Eye blinks Detection using Facial Landmarks

Tuan-Anh L.Dang
University of Science, VNU-HCM
Faculty of Information Technology
Advanced Program in Computer Science
dltanh@apcs.vn

Tan-Dat Huynh
University of Science, VNU-HCM
Faculty of Information Technology
Advanced Program in Computer Science
htdat@apcs.vn

Hoang-Dat Le
University of Science, VNU-HCM
Faculty of Information Technology
Advanced Program in Computer Science
lhdat@apcs.vn

Trung-Duc Tran
University of Science, VNU-HCM
Faculty of Information Technology
Advanced Program in Computer Science
ttduc@apcs.vn

***Abstract*—Real-time eye blink detection is used for detecting driver's drowsiness in Intelligent Traffic System. The authors use an EAR threshold based algorithm to detect eye blinks. The result is evaluated based on a new method proposed by Andrej Fogelton and Wanda Benesova in even November–December 2018. The F1-score achieves 51.53% on a challenging Eyeblink8 dataset. The authors evaluate the method.**

I. Introduction

Blinking is a regular and important action for each human being. Base on the number of blinks in an exceptional time period, we can determine some crucial information about a person health or status. Therefore, there has been many papers on this interesting subject in different programming languages. In these paper, methods are either active or passive. Active methods are reliable but use special hardware, often expensive and intrusive but very convenient, e.g. infrared cameras and illuminators [2], wearable devices, glasses with a special close-up cameras to observe the human eyes. However, the passive systems still rely on a standard remote camera only. But, most of them were based on a same idea which is the facial landmark to determine the location and tracking motion of the eyelids for counting eye blinks.

Recently, in many research in computer vision, highly-developed real-time facial landmark detectors can capture most of the characteristic points on a human face image, including eye corners and eyelids. Most of the state-of-the-art landmark detectors formulate a regression problem, where a mapping from an image into landmark positions [29] or into other landmark parameterization [1] is learned. These modern landmark detectors are trained on "in-the-wild datasets" and they are thus robust to varying illumination, various facial expressions, and moderate non-frontal head rotations. An average error of the landmark localization of a state-of-the-art detector is usually below five percent of the inter-ocular distance. Recent methods run even significantly super real-time [23]. Therefore, we propose a simple but efficient algorithm to detect eye blinks by using a recent facial landmark detector. A single scalar quantity that reflects a level of the eye opening is derived from the landmarks. Finally, having a per-frame sequence of the eye opening estimates, the eye blinks are found by an SVM classifier that is trained on examples of blinking and non-blinking patterns.

II. Background & Related Works

Analyzing blinks and obtaining blinks statistics give us lots of potential to build up a technology using eye blink. Recently, there has been some attention point toward eye blink detection mostly because of it its use in face liveliness detection [19] [26]. Eye blinks are a way of interaction between disabled people [12] and computers. Eye blink frequency and duration are important signs of sleepiness [28] that can be used to detect driver's drowsiness [7] and eventually to prevent them to cause any accident.

The tear film is a microscopic protecting coat of the eye against dust and microorganisms. It lubricates the ocular surface and protects it from evaporation. The tear film consists of three layers, organized from the ocular surface: mucin, water, and lipid layer. The first two layers provide moisture to the eye and the third layer prevents their evaporation. The lipid layer consists of meibum produced by Meibomian glands and it is spread on the eye surface during the complete blink only. Meibomian glands are placed at the rim of eyelids inside the tarsal plate. During eye blink, eyelids need to touch, so the meibum can be spread over the ocular surface to protect it from evaporation [5]. Portello et al. [13] observe the negative influence of the incomplete blinks on dry eye. One of the main causes of dry eye syndrome is also low blink rate. A healthy human blinks 10 to 15 times per minute. 70% of computer users have decreased blink rate up to 60% [3].

