Link to data:
https://atreus.informatik.uni-tuebingen.de/seafile/d/8e2ab8c3fdd444e1a135/?p=%2F&mode=list

BORE: Boosted-oriented edge optimization for robust, real time remote pupil center detection

Wolfgang Fuhl
University Tuebingen, Perception Engineering
Tuebingen, Baden-Wuerttemberg
wolfgang.fuhl@uni-tuebingen.de

Shahram Eivazi
University Tuebingen, Perception Engineering
Tuebingen, Baden-Wuerttemberg
shahram.eivazi@uni-tuebingen.de

Benedikt Hosp
University Tuebingen, Perception Engineering, Institute of Sports Science
Tuebingen, Baden-Wuerttemberg
benedikt.hosp@uni-tuebingen.de

Anna Eivazi
University Tuebingen, Perception Engineering
Tuebingen, Baden-Wuerttemberg
anna.eivazi@mnf.uni-tuebingen.de

Wolfgang Rosenstiel
University Tuebingen, Technical Computer Science
Tuebingen, Baden-Wuerttemberg
Wolfgang.Rosenstiel@uni-tuebingen.de

Enkelejda Kasneci
University Tuebingen, Perception Engineering
Tuebingen, Baden-Wuerttemberg
Enkelejda.Kasneci@uni-tuebingen.de

ABSTRACT

Undoubtedly, eye movements contain an immense amount of information, especially when looking to fast eye movements, namely time to the fixation, saccade, and micro-saccade events. While, modern cameras support recording of few thousand frames per second, to date, the majority of studies use eye trackers with the frame rates of about 120 Hz for head-mounted and 250 Hz for remote-based trackers. In this study, we aim to overcome the challenge of the pupil tracking algorithms to perform real time with high speed cameras for remote eye tracking applications. We propose an iterative pupil center detection algorithm formulated as an optimization problem. We evaluated our algorithm on more than 13,000 eye images, in which it outperforms earlier solutions both with regard to runtime and detection accuracy. Moreover, our system is capable of boosting its runtime in an unsupervised manner, thus we remove the need for manual annotation of pupil images.

CCS CONCEPTS

• **Computing methodologies** → **Image processing**; *Feature selection*; *Shape analysis*;

KEYWORDS

Remote eye tracking, pupil detection, pupil center, boosting

ACM Reference Format:
Wolfgang Fuhl, Shahram Eivazi, Benedikt Hosp, Anna Eivazi, Wolfgang Rosenstiel, and Enkelejda Kasneci. 2018. BORE: Boosted-oriented edge optimization for robust, real time remote pupil center detection. In *ETRA '18: ETRA '18: 2018 Symposium on Eye Tracking Research and Applications, June 14–17, 2018, Warsaw, Poland.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3204493.3204558

ETRA '18, June 14–17, 2018, Warsaw, Poland

ACM ISBN 978-1-4503-5706-7/18/06…$15.00
https://doi.org/10.1145/3204493.3204558

1 INTRODUCTION

Visual gaze behavior is in the focus of various scientific fields such as medicine, driving, psychology, human cognition, human-computer interaction, and many more. Approaches for the extraction of human gaze data include brain computer interface (BCI) based on EEG [Tan et al. 2013], magnetic fields [Robinson 1963], and video-based eye feature extraction. Video-based gaze tracking systems are widely used due to their comfort and availability at low costs. However, efforts need to be applied to signal and image processing in order to deal with the challenges such as reflections in image-based techniques. Nevertheless, video-based eye tracking becomes more popular because it is less invasive than the other approaches.

Video-based eye tracking can be divided into two main groups; remote and head-mounted. In remote eye tracking, the subject is recorded using one or more external cameras. The gaze location of a subject can be derived after varying processing steps: namely face, eye, pupil center detection, and head orientation estimation. For head-mounted eye trackers, the processing step consists mainly of pupil localization in the eye images. Since head-mounted eye trackers are worn by subjects, meaning the field camera moves with the subject's head, we have an implicit compensation for head movements. Therefore, the only information which has to be extracted is the pupil center in the eye cameras. While modern devices constantly improve their design to be lightweight and as non-intrusive as possible [Kassner et al. 2014], they are still worn on the head. Therefore, they become uncomfortable after a certain time and the cameras remain in the user's peripheral field of view.

