Brightness- and Motion-Based Blink Detection for Head-Mounted Eye Trackers

Tobias Appel
Perception Engineering
University of Tübingen, Germany
tobias.appel@student.uni-tuebingen.de

Thiago Santini
Perception Engineering
University of Tübingen, Germany
thiago.santini@uni-tuebingen.de

Enkelejda Kasneci
Perception Engineering
University of Tübingen, Germany
enkelejda.kasneci@uni-tuebingen.de

Abstract
Blinks are an indicator for fatigue or drowsiness and can assist in the diagnose of mental disorders, such as schizophrenia. Additionally, a blink that obstructs the pupil impairs the performance of other eye-tracking algorithms, such as pupil detection, and often results in noise to the gaze estimation. In this paper, we present a blink detection algorithm that is tailored towards head-mounted eye trackers and is robust to calibration-based variations like translation or rotation of the eye. The proposed approach reached 96,35% accuracy for a realistic and challenging data set and in real-time even on low-end devices, rendering the proposed method suited for pervasive eye tracking.

Author Keywords
Blink detection; Pervasive eye tracking, Real time; Image processing

ACM Classification Keywords
I.5.4 [PATTERN RECOGNITION]: Applications; I.4.8 [IMAGE PROCESSING AND COMPUTER VISION]: Object recognition,Shape; I.4.9 [IMAGE PROCESSING AND COMPUTER VISION]: Applications

Introduction
A blink is a rapid closing and opening of the eyelids that falls within three classes: endogenous, reflex, and volun-

Figure 1: Eyelid movement during an endogenous blink, which typically lasts for 75 to 400 ms [4]. For a low-end eye tracker (≈ 25 frames per second), this results in approximately 4 to 16 frames.

tary movements [21]. Endogenous (or spontaneous) blinks serve to spread the tear film over the cornea and remove irritants.

Reflex blinks originate from the startle reflex to protect the eye from external stimuli. For these two classes, blinks usually range from 75 to 400 ms [4]. In contrast, voluntary blinks are performed consciously, can be used for interaction – e.g., in human-computer interfaces (HCI) [10] – and have no determined duration patterns. Apart from these biological functions, unusual blink patterns are also indicative of a person's state of vigilance, fatigue, and drowsiness [15, 25, 22]. Such states are specially important in situations that require quick reactions, e.g., during driving [3]; in this context, real-time blink detection combined with pervasive eye tracking has the potential to prevent dangerous and life-threatening circumstances. Furthermore, blinks are a significant source of noise for eye-tracking algorithms. For instance, there is a trade-off between detecting pupils in realistic and challenging scenarios and false positives during blink (when no pupil is visible). Moreover, mid-blink the pupil becomes partially occluded causing pupil detection algorithms to bias the pupil center towards the still visible part; as a result, blinks must be taken into account during the automatic classification of eye movements [19]. Thus, a robust and accurate blink detection algorithm enables not only the employment of blink-related data (e.g., frequency) but also circumvents the noise introduced by blinks in other eye-tracking algorithms.

Due to the many advantages that head-mounted eye trackers offer – e.g., mobility and unintrusivnes – they are promising candidates for pervasive eye tracking, and, thus, this work focuses on these eye trackers. An algorithm for use in head-mounted eye trackers has different requirements than one for remote eye tracking. There is no need for head or eye localization, but the exact location, alignment, and angle of the eye in the video depends on the eye camera position, which varies significantly from subject to subject. This makes it uncertain, where to expect the pupil or eyelids and mostly prohibits the use of any priors in an algorithm. Motion bluring and frame skips also pose problems. The former renders blink detection based on edges almost impossible, and the latter can significantly distort a blink sequence. Further challenges are added by reflections, uncommon angles, and illumination changes.

