Exploiting the potential of eye movements analysis in the driving context

Christian Braunagel, Wolfgang Stolzmann, Daimler AG

Enkelejda Kasneci, Thomas C. Kübler, Wolfgang Fuhl, Wolfgang Rosenstiel
Computer Engineering, University of Tübingen

1 Abstract

Driving is a complex and highly visual task. With the development of high-end eye-tracking devices, numerous studies over the last two decades have investigated eye movements of the driver to identify deficits in visual search patterns and to derive assistive, informative, and entertainment systems. However, little is known about the visual behavior during autonomous driving, where the driver can be involved in other tasks but still has to remain attentive in order to be able to resume control of the vehicle.

This work aims at exploiting the potential of eye movement analysis in the autonomous driving context. In a pilot study, we investigated whether the type of the secondary task in which the driver is involved, can be recognized solely from the eye movement parameters of the driver. Furthermore, we will discuss several applications of eye movement analysis to future autonomous driving approaches, e.g., to automatically detect whether the driver is being attentive and – when required – to guide her visual attention towards the driving task.

2 Introduction

The eye movements of the driver have been investigated in numerous studies with different purposes [Bergasa et al. (2006), Kasneci (2013), Mourant and Rockwell (1972), Pérez et al. (2010), Smith et al. (2003), Underwood et al. (2003)]. However, little is known about the visual behavior of the driver during fully automated driving, where the automated driving system is expected to performs the driving task [Brandenburg and Skottke (2014)]. In such a scenario, the driver can be engaged with other tasks, but has to remain attentive in order to be able to resume control of the vehicle within shortest time [Jamson et al. (2013), Merat et al. (2014)]. According to a recent study [Merat et al. (2014)] employing eye movements analysis, vehicle and eye-tracking measures drivers need ~15 seconds to resume control and up to 40 seconds to stabilize vehicle control. Other authors have reported that the driver is able to resume control over the vehicle within 10 seconds [Petermann-Stock et al. (2013)]. Clearly, take-over time depends on the driver's current attentive state at the time of the take-over request. Drivers that are aware of the current driving situation are likely to resume control faster. Although there is no consensus on the take-over time during fully automated driving, the above findings clearly underline the need for new paradigms on how to keep the drivers engaged and attentive to the driving task and on how to inform them of their obligation to resume control over the vehicle.

This work focuses on the question whether eye movement measures can be used to automatically detect the type of secondary task in which the driver is being involved.

This knowledge can be then used to estimate the driver's level of attentiveness and, consequently, as input to an informative system that recaptures the driver's visual attention and directs it towards the driving task.

This paper is organized as follows. Section 3 introduces a workflow for the automated recognition of the type of the secondary task based solely on the eye movements of the driver. This workflow was evaluated on eye-tracking data derived from an autonomous driving experiment in a driving simulator. The evaluation results are presented in Section 4. Section 5 discusses the developed method regarding its limitations and potential and gives an overview over other application scenarios. Section 6 concludes this paper.

3 Methods

To automatically detect secondary tasks during autonomous driving, we developed the work-flow presented in Figure 1. [Bulling et al. (2011)] proposed a similar work-flow to distinguish between five different tasks: copying a text, reading a printed paper, taking handwritten notes, watching a video and browsing the Web.

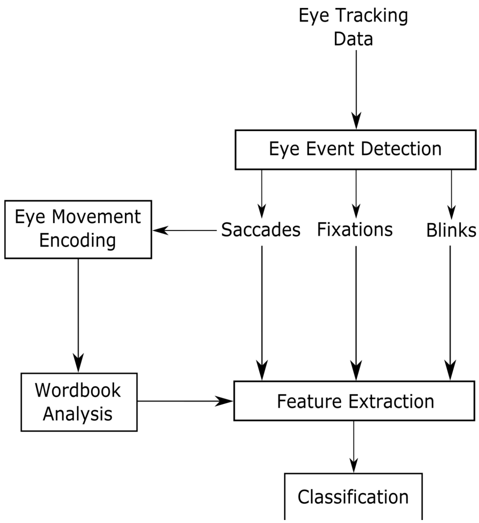

Figure 1: Adapted architecture for detecting secondary tasks

In contrast to the approach in [Bulling et al. (2011)], where electrooculography (EOG) was used to record eye movements of a subject, we recorded the eye movements by means of a mobile Dikablis video based eye tracker (Ergoneers GmbH, Manching). Compared to the EOG recording technique, video based eye-trackers are more comfortable to wear, but they come along with challenges concerning the quality of the eye-tracking signal. Poor signal quality becomes a bottleneck, especially during on-road driving, where due to changing illumination conditions robust gaze position estimation of the driver is challenging. The reliable, image-based detection of the pupil is, however, an essential prerequisite for the development of gaze-based assistive or informative systems.

