Saliency Detection in a Virtual Driving Environment for Autonomous Vehicle Behavior Improvement

Csaba Antonya, Florin Gîrbacia(✉), Cristian Postelnicu, Daniel Voinea, and Silviu Butnariu

Transilvania University of Brasov, 29 Eroilor, 500036 Brasov, Romania
{antonya,garbacia,cristian-cezar.postelnicu,daniel.voinea,butnariu}@unitbv.ro

Abstract. To make the best decisions in real-world situations, autonomous vehicles require learning algorithms that process a large number of labeled images. This paper aims to compare the automatically generated saliency maps with attention maps obtained with an eye-tracking device in order to provide automated labeling of images for the learning algorithm. To simulate traffic scenarios, we are using a virtual driving environment with a motion platform and an eye-tracking device for identifying the driver's attention. The saliency maps are generated by post-processing the driver's view provided by the front camera.

Keywords: Driving simulator · Autonomous vehicle · Saliency map

1 Introduction

Autonomous vehicle development in the last decade has increased substantially. All major car manufacturers have launched their own semi o fully autonomous vehicles and they continuously work on extending the range, reducing energy consumption, reducing the time to market and also developing new simulation techniques. Software and hardware-in-the-loop simulations of autonomous vehicles are often overlooking the human factor [1]. A driver is expected to react based on his experience to different tracks and environmental conditions (variations in weather, tire degradation, fuel consumption). The testing of virtual cars in a multi-modal virtual environment is an important step in the validation process of new concepts and technologies. Virtual driving environments (VDE) are providing the user realistic feedback regarding the required visual, auditory, haptic, and kinesthetic information. The most common way of imposing the motion of the driving simulator is by the 6 degrees of freedom Stewart hexapod platform. A driving simulator with realistic interaction, operating environment and feedback eliminates the difficulties of the road test but allows the understanding of driving behavior, testing driver assistant systems, and traffic research.

Driving and especially safe driving is a collection of competencies that are acquired, refined, automated and maintained. Driving behavior can also be influenced by the

L. T. De Paolis et al. (Eds.): AVR 2021, LNCS 12980, pp. 511–518, 2021.
https://doi.org/10.1007/978-3-030-87595-4_37

driver's desire for smooth driving [2]. The main parameters proposed in the literature to assess driver behavior are the longitudinal and the lateral accelerations [3], which can be reproduced accurately with the VDE. Driver behavior can be modeled as a dynamic system in a phase transition framework as changes in the physiological system [4] and is also correlated with age, gender and sensation seeking [5]. The evaluation of driving scenarios is complex because it is subject on many closely interconnected variables depending not only on the different types of drivers but also on the road environment, the traffic characteristics and the categories of road infrastructure.

Machine learning and artificial intelligence are the cornerstones of autonomous vehicle development. The data for the construction of the model of the environment is provided by sensors like cameras, radar, and lidar. Complex driving maneuvers require a detailed model of the environment, and deep-learning algorithms based on image processing are of great importance. For training the neural networks, labeled images are required, which can be obtained manually or automatically from a driving simulator. Attention allocation analysis and prediction of the user on the driving scene can help the image labeling process [6].

In this paper, we are proposing a verification metric for the automated labeling of images using the saliency map comparing it with the attention map obtained with an eye-tracking device. For this, we are proposing a VDE with a motion platform and an eye-tracking device. The front-camera image (the driver's view) is post-processed for obtaining the saliency map and this is compared with the attention map. We are interested in the accuracy of the automated salient region detection in case the decision made by the human driver in a specific traffic situation.

2 The Virtual Driving Environment

The proposed virtual driving environment is composed by a motion platform, a driving seat with pedals, a steering wheel and a Tobii eye tracking device [7]. The Stewart platform is the MOOG 6 DOF 2000E, which is a six degree of freedom motion platform (Fig. 1). The dynamic model of the platform was developed, analyzed and a co-simulation environment was proposed in [8]. The performance of a driving simulator is defined by the Motion Cueing Algorithm, which is a system of filters that takes into account the limits of the simulator as well as the threshold of the driver's motion perception to reproduce simulated vehicle acceleration.