In this work, we propose a brightness and motion based algorithm that runs in real-time in systems equivalent to those used in state-of-the-art eye tracking systems and does not rely on prior information. The fact that a sequence is analyzed in contrast to a frame-by-frame approach makes our algorithm robust to frame skips and the brightness-based detection does not rely on edges of any kind. Furthermore, we introduce a new labeled data set of realistic and challenging images from on-road driving experiments. To foster further research in the field and effortless replication of our result, we contribute the algorithm implementation and data sets openly at:
www.ti.uni-tuebingen.de/perception

Related work

The great majority of video-based blink identification concerns remote eye trackers or regular cameras. In a first stage, a plethora of methods, such as Viola-Jones [24] and KLT trackers [23], are used to identify and track the subject's face and eyes region. As previously mentioned, in this work we focus on head-mounted eye trackers. On one hand, these devices impose extra constraints on the blink identification task. For instance, since near-infrared images are employed, no color information is available. Furthermore, the orientation of the eye image and eye corner positions are not known a priori. On the other hand, head-mounted devices do not require the aforementioned stage; thus, henceforth we discuss related work assuming this initial stage is performed appropriately and eye boxes have been identified correctly.

Smith et al. [20] first identify the *eye-white color* as the brightest pixel in the eye region on an initial frame; further frames are classified on whether *eye-white pixels* exist in the eye region (non-blinks) or not (blinks). Grauman et al. [7] employ an open eye template and correlation scores to determine whether a frame contains an open or closed eye (based on a predetermined threshold). Ito et al. [9] divide the input image into vertical sections and, for each section, find the pair of maximal and minimal intensity derivatives most distant from the darkest point in the section. The candidates from five sections are grouped, and two groups are estimated to represent the upper and lower eyelid; the average distance between the upper and lower eyelid points is used to measure the degree of closure. A threshold then discriminates between blinks and open eyes. Moriyama et al. [16] rely on the average illumination intensity for the upper and lower halves of the eye region. Crossings between these values are employed to determine when the eyelid crosses the line separating these regions and, thus, blinks. Morris et al. [17] use a mean image and variance map to detect blinks; these are updated with each new frame, and the resulting variance maps is thresholded. If the number of pixels remaining is larger in relation to the eye box, a blink is assumed. Lalonde et al. [13] employ scale-invariant feature transform to identify tracking points. The optical-flow of the points inside the eye region is then employed to identify when the eyelid descends and ascends. Bavivarov et al. [1] models the eyes through an active appearance model and define a blink criteria based on the ratio between the resulting eye width and aperture. Lee et al. [14] first normalize the eye region illumination to account for illumination variations. The cumulative difference of black pixels in a binarized version of the normalized eye region and the ratio between the eye height/width are used as features for a support vector machine classifier that discriminates between open and closed eyes. Drutarovsky and Fogelton [5] employ a flock of KLT trackers within the eye region, which are used to evaluate the average motion of nine equally sized cells in the region. The variance of the superior six cells drive a finite state machine that identifies downward and upward eyelid movements. Jiang et al. [11] consider images provided by head-mounted eye trackers. Their approach consists with thresholding the difference between two subsequent frames. The resulting image is then morphologically opened, and pupil and eyelids identified. Blink onset is determined based on eyelid position changes between the two consecutive frames, whereas blink offset is identified based on pupil size changes during an ongoing blink.

Method

The nature of our data leads to an approach that does not rely on edge detection or prior knowledge about the location of the pupil or shape of the eye. It is based on two simple assumptions: the pupil is dark and gets at least partly obstructed by the eyelid during a blink. Those two facts can be exploited if we look at the brightness of consecutive frames. During the blink onset, the frame brightness steadily increases, reaching its maximum at the blink apex. Afterwards, frame brightness decreases during the blink offset until it approaches a level similar to the one prior the blink (see Figure 2).