On the other hand, remote-based eye trackers are non-invasive and modern cameras are able to record with high frame rates. Both properties are important advantages over head-mounted eye trackers. The high frame rates are useful in medical and psychological studies and eye movement research [Holmqvist et al. 2011]. When driving, for example, it is also necessary that the system used does not affect the driving performance and the driver is not disturbed.

This is the case for driver monitoring [Braunagel et al. 2015; Liu et al. 2002] or gaze-based assistance systems [Kasneci 2013].

With this paper, we develop a novel pupil center detection algorithm which is based on oriented edge optimization. We believe, this is an important contribution to the eye-tracking community, because for high-speed cameras processing time is critical; for instance, a camera with 500 Hz has only 2 milliseconds per frame. Therefore, we propose an unsupervised boosting approach (no annotations needed) that is capable of decreasing the run time drastically. The algorithm is compared to other state-of-the-art pupil detection approaches based on the public data sets BioID [Jesorsky et al. 2001], [Fuhl et al. 2016a], GI4E [Villanueva et al. 2013], and MUCT [Milborrow et al. 2010].

2 RELATED WORK

There has been extensive research in the machine vision and eye tracking communities on pupil detection algorithms. Asadifard and Shanbezadeh [Asadifard and Shanbezadeh 2010] used a cumulative distribution function together with a threshold. First, The algorithm extracts the iris region and then separates the pupil out of this area. Before the pupil is extracted, the threshold region is enhanced using morphologic erosion to remove artifacts produced by eyelashes. The pupil extraction step is performed by thresholding the iris region with its mean intensity value. Finally, the center of mass is used as the pupil center.

Droege and Paulus [Droege and Paulus 2010] employed a band-pass filter on gradients. This filter reduces the amount of possible pupil edge candidates. The gradients are interpreted as lines by using their position and the gradient direction information. The intersection point of all these lines is computed using an M-Estimator.

Timm and Barth [Timm and Barth 2011] used the angle difference between each gradient and all image positions for pupil center estimation. Therefore, the angle between all gradient directions and the vector from each image position to the gradient location is computed. The result is weighted by the inverted image intensity to increase the response of dark image locations. The final pupil center candidate with the highest response is chosen.

Fuhl et al. [Fuhl et al. 2016a] used the second part of the ElSe [Fuhl et al. 2016b] algorithm. It consists of a center surround difference filter weighted with an inverted mean filter. The maximal response is used as pupil center.

Taking a different approach by [Skodras and Fakotakis 2015], the pupil detection algorithm used a luminance map of the eye region in combination with the fast radial symmetry transform proposed in [Loy and Zelinsky 2003]. The luminance map is computed using the YCbCr color space together with gray-scale dilation and erosion. In the fast radial symmetry transform, each image gradient votes for possible circle center candidates. Each candidate vote is computed based on the gradient magnitude and is spread to its surrounding pixels by using a Gaussian distribution. The luminance map is then added to the result from the radial symmetry transform and the maximum is selected as pupil center.

More recently, George and Routray [George and Routray 2016] applied size invariant circle detection [Atherton and Kerbyson 1999]. For coarse positioning, the orientation annulus is used with a vertically stronger weighting. Outgoing from this position, the gradients in a pre-specified range are collected and outliers are removed based on their angle and magnitude. The second outlier removal step is based on median filtering of the remaining gradient candidates. The final pupil center is found using RANSAC ellipse fitting.

Additionally, there are several approaches developed for head-mounted eye trackers. These approaches were developed for close-up eye images, different from those captured with remote setups and, thus, are not evaluated in this work. For a comprehensive review of such methods, we refer the reader to [Fuhl et al. 2016c] and [Santini et al. 2018].

3 METHOD

The proposed algorithm is based on the oriented edge optimization formulation from [Fuhl et al. 2017]. Their method was for eyelid detection; it used different heuristics to find initial positions before the optimization was applied to reduce computational costs. In the original formulation, two polynomials were refined by optimizing their overall oriented edge response. Therefore, the equation shifted three points per polynomial, where two points were shared by both polynomials (eye corners).