Before proposing a new method for this subject, we would like to have a look on some existed methods and paper in order to find their weaknesses and strengths. We approach a common based method, which is a Viola– Jones type algorithm to detect the face and eyes e.g [6] [11]. In different situations, the detector sometimes is not able to detect the frontal face or eyes, which can be compensated with region tracking [16] [4]. Many recently proposed methods are based on motion tracking with eyes region [8] All of these methods were classified into two basic groups, appearance based and sequential based. In appearance based methods, the state of the eye (open, closed [16] or the eye closure [6]) is estimated for individual frames to determine whether there is an eye blink or not. On the other hand, the sequential based methods try to find the difference between each frame (pixels values [15], descriptors [17], etc.). Take a look further in these existed methods, a few of them have the right constraints for eye blink detection. For example, Morris et al. [18] proposed a blink detection system based on variance map calculation and eye corner analysis. They receive good result (95% true positive) but head movement greatly effect variance map calculation and cause a sharp drop in performance. Divjak & Bischof [8] detect blinks base on eyelid movements. By using FAST [24] and tracked with Lucas–Kanade tracker [27], the features of the eyelid can be detected. These features are classified using their location; face, left and right eye. Based on these features, the eye and face regions are tracked. The authors calculate a normal flow of the regions in the direction of intensity gradients. Eyelid motion come along with head movements, so compensation based on the already extracted head movement takes place. Dominant orientations of the local motion vectors for the individual classes are extracted from a histogram of orientations, due to which partial invariance to eye orientation is achieved. To filter the eyelid motion, only the flow in the direction perpendicular to the line segment between the eyes is considered. The angle between this line and the horizon is calculated and flow vectors are transformed correspondingly. Corrected and normalized flow is used to calculate an average flow magnitude of the eye regions. The dominant flow direction is recognized based on the individual orientation of local motion vectors (optical flow) in a histogram with 36 bins, each bin representing 10 degrees. Normal flow orientation and magnitude are used as the input parameter for a state machine. Radlak & Smolka [21] introduce the Weighted Gradient Descriptor (WGD) which is based on computing of partial derivatives per each pixel in the eye region over time. Feature vectors are averaged in two orientations ("up" and "down") based on location. The vertical distance between their points of origin is used as the feature. Closing and opening of the eye is represented by negative as well as positive peak within the feature vector. After noise filtering, zero-crossing point between the local maximum and minimum represents the detected eye blink. In this modification Gauss weighting is used to suppress eyebrow movements often falsely detected as eye blinks. The maximum and minimum of the entire feature vector (for given video) are found and used to estimate proper thresholds, which reduces the usability for cameras. A new dataset of 5 people recorded by Basler 100 fps camera is introduced, which is part of Silesian Deception Database [20]. We will refer to it as Silesian5 dataset. The reported detection rate on Silesian5 is around 90% and 98.8% on ZJU dataset. In evaluation, only the right eye of the subject is used. The authors report the best obtained results for given datasets while tuning parameters. Template matching using histogram of Local Binary Pattern (LBP) can be used to detect eye blinks [17]. First, the initialization takes place. An open eye template is calculated from several initial images where eye is open and not moving. For each image in a sequence, LBP histogram is computed from the eye region and compared with the template using the Kullback–Leibler divergence measure. The output is a curve where noise is filtered out using Savitzky–Golay filter and the top hat operator. Afterward, peaks are detected using Grubb test and considered as eye blinks. Detection rate of 99% is reported on ZJU and Silesian5 dataset (different parameters are used for each dataset). Because of the Grubb test, this method is not suitable for a real-time stream from camera. Different evaluation procedures take place over the mentioned methods. Often, they are not even specified. For example false positive rate and mean accuracy. We assume (because it is not specified) that the number of images with open eyes is used as Negatives (N). This is not proper because blink usually consists of 7 frames in average at 30 fps. Many of the mentioned papers do not even define what is considered as true positive blink detection. Whether they use per frame annotation or just threshold a distance of the detected blink from the ground truth. Based on the overview of the related work, we observe superior performance of motion based methods. Divjak & Bischof [8] achieve interesting results. The authors try to compensate head movement by subtracting the average face motion vector from the individual eye motion vectors. The problem is that this motion vector is affected by facial mimics and hand touches over the face. Instead of compensating the head move we analyze the behavior of motion vectors.