Percentile values serve as a measure of brightness in our algorithm. In addition, differences between consecutive frames are used to determine if there is enough change to

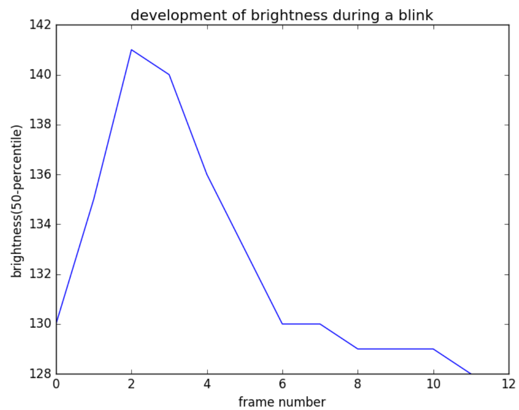

Figure 2: The brightness development for a typical blink.

classify a sequence as a blink or not. For this, the frames f_i and f_{i-1} were blurred to reduce noise, and the absolute difference between them are calculated and summed up. Both together form the features of a single frame. In the following formula, c and r denote column number and row number respectively, and $b(f_i)$ is the blurred version of frame f_i.

$$\mathrm{diff}_{i,i-1} = \sum_{r,c} |b(f_i)(r,c) - b(f_{i-1})(r,c)| \quad (1)$$

$$\mathrm{feature}_i = (P_{percentile}(f_i), \mathrm{diff}_{i,i-1}) \quad (2)$$

These features are calculated for k consecutive frames, which together make up the feature vector for a window of size k. Figure 3 illustrates the feature extraction procedure.

$$\mathrm{feature}_{i-k,i} = (\mathrm{feature}_{i-k}, \ldots, \mathrm{feature}_i) \quad (3)$$

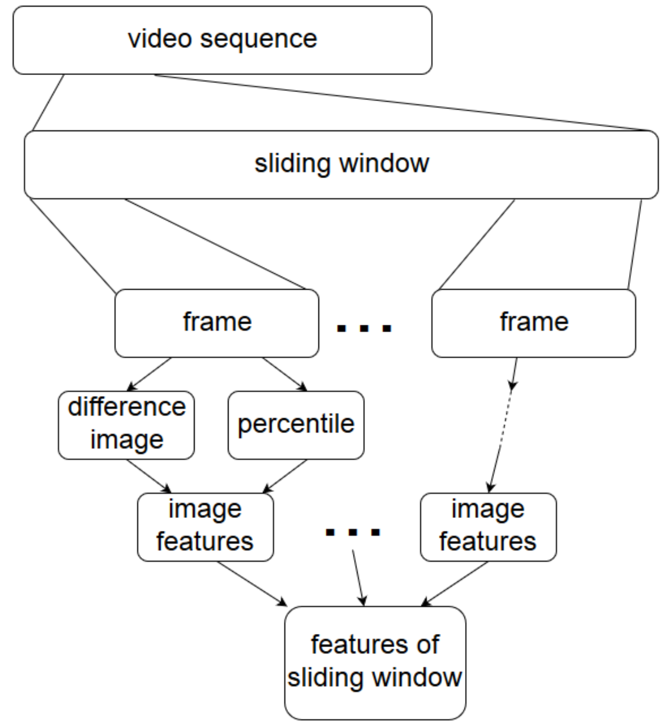

Figure 3: Feature extraction process

Even though blinks may vary in length, their basic structure remains similar. For a small amount of frames the eyelid descends, whereas ascension takes a larger amount of frames since eyelid velocity is higher during blink onset. Thus, it is reasonable to assume a fixed window size that appropriately models this behavior exists. Different choices for the chosen percentile and the window size k are discussed in the evaluation section.