In the VDE, the dynamic model of the vehicle and the visual feedback is provided by the CARLA simulator. CARLA is an open-source software platform, which is intended to be a system that includes individual projects developed to smooth the process of development, training and validation of autonomous management systems. The CARLA simulator consists of a scalable client-server architecture in which the server manages the simulation itself: sensor playback, physics calculation, updates on the state of the world and its actors and connects to client modules that control the logic of the actors on stage and set the conditions of the world, using as programming environments Python or C++ [9]. The basic structure of the CARLA simulator is composed of traffic management subsystem, sensors, recording subsystem, simulator integration subsystem in other learning environments, various libraries with maps, weather conditions and sets of actors, as well as a series of predefined routes and scenarios [9].

To highlight the simulation capabilities of the CARLA platform, in [10] three types of autonomous leadership are analyzed: a classic modular pipeline, an end-to-end model trained by imitation learning and an end-to-end model. to-end trained through hardened learning. Driving software was tested, testing various sensors (camera and LiDAR), using a real-time hardware-in-the-loop simulation system without constraints based on the CARLA platform [11]. The paper [12] proposes a complex method, which can achieve a self-driving scale-ball, which can manage massive car traffic scenarios (over 7,000 km traveled) using a high-fidelity driving simulator, respecting traffic rules and in a wide variety of environments (urban, rural, highway, narrow roads, roundabouts and pedestrian crossings).

In [13] it is proposed to study the quality of the results obtained artificially compared to the results obtained with the help of real sensors, in the field of object detection with LiDAR. Vehicle control activities were analyzed by precise decoding of motion intention using the BMI-VCS method - Integration of brain-machine interface (BMI) neurotechnology with vehicle control systems (VCS). [14] is studying the possibility that defective autonomous vehicles can be driven, in the event of a breakdown, with the help of tele-driving. This system is an extension of the CARLA open-source simulator, responsible for rendering the driving environment and ensuring an evaluation of the reproducible scenario.

Fig. 1. The proposed virtual driving environment

3 Visual Saliency Detection and Gaze Tracking

For the driver's visual system, certain parts of the driving scene present crucial information. These are the perceptually salient regions that contain semantically meaningful information.

Saliency detection is a process of location of important objects or regions in an image. The quality of the salient region labeling is very important since it will control the accuracy and the capability of the autonomous vehicle to find the right path in its surroundings. Image labeling and semantic segmentation can be completed manually or can be automatized. Manual image segmentation and labeling is a time-consuming process, because of the high volume of images in different driving scenarios. Automatic salient region detection can be bottom-up and top-down modes [15]. The bottom-up approach is fast, data-driven, and task-independent, while the top-down is based on supervised learning and are task-oriented.

Salient region retrieval is used in various filed like automotive or robotics. It is used in vehicle headlights detection with the region-of-interest segmentation method together with the pyramid histogram of oriented gradients features detection in a support vector machine classifier [16]. In [17] the saliency map is used in the prediction for making braking decisions. This application is using a deep neural network to predict salient features, then relate these with driving decisions. Dang et al. is proposing a visual saliency–aware receding horizon exploration for path planning of aerial robots with a two-step optimization paradigm [18].

Eye-tracking in the VDE is used to obtain the point of gaze, the spot on which the user is focusing. Eye-tracking devices were successfully used in drivers' testing in perceiving objects in the visual field [19] and to determine fatigue driving state [20].

Studies in eye movement during driving are showing that there are different salient regions on which the users are focusing their attention. In a survey on 40 subjects, Deng et al analyzed the eye-tracking data when viewing traffic images [21]. They concluded that the driver's attention was mostly concentrated on the end of the road in front of the vehicle.

