzy logic to classify these obstacles. In Liu et al. (2014) the authors applied a segmentation in the Disparity Map and also use width and height features from the detected obstacles to make the classification.

Stereo Disparity map is also used in Chen et al. (2012) together with Histogram of Oriented Gradient (HOG) to extract the obstacles features. Finally, the classification of obstacles is made through a Support Vector Machine.

To predict future vehicle localization the authors from Yao et al. (2018) use a recurrent neural network (RNN) with a dense optical flow incorporation. In Deo et al. (2018) the authors also shown how to predict other vehicles actions using hidden Markov model (HMM), interacting multiple model (IMM) and variational Gaussian mixture models (VGM). Also to predict the trajectories from other vehicles the Jawed et al. (2019) presents an approach which use a Convolutional neural network (CNN).

These works used different combinations of methods and techniques and are focused on classifying the types of obstacles or, at most, tracking some of the obstacles. Not focusing on extracting behavioral features from obstacles in relation to the moving vehicle.

3 Material and methods

We employed two different data-sets in our experiments, both presenting stereo images from urban vehicle navigation scenarios, but in different contexts (Germany and Brazil). Both data-sets are presented in Section 3.1. In Section 3.2, 3.3 and 3.4 we provide a brief descriptions about each method we used in our model.

3.1 Data-sets

The two data-sets used in our experiments were: KITTI data-set\textsuperscript{1} (Geiger et al. 2013) and CaRINA data-set\textsuperscript{2} (Shinzato et al. 2016). Both provide high-quality stereo images in vehicle navigation scenarios. KITTI uses a PointGray Flea2 cameras and CaRINA uses a Bumblebee XB3 camera.

Created by the Mobile Robot Laboratory group (ICMC / USP - Sao Carlos) filmed in Brazil, more specifically in the city of So Carlos in So Paulo state, the CaRINA data-set aimed to provide images for experiments in autonomous navigation visual perception in emerging countries scenarios, containing low quality roads. There are few pedestrian situations (almost none), but contains other vehicles in the scene (e.g.: cars, motorbikes, trucks).

In contrast, the data-set provided by KITTI contains a considerable amount of pedestrian and cyclist situations in the scene, in addition to other vehicles. KITTI was created by the Karlsruhe Institute of Technology in Karlsruhe city, Germany. It is probably one of the most commonly used data-sets in visual perception works for vehicle navigation tasks, including for path detection and obstacle detection.

---
\textsuperscript{1} http://www.cvlibs.net/datasets/kitti/raw_data.php
\textsuperscript{2} http://www.lrm.icmc.usp.br/dataset
3.2 Mask R-CNN

In [He et al. (2017)] the authors present a framework for object instance segmentation. The Mask R-CNN, in addition to detecting and classifying objects in the scene, also applies a segmentation mask to each detected object (eg.: Figure 1). According to the authors, Mask R-CNN is an extension of Faster R-CNN [Ren et al. (2015)].

![Fig. 1 Mask R-CNN example on KITTI data-set.](Image)

We use this framework with pre-trained models in the Inception backbone architecture [Szegedy et al. (2016)], which has good classification accuracy and is faster than many other architectures. Also, our experiments runs in a model trained with MSCOCO data-set [Lin et al. (2014)], which is a data-set specific for object detection and segmentation.

3.3 Disparity map

The disparity is the difference that the same pixel has between two images, this difference takes into account the position of the same pixel in each images. It is common to use disparity as a synonym of depth [Bleyer (2013)]. The ideal for Stereo Vision works is that the images are perfectly rectified on the y-axis, allowing the scanning by checking the corresponding pixels and their respective differences to occur only on the x-axis:

$$D = x_l - x_r$$  \hspace{1cm} (1)

where $x_l$ is the specific pixel coordinate in left image, $x_r$ is the coordinate of the same specific pixel in the right image and $D$ is the disparity value between these points. Both data-sets used in our experiments have perfectly rectified images.

The Disparity Map is the image that represents the pixel disparity values as an intensity image, where high intensity values represent high disparities and low intensity values represent lower disparities [Bleyer (2013)]. Normally the Disparity Map is displayed as grayscale image, we applied a simple color conversion for a better visualization, but the intensity information is the same (eg.: Figure 2).

![Fig. 2 Disparity Map example on KITTI data-set. Original left image on top, Disparity Map on bottom.](Image)

3.4 Optical flow

The goal of Optical Flow (OF) is to identify the displacement of intensity patterns in the image along sequential frames. This movement information can be very useful in computer vision because it also allows the identification of certain patterns in the scene [Fleet and Weiss (2005)].