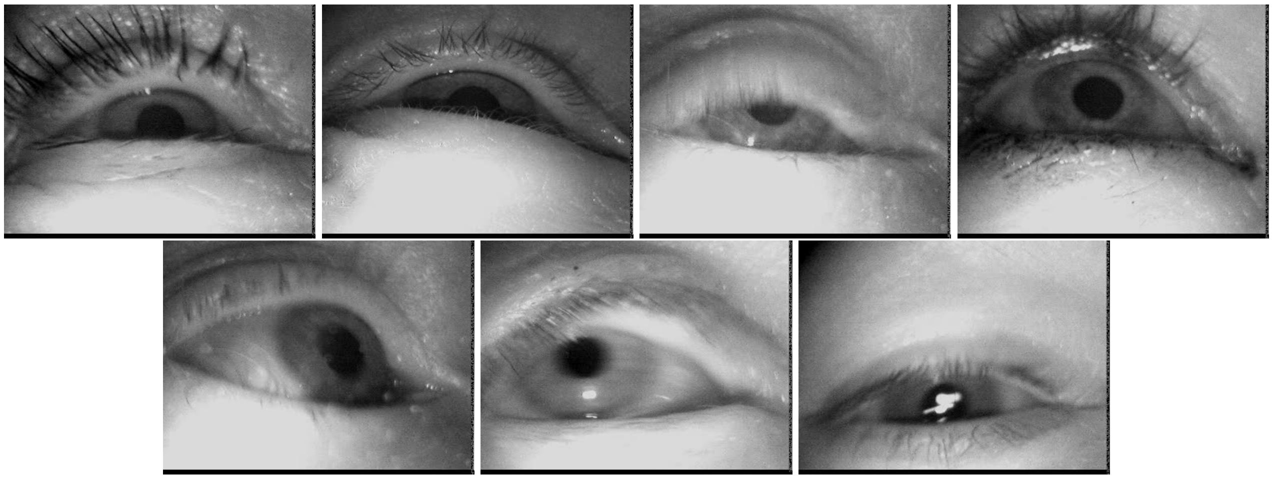

Figure 4: Challenging examples due to bad angle, incomplete blink, make-up, motion blurring and reflections

Duration	Blinks
<5	5
5	69
6	316
7	699
8	683
9	349
10	156
11	65
12	33
13	19
>13	16

Table 1: Duration distribution in terms of frames for all recorded blinks. Each frame encompasses $\approx$ 40 ms.

Based on these features, a Random Forest Classifier [8] with 100 trees is trained. If more computational power is available, the amount of trees can be increased to increase accuracy. In contrast, the number of trees can be scaled down to allow a faster evaluation. Random Forests are the method of choice, because they are quick to train and very fast to evaluate in addition to being able to handle non-linearity. The possibility to parallelize both processes increases the speed further. In addition to the already mentioned advantages, Random Forests do not need scaling for their input and, thus, effectively handle the combined feature vector of summed differences and brightness changes. Furthermore, these classifiers are resilient to outliers, so eccentric blinks do not influence its ability to generalize.

Evaluation

The proposed algorithm was evaluated using a data set of 20 video sequences of 5 minutes each extracted from a on-road driving experiment [12]; thus, the data contains endogenous and, possibly, reflex blinks. Each video corresponds to a different subject. The videos were recorded using a Dikablis Essential eye tracker at a sampling rate of 25Hz and a resolution of 384 x 288 pixels. All blink sequences were annotated, amounting to a total of 2410 blinks; their duration distribution is shown in 1. Every blink sequence starts with a completely open eye, is followed by the blink apex, and continues until roughly the same degree of openness that the eye had before the blink is reached. It is worth noticing that this data set provides realistic and challenging eye images, including quick changing illumination, blurring, reflections, and makeup. In contrast, related work usually employs data from indoor scenarios, which differ little from laboratory settings and are not realistic in the context of pervasive eye tracking. This data set is available for download at:
www.ti.uni-tuebingen.de/perception