After research many existed methods, we decided to adapt a eye blink detection with OpenCV, Python, and dlib by Adrian Rosebrock on April 24, 2017 in dlib, Facial Landmarks, Tutorials.

III. Method

A. Study process

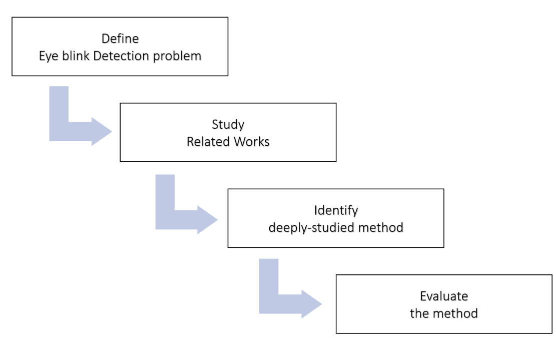


Fig. 1. Study progress

Our study is processed with 4 stages (Fig. 1). The important thing, first, is to identify correctly what problem studied. Eye blink detection is misunderstood with eye blink counting that retrieves the number of eye blinks in a piece of frames, rather than the information when blinks is performed. Scientific works, mentioned in details in section 2, is made to bring the problem to the state-of-the-art. Studying on it is necessary for understand comprehensively the problem before deriving our studied method. A dataset that is free, standard, and challenging is an important material for testing the method. We got Eyeblink8 dataset to evaluate.

B. Eye blinks detection using Facial Landmarks

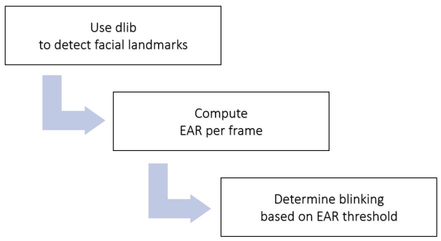


Fig. 2. Overview of the method

1) Dissecting dlib's facial landmarks detector: dlib is a robust support-library in computer vision. We exploit 19.16.0 version of dlib.

Real-time facial landmarks detector in dlib is implemented based on an algorithm from "One Millisecond Face Alignment with an Ensemble of Regression Trees" paper written by Vahid Kazemi and Josephine Sullivan [14].

The algorithm uses a cascade of T regressors to predict the position of facial landmarks. Each regressor, denoted by r_t, consists of K binary tree based regressors (ferns) (Fig. 3).

"Assume we have training data $(I_i, \mathbf{S}_1), ..., (I_n, \mathbf{S}_n)$ where each I_i is a face image and $\mathbf{S}_i$ is its shape vector" [14] The task of each progressor is to produce a Δlandmark (Fig. 8) to improve the current shape.

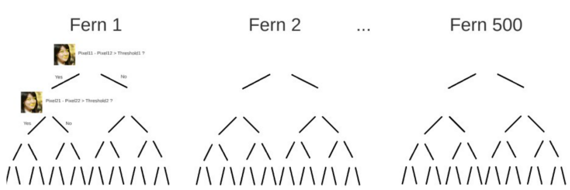


Fig. 3. A regressor of K = 500 ferns

Each regressor is training on a labeled-facial-landmarks dataset (Fig. 4).

Suppose the face recognition is performed that returns a rectangle bounding box. A mean shape $\mathbf{S}_0$ is initialized for the 1^{st} regressor by centering and scaling to the rectangle (Fig. 5). To obtain the mean shape, we sum of the labeled faces in the training dataset and divide by the amount of the faces.