3.1 Pupil Detection

The multitude of different sources of noise that are present during driving render accurate signal acquisition a challenging problem, e.g., changing illumination conditions, reflections on driver's glasses, individual features of the eye, etc. Our approach is based on edge filtering and refinement.

The algorithm is based on decision rules and consists of two main calculation paths. One path handles images in which the pupil is expected to be bright (Step 2.2 in Figure 3). The second path handles dark images (Step 2.3 in Figure 3).

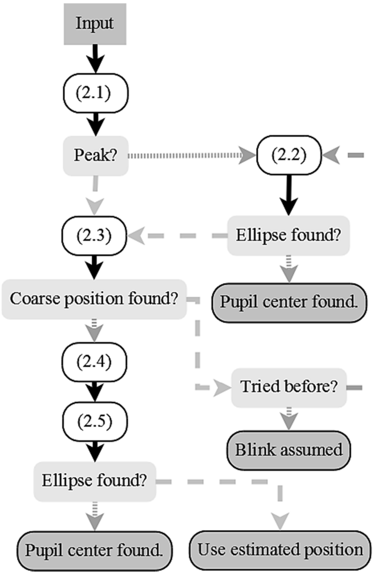

Figure 2: The work-flow for pupil detection. Decision points are represented by light gray cells, dark gray stands for termination, and white for processing steps. Dashed arrows represent a NO decision, others a YES.

The decision of the calculation path is based on an intensity histogram of the image (Step 2.1 in Figure 3). The ratio of bright bins (intensity value >200, where the threshold value 200 is set empirically) and the remaining bins of the histogram is first calculated. If the algorithm expects the pupil to be bright, it tries to find an edge belonging to the pupil. Therefore, a Canny edge filter is applied to the image (Figure 2 (a)).

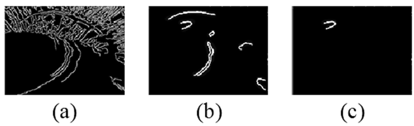

Figure 3: (a) Canny edge filtered image, (b) remaining curved edges and (c) selected curved edge.

All thin edges are then removed by analyzing each edge pixel's angle relative to its neighbor edge pixels. The remaining edges are thinned and orthogonal superimposed edges are separated using morphologic pixel operations. All remaining edge pixels now belong either to a curved or to a straight line. We expect edges belonging to the pupil border to be strongly curved. To remove straight lines, the distance of each edge pixel to the edge's centroid is calculated. Edges with a small distance towards their centroid are straight and will be removed. From the remaining curved lines (Figure 2(b)), the longest line with the lowest interior intensity value is chosen (Figure 2(c)).

For pupil center estimation, a least squares ellipse fit is calculated using the positions of the line pixels. In case no adequate pupil boundary can be determined, the second calculation path is applied (Step 2.3 in Figure 3). For images of normal and dark average intensity the algorithm calculates an intensity threshold using the variance and mean of the intensity values. The threshold is then applied on the image and four intensity histograms rotated by 0, 45, 90, and 135 degree are calculated (via angular integral projection function [Mohammed et al. (2012)]). For each histogram, the highest value over a range of bins closest to the center of the image is chosen. The intersections between the histograms for 0 and 90 degree as well as 45 and 135 degree are calculated. A coarse pupil center position is estimated as the mean between these intersections. To refine this position, a small area around the coarse position is extracted from the image. The centroid for all pixels in this area with intensity lower or equal to their neighbors is calculated and used as updated position estimation (Step 2.4 in Figure 3).

(a) (b) (c) (d) (e) (f)

Figure 4: (a) Threshold image, (b) canny edge filtered image, (c) threshold border, (d) edge pixels near threshold border, (e) remaining curved lines, and (f) curved line with ray hits as white dots.