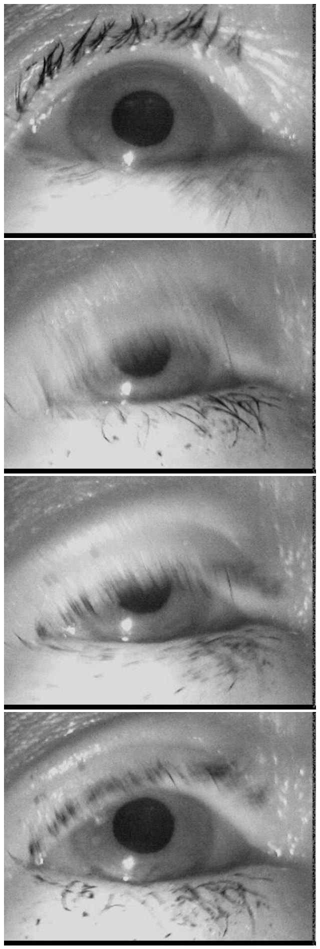

Figure 5: Subject 17 showed only little occlusion of the pupil during most blinks. This leads to a relativly high false negative rate.

We investigate the algorithm behaviour for window sizes that encompass expected durations of endogenous and reflexive blinks, ranging from $\approx$ 80 ms (i.e., $k = 2$) to $\approx$ 600 ms (i.e., $k = 15$). Blinks durations vary inside this range. However, in order to train a classifier, we need a fixed amount of features; thus, it is necessary to clip or extend blink sequences in order to fit the selected window size. In the case that our blink sequence was too short to fit the window, subsequent frames were added to the blink sequence until it was the same length as the window. Since mostly the start of a blink is needed for classification, extending the sequence at the end only influences the classification negatively if there are blinks in extremely rapid succession – i.e., the sequence covers two or more blinks. If a blink sequence was to long, we trimmed it at the end. The second parameter under investigation is the percentile used as brightness indicator. For this, we investigate a wide range, from 5 up to 90-percentile. The method was evaluated through leave-one-out cross-validation. In other words, results were obtained by training on the other 19 subjects and evaluating on the remaining one; this procedure was performed for each subject. Negative training samples were chosen at random from periods not encompassing blinks, and the amount of negative sequences was chosen as to equal the amount of positive (blink) sequences for each subject.

Figure 8 reports the average positive and negative predictive value as the window size and brightness percentile change. This figure clearly shows that the detection rate of a blink increases with the window size. This is to be expected because by clipping blink sequence information is lost. Nonetheless, since blinks usually do not last longer than 400ms, there is a point where increasing the window size further does not further improve detection. Moreover, larger windows enables a more clear distinction between

Subject	**Accuracy (%)**			
	5 Percentile		50 Percentile	
	Blink	Non-Blink	Blink	Non-Blink
1	96.0	97.0	97.0	95.0
2	100.0	53.1	96.9	89.0
3	100.0	98.5	98.5	100.0
4	95.0	95.0	95.6	95.6
5	97.0	99.0	96.0	99.0
6	96.6	97.8	97.8	97.8
7	100.0	92.3	100.0	89.3
8	97.8	98.9	96.6	97.8
9	92.0	96.6	92.5	94.3
10	100.0	100.0	100.0	100.0
11	92.1	99.5	91.6	98.0
12	96.7	97.5	96.7	95.0
13	91.1	97.8	91.7	97.8
14	100.0	91.3	100.0	95.7
15	96.0	98.0	97.0	93.9
16	97.6	99.2	96.8	100.0
17	90.0	100.0	83.8	100.0
18	93.3	98.7	88.7	100.0
19	91.0	99.6	91.9	97.9
20	96.5	100.0	94.8	99.1

