Person Independent, Privacy Preserving, and Real Time Assessment of Cognitive Load using Eye Tracking in a Virtual Reality Setup

Efe Bozkir*
Perception Engineering
University of Tübingen

David Geisler†
Perception Engineering
University of Tübingen

Enkelejda Kasneci‡
Perception Engineering
University of Tübingen

Abstract

Eye tracking is handled as key enabling technology to VR and AR for multiple reasons, since it not only can help to massively reduce computational costs through gaze-based optimization of graphics and rendering, but also offers a unique opportunity to design gaze-based personalized interfaces and applications. Additionally, the analysis of eye tracking data allows to assess the cognitive load, intentions and actions of the user. In this work, we propose a person-independent, privacy-preserving and gaze-based cognitive load recognition scheme for drivers under critical situations based on previously collected driving data from a driving experiment in VR including a safety critical situation. Based on carefully annotated ground-truth information, we used pupillary information and performance measures (inputs on accelerator, brake, and steering wheel) to train multiple classifiers with the aim of assessing the cognitive load of the driver. Our results show that incorporating eye tracking data into the VR setup allows to predict the cognitive load of the user at a high accuracy above 80%. Beyond the specific setup, the proposed framework can be used in any adaptive and intelligent VR/AR application.

Keywords: Eye tracking, cognitive load recognition, virtual reality, driving simulation.

Index Terms: Computing methodologies—Computer graphics—Graphics systems and interfaces—Virtual reality, Perception; Computing methodologies—Machine learning—Machine learning approaches—Classification and regression trees, Kernel methods; Human-centered computing—Human computer interaction (HCI)—Empirical studies in HCI

1 Introduction

Cognitive load is referred to as the amount of information processing activity that is imposed on working memory [6]. Cognitive load recognition is important and beneficial for many applications. It has been studied extensively in various domains, such as in education, psychology, or driving, since information on the cognitive load of an individual can be helpful to design user-adaptive interfaces. Various ways have therefore been proposed to assess the cognitive load of a subject, such as by means of N-back tasks (e.g., Appel et al. [2]), through the analysis of electroencephalography (EEG) signals (e.g., Zarjam et al. [22], Walter et al. [20]), by means of eye movements studies or through assessment of facial expressions (e.g., Hussain et al. [11]). Eye tracking offers a particularly non-invasive way of cognitive load assessment, especially through the measurement and analysis of the pupil diameter.

Meanwhile, eye tracking has also found its way into the driving domain, not only as a means to study driving behavior, but also as a powerful input modality for advanced driver assistance systems (e.g., Kübler et al. [14]) or even as a means of driver observation on context of automated driving (e.g. Braunagel et al. [3, 5]). Modern cars are already capable of tasks such as lane following, traffic sign and light detection, automated parking, and collision warning. However, the full autonomous driving task is still too complex without human input and guidance. For this reason, current cars employ a variety of multi-modal warning systems for many different purposes to ensure driving safety and provide smooth driving experience. Augmented reality (AR) and head-up-display (HUD) technologies have been used as interfaces to such systems both in practice and driving simulation research. In the following, we will briefly review related work in this context.

Many driving simulation studies have been conducted in driving simulators or virtual reality (VR) environments in order to analyze driving behavior, safety, performance and training using HUDs or virtual warnings. For example, HUDs for blind spot detection and warning were discussed in a related work by Kim et al. [12]. Tran et al. [19] addressed the usage and benefits of HUDs during left turns. Moreover, benefits and improvement of driving behavior for lane keeping using adaptive warnings were discussed by Dijksterhuis et al. [7]. The effect of improving bad driving habits using VR in a user-study was discussed by Lang et al. [15]. In the context of eye tracking and driving, there are several studies with various goals. For example, Konstantopoulos et al. [13] studied eye movements during day, night, and rainy driving in a driving simulator. Braunagel et al. [4] introduced a novel approach for driver activity recognition using head pose and eye tracking data. Furthermore, Braunagel et al. [5] proposed a classification scheme to recognize driver take-over readiness using gaze, traffic, and a secondary task in conditional automated driving. Pomarjanschi et al. [17] showed that gaze guidance reduced the number of pedestrian collisions in a driving simulator.