A good property is that the model predicts the facial landmarks position directly from pixel intensity values of the image detected. P random pixels is sampled in the mean shape. For an arbitrary shape $\mathbf{S}_i$, the similarity transform T_s is obtained by subtract $\mathbf{S}_0$ from $\mathbf{S}_i$. P pixels is extracted in $\mathbf{S}_i$ using the similarity transform T_s (Fig. 6).

Fig. 4. A sample of HELEN dataset, including 194 labeled points on the face

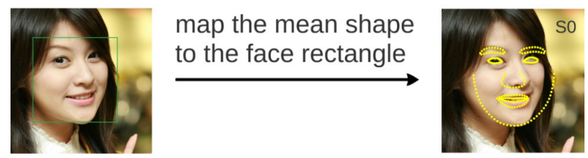


Fig. 5. The mean shape $\mathbf{S}_0$ is mapped to the face rectangle

The facial landmark detector localizes the eyes, eyebrows, nose, mouth and jawline on the face; therefore, pixels on the cheek are not helpful rather than those on or near the eyes, eyebrows, nose, mouth and jawline. A exponential prior is used to improve this limit (Fig. 7).

The intensity values of extracted pixels set is inputted into the regressor. At each level of the fern, the difference between intensities of two pixels is compared to a threshold. At the leaf of regression tree, a delta vector is derived (Fig. 8). The process is iterated and the shape towards to the ground truth (Fig. 9). dlib's facial landmark detector is trained on iBUG 300-W dataset (Fig. 10) with a cascade of 10 regressors, each regressor has 500 ferns with 5 levels, derive Δlandmarks by extracting 400 pixels.

2) Eye Aspect Ratio: Eye Aspect Ratio (EAR) [25], is obtained by the equation (1). In the equation, p_i, i $\in$ {1, 2, 3, 4, 5, 6} is valued from 6 facial landmarks on the eyes (Fig. 11). The EAR mostly is a constant if no blink occurs; otherwise it falls dramatically into zero (Fig. 11). Basing on the property of the metric EAR, a EAR threshold *EAR_THRESHOLD* is used to determine whether eyes blink or not. The EAR is computed for each frame, then using a EAR threshold *EAR_THRESHOLD* to compare with. In the case that the per-frame EAR for *CONTINUOUS_FRAMES* continuous frames is less than *EAR_THRESHOLD*, a blink action is performed. The information about frames that its EAR less than the EAR threshold is outputted into a file (Fig. 12).

$$EAR = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\|p_1 - p_4\|} \tag{1}$$

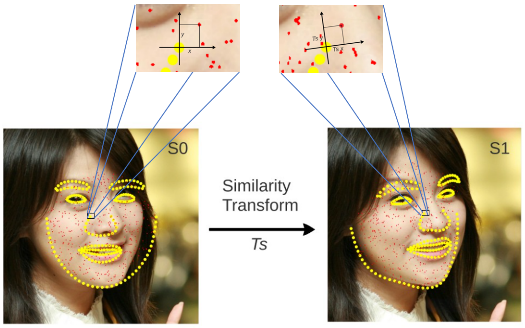
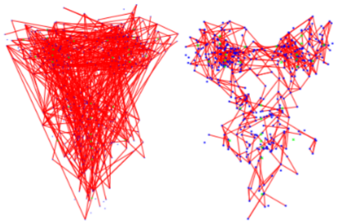


Fig. 6. Similarity transform is performed to extract pixels



(a) Uniform prior (b) Exponential prior

Fig. 7. Difference between uniform and exponential priors in extracted pixels

IV. Evaluation

While mentioning the related works, we realize that most of them are not bring out the evaluation procedure in a specific way. Often this process based on Accuracy but not show the formula also the True Positive and False Positive while they are principal. After making inquiries about the evaluate performance of eye blink detection, we see that there is no uniform way to measure the result: Radlak & Smolka consider that the blink is detected by the peak frame between the beginning and ending frame [22] while Andrej Fogelton base on an interval of it [10] over a video. And we choose to evaluate the result according to Andrej Fogelton. We use the Intersection Over Union (2) that is introduced in PASCAL VOC challenge [9].