Table 2: Individual results for the subjects obtained with a window size of 12

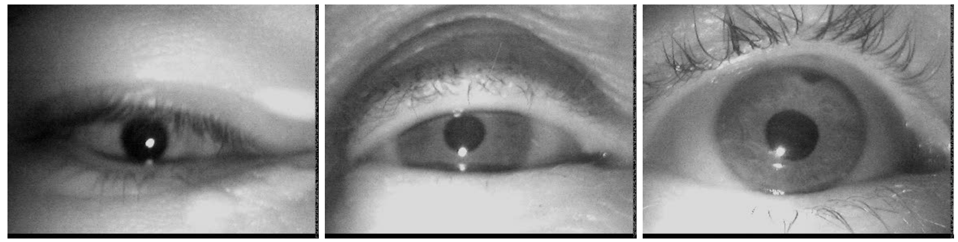

Figure 6: Difficult cases in terms of false positives: subjects 2, 7 and 14.

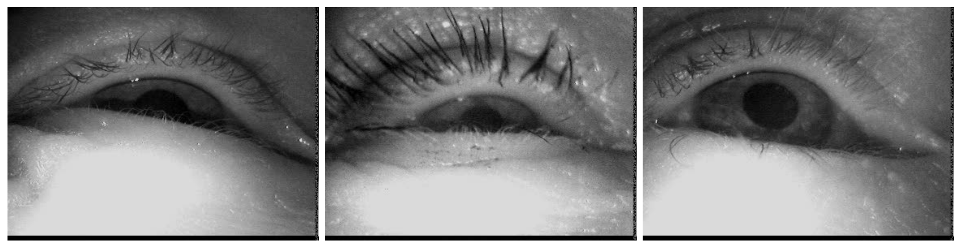

Figure 7: Difficult cases in terms of false negatives: subjects 9, 11 and 13

blinks and non-blinks, because the features of a non-blink sequence are less likely to match those of blink sequences by chance. However, problems arise when there are several blinks in rapid succession that are to short to fill the window. Oftentimes, this leads to misclassification as non-blink. The brightness percentile affects both accuracies of blinks and non-blinks. The smaller the percentile, the better in terms of blink accuracy. With the choice of the 5-percentile only changes from very dark pixels to brighter ones are measured, and, thus, mainly changes in pupil pixels occur. This increases the accuracy of blink detection. However, if a subject has distinctly dark eyelashes, these have roughly the same pixel intensity as the pupil; as a result, the algorithm responds to every sequence, classifying it as a blink. This is especially true for small narrow eyes.

Per subject results are reported in Table 2 for a window size of 12 frames. Averaged over all evaluated percentile, a window size of 12 yielded the best F1-score. As can be seem in this table, only one subject (subject 2) presented such distinctly dark eyelashes. Remaining subjects were classified properly with the 5-percentile or equally well with both percentiles. With 96,3795% the overall best F1-score was achieved using the 50th percentile and a window size of 11.

Figure 6 illustrates the problems with subject 2 as well as subject 7, who too had a small pupil and wore make-up. In addition to that, subject 7 had a partly occluded pupil even when not blinking. Subject 14 suffered from an iris defect that can be misinterpreted as a pupil and in the process of looking downwards it gets obscured, mimicking a blink.