Fig. 2. The virtual driving scenario implemented using Carla simulator

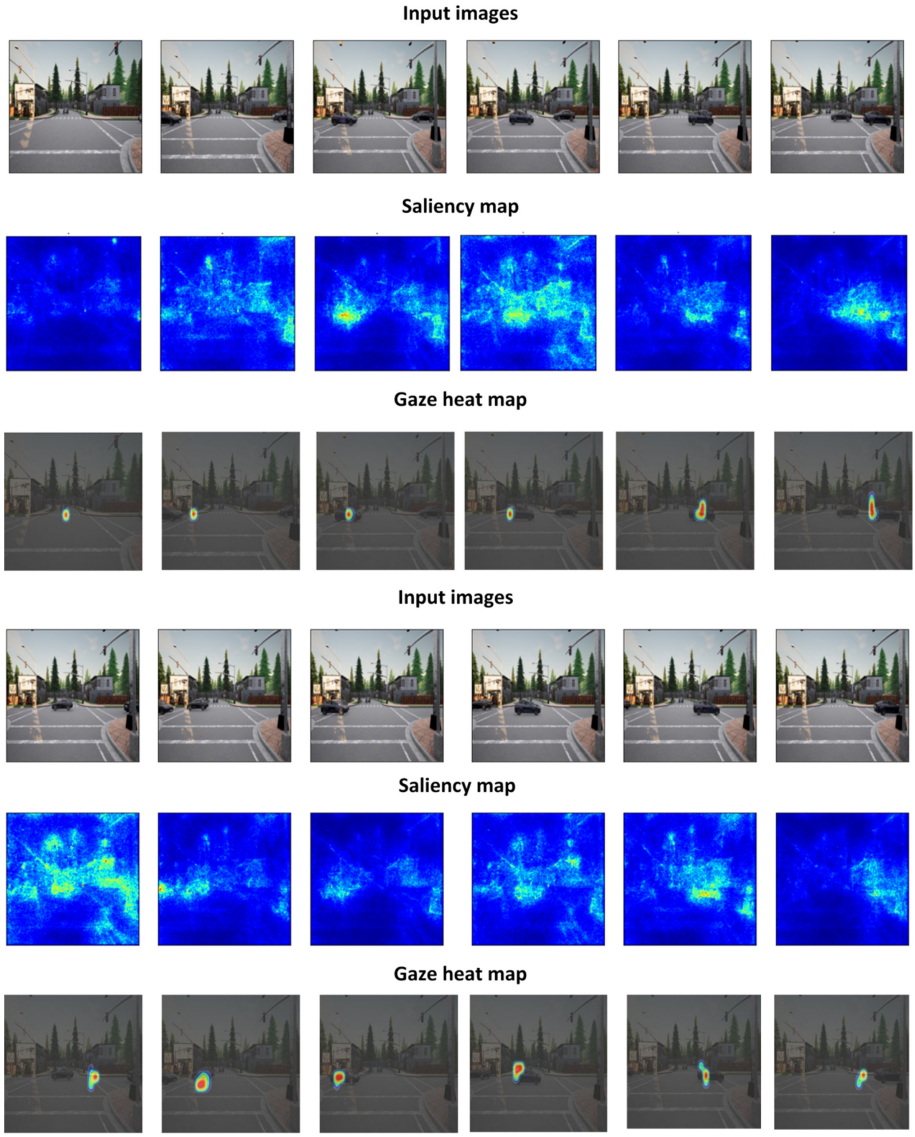

Fig. 3. Examples of saliency maps obtained using SmoothGrad and gaze heat maps generated from the input images recorded from virtual driving scenario

4 Experimental Setup for the Saliency Detection in Driving Scenario

In the Carla simulator we implemented the following scenario: the user is driving the ego-car along a secondary road, then the ego-car is reaching an intersection where the user is waiting for the possibility to turn right on the main road (Fig. 2).