$$IOU = \frac{A \cap B}{A \cup B} \tag{2}$$

,where A and B are the interval of eye blink in experiment and ground truth

Andrej Fogelton proposed the True Positive is defined whenever the IOU is larger than or equal to 0.2. The number of False Negative is determined if the IOU less than 0.2 and the number of False Positive is equal the times detected blink but it does not belong to ground truth. Each ground truth and detected blink is just counted once because they can overlap together.

We evaluate the result based on Precision, Recall and F1-score:

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

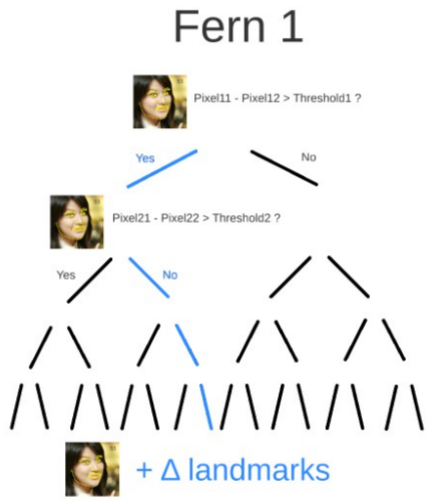
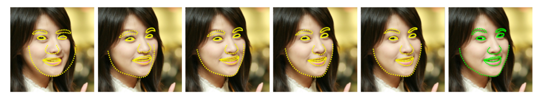


Fig. 8. A fern working



(a) $T = 0$ (b) $T = 1$ (c) $T = 2$ (d) $T = 3$ (e) $T = 10$ (f) Ground truth

Fig. 9. Resutl estimate facial landmarks for T=10 regressors

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{5}$$

V. Experimental Results

We test the method with *CONTINUOUS_FRAMES* = 3 and on different EAR thresholds from 0.00 to 0.30 with the step of 0.05 (Fig. 13). The experiment with the *EAR_THRESHOLD* = 0.30 achieves: F1-score = 40.54%, precision = 27.27%, and recall = 78.95%. F1-score, precision, and recall is 0.00 if *EAR_THRESHOLD* = {0.00, 0.05}. The best result is 51.53% of F1-score at *EAR_THRESHOLD* = 0.25.

The drawback of the method is affected by the size of the eyes. This results in that no fixed *EAR_THRESHOLD* is able to used.

VI. Conclusion

References

[1] Akshay Asthana, Ashish Asthana, Stefanos Zafeiriou, Shiyang Cheng, and Maja Pantic. Incremental face alignment in the wild. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014*, pages 1859–1866, United States, 6 2014. IEEE Computer Society. eemcs-eprint-25818.

[2] Luis Bergasa, Jesús Nuevo, Miguel-Angel Sotelo, Rafael Barea, and María Guillén. Real-time system for monitoring driver vigilance. *IEEE Transactions on Intelligent Transportation Systems*, 7:63–77, 01 2006.

[3] Clayton Blehm, Seema Vishnu, Ashbala Khattak, Shrabanee Mitra, and Richard W. Yee. Computer vision syndrome: A review. *Survey of Ophthalmology*, 50(3):253 – 262, 2005.

[4] Thomas Brandt, Ralf Stemmer, and Andry Rakotonirainy. Affordable visual driver monitoring system for fatigue and monotony. *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*, 7:6451–6456 vol.7, 2004.

[5] A.J. Bron, J.M. Tiffany, S.M. Gouveia, N. Yokoi, and L.W. Voon. Functional aspects of the tear film lipid layer. *Experimental Eye Research*, 78(3):347 – 360, 2004. Special issue in honour of David Maurice.