In the driving context, there are many studies that focus on cognitive load and driving. Engström et al. [8] analyzed the effect of cognitive load on driving performance and found out that the effects of cognitive load on driving are task dependent. Yoshida et al. [21] proposed an approach to classify driver cognitive load to improve in-vehicle information service using real world driving data. Gabaude et al. [10] conducted a study in a driving simulator to understand the relationship between mental effort and driving performance using cardiac activity, driving performance and subjective data measurements. Mizoguchi et al. [16] proposed an approach to identify cognitive load of the driver using inductive logic programming with eye movement and driver input measurements in real driving situations. Fridman et al. [9] proposed a scheme to estimate cognitive load in a 3-class problem in the wild for driving scenarios using convolutional neural networks.

Driving simulation studies for safety critical situations using warnings and cognitive load recognition in driving exist in the literature. However, it is still an open question whether it is possible to recognize cognitive load of the driver in safety-critical situations and especially when the driver is confronted with visual gaze-aware warnings. In order to tackle this issue, we used the data collected using a VR setup from our previous work [1] where drivers encountered a dangerously crossing pedestrian in an urban road. In order to keep

*e-mail: efe.bozkir@uni-tuebingen.de
†e-mail: david.geisler@uni-tuebingen.de
‡e-mail: enkelejda.kasneci@uni-tuebingen.de

2019 IEEE Conference on Virtual Reality and 3D User Interfaces
23-27 March, Osaka, Japan
978-1-7281-1377-7/19/$31.00 ©2019 IEEE

the situation safety critical, Time-to-Collision (TTC) between driving vehicle and crossing pedestrian was kept $1.8sec < TTC < 5sec$. Rasouli et al. [18] discussed that in this range of TTC, there is a high likelihood that pedestrian or joint attention between driver and pedestrian happens. However, if it does not happen, the outcome can be fatal. Our study was conducted with 16 participants. Half of them received gaze-aware pedestrian warning cues, whereas the other half did not receive any cue.

In our proposed scheme, the cognitive load of the drivers are recognized using critical and non-critical time frames of driving for each participant. Since critical time frames are very short, we kept non-critical time frames also short in order to have a uniform distribution in the training data. We trained multiple classifiers and evaluated them leave-one-person-out fashion in order to obtain person-independent results. Furthermore, since the time frames that are used in training and testing are very short, they do not reflect the complete intention of driver during driving. Therefore, we obtained a privacy-preserving scheme. In addition to person-independence and privacy-preserving features, our system also works in real time, which brings the opportunity to implement the same system in real life.

In general, when the cognitive load of the driver is recognized in a safety critical situation, visual cues and support can be adapted accordingly in order to provide safer, smoother, and less stressful driving experience even in very risky situations. In this work, we focused on a proof-of-concept in the driving scenario due to its highly dynamic and uncertain nature. However, our results show that the same methodology can be applied to any adaptive and gaze-aware application, especially in VR/AR.

2 Proposed Approach

Since the proposed system depends on the driving data which were collected using a VR setup, Section 2.1 describes first the VR setup and the collected data from our previous work [1]. Then in Section 2.2, data processing, training, and testing procedures are discussed.

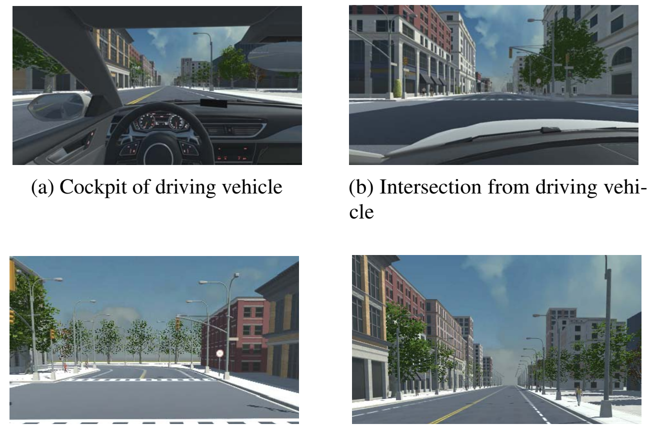
2.1 VR Setup and Environment

In our previous work [1], we conducted a user-study to evaluate safety during driving in VR.