A total of nine users were requested to perform the experiment. The participants ages were between 25 and 41 years old and none of them wore glasses during the experiment.

At the beginning of the experiment, they were asked to calibrate the Tobii eye tracking device by looking at 5 predefined points. Then each of them performed the experiment scenario. The eye movement was recorded for 10–20 s. On the main road, there is traffic from the main direction and also from the front road. The user's view (front camera image of the car) is recorded at the speed of 10 frames/s.

5 Comparative Evaluation of Saliency and User's Gaze

Visualizing saliency maps is used in order to detect relevant image regions (in our scenario: the cars in the traffic). To perform this stage of evaluation, the tf-keras-vis visualization toolkit [22] was used. This framework allows obtaining two types of saliency maps: vanilla saliency or SmoothGrad. SmoothGrad was used in this paper because the results obtained with the vanilla saliency map were noisy. SmoothGrad improves the visibility by sharpening the gradient-based saliency maps [23]. For the implementation a Convolutional Neural Network (CNN) model was used based on the pre-trained popular VGG16 model [24]. The input images had a fixed size of 224 $\times$ 224 pixels.

In order to display the distribution of each user's gaze fixations, we used heat maps generated by Eye Movements Metrics and Visualizations Toolbox [25].

After analyzing the saliency and heat maps (Fig. 3), we obtained an accuracy of 83.3% regarding the overlap of predicted relevant regions with the user's gaze fixation. In some cases, where the overlap did not occur, the car was absent or partially present in the image or there were reflecting lights that focused the user's gaze.

6 Conclusions

To prepare the future autonomous vehicles to deal with real-word situation, the learning algorithms require tremendous number of labeled images. Because different objects and subjective factors are present in images, one way of extracting the meaningful content of an image is to use an automated salient extraction algorithm. This is important also in advanced driver assistance systems for warning generation by situation awareness models. Drivers are using visual perception and are usually focusing their attention for decision-making on the main features of the scenery ahead, like the curvature of the road ahead, neighboring vehicles, bicycles, pedestrians and other obstacles. We used the popular pretrained VGG16 model and SmoothGrad for saliency map generation, which accomplished 83.3% accuracy. The interpretation of scene from an event-reasoning point of view using automated salient region detection is an important step in image labeling for autonomous vehicle's behavior training and improvement. In the future, we will apply the discussed method to create datasets for driving decisions based on saliency maps.

Acknowledgement. This work was supported by a grant of the Ministry of Research, Innovation and Digitization, CNCS/CCCDI – UEFISCDI, project number PN-III-P2-2.1-PED-2019-4366 within PNCDI III (431PED).