Fig. 10. iBUG-300W's annotation with 68 facial landmark positions

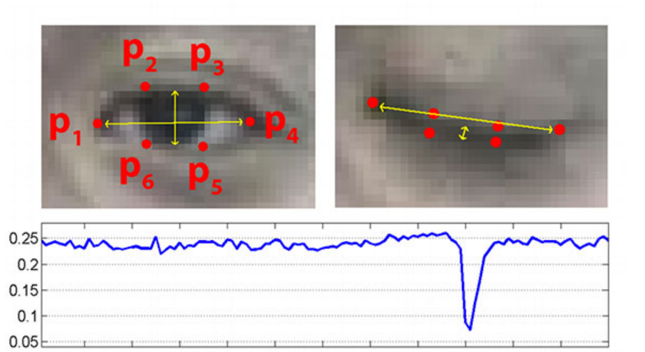


Fig. 11. Relationship between the state of eye (blink or not blink) and EAR

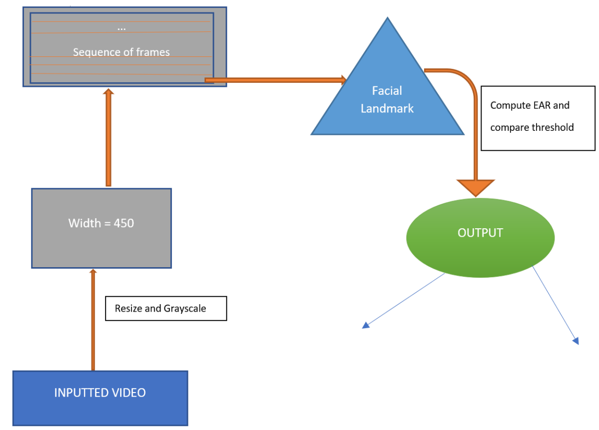


Fig. 12. Method in details

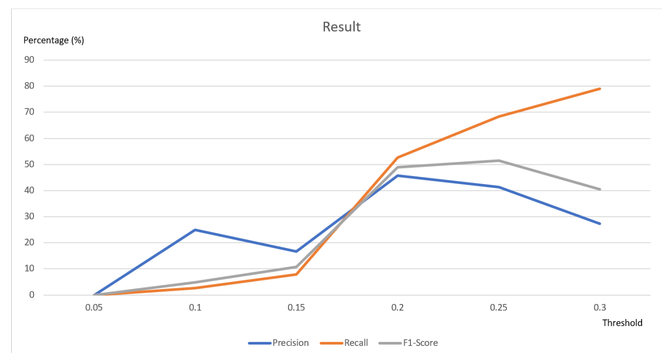


Fig. 13. Precision, Recall, and F1-Score is related to EAR

[6] Taner Danisman, Ian Marius Bilasco, Chabane Djeraba, and Nacim Ihaddadene. Drowsy driver detection system using eye blink patterns. *2010 International Conference on Machine and Web Intelligence*, pages 230–233, 2010.

[7] Taner Danisman, Ioan Marius Bilasco, Chabane Djeraba, and Nacim Ihaddadene. Drowsy driver detection system using eye blink patterns. *2010 International Conference on Machine and Web Intelligence, ICMWI 2010 - Proceedings*, 10 2010.

[8] Matjaz Divjak and Horst Bischof. Eye blink based fatigue detection for prevention of computer vision syndrome. In *MVA*, 2009.

[9] Mark Everingham, Luc Gool, Christopher K. Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *Int. J. Comput. Vision*, 88(2):303–338, June 2010.

[10] Andrej Fogelton and Wanda Benesova. Eye blink completeness detection. *Computer Vision and Image Understanding*, 176-177:78–85, 2018.

[11] I Garcia, Sebastian Bronte, Luis Bergasa, Javier Almazán, and José Yebes. Vision-based drowsiness detector for real driving conditions. pages 618–623, 06 2012.