The hardware setup was created using HTC-Vive, Logitech G27 Steering Wheel and Pedals, Phillips headphones and Pupil-Labs HTC-Vive Binocular Add-on. Figure 1 shows the dedicated setup.

Figure 1: Experimental setup for VR

The hardware setup was used in a virtual environment created using Unity3D. We used 3D models from Urban City Pack, City Park Exterior, and Traffic Sign Sets packages to design the virtual city. Since we had not only critically crossing pedestrians, but also other pedestrians, we used Modern People asset packages for pedestrian models. Vehicle models and helper scripts were obtained from Realistic Car HD02, Traffic Cars, and Realistic Car Controller asset packages. Lastly, Playmaker and Simple Waypoint System packages were used to make pedestrian and vehicle movements smoother. For the eye tracking measures, Pupil Service version 1.7 of open source hmd-eyes[1] from Pupil-Labs was used. Examples scenes from VR environment are shown in Figure 2.



(a) Cockpit of driving vehicle

(b) Intersection from driving vehicle

(c) Intersection

(d) Main road

Figure 2: Example scenes from VR environment

The user-study consisted of acclimation and data collection phases. In the acclimation phase, no data was collected. In the data collection phase, participants encountered with a dangerously crossing pedestrian. Two critical pedestrians on the side walk of main road were generated. In the beginning of the experiment, one of them was marked as crossing pedestrian. This critical pedestrian started crossing the road when the distance between driving vehicle and pedestrian was ($d_{critical} \approx 45m$). Every participant encountered with critically crossing pedestrian due to the start position of the vehicle in the data collection phase. They had the opportunity to speed up or slow down until the pedestrian crossing. Half of the participants started observing critical pedestrian warning cues around the pedestrian model with red color ≈ 32 meters in advance to pedestrian crossing. These parameters helped to keep TTC as $1.8s < TTC < 5s$, since the speed limit of main road was $90km/h$. Participants were supposed to realize the speed limit via speed signs on the road. Otherwise, the vehicle was equipped with maximum speed warning on a small in-car board. The pedestrian cues were made gaze-aware and were deactivated when gaze signal of the driver was closer than 5 meters to pedestrian for ≈ 0.85 seconds. Gaze signal on 2D canvas was obtained from Pupil Service from Pupil-Labs and then mapped from 2D to 3D with the help of ray-casting and Unity colliders. The hyper-parameters were determined empirically. The measurements, which changed over the time, were recorded in real time and were available for offline analysis. Since the pupil diameter values are very important for recognizing cognitive load, we post-processed pupil diameter measurements to remove the noise and normalize the data. For smoothing and normalization, we applied Savitzky-Golay filter and divisive baseline correction using a baseline duration of 0.5 seconds respectively.

Corresponding setup and experiments were run on a PC equipped with NVIDIA Titan X graphics card with 12GB memory, a 3.4GHz Intel i7-6700 processor and 16GB of RAM.

[1]https://github.com/pupil-labs/hmd-eyes

2.2 Cognitive Load Recognition

The data we obtained from the experiment (mentioned in Section 2.1) is not annotated with regard to the cognitive load levels. Therefore, we first annotated our data with two levels of cognitive load: Low and high. We set $t_{critical}$ for both with-and without-pedestrian cue scenarios. The purpose of $t_{critical}$ is to separate the time domain into low and high cognitive load levels. It is taken as $t_{warning}$ and $t_{movement}$ for with-and without-warning scenarios respectively. The reason of taking two different $t_{critical}$ values is that cognitive load of the drivers who receive critical pedestrian cues starts increasing from $t_{warning}$, whereas cognitive load of others who do not receive any cue increases after the start of pedestrian movement.