This optimization step is important because the following calculations require the position estimation to be contained within the pupil area. A threshold and an edge filter image is created for the area around the pupil center estimation (Figure 4(a) and (b)) using a higher threshold than in the previous steps. For the threshold image the border is calculated (Figure 4(c)) by inspecting the neighbors of each pixel. These are used to remove edge pixels in the edge image which are not close to the border (Step 2.5 in Figure 3). The resulting edge image (Figure 4(d)) is filtered using the steps described in the calculation path for bright images except for the best edge selection step. All remaining edge pixels belong to curved lines describing the pupil boundary (Figure 4(c)). Starting from the coarse position eight rays are cast out with an angular step width of 45 degree (this step is inspired by the Starburst algorithm [Li et al. (2005)]). If the rays hit a curved line (Figure 4 (f)) all pixel positions from this line are collected and the ray is stopped. All collected pixel positions are used to calculate a least squares ellipse fit and therefore estimate the pupil center.

3.2 Eye Event Detection

Among all types of eye movements, fixations, saccades, and smooth pursuits are the most studied in driving scenarios [Kasneci et al. (2015)]. During a fixation the gaze position is kept relatively stable on an area of interest (AOI). Fixations usually have a duration of about 200−300 ms, although much longer fixations are possible. The duration of fixations varies depending on the visual task. Furthermore, fixation durations show inter- and intra-subject variability. In contrast, saccades correspond to rapid eye movements enabling the retinal part of sharpest vision (fovea) to fixate different areas of the scene. They occur at a maximum frequency of about 4Hz (e.g., during reading) and maximum velocity of approximately 500°/s [Land and Tatler (2009)]. Smooth pursuits occur whenever the eye follows a moving target, usually at velocities of approximately 15°/s.

The reliable detection of fixations, saccades, and smooth pursuits from the eye-tracking protocol is a crucial step towards the development of assistive and informative systems that are based on the driver's gaze. Before extraction of such events, the eye-tracking signal is preprocessed, i.e., blinks and invalid data points are not considered for event detection.

Eye blinks are usually not explicitly detected but modeled from the data. Sequences where the eye-tracker was unable to detect a pupil can be labeled as a blink event of a certain duration t if the following threshold criteria is fulfilled:

$$th_{min} \leq t \leq th_{max} \; with \; th_{min} = 0.1\, s\, , th_{max} = 0.4\, s \qquad \textbf{(1)}$$

The threshold values from (1) were chosen according to [Milo et al. (2010)] and represent the average minimum and maximum duration of an eye blink.

The event detection algorithms, i.e., the automated recognition of fixations, saccades, and smooth pursuits, consists of two steps: In the first step, a Bayesian online mixture model is employed to distinguish saccades from data points that might represent fixations and smooth pursuits [Tafaj et al. (2012), Kasneci et al. (2014)]. In a second step, a fine-grained Principal Component Analysis (PCA), revealing the main axes of variance, is conducted to distinguish fixations from smooth pursuits [Tafaj et al. (2013)].

The sequence of eye movements performed during driving is also known as visual scanpath. Visual scan patterns are highly individual and vary with the task and scene.

3.3 Feature Extraction and Classification

Since our aim was to produce independent data samples, we selected the same 90 features as in [Bulling et al. (2011)] but without the use of a sliding window. Instead, we calculated the features over each complete sequence of the performed secondary tasks. The features fall into four categories: saccade features, blink features, fixation features, and wordbook features (i.e., sequences of saccades). All features were normalized to the range [0 … 1].

To classify the type of the secondary task, we employed a Support Vector Machine (SVM) as implemented by the library LIBSVM using a radial basis function kernel [Chang and Lin (2011)]. To perform a multi-class classification, LIBSVM uses the "one-against-one" approach [Knerr et al. (1990)].

4 Experimental Evaluation

4.1 Data Collection

We evaluated our approach on data collected during an autonomous driving experiment with 57 drivers (female ⁄ male=27⁄30, mean age=37.7 years) in the Mercedes-Benz Driving Simulator. The drive of 30 minutes was highly autonomous on a two-lane highway. Furthermore, the driver was instructed to perform secondary tasks, i.e., listening to music, reading news articles and watching movies, between four different take-over situations. The order and number of tasks varies and can be seen as randomly distributed among the drivers. To provide a more comfortable interface, the news articles and movies were presented on a touchscreen integrated in the driver's cabin. Eye movements of the driver were recorded by means of a Dikablis eye tracker at a sampling rate of 25Hz. Three additional video cameras installed in the driver's cabin recorded every action of the driver and the current content of the touchscreen.