In the literature there are examples of OF obtained through Neural Networks [Dosovitskiy et al. (2015), Ilg et al. (2017)] and also through traditional numeric methods [Farnebäck (2003), Lucas and Kanade (1981)]. Neural OF methods may be a more recent tendency, but they also require more computational resources. In our work, we already performed the detection and segmentation of the obstacles with the use of a CNN and we only need to apply the OF calculations to the detected objects. For this reason we opted to perform a post-processing employing a traditional OF approach. In addition, this approach provides us with explicit vector data which could be later used by a vehicle for threat assessment, which is not possible with the present CNN-based OF approaches.

In our approach we used the Gunnar-Farneback algorithm [Farneback (2003)], which produces a dense OF working on a grid of points. In this algorithm, the movement vector value is extracted through information obtained from two consecutive frames [Farneback (2003)]. As this algorithm calculates the OF for each pixel in the image, it performs a good motion estimation of the regions encompassing the detected objects. An example with flow vectors is shown in Figure 3, where a pedestrian is crossing the street in front of an awaiting vehicle.
4 Our approach

In Figure 4 we present a schematic overview of our approach, which consists in combining the techniques described in Section 3. We integrate the obstacle detection and SS results obtained by Mask R-CNN with the results from OF and the Disparity Map. In this way, it is possible to extract the OF and disparity values from specific pixels in each object, which allows us to generate a detailed analysis of depth and movement for each object in the scene.

Both KITTI and CaRINA data-sets provide stereo images. For the OF calculation and the CNN object detection we employ only the left-captured (driver-side) images from the data-sets. For the calculation of the disparity maps we employ the whole stereo data.

With the disparity values obtained from each object segmentation, it is possible to generate an average disparity value for each segmented object. Thus, in the final analysis, we defined some depth labels with pre-set thresholds. We defined four depth labels: very close, close, far and very far.

In the same way, we used the Optical Flow values from each detected object to generate average movement values, computing their means as the resulting direction and intensity motion vector. We collected the direction values from the OF vector on the x-axis to label whether the vehicle is stationary, or going from right to left or from left to right. We defined the direction labels on the x-axis as: left to right, right to left and stable direction.

The direction values from the OF vector on the y-axis indicate to us whether the vehicle is approaching, moving away or maintaining a stable distance. We defined three labels for the y-axis being: approaching, moving away and stable distance.

The greater the displacement of a pixel between two frames, the greater will be the vector representing that displacement. This value allows us to have a sense whether the detected object is moving fast or slow. We obtained the displacement value from each object by multiplying the average values of the x-axis and y-axis from each object:

\[
x_M = \frac{1}{n} \sum_{i=1}^{n} x_i
\]

\[
y_M = \frac{1}{n} \sum_{i=1}^{n} y_i
\]

\[
VL = x_M \times y_M
\]

where \(x_M\) is the x-axis average value in an object, \(y_M\) is the y-axis average value in the same object, and \(VL\) is the vector intensity value from that object. We defined five labels to represent the movement intensity: stopped, slow, average speed, fast and very fast.

5 Results

We compared the results obtained with manual annotations made in a total of 415 obstacles over 100 frames, being part from the CaRINA dataset and part from the KITTI dataset. 20 sequences of frames were selected containing 5 frames each sequence. In Table 1 it is presented the general accuracy for each task in the extraction and analysis of the obstacles positioning and movement. A more individual analysis is possible by the Tables 2, 3, 4 and 5 where we presented, through the confusion matrices, the detailed results of each task and their labels.

In Table 2 the confusion matrix of the distance (depth) analysis of the obstacles in the scene, it is noticed that the worst results were with the labels “very-close” and “far”. However, it is also possible to verify that the biggest errors in both classes were in neighboring labels. Still, the “very-close” label featured a considerable amount of errors as “far”. This occurs in situations when there are objects close between each other
Table 1 Accuracy results.

| task                  | accuracy |
|-----------------------|----------|
| Depth                 | 81.75%   |
| x-axis Direction      | 89.51%   |
| y-axis Direction      | 83.57%   |
| Movement Intensity    | 80.96%   |

Table 2 Depth Confusion Matrix

|               | very close | close | far  | very far |
|---------------|------------|------|------|----------|
| very close    | 92.52%     | 17.48%| 9.71%| 0.00%    |
| close         | 1.00%      | 86.00%| 6.00%| 7.00%    |
| far           | 0.00%      | 0.00% | 67.00%| 33.00%   |
| very far      | 0.00%      | 0.00% | 0.00% |100.00%   |