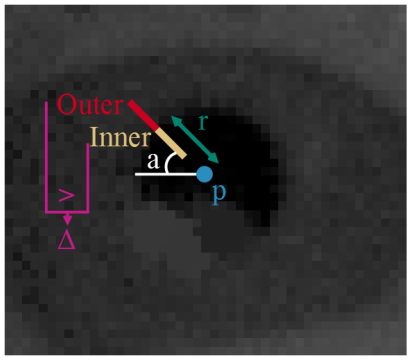


Figure 1: Explanatory example for the variables in equation 1.

In contrast, our approach uses edges on multiple circular shapes. For each shape, the oriented edge value is computed. The main advantage of our formulation is that multiple shapes contribute to the final result. Meaning, ellipses are contained in multiple circle shapes. This step reduces the computational costs in contrast to evaluating each elliptical shape separately. Furthermore, our approach can deactivate single oriented edges, which reduces the runtime drastically and will be described in the unsupervised boosting section.

However, we use circles as the base function ($C()$). At each point on a circle, we compute the gradient orthogonal to the tangent (Δ). These points are described by the center point of the circle (p), the radius of the circle (r), and an angle (a). These variables together with an explanatory visualization of equation 2 are shown in Figure 1. This oriented-edge optimization is described as

$$argmax_p \int_{r=r_{min}}^{r_{max}} \int_{a=0}^{2\pi} L(\Delta C(p,r,a))\, dr\, da, \tag{1}$$

where r_{min} and r_{max} are the minimal and maximal radii of circles; two and twenty four in our implementation, respectively. The function $L()$ evaluates the gradient as

$$L = \begin{cases} inner < outer & 1 \\ else & 0, \end{cases} \quad (2)$$

where *inner* and *outer* are the inner and outer pupil regions for a particular direction, as exemplified in Figure 1. The pupil center is not detected based on the gradient magnitude, but rather on the sum of gradients supporting the center. Consequently, this approach is less sensitive to low contrast images and reflections. In Equation 2, *inner* and *outer* stand for the sum of intensity values collected inside and outside of the circle per gradient.

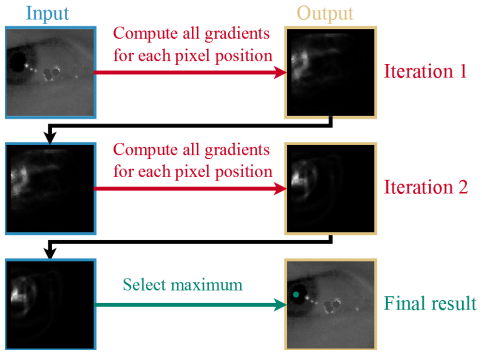


Figure 2: The algorithmic work flow of the optimization for two iterations.

Figure 2 shows the workflow of our approach. In the first iteration, all gradients for each image pixel position are computed (Equation 1). The resulting probability map is then used as input for the second iteration. This again results in a probability map where most of the noise is removed. The entire iterative process can be understood as a circle shrinking procedure. The final pupil center is obtained by selecting the pixel position with the highest probability.

The process of computing all gradients for each pixel position is computationally demanding. To overcome these computational costs, often a coarse positioning is proposed (e.g. Haar cascades [Świrski et al. 2012]), where only the surrounding region is evaluated. Although this approach is applicable to our algorithm, we propose to use unsupervised boosting on a recorded eye video without further annotations.

3.1 Unsupervised Boosting

The particular property behind this approach is to select only the most important positions on the circle outline for computation (Figure 3(b) red lines). Therefore, we evaluate each frame in a video and increase the score for each oriented gradient that voted for this center position (Histogram in Figure 3(a)). Finally, only a percentage is enabled for the final detector (Figure 3(b) red lines are enabled and the others are disabled).