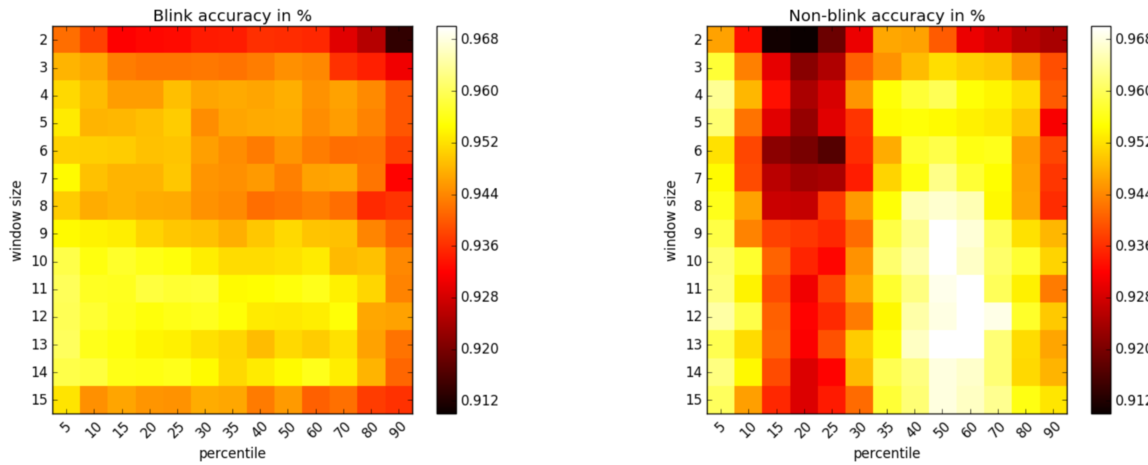

Figure 8: Average predictive value across subjects for blink (left) and non-blink (right) sequences as the window size k and brightness percentile change.

Figure 7 shows three subjects that had a false negative rate above average. This can stem from a bad angle or illumination. Both can lead to a low visibility of the pupil and lessen the brightness change, which ultimately results in a lower detection rate. In the case of subject 17 the lower detection rate arises from the fact that the pupil does not get occluded significantly during a blink (see figure 5). Double blinks only occurred in the video of subject 19. The rapid sequence of blinks lead to some misclassification and a lower detection rate.

The algorithm was implemented in *Python* using the *OpenCV* [2] and *Scikit-learn* [18] libraries. Testing was done with an Intel® Core™ i7-4790 at 3.60GHz and with 12GB of RAM. This system is consistent with those used by state-of-the-art eye trackers (e.g., Dikablis [6]). The mean runtime of the feature extraction process was 0.6630ms and predicting the class of all 4820 training samples took 0.0264ms on average. This amounts to a processing rate of 1450.54 frames per second once the images are loaded.

Conclusion

We presented an approach that is fast and has very high detection rate for blinks. For further elaboration of the proposed algorithm, we plan to broaden the spectrum of subjects to have a representative data base to train, which will significantly decrease false positives. A greater data base would also allow us to train different classifiers for different window sizes and would enable us to narrow down the time span of a blink. This way the blink duration can be estimated. Additionally a calibration phase can be integrated at the start of an experiment where the subject is asked to open and close its eyes to have samples to construct blink

sequences of different lengths and estimate the expected brightness change during a blink, which allows for normalization. To foster further research in the field and effortless replication of our result, we contribute the algorithm implementation and data sets openly at:
www.ti.uni-tuebingen.de/perception