Table 3 x-axis Direction Confusion Matrix

|               | left to right | right to left | stable direction |
|---------------|---------------|---------------|------------------|
| left to right | 86.90%        | 8.28%         | 4.83%            |
| right to left | 5.11%         | 92.05%        | 2.84%            |
| stable direction | 11.24%    | 0.00%         | 88.76%           |

Table 4 y-axis Direction Confusion Matrix

|               | approaching | moving away | stable distance |
|---------------|-------------|-------------|-----------------|
| approaching   | 78.40%      | 8.45%       | 13.15%          |
| moving away   | 0.00%       | 89.13%      | 10.87%          |
| stable distance| 5.16%      | 5.81%       | 89.03%          |

Table 5 Movement Intensity Confusion Matrix

|               | stopped | slow | average speed | fast | very fast |
|---------------|---------|------|---------------|------|-----------|
| stopped       | 90.41%  | 5.48%| 4.11%         | 0.00%| 0.00%     |
| slow average speed | 13.79%| 58.62%| 24.14% | 0.00%| 0.00%     |
| fast          | 0.00%   | 1.49%| 77.61%        | 20.90%| 0.00%     |
| very fast     | 0.00%   | 0.00%| 4.00%         | 26.00%| 70.00%    |

in the scene making it difficult to analyze as individual objects. The same occurs with the “approaching” label in Table 4, which has 8.45% being as “moving-away”. In Table 5 the worst result was in label “slow”, yet in all classes have errors occurring as being from neighboring labels.

In Figure 5 and Figure 6 we present some of the results obtained by our approach. In the left column we show the combined results obtained with the CNN, together with the Disparity Map and OF patterns. In the right column we present the labels on the objects based on the analysis of the depth and movement patterns of each object.

In the first row of Figure 5 the vehicle responsible for the capture of the images (capture source, CS) is stationary, and four cars are passing through the right lane. Here, using the patterns obtained by Disparity Map and OF it is possible to verify their behavior and how distant these four vehicles are, as highlighted in Figure 7. In comparison, traffic lights are identified as static. The second row of Figure 5 presents the continuity of the first row, with the CS still stationary and vehicles passing through the right lane.

The vehicles passing through the right lane, present as a result of direction the label right to left because even though it is not crossing abruptly in front of the CS, it is not an exactly parallel movement because the image perspective, by surpassing the CS it is like moving in the x-axis, from right to left. Considering this perspective, as lines going to meet at the vanishing point.

The third, fourth and fifth rows of Figure 5 are a sequence and presents vehicles with a trajectory that will generate an actual direct crossing. These vehicles are further away while the CS is standing behind another, nearby vehicle, which is also stationary (Figure 5 a).

The figures in the fourth and fifth rows of Figure 5 show a truck crossing the front of the vehicle from right to left. In the sixth row, we show the extraction and analysis of the patterns on a pedestrian very close to the CS and another vehicle more distant, both crossing the front of the CS in opposite directions (Figure 5 b). Different objects with different distances are shown on the seventh row.

In the last row of Figure 5 we present a sequence where the CS is slowing down, almost stopping, while several pedestrians begin to cross with similar but not synchronized behavior, resulting in some data variation.

In the five rows of Figure 6 showing results from Carina data-set, the CS is in motion and, although it presents images with less movement than in the KITTI data-set, it is still possible to observe the patterns of movement and distance from the detected objects. Mainly from first to third row, which correspond to a sequence.

6 Conclusion and discussion

Obstacle detection and recognition focused on ADAS and/or on Autonomous Vehicles navigation has made a major breakthrough in the state of the art in recent years, especially considering the advances in CNNs. The approach we present in this paper focuses on the next step after the detection and recognition of obstacles: the extraction of the depth and movement patterns of the detected objects.

We understand that identifying these patterns will allow a smarter and safer decision making in an ADAS or in an Autonomous Vehicle, helping to identify potential threats. Both providing for a more precise alert for
Fig. 5 Examples of results in the KITTI data-set.
a human driver, as well as passing more data to an intelligent agent module responsible for making decisions in an Autonomous Vehicle.

In our approach, we combine CNN-based detection and object recognition results with the depth patterns by a Disparity Map and movement patterns (direction and velocity) by an Optical Stream. The results obtained are promising and motivate the continuity of this research.

Fig. 6 Examples of results in the CaRINA data-set.