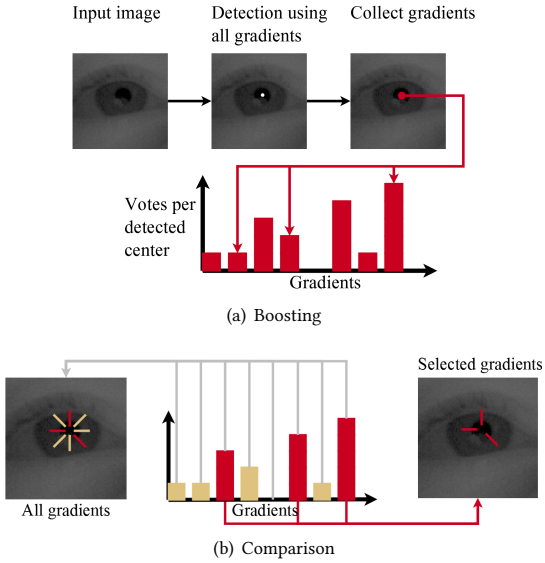


Figure 3: In (a) the computation of the gradient score is shown. The red bars in the histogram represent the score per gradient. The selection process, in comparison to the usage of all gradients is shown in (b). Scores colored in ochre are disabled.

$$S(r, a, p, fp) = \begin{cases} L(\Delta C(p, r, a)) > 0, & \\ |p - fp| < d_{th} & \frac{1.0}{\frac{|p-fp|}{3.0} + \epsilon} \\ else & 0 \end{cases} \quad (3)$$

Equation 3 formalizes the scoring function, in which fp represents the pupil center position found by the proposed method (white dot in Figure 3(a)). With $||$, we refer to the Euclidean distance, where ϵ denotes a small constant to avoid division by zero. d_{th} determines the size of the region is used for collecting the score of each gradient (red circle in Figure 3(a)), which we have empirically set to three.

The main advantage of our approach is that the resulting performance of the algorithm is adjustable to the needs of the user and to the available hardware. In addition, the algorithm can be calibrated to each user during the calibration procedure for the gaze estimation. In contrast, the boosting relies on the accurate detection of the entire approach. Therefore, if the algorithm has a low overall detection rate on the video, the boosting step would fail as well.

4 EVALUATION

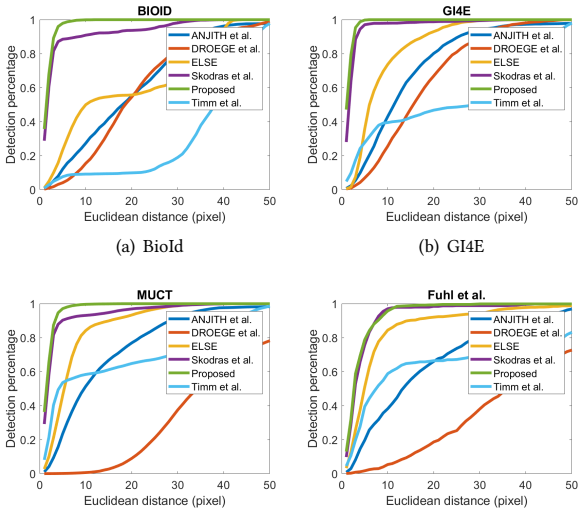
Detecting the pupil center in remote images is a non-trivial task. The detected regions vary regarding their size and position. Therefore, we increased the size of each annotated eye box by 20% in each direction to provide a similar condition for all cases. This step influences algorithm runtime and adds an extra challenge for the detection algorithms. We compared the proposed approach against five state-of-the-art algorithms [Droege and Paulus 2010; Fuhl et al.

2016b; George and Routray 2016; Skodras and Fakotakis 2015; Timm and Barth 2011] on the above mentioned publicly available data sets. The error is reported in pixels, measured as the Euclidean distance between the annotated and the algorithmically detected pupil center.

4.1 Data sets

For evaluation, we used four publicly available datasets. Those are BioID [Jesorsky et al. 2001], GI4E [Villanueva et al. 2013], [Fuhl et al. 2016a] and MUCT [Milborrow et al. 2010]. The BioID [Jesorsky et al. 2001] dataset contains 1521 grayscale images from 23 subjects recorded in an office environment. Each frame has a resolution of 384×286 pixels. For the GI4E [Villanueva et al. 2013] data set each frame has a resolution of 800×600 pixels and it contains 1236 RGB images from 103 subjects. These images were recorded using a standard web camera and include challenges such as reflections on the subject's glasses. The data set from [Fuhl et al. 2016a] contains 445 RGB and near infrared images of two subjects with a resolution of 1280×960 pixels. These images were recorded inside an office with a pan tilt zoom camera. The main challenges in this data set are different distances, head rotations, blinks, and reflections. The last data set is MUCT [Milborrow et al. 2010] with 3755 images. Each frame has a resolution of 480×640 pixels and is in RGB. The main challenges in this data set are changes in the lighting conditions, reflections, glasses, and different distances of the subject. In addition, the subjects are of different age and ethnicity.