REFERENCES

1. I. Bacivarov, M. Ionita, and P. Corcoran. 2008. Statistical models of appearance for eye tracking and eye-blink detection and measurement. *Consumer Electronics, IEEE Transactions on* 54, 3 (2008), 1312–1320.
2. G. Bradski and others. 2000. The opencv library. *Doctor Dobbs Journal* 25, 11 (2000), 120–126.
3. C. Braunagel, E. Kasneci, W. Stolzmann, and W. Rosenstiel. 2015. Driver-Activity Recognition in the Context of Conditionally Autonomous Driving. In *2015 IEEE 18th International Conference on Intelligent Transportation Systems*. 1652–1657.
4. P. P Caffier, U. Erdmann, and P. Ullsperger. 2003. Experimental evaluation of eye-blink parameters as a drowsiness measure. *European journal of applied physiology* 89, 3-4 (2003), 319–325.
5. T. Drutarovsky and A. Fogelton. 2014. Eye Blink Detection Using Variance of Motion Vectors. In *Computer Vision-ECCV 2014 Workshops*. Springer, 436–448.
6. Ergoneers. 2016. Dikablis Glasses. `http://www.ergoneers.com/`. (2016).
7. K. Grauman, M. Betke, J. Gips, and G. R Bradski. 2001. Communication via eye blinks-detection and duration analysis in real time. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, Vol. 1. IEEE, I–1010.
8. T. K. Ho. 1995. Random decision forests. In *Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on*, Vol. 1. IEEE, 278–282.
9. T. Ito, S. Mita, K. Kozuka, T. Nakano, and S. Yamamoto. 2002. Driver blink measurement by the motion picture processing and its application to drowsiness detection. In *Intelligent Transportation Systems, 2002. Proceedings. The IEEE 5th International Conference on*. IEEE, 168–173.
10. R. J. Jacob and K. S. Karn. 2003. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind* 2, 3 (2003), 4.
11. X. Jiang, G. Tien, D. Huang, B. Zheng, and M. S. Atkins. 2013. Capturing and evaluating blinks from video-based eyetrackers. *Behavior research methods* 45, 3 (2013), 656–663.
12. E. Kasneci, K. Sippel, K. Aehling, M. Heister, W. Rosenstiel, U. Schiefer, and E. Papageorgiou. 2014. Driving with Binocular Visual Field Loss? A Study on a Supervised On-road Parcours with Simultaneous Eye and Head Tracking. *Plos One* 9, 2 (2014), e87470.
13. M. Lalonde, D. Byrns, L. Gagnon, N. Teasdale, and D. Laurendeau. 2007. Real-time eye blink detection with GPU-based SIFT tracking. In *Computer and Robot Vision, 2007. CRV'07. Fourth Canadian Conference on*. IEEE, 481–487.
14. W. O. Lee, E. C. Lee, and K. R. Park. 2010. Blink detection robust to various facial poses. *Journal of neuroscience methods* 193, 2 (2010), 356–372.

15. L. K. McIntire, R. A. McKinley, C. Goodyear, and J. P. McIntire. 2014. Detection of vigilance performance using eye blinks. *Applied ergonomics* 45, 2 (2014), 354–362.

16. T. Moriyama, T. Kanade, J. F Cohn, J. Xiao, Z. Ambadar, J. Gao, and H. Imamura. 2002. Automatic recognition of eye blinking in spontaneously occurring behavior. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, Vol. 4. IEEE, 78–81.

17. T. Morris, P. Blenkhorn, and F. Zaidi. 2002. Blink detection for real-time eye tracking. *Journal of Network and Computer Applications* 25, 2 (2002), 129–143.

18. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

19. T. Santini, W. Fuhl, T. Kübler, and E. Kasneci. 2016. Bayesian Identification of Fixations, Saccades, and Smooth Pursuits. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications (ETRA '16)*. ACM, New York, NY, USA, 163–170. DOI: http://dx.doi.org/10.1145/2857491.2857512

20. P. Smith, M. Shah, and N. da Vitoria Lobo. 2000. Monitoring head/eye motion for driver alertness with one camera. In *icpr*. IEEE, 4636.

21. J. A. Stern, L. C. Walrath, and R. Goldstein. 1984. The endogenous eyeblink. *Psychophysiology* 21, 1 (1984), 22–33.

22. M. Suzuki, N. Yamamoto, O. Yamamoto, T. Nakano, and S. Yamamoto. 2006. Measurement of driver's consciousness by image processing-a method for presuming driver's drowsiness by eye-blinks coping with individual differences. In *Systems, Man and Cybernetics, 2006. SMC'06. IEEE International Conference on*, Vol. 4. IEEE, 2891–2896.

23. C. Tomasi and T. Kanade. 1991. *Detection and tracking of point features*. School of Computer Science, Carnegie Mellon Univ. Pittsburgh.

24. P. Viola and M. J. Jones. 2004. Robust real-time face detection. *International journal of computer vision* 57, 2 (2004), 137–154.

25. F. Yang, X. Yu, J. Huang, P. Yang, and D. Metaxas. 2012. Robust eyelid tracking for fatigue detection. In *2012 19th IEEE International Conference on Image Processing*. 1829–1832.