4.2 Tasks

Aim of this study was to evaluate whether three different tasks, namely watching a video, listening to music, and reading a document, can be classified based on the eye movements of the driver. Indeed, such tasks are considered as typical tasks that might be performed by the driver while driving highly autonomous. Although watching a video and reading a document are both visual tasks, we consider them separately since the level of engagement differs between these tasks. On the other hand, listening to music is an auditory task, where we do not expect any special patterns in the eye movements. Thus, listening to music could also be considered as "no activity", in terms of no specific visual processing is required.

4.3 Evaluation and Results

Before training the model and evaluating the proposed architecture we preprocessed the data samples. It needs to be ensured that we use a balanced data set for training and a driver independent evaluation for testing. For this purpose, we applied the approach outlined in Figure . We evaluated in terms of the leave-one-out cross-validation technique, i.e., by taking one driver for the testing phase and the remaining 56 drivers to train the classifier.

Since the number of performed tasks varied over all subjects, i.e., there is not the same duration of video watching, listening to music, and reading sequences, we obtained an imbalanced data set. Since this would strongly influence the feature calculation, all recorded situations with duration less than 120 second were not considered further. On the other hand, sequences with a duration longer than 120 seconds were cut down to 120 seconds. In the next step, we equalized the amount of sequences per task by setting the number of sequences per task to the minimum over all tasks. Following these steps, we obtained a balanced training data set with an average number of 100 sequences per task with equal durations for training our classifier.

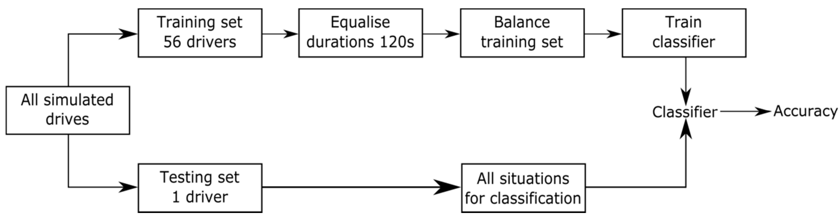

Figure 5: Structure for a balanced data sample and a driver independent evaluation

The result of the evaluation in terms of the detection accuracy is presented in Table 1. The proposed algorithmic workflow is capable of detecting reading sequences at an average accuracy of over 80%. This is associated with the structure of visual patterns during reading, namely repetitive patterns of short fixations followed by saccades of short amplitude. The recognition of the music and movie tasks is in contrast less accurate. While “listening to music” could be detected at an accuracy of 73%, the task of “watching a movie” showed a very low detection rate of 59% accuracy. This is due to a large amount of movie sequences misclassified as music sequences. The reason behind it becomes clear from the analysis of the additional driver recordings in the driving simulator. Many subjects performed control fixations, i.e., fixations towards the road in order to check the current driving scene and, consequently, did not focus on the presented movie for a relevant amount of time. Thus, their eye movement patterns during “listening to music” and “watching a movie” are more similar the longer and more frequent the control fixations occur.

	Predicted Movie Sequence	Predicted Music Sequence	Predicted Reading Sequence
Movie	59%	34%	7%
Music	19%	73%	8%
Reading	10%	8%	82%

Table 1: Average accuracy of secondary task detection for 57 subjects.

To separate the movie and music sequence more reliably, we performed an additional analysis based on these control fixations. More specifically, we looked at the saccadic amplitudes during the secondary tasks, which show a distribution into the three distinct clusters as depicted in Figure .

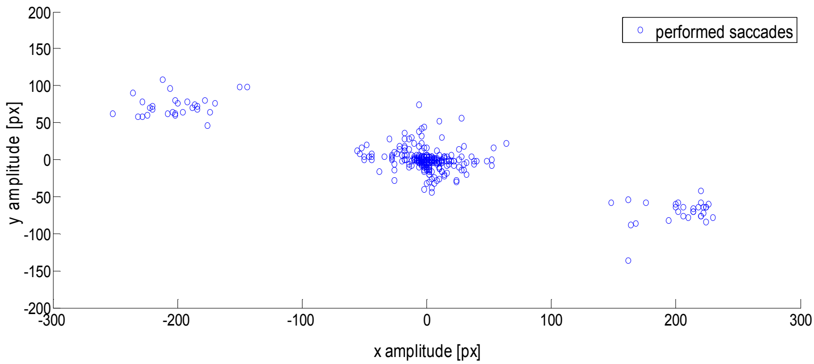

Figure 6: Saccade amplitude while watching a movie

The two outer clusters in Figure consist of large saccades. The saccades of the top left cluster can be assigned to saccades starting at the touchscreen and ending on the street. Saccades at the bottom right cluster redirect the driver’s gaze from the driving scene towards the touchscreen.