Fig. 7 Highlighted pattern analysis results.
6.1 Future work

One of our next steps with this approach consists of applying that same model, with the same combined methods in a NVIDIA Jetson card, provided by NVIDIA for our project through the NVIDIA GPU Grant Program, in a vehicle with real time image capture and a specially developed stereo rig. Thus, improving the performance of the current proposed flow.

In addition, one of the possibilities opened by the extraction of the distance, trajectory and movement patterns we are performing, is to try to predict the actions of the participants in the scene, such as other vehicles, cyclists, pedestrians and animals, performing threat assessment, which is a project that is already underway in our group.

In the context of these possible next steps we are also investigating the possibilities associated to the analysis of the obtained patterns, studying the creating of potential new behavior labels. During this next phases we plan to perform more experiments related to the obtained patterns analysis, differentiating for example the direction label in order to differentiate situations where in fact there will occur some crossing in front of the capture source from when it is a possible lateral overtaking.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

Bleyer M (2013) VU Stereo Vision. Karlsplatz, Vienna, Austria
Bouchafa S, Zavidovique B (2011) Obstacle detection “for free”, in the c-velocity space. In: 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC), pp 308–313, DOI 10.1109/ITSC.2011.6082872
Chanawangsa P, Chen CW (2013) A novel video analysis approach for overtaking vehicle detection. In: 2013 International Conference on Connected Vehicles and Expo (ICCVE), pp 802–807, DOI 10.1109/ICCVE.2013.6799901
Chen YS, Tsai AC, Lin TT (2012) Road environment recognition method in complex traffic situations based on stereo vision. In: 2012 12th International Conference on ITS Telecommunications, pp 180–184, DOI 10.1109/ITST.2012.6425161
Commission IE (2001) Safety of laser products. part 1: Equipment classification, requirements and user’s guide
Deo N, Rangesh A, Trivedi MM (2018) How would surround vehicles move? A unified framework for maneuver classification and motion prediction. CoRR abs/1801.06523, URL http://arxiv.org/abs/1801.06523
Dosovitskiy A, Fischer P, Ilg E, Husser P, Hazirbas C, Golkov V, v d Smagt P, Cremers D, Brox T (2015) Flownet: Learning optical flow with convolutional networks. In: 2015 IEEE International Conference on Computer Vision (ICCV), pp 2758–2766, DOI 10.1109/ICCV.2015.316
Farnebäck G (2003) Two-frame motion estimation based on polynomial expansion. In: Bigun J, Gustavsson T (eds) Image Analysis: 13th Scandinavian Conference, SCIA 2003 Halmstad, Sweden, June 29 – July 2, 2003 Proceedings, Springer Berlin Heidelberg, Berlin, Heidelberg, pp 363–370, DOI 10.1007/3-540-45103-X_50, URL https://doi.org/10.1007/3-540-45103-X_50
Fernandes LC, Souza JR, Pessin G, Shinzato PY, Sales D, Mendes C, Prado M, Klaser R, Magalhes AC, Hata A, Pigatto D, Branco KC, Jr VG, Osorio FS, Wolf DF (2014) Carina intelligent robotic car: Architectural design and applications. Journal of Systems Architecture 60(4):372 – 392, DOI 10.1016/j.sysarc.2013.12.003, URL http://www.sciencedirect.com/science/article/pii/S1383732313002841
Fleet DJ, Weiss Y (2005) Optical flow estimation
Geiger A, Lenz P, Stiller C, Urtasun R (2013) Vision meets robotics: The kitti dataset.
Giosan I, Nedevschi S (2014) Superpixel-based obstacle segmentation from dense stereo urban traffic scenarios using intensity, depth and optical flow information. In: 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), pp 1662–1668, DOI 10.1109/ITSC.2014.6957932

Gupta K, Upadhyay S, Gandhi V, Krishna KM (2017) Small obstacle detection using stereo vision for autonomous ground vehicle. In: Proceedings of the Advances in Robotics, ACM, New York, NY, USA, AIR ’17, pp 25:1–25:6, DOI 10.1145/3132446.3134889, URL http://doi.acm.org/10.1145/3132446.3134889

Huang Y, Liu S (2016) Multi-class obstacle detection and classification using stereovision and improved active contour models. IET Intelligent Transport Systems 10(3):197–205, DOI 10.1049/iet-its.2014.0308

Hue C, Heng L, Lee GH, Fraundorfer F, Furgale P, Sattler T, Pollefeys M (2017) 3d visual perception for self-driving cars using a multi-camera system: Calibration, mapping, localization, and obstacle detection. Image and Vision Computing 68:14 – 27, DOI 10.1016/j.imavis.2017.07.003, URL http://www.sciencedirect.com/science/article/pii/S0262885617301117