In order to find the time frames to annotate exactly, we applied T-test using the pupil diameter data of each participant between $[t_{critical} - \delta t, t_{critical}]$ and $[t_{critical}, t_{critical} + \delta t]$. We used pupil diameter measurements due to the fact that pupil diameter is one of the main indicators of cognitive load. Once a significant difference in T-test was found with $p < 0.05$, we assumed that we found a proper δt value. In order to keep the distributions significantly different but rather close to each other, we did not accept distributions where $p < 0.01$. Since cognitive load also depends on biological factors, which do not happen immediately, we shifted $t_{critical}$ by $+\theta t_{shift} = 0.8s$. In the end, we annotated each frame in the dedicated time frames with low or high cognitive load as it is shown in Table 1:

Table 1: Cognitive load annotations for time-frames

Time Frame	**Cognitive Load**
$[t_{critical} + \theta t_{shift} - \delta t, t_{critical} + \theta t_{shift}]$	Low
$[t_{critical} + \theta t_{shift}, t_{critical} + \theta t_{shift} + \delta t]$	High

In order to recognize cognitive load of the driver, we trained different classifiers including Support Vector Machines (SVM), decision trees, random forests, and k-Nearest Neighbors (k-NN) using each frame. For the feature set, we used pupil diameters, and driver inputs on accelerator and brake pedals and steering wheel. Min-max normalization was applied to input data. In order to make our approach person-independent, we evaluated the data of each driver against the trained model using rest of the drivers. For example, in order to evaluate the first participant, we trained classifiers with other 15 participants and then evaluated the first participant using the data and its labels. This approach assures that we obtain person-independent results in the end.

Offline analyses offer many insights from the collected data. However, real time working capability is as important as the accuracy of the system especially in VR/AR fields. With this motivation, we evaluated whether our proposed scheme is capable of working in real time.

3 Results

In the following, we report results of our automated cognitive load recognition and its real time working capabilities that was conducted using MATLAB on a PC which is equipped with NVIDIA GeForce GTX 1070 mobile graphics card with 8GB of RAM, a 2.2GHz Intel i7-8750 processor, and 32GB of RAM.

In our dataset, there are 1171 frames in total and from each frame, maximum four features were used in training and testing. In addition, there are ≈ 73 frames (Mean) per participant (SD=12.5). We trained SVM, decision trees, random forests, and k-NN and tested according to the discussed setup in Section 2.2 and used different combinations of features along with pupil diameter. We observed that using steering wheel input of driver did not lead to more accurate recognition. Since participants did not need to change steering wheel angle too much during the encountered scenarios, it is acceptable. Taking into account that cognitive load does not change very dramatically in short amount of time and each participant was evaluated against the trained models using the rest of the participants, the cognitive load recognition results are reasonable. The highest accuracy of 80.7% was achieved by SVM. Adding more training data and participants has a great potential to increase the accuracy of predictions. Accuracy, precision, recall, and F1-score results which were obtained using these classifiers and feature set of pupil diameter and driver inputs on accelerator and brake are shown in Table 2.

Table 2: Results of cognitive load recognition

Method	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Support Vector Machine	0.8070	0.7671	0.8574	0.8098
Decision Tree	0.7344	0.7332	0.7005	0.7165
Random Forest	0.7436	0.7372	0.7230	0.7299
1-Nearest Neighbor	0.6968	0.6846	0.6809	0.6828
5-Nearest Neighbor	0.7566	0.7473	0.7433	0.7453
10-Nearest Neighbor	0.7882	0.7947	0.7522	0.7729

During the training of classifiers which are mentioned in Table 2, we set some hyper-parameters. For SVM, we used linear kernel function. For k-NN approach, we evaluated 1-NN, 5-NN and 10-NN. The accuracy results increase by increasing the k value. For random forest classifier, we used five trees to train for classification purposes.

Since it is important to apply the proposed approach in real life scenarios, we evaluated whether cognitive load recognition can be done in real time. Table 3 shows the mean time spent for one prediction in each method.