We analyzed all data samples that were originally classified as music sequences and tried to extract the cluster patterns by means of the DBSCAN (density-based spatial clustering of applications with noise) algorithm [Ester et al.(1996)]. If such a pattern was found in the extracted clusters, we re-classified the current music sequence as a movie sequence. The results are shown in Table 2.

	Predicted Movie Sequence	**Predicted Music Sequence**	**Predicted Reading Sequence**
Movie	69%	24%	7%
Music	22%	70%	8%
Reading	10%	8%	82%

Table 2: Classification accuracy with additional clustering.

The percentage of correctly classified movie sequences increased from 59% to 69% while the correctly classified music sequences decreased only by 3%. In total 119 music sequences, 148 movie sequences and 98 reading sequences have been classified correctly. By means of the above clustering, we were able to correctly classify 15 additional movie sequences while misclassifying four music sequences.

5 Discussion

Especially during the last two decades, several studies have investigated the visual behavior of drivers to derive driver assistance or information systems. However, less is known about the visual behavior when driving highly autonomous. In a highly autonomous driving scenario, the driver will be able to draw his attention to secondary tasks, such as reading, writing, etc. Such tasks, however, will bind the driver's cognitive resources and may interfere with the capability of the driver to resume control over the vehicle within shortest time, e.g., in case of a system failure.

An arising challenge in this context will be the question on how to keep the driver attentive to the driving task and scene without disturbing the secondary task. We believe that the investigation of the visual behavior of the driver can give precious insights to face this challenge. In the present work we studied whether eye movement measures can be used to automatically detect the type of secondary task in which the driver is being involved.

We presented a workflow to best process eye-tracking data and classify secondary tasks based on eye movement parameters. We found that the analysis of eye movement parameters allows an accurate recognition of the type of secondary task. The presented method was evaluated on a pool of 57 drivers. Despite the large number of sessions where secondary tasks were performed, an evaluation with more subjects is still required. Beyond this, for the detailed analysis of further eye movement parameters (e.g., frequency of smooth pursuits) mobile eye trackers with higher sampling rates would have to be employed. Although the driving simulator allows a very realistic driving experience, further evaluation in on-road experiments will also give better insights into the present research question. Overall, our results are quite promising and underlie the need for more investigations in this realm.

Future research will extend the variety of secondary tasks and also include session where no defined secondary task is conducted by the driver. Hence, we will be able to study whether there is a difference in gaze patterns during “hearing music” and “no activity”. Furthermore, in our future work we will including features that are not time-integrated but derived from complex, repetitive gaze patterns, e.g., shoulder checks or repeated short saccades towards the right side during reading a line of text, followed by a large saccade to the left when jumping to the next line [Kübler et al. (2014)]. Such features are likely to contribute to higher classification rates, especially with regard to a broader range of secondary tasks.

6 Conclusion

Our findings indicate that observing the eye movements of the driver can be a powerful means towards the automated estimation of the driver’s level of attention in fully automated driving. Although preliminary in nature, our results show that the secondary task with which the driver is engaged can be recognized by analyzing gaze-related parameters. The current challenge is to efficiently use such knowledge to derive new paradigms of informative systems, especially on how to best recapture the driver’s visual attention and direct it towards the driving task.

References

1. Bergasa, L. M., Nuevo, J., Sotelo, M. A., Barea, R., and Lopez, M. E. (2006). Real-time system for monitoring driver vigilance. *Intelligent Transportation Systems, IEEE Transactions on*, *7*(1), 63–77.

2. Brandenburg, S., and Skottke, E. M. (2014). Switching from manual to automated driving and reverse: Are drivers behaving more risky after highly automated driving? In *Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on* (pp. 2978–2983). IEEE.

3. Bulling, A., Ward, J. A., Gellersen, H., and Troster, G. (2011). Eye movement analysis for activity recognition using electrooculography. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *33*(4), 741–753.