Ilg E, Mayer N, Saikia T, Keuper M, Dosovitskiy A, Brox T (2017) Florenet 2.0: Evolution of optical flow estimation with deep networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 1647–1655, DOI 10.1109/CVPR.2017.179

Jawed S, Boumaiza E, Grabocka J, Schmidt-Thieme L (2019) Data-driven vehicle trajectory forecasting. CoRR abs/1902.05400, URL http://arxiv.org/abs/1902.05400

Kitchenham B, Charters S (2007) Guidelines for performing systematic literature reviews in software engineering

Li J, Chen M (2014) On-road multiple obstacles detection in dynamical background. In: 2014 Sixth International Conference on Intelligent Human-Machine Systems and Cybernetics, vol 1, pp 102–105, DOI 10.1109/IHMSC.2014.33

LiDAR-UKcom (2015) LiDAR. UK, URL http://www.lidar-uk.com/

Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL (2014) Microsoft coco: Common objects in context. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T (eds) Computer Vision – ECCV 2014, Springer International Publishing, Cham, pp 740–755

Liu S, Huang Y, Zhang R (2014) Obstacle recognition for adas using stereovision and snake models. In: 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), pp 99–104, DOI 10.1109/ITSC.2014.6957673

Lucas BD, Kanade T (1981) An iterative image registration technique with an application to stereo vision. In: Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, IJCAI’81, pp 674–679, URL http://dl.acm.org/citation.cfm?id=1623264.1623280

Mitzel D, Floros G, Sudowe P, van der Zander B, Leibe B (2011) Real time vision based multi-person tracking for mobile robotics and intelligent vehicles. In: Jeschke S, Liu H, Schilberg D (eds) Intelligent Robotics and Applications, Springer Berlin Heidelberg, Berlin, Heidelberg, pp 105–115

Prabhakar G, Kailath B, Natarajan S, Kumar R (2017) Obstacle detection and classification using deep learning for tracking in high-speed autonomous driving. In: 2017 IEEE Region 10 Symposium (TENSYMP), pp 1–6, DOI 10.1109/TENCONSpring.2017.8069972

Rateke T, von Wangenheim A (2018) Systematic literature review for passive vision road obstacle detection. Tech. rep., Brazilian Institute for Digital Convergence - INCoD, DOI 10.13140/RG.2.2.10198.14408

Rateke T, von Wangenheim A (2020) Passive vision road obstacle detection: a literature mapping. International Journal of Computers and Applications 0(0):1–20, DOI 10.1080/1206212X.2020.1758877, URL https://doi.org/10.1080/1206212X.2020.1758877

Ren S, He K, Girshick R, Sun J (2015) Faster r-cnn: Towards real-time object detection with region proposal networks. In: Proceedings of the 28th International Conference on Neural Information Processing Systems.
Shinzato PY, dos Santos TC, Rosero LA, Ridel DA, Massera CM, Alencar F, Batista MP, Hata AY, Osorio FS, Wolf DF (2016) Carina dataset: An emerging-country urban scenario benchmark for road detection systems. In: 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), pp 41–46, DOI 10.1109/ITSC.2016.7795529

STANDARD AN (2005) American National Standard for Safe use of Lasers Outdoors. Orlando, FL

Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 2818–2826, DOI 10.1109/CVPR.2016.308

Urmson C, Anhalt J, Bae H, Bagnell JAD, Baker CR, Bittner RE, Brown T, Clark MN, Darms M, Demitrish D, Dolan JM, Duggins D, Ferguson D, Galatali T, Geyer CM, Gittleman M, Harbaugh S, Hebert M, Howard T, Kolski S, Likhachev M, Litkouhi B, Kelly A, McNaughton M, Miller N, Nickolaou J, Peterson K, Pilnick B, Rajkumar R, Rybski P, Sadekar V, Salesky B, Seo YW, Singh S, Snider JM, Struble JC, Stentz AT, Taylor M, Whittaker WRL, Wolkowicki Z, Zhang W, Ziglar J (2008) Autonomous driving in urban environments: Boss and the urban challenge. Journal of Field Robotics - 2007 DARPA Urban Challenge 25(8):425–466

Yao Y, Xu M, Choi C, Crandall DJ, Atkins EM, Darioush B (2018) Egocentric vision-based future vehicle localization for intelligent driving assistance systems. CoRR abs/1809.07408, URL http://arxiv.org/abs/1809.07408