Table 3: Evaluation of mean prediction durations

Method	**Mean Prediction Duration (ms)**
Support Vector Machine	0.319
Decision Tree	0.305
Random Forest	5.42
1-Nearest Neighbor	0.741
5-Nearest Neighbor	0.742
10-Nearest Neighbor	0.764

It is clear that all methods can be used for real time purposes. However, under this setup, it is reasonable to use SVM due to its higher accuracy and low prediction duration. In addition, if the dataset size increases, the real time working capability of k-NN is affected negatively. The same applies when the number of trees in random forests is increased.

4 Conclusion and Discussion

We proposed a scheme to recognize cognitive load of the drivers in safety critical situations using data collected during a driving study in VR. The scheme is person-independent because it generalizes well cross-subject. With more training data, there is a high potential for this scheme to work in a generic way. If person-specific setup is requested, the same scheme can be applied by adjusting the training data. In this case, even a more accurate cognitive load recognition can be obtained.

Due to the fact that we concentrated on very short time frames, complete driving data of participants were not exposed and only small amount of frames was used in training and testing. Only, the pupil diameter measurements were baseline-corrected using the first 0.5 seconds of driving. Therefore, it is a privacy-preserving scheme. Lastly, our scheme is capable of working real time. This outcome is very important and means that same scheme can be used in real driving studies and vehicles. It will enable more adaptive and intelligent feedbacks and inputs in driver warning systems; and eventually lead to safer and smoother driving experiences. We strongly suggest that similar schemes should be applied to real vehicles.

While this study is in driving domain, the outcome shows that our approach can be applied in similar adaptive user studies in VR and AR fields. The results indicate that there is a unique opportunity to design eye-tracking enabled interfaces and applications. Since we think that eye tracking has a great potential to transform VR and AR into another level, the outcome is valuable.

Despite the advantages and reasonable outcomes, there are some limitations as well. Firstly, since data were collected under VR setup, there is a likelihood that drivers do not behave naturally in VR. Virtual environment, weight of Head-Mounted-Display (HMD), or different dynamics of pedals or steering wheels can cause different behaviors than the real life. While we assume that participants became familiar with these in the acclimation phase, one should not ignore this possibility. Secondly, since the safety critical situations during driving happen in very short amount of time, it is difficult to collect big data in this context both using simulations or in real world.

As future work, more data and features can be used. There is a high likelihood that the accuracy of cognitive load recognition increases with more data. The same scheme can be applied to real driving simulators along with safety critical scenarios. Therefore, the findings in VR experiment can be compared with the future driving simulator experiments in terms of cognitive load recognition. Secondly, using raw eye videos along with other extracted features can be used to train deep models to estimate cognitive load. Furthermore, markov models or recurrent neural networks can be used to predict the cognitive load since they are suitable for time dependent data.

References

[1] Author names and title omitted due to the blind review. *Under submission.*

[2] T. Appel, C. Scharinger, P. Gerjets, and E. Kasneci. Cross-subject workload classification using pupil-related measures. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, ETRA '18, pp. 4:1–4:8. ACM, New York, NY, USA, 2018. doi: 10.1145/3204493.3204531

[3] C. Braunagel, D. Geisler, W. Rosenstiel, and E. Kasneci. Online recognition of driver-activity based on visual scanpath classification. *IEEE Intelligent Transportation Systems Magazine*, 9(4):23–36, 2017.

[4] C. Braunagel, E. Kasneci, W. Stolzmann, and W. Rosenstiel. Driver-activity recognition in the context of conditionally autonomous driving. In *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, pp. 1652–1657, Sept 2015. doi: 10.1109/ITSC.2015.268

[5] C. Braunagel, W. Rosenstiel, and E. Kasneci. Ready for take-over? a new driver assistance system for an automated classification of driver take-over readiness. *IEEE Intelligent Transportation Systems Magazine*, 9(4):10–22, winter 2017. doi: 10.1109/MITS.2017.2743165

[6] C. Cavanaugh. *Augmented Reality Gaming in Education for Engaged Learning*. 01 2010. doi: 10.4018/9781609601959.ch103