[12] Kristen Grauman, Margrit Betke, J. Lombardi, James Gips, and Gary R. Bradski. Communication via eye blinks and eyebrow raises: video-based human-computer interfaces. *Universal Access in the Information Society*, 2:359–373, 2003.

[13] Joan K Portello, Mark Rosenfield, and Christina A Chu. Blink rate, incomplete blinks and computer vision syndrome. *Optometry and vision science : official publication of the American Academy of Optometry*, 90, 03 2013.

[14] Vahid Kazemi and Josephine Sullivan. One millisecond face alignment with an ensemble of regression trees. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1867–1874, 2014.

[15] Y. Kurylyak, F. Lamonaca, and G. Mirabelli. Detection of the eye blinks for human's fatigue monitoring. In *2012 IEEE International Symposium on Medical Measurements and Applications Proceedings*, pages 1–4, May 2012.

[16] Won Oh Lee, Eui Chul Lee, and Kang Ryoung Park. Blink detection robust to various facial poses. *Journal of Neuroscience Methods*, 193(2):356 – 372, 2010.

[17] Krystyna Malik and Bogdan Smolka. Eye blink detection using local binary patterns. *2014 International Conference on Multimedia Computing and Systems (ICMCS)*, pages 385–390, 2014.

[18] T. Morris, P. Blenkhorn, and Farhan Zaidi. Blink detection for real-time eye tracking. *J. Netw. Comput. Appl.*, 25(2):129–143, April 2002.

[19] Gang Pan, Lin Sun, Zhaohui Wu, and Shihong Lao. Eyeblink-based anti-spoofing in face recognition from a generic webcamera. *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8, 2007.

[20] Krystian Radlak, Maciej Bożek, and Bogdan Smolka. Silesian deception database: Presentation and analysis. pages 29–35, 11 2015.

[21] Krystian Radlak and Bogdan Smolka. Blink detection based on the weighted gradient descriptor. In Robert Burduk, Konrad Jackowski, Marek Kurzynski, Michał Wozniak, and Andrzej Zolnierek, editors, *Proceedings of the 8th International Conference on Computer Recognition Systems CORES 2013*, pages 691–700, Heidelberg, 2013. Springer International Publishing.

[22] Krystian Radlak and Bogdan Smolka. Blink detection based on the weighted gradient descriptor. In Robert Burduk, Konrad Jackowski, Marek Kurzynski, Michal Wozniak, and Andrzej Zolnierek, editors, *CORES*, volume 226 of *Advances in Intelligent Systems and Computing*, pages 691–700. Springer, 2013.

[23] Shaoqing Ren, Xudong Cao, Yichen Wei, and Jian Sun. Face alignment at 3000 fps via regressing local binary features. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1685–1692, 2014.

[24] Edward Rosten and Tom Drummond. Machine learning for high-speed corner detection. In Aleš Leonardis, Horst Bischof, and Axel Pinz, editors, *Computer Vision – ECCV 2006*, pages 430–443, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[25] Tereza Soukupová and Jan Cech. Real-time eye blink detection using facial landmarks. 2016.

[26] Mariusz Szwoch and Paweł Pieniażek. Eye blink based detection of

liveness in biometric authentication systems using conditional random fields. In Leonard Bolc, Ryszard Tadeusiewicz, Leszek J. Chmielewski, and Konrad Wojciechowski, editors, *Computer Vision and Graphics*, pages 669–676, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[27] Carlo Tomasi and Takeo Kanade. Detection and tracking of point features. Technical report, International Journal of Computer Vision, 1991.

[28] Udo Trutschel, Bill Sirois, David Sommer, Martin Golz, and David Edwards. Perclos: An alertness measure of the past. pages 172–179, 10 2017.

[29] Xuehan Xiong and Fernando De la Torre. Supervised descent method and its applications to face alignment. pages 532–539, 06 2013.