4.2 Results



(a) BioId (b) GI4E

(c) MUCT (d) [Fuhl et al. 2016a]

Figure 4: Evaluation results for all algorithms separated per data set. The detection percentage of 1 corresponds to 100%.

Figure 4 shows the results of the evaluated algorithms for each data set separately. As can be seen from the Figure, the proposed approach outperforms the state-of-the-art for both accuracy and detection rate. For the data set proposed in [Fuhl et al. 2016a], (Figure 4 (d)) our algorithm has lower accuracy with regard to the Euclidean distance ranging from seven to eleven pixels compared to the algorithm in [Skodras and Fakotakis 2015]. However, we argue that for a pixel distance of four, the error is already too large for the gaze estimation to be useful in remote scenarios. The influence of the enlarged region for the pupil center detection can be seen especially in Figure 4(a). Here, all algorithms evaluated in [Fuhl et al. 2016a] show greatly reduced detection rates. The runtime of all algorithms is shown in Table 1. As can be seen, the method from Droege [Droege and Paulus 2010] is the fastest, followed by ElSe [Fuhl et al. 2016a]. The proposed approach needed fifteen milliseconds for all circle evaluations. However, boosting the percent of oriented edges to either ten or five percent reduces the runtime to approximately two milliseconds. The final performance boost to reach a runtime below one millisecond is obtained by storing all indexes for each image position and removing all online memory allocations in the code.

Table 1: The runtime of each algorithm measured on the GI4E data set. 10% mean boosted so that only this percentage of gradients is remaining and PI denotes precomputed indexes.

Droege	Timm	Anjith	ElSe	Skodras	Proposed
0.9ms	881ms	51ms	8ms	65ms	15ms 1.9ms (10%) 0.8ms (PI)

Table 2 reports the impact of the boosting percentage to the accuracy and the runtime. For a small percentage, the accuracy is increased. Deactivating larger amounts of gradients reduces the accuracy, but also the runtime.

Table 2: The runtime and accuracy per boosting percentage on the GI4E data set with precomputed indexes for one iteration.

Boosting (%)	0	20	40	60	80	90
Runtime (ms)	5.4	4.6	3.6	2.4	1.1	0.6
Detection (3px,%)	93.4	93.7	93.4	93.0	90.6	83.9

5 CONCLUSION

In this work, we proposed an adaptable remote pupil detection algorithm. It iteratively improves the result and is formulated as an optimization equation. It outperforms the state-of-the-art algorithms in terms of accuracy, detection rate, and runtime. The main advantage of the proposed approach is the unsupervised runtime adaption, which makes it usable for micro-controllers or other low performance computers. Further research will outline extraction of the detected pupil as well as evaluate the applicability to head mounted images. A compiled library with example code together with a version that applies an ellipse fit can be downloaded from: http://www.ti.uni-tuebingen.de/Pupil-detection.1827.0.html.

REFERENCES

Mansour Asadifard and Jamshid Shanbezadeh. 2010. Automatic adaptive center of pupil detection using face detection and cdf analysis. In *Proceedings of the International MultiConference of Engineers and Computer Scientists*, Vol. 1. 3.

T. J. Atherton and D. J. Kerbyson. 1999. Size invariant circle detection. *Image and Vision computing* 17, 11 (1999), 795–803.

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. https://doi.org/10.1109/ITSC.2015.268

D. Droege and D. Paulus. 2010. Pupil center detection in low resolution images. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*. ACM, 169–172.