[7] C. Dijksterhuis, A. Stuiver, B. Mulder, K. A. Brookhuis, and D. de Waard. An adaptive driver support system: User experiences and driving performance in a simulator. *Human Factors*, 54(5):772–785, 2012. PMID: 23156622. doi: 10.1177/0018720811430502

[8] J. Engström, G. Markkula, T. Victor, and N. Merat. Effects of cognitive load on driving performance: The cognitive control hypothesis. *Human Factors*, 59(5):734–764, 2017. PMID: 28186421. doi: 10.1177/0018720817690639

[9] L. Fridman, B. Reimer, B. Mehler, and W. T. Freeman. Cognitive load estimation in the wild. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pp. 652:1–652:9. ACM, New York, NY, USA, 2018. doi: 10.1145/3173574.3174226

[10] C. Gabaude, B. Baracat, C. Jallais, M. Bonniaud, and A. Fort. Cognitive load measurement while driving. in : Human factors : a view from an integrative perspective. 10 2012.

[11] M. S. Hussain, R. A. Calvo, and F. Chen. Automatic cognitive load detection from face, physiology, task performance and fusion during affective interference. *Interacting with Computers*, 26:256–268, 2014.

[12] H. Kim, X. Wu, J. L. Gabbard, and N. F. Polys. Exploring head-up augmented reality interfaces for crash warning systems. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '13, pp. 224–227. ACM, New York, NY, USA, 2013. doi: 10.1145/2516540.2516566

[13] P. Konstantopoulos, P. Chapman, and D. Crundall. Driver's visual attention as a function of driving experience and visibility. using a driving simulator to explore drivers' eye movements in day, night and rain driving. *Accident Analysis & Prevention*, 42(3):827 – 834, 2010. Assessing Safety with Driving Simulators. doi: 10.1016/j.aap.2009.09.022

[14] T. C. Kübler, E. Kasneci, W. Rosenstiel, U. Schiefer, K. Nagel, and E. Papageorgiou. Stress-indicators and exploratory gaze for the analysis of hazard perception in patients with visual field loss. *Transportation Research Part F: Traffic Psychology and Behaviour*, 24:231–243, 2014.

[15] Y. Lang, L. Wei, F. Xu, Y. Zhao, and L. Yu. Synthesizing personalized training programs for improving driving habits via virtual reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 297–304, March 2018. doi: 10.1109/VR.2018.8448290

[16] F. Mizoguchi, H. Ohwada, H. Nishiyama, and H. Iwasaki. Identifying driver's cognitive load using inductive logic programming. In F. Riguzzi and F. Železný, eds., *Inductive Logic Programming*, pp. 166–177. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[17] L. Pomarjanschi, M. Dorr, and E. Barth. Gaze guidance reduces the number of collisions with pedestrians in a driving simulator. *ACM Trans. Interact. Intell. Syst.*, 1(2):8:1–8:14, Jan. 2012. doi: 10.1145/2070719.2070721

[18] A. Rasouli, I. Kotseruba, and J. K. Tsotsos. Agreeing to cross: How drivers and pedestrians communicate. *CoRR*, abs/1702.03555, 2017.

[19] C. Tran, K. Bark, and V. Ng-Thow-Hing. A left-turn driving aid using projected oncoming vehicle paths with augmented reality. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '13, pp. 300–307. ACM, New York, NY, USA, 2013. doi: 10.1145/2516540.2516581

[20] C. Walter, W. Rosenstiel, M. Bogdan, P. Gerjets, and M. Spüler. Online eeg-based workload adaptation of an arithmetic learning environment. *Frontiers in Human Neuroscience*, 11:286, 2017. doi: 10.3389/fnhum.2017.00286

[21] Y. Yoshida, H. Ohwada, F. Mizoguchi, and H. Iwasaki. Classifying cognitive load and driving situation with machine learning. *International Journal of Machine Learning and Computing*, 4:210–215, 06 2014. doi: 10.7763/IJMLC.2014.V4.414

[22] P. Zarjam, J. Epps, and F. Chen. Characterizing working memory load using eeg delta activity. In *2011 19th European Signal Processing Conference*, pp. 1554–1558, Aug 2011.