4. Chang, C. C., and Lin, C. J. (2011). LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *2*(3), 27.

5. Ester, M., Kriegel, H. P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *KDD* 96, pp. 226–231).

6. Jamson, A. H., Merat, N., Carsten, O. M., and Lai, F. C. (2013). Behavioural changes in drivers experiencing highly-automated vehicle control in varying traffic conditions. *Transportation Research Part C: Emerging Technologies*, *30*, 116–125.

7. Kasneci, E. (2013). Towards the Automated Recognition of Assistance Need for Drivers with Impaired Visual Field. *PhD thesis, University of Tübingen, Wilhelmstr. 32, 72074,* Tübingen.

8. Kasneci, E., Kasneci, G., Kübler, T. C., and Rosenstiel, W. (2014). The applicability of probabilistic methods to the online recognition of fixations and saccades in dynamic scenes. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 323–326). ACM.

9. Kasneci, E., Kasneci, G., Kübler, T. C., and Rosenstiel, W. (2015). Online Recognition of Fixations, Saccades, and Smooth Pursuits for Automated Analysis of Traffic Hazard Perception. In *Artificial Neural Networks* (pp. 411–434). Springer International Publishing.

10. Knerr, S., Personnaz, L., and Dreyfus, G. (1990). Single-layer learning revisited: a stepwise procedure for building and training a neural network. In *Neurocomputing* (pp. 41–50). Springer Berlin Heidelberg.

11. Kübler, T. C., Kasneci, E., and Rosenstiel, W. (2014). SubsMatch: Scanpath similarity in dynamic scenes based on subsequence frequencies. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 319–322). ACM.

12. Land, M. F. and Tatler, B. W. (2009).*Looking and acting: vision and eye movements in natural behaviour*. Oxford University Press.

13. Li, D., Winfield, D., and Parkhurst, D. J. (2005). Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches. In *Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on* (pp. 79–79). IEEE.

14. Merat, N., Jamson, A. H., Lai, F. C., Daly, M., and Carsten, O. M. (2014). Transition to manual: Driver behaviour when resuming control from a highly automated vehicle. *Transportation researchpart F: Traffic psychology and behaviour*, *27*, 274–282.

15. Milo, R., Jorgensen, P., Moran, U., Weber, G., and Springer, M. (2010). BioNumbers—the database of key numbers in molecular and cell biology. *Nucleicacids research*, *38*(suppl 1), D750–D753.

16. Mohammed, G. J., Hong, B. R., and Jarjes, A. A. (2012). Accurate pupil features extraction based on new projection function. *Computing and Informatics*, *29*(4), 663–680.

17. Mourant, R. R., and Rockwell, T. H. (1972). Strategies of visual search by novice and experienced drivers. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *14*(4), 325–335.

18. Pérez, A., Garcia, M. I., Nieto, M., Pedraza, J. L., Rodríguez, S., and Zamorano, J. (2010). Argos: an advanced in-vehicle data recorder on a massively sensorized vehicle for car driver behavior experimentation. *Intelligent Transportation Systems, IEEE Transactions on, 11*(2), 463–473.

19. Petermann-Stock, I., Hackenberg, L., Muhr, T., and Mergl, C. (2013). Wie lange braucht der Fahrer --eine Analyse zu Übernahmezeiten aus verschiedenen Nebentätigkeiten während einer hochautomatisierten Staufahrt. In *6. Tagung Fahrerassistenzsysteme. Der Weg zum automatischen Fahren.*TÜV SÜD Akademie GmbH.

20. Smith, P., Shah, M., and da Vitoria Lobo, N. (2003). Determining driver visual attention with one camera. *Intelligent Transportation Systems, IEEE Transactions on*, *4*(4), 205–218.

21. Tafaj, E., Kasneci, G., Rosenstiel, W., and Bogdan, M. (2012). Bayesian online clustering of eye movement data. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 285–288). ACM.

22. Tafaj, E., Kübler, T. C., Kasneci, G., Rosenstiel, W., and Bogdan, M. (2013). Online classification of eye tracking data for automated analysis of traffic hazard perception. In *Artificial Neural Networks and Machine Learning–ICANN 2013* (pp. 442–450). Springer Berlin Heidelberg.

23. Underwood, G., Chapman, P., Brocklehurst, N., Underwood, J., and Crundall, D. (2003). Visual attention while driving: sequences of eye fixations made by experienced and novice drivers. *Ergonomics*, *46*(6), 629–646.