Wolfgang Fuhl, David Geisler, Thiago Santini, Wolfgang Rosenstiel, and Enkelejda Kasneci. 2016a. Evaluation of State-of-the-art Pupil Detection Algorithms on Remote Eye Images. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct (UbiComp '16)*. ACM, New York, NY, USA, 1716–1725. https://doi.org/10.1145/2968219.2968340

W. Fuhl, T. Santini, and E. Kasneci. 2017. Fast and Robust Eyelid Outline and Aperture Detection in Real-World Scenarios. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 1089–1097. https://doi.org/10.1109/WACV.2017.126

W. Fuhl, T. C. Santini, T. Kübler, and E. Kasneci. 2016b. ElSe: Ellipse Selection for Robust Pupil Detection in Real-world Environments. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications (ETRA '16)*. ACM, New York, NY, USA, 123–130.

Wolfgang Fuhl, Marc Tonsen, Andreas Bulling, and Enkelejda Kasneci. 2016c. Pupil detection for head-mounted eye tracking in the wild: an evaluation of the state of the art. *Machine Vision and Applications* 27, 8 (2016), 1275–1288.

A. George and A. Routray. 2016. Fast and Accurate Algorithm for Eye Localization for Gaze Tracking in Low Resolution Images. *arXiv preprint arXiv:1605.05272* (2016).

Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost Van de Weijer. 2011. *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.

O. Jesorsky, K. J Kirchberg, and Robert W. F. 2001. Robust face detection using the hausdorff distance. In *Audio-and video-based biometric person authentication*. Springer, 90–95.

E. Kasneci. 2013. *Towards the Automated Recognition of Assistance Need for Drivers with Impaired Visual Field*. Ph.D. Dissertation. University of Tübingen, Wilhelmstr. 32, 72074 Tübingen. http://tobias-lib.uni-tuebingen.de/volltexte/2013/7033

Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-based Interaction. In *Adjunct Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14 Adjunct)*. ACM, New York, NY, USA, 1151–1160. https://doi.org/10.1145/2638728.2641695

X. Liu, F. Xu, and K. Fujimura. 2002. Real-time eye detection and tracking for driver observation under various light conditions. In *Intelligent Vehicle Symposium, 2002. IEEE*, Vol. 2. IEEE, 344–351.

Gareth Loy and Alexander Zelinsky. 2003. Fast Radial Symmetry for Detecting Points of Interest. *IEEE Trans. Pattern Anal. Mach. Intell.* 25, 8 (Aug. 2003), 959–973. https://doi.org/10.1109/TPAMI.2003.1217601

S. Milborrow, J. Morkel, and F. Nicolls. 2010. The MUCT Landmarked Face Database. *Pattern Recognition Association of South Africa* (2010). http://www.milbo.org/muct.

David A Robinson. 1963. A method of measuring eye movemnent using a scieral search coil in a magnetic field. *Bio-medical Electronics, IEEE Transactions on* 10, 4 (1963), 137–145.

Thiago Santini, Wolfgang Fuhl, and Enkelejda Kasneci. 2018. PuRe: Robust pupil detection for real-time pervasive eye tracking. *Computer Vision and Image Understanding* (Feb 2018). https://doi.org/10.1016/j.cviu.2018.02.002

Evangelos Skodras and Nikos Fakotakis. 2015. Precise Localization of Eye Centers in Low Resolution Color Images. *Image Vision Comput.* 36, C (April 2015), 51–60. https://doi.org/10.1016/j.imavis.2015.01.006

L. Świrski, A. Bulling, and N. Dodgson. 2012. Robust real-time pupil tracking in highly off-axis images. In *Proceedings of the Symposium on Eye Tracking Research & Applications (ETRA)*. ACM, 173–176. https://doi.org/10.1145/2168556.2168585

Te Tan, Jan Philipp Hakenberg, and Cuntai Guan. 2013. Estimation of glance from EEG for cursor control. In *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*. IEEE, 2919–2923.

F. Timm and E. Barth. 2011. Accurate Eye Centre Localisation by Means of Gradients. *VISAPP* 11 (2011), 125–130.

A. Villanueva, V. Ponz, L. Sesma-Sanchez, M. Ariz, S. Porta, and R. Cabeza. 2013. Hybrid method based on topography for robust detection of iris center and eye corners. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 9, 4 (2013), 25.