References

1. Riener, A., Jeon, M., Alvarez, I., Frison, A.K.: Driver in the loop: Best practices in automotive sensing and feedback mechanisms. In: Meixner, G., Müller, C. (eds.) Automotive user interfaces. HIS, pp. 295–323. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-49448-7_11
2. Wang, J., Sun, F., Ge, H.: Effect of the driver's desire for smooth driving on the car-following model. Physica A: Stat. Mech. Appl. **512**, 96–108 (2018)
3. Vaiana, R., et al.: Driving behavior and traffic safety: an acceleration-based safety evaluation procedure for smartphones. Mod. Appl. Sci. **8**(1), 88 (2014)
4. Mirman, J.H.: A dynamical systems perspective on driver behavior. Transp. Res. F: Traffic Psychol. Behav. **63**, 193–203 (2019)
5. Witt, M., Kompaß, K., Wang, L., Kates, R., Mai, M., Prokop, G.: Driver profiling–data-based identification of driver behavior dimensions and affecting driver characteristics for multi-agent traffic simulation. Transp. Res. F: Traffic Psychol. Behav. **64**, 361–376 (2019)
6. Deng, T., Yan, H., Qin, L., Ngo, T., Manjunath, B.S.: How do drivers allocate their potential attention? Driving fixation prediction via convolutional neural networks. IEEE Trans. Intell. Transp. Syst. **21**(5), 2146–2154 (2019)
7. Tobii homepage. https://www.tobii.com/. Accessed 20 Feb 2021
8. Antonya, Cs., Irimia, C., Grovu, M., Husar, C., Ruba, M.: Co-simulation environment for the analysis of the driving simulator's actuation. In: 7th International Conference on Control, Mechatronics and Automation (ICCMA), Delft, Netherlands, pp. 315–321 (2019)
9. CARLA - Open-source simulator for autonomous driving research, homepage. https://carla.org/. Accessed 5 May 2021
10. Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., Koltun, V.: CARLA: an open urban driving simulator. In: Conference on Robot Learning PMLR, pp. 1–16 (2017)
11. Brogle, C., Zhang, C., Lim, K.L., Bräun, T.: Hardware-in-the-loop autonomous driving simulation without real-time constraints. IEEE Trans. Intell. Veh. **4**(3), 375–384 (2019)
12. Cai, P., Wang, H., Sun, Y., Liu, M.: Learning scalable self-driving policies for generic traffic scenarios. arXiv preprint arXiv:2011.06775 (2020)
13. Dworak, D., Ciepiela, F., Derbisz, J., Izzat, I., Komorkiewicz, M., Wójcik, M.: Performance of LiDAR object detection deep learning architectures based on artificially generated point cloud data from CARLA simulator. In: 24th International Conference on Methods and Models in Automation and Robotics (MMAR), pp. 600–605 (2019).
14. Hofbauer, M., Kuhn, C.B., Petrovic, G. Steinbach, E.: TELECARLA: an open source extension of the CARLA Simulator for tele-operated driving research using off-the-shelf components. In: IEEE Intelligent Vehicles Symposium (IV), Las Vegas, USA (2020)
15. Xue, J.R., Fang, J.W., Zhang, P.: A survey of scene understanding by event reasoning in autonomous driving. Int. J. Autom. Comput. **15**(3), 249–266 (2018)
16. Shang, J., Guan, H.P., Liu, Y., Bi, H., Yang, L., Wang, M.: A novel method for vehicle headlights detection using salient region segmentation and PHOG feature. Multimedia Tools Appl. 1–21 (2021)
17. Aksoy, E., Yazıcı, A., Kasap, M.: See, attend and brake: an attention-based saliency map prediction model for end-to-end driving. arXiv preprint arXiv:2002.11020 (2020)
18. Dang, T., Papachristos, C., Alexis, K.: Visual saliency-aware receding horizon autonomous exploration with application to aerial robotics. In: 2018 IEEE International Conference on Robotics and Automation (ICRA), pp. 2526–2533 (2018)
19. Xu, J., Min, J., Hu, J.: Real-time eye tracking for the assessment of driver fatigue. Healthc. Technol. Lett. **5**(2), 54–58 (2018)

20. Kapitaniak, B., Walczak, M., Kosobudzki, M., Jozwiak, Z., Bortkiewicz, A.: Application of eye-tracking in drivers testing: a review of research. Int. J. Occup. Med. Environ. Health **28**(6), 941 (2015)
21. Deng, T., Yang, K., Li, Y., Yan, H.: Where does the driver look? Top-down-based saliency detection in a traffic driving environment. IEEE Trans. Intell. Transp. Syst. **17**(7), 2051–2062 (2016)
22. tf-keras-vis toolkit. https://github.com/keisen/tf-keras-vis. Accessed 11 Apr 2021
23. Smilkov, D., Thorat, N., Kim, B., Viégas, F., Wattenberg, M.: Smoothgrad: removing noise by adding noise. arXiv preprint arXiv:1706.03825 (2017)
24. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
25. Krassanakis, V., Filippakopoulou, V., Nakos, B.: EyeMMV toolbox: an eye movement post-analysis tool based on a two-step spatial dispersion threshold for fixation identification. J. Eye Mov. Res. **7**(1) (2014)