

# Pupil Localization Using Geodesic Distance

No Author Given

No Institute Given

**Abstract.** The main contributions of the presented paper can be summarized as follows. Firstly, we introduce a unique and robust dataset of human eyes that can be used in many detection and recognition scenarios, especially for the recognition of driver drowsiness, gaze direction, or eye-blinking frequency. The dataset consists of approximately 85,000 different eye regions that were captured using various near-infrared cameras, various resolutions, and various lighting conditions. The images are annotated into many categories. Secondly, we present a new method for pupil localization that is based on the geodesic distance. The presented experiments show that the proposed method outperforms the state-of-the-art methods in this area.

**Keywords:** Pupil detection, object detection, shape analysis, geodesic distance

## 1 Introduction

In this paper, we focus on the eye detection and pupil localization, especially for recognition of driver behavior inside the car cockpit (e.g. driver fatigue, drowsiness, gaze direction, or eye-blinking frequency). Driving safety is a very important topic and a robust drowsiness detection system can save a lot of human lives. Similar systems can be used for controlling various devices using the eyes as well. Many of the micro-sleep or drowsiness scenarios occur during late-night driving, and it is difficult to recognize them using classical RGB images and sensors which have big problems in low-light conditions. The use of the near-infrared cameras and images may represent an appropriate solution. These images are not so sensitive to different lighting conditions as the day and night images looks relatively similar thanks to IR illuminator, which is part of many IR cameras. Therefore, the quality of night images can be relatively good. However, lots of the state-of-the-art recognition approaches have to be tested or trained on a large number of different eye regions captured by infrared sensors. Obtaining such a large number images is difficult.

Therefore, we believe that the contribution of the paper is twofold. We offer the big dataset that was created using the images of various people using various IR cameras under many different lighting conditions. We proposed a new method for localization of pupil based on the geodesic distance. We tested the method on the two datasets (dataset [1] proposed by us and BioID dataset [10]).

The rest of the paper is organized as follows. The proposed dataset is described in Section 2. The previously presented papers from the area of eye analysis and pupil localization are mentioned in Section 3. In Section 4, the main ideas of proposed method are described. In Section 5, the results of experiments are presented showing the properties of the new method.

## 2 Proposed Dataset



**Fig. 1.** Examples of images captured by an in-car infra-red camera.

Many datasets that contain eye images have been introduced in recent years. These datasets are usually recorded in good conditions and contain images in a high resolution, which makes them suitable for pupil detection, iris detection, eye tracking, or gaze detection. For the tasks such as pupil detection, high-resolution images are required to achieve precise detection. For example, the dataset focused on pupil tracking and the gaze detection was introduced in [12]. This dataset contains  $720 \times 480$  eye images recorded by an infrared camera placed very close to the eye. The dataset presented in [14] is designed for the iris detection and contains the color and infra-red eye images of the size of  $900 \times 800$  pixels, and the iris images of the size of  $400 \times 300$  pixels. In [2], the authors introduced the dataset for eye tracking. This dataset contains videos of various people recorded in the resolution of  $1280 \times 720$  by an RGB camera placed approximately 80 cm in front of the person. This setup makes the detection more challenging since the eye images have lower resolution, but it contains only the color images in which no IR reflections occur. The Closed Eyes In The Wild (CEW) dataset is mentioned in [15]. This dataset consists of approximately 5,000 eye images that are obtained from the Labeled Face in the Wild database [7]. However, the images are without IR reflections. In [13], the authors proposed the ZJU eyeblink database that contains indoor images (also without IR reflections). The eye images can also be obtained from the BioID Face database [9] (1521 gray level images), or the GI4E database [17] (1236 images acquired by a usual camera).

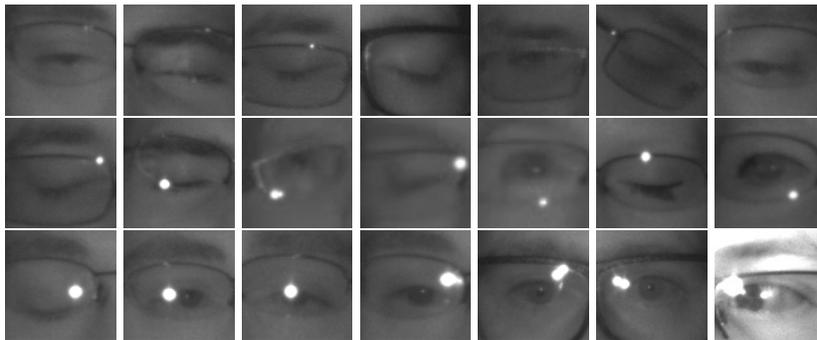
The quality of images that are included in the majority of the mentioned datasets usually does not correspond to the conditions that occur in the real environment when the ambient light or the distance of the person from camera is changing. We can suppose that (in near future) a lot of vehicles will be equipped



**Fig. 2.** Examples of eye detection using the HOG-SVM detector that we created for automatic eye region detection. The images were obtained using different camera settings (e.g. different distances of the person). The eye images presented in the proposed dataset can be used to train the eye detector.

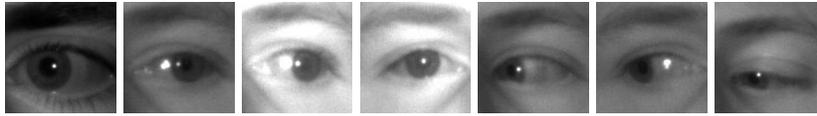
with in-car cameras that watch the driver. Driver’s fatigue or the gaze direction can be determined from these cameras. To reduce the cost of such equipments, low-cost cameras are often used; sample images are presented in Fig. 1. It is seen that the eye covers only a small region of the image. Unfortunately, the extracted eye images are usually of a lower quality, which makes the eye state estimation or the eye parts more difficult than it would be in the previously mentioned eye datasets. However, many detection methods are based on the training and testing process and it is difficult to acquire a big training dataset that consists of suitable images. Therefore, in this paper, we propose a new big dataset of human eye images captured by various near-infrared (NIR) cameras (e.g. Intel Realsense, IDS Imaging cameras).

## 2.1 The Eye Regions



**Fig. 3.** Examples of eyes with glasses and reflections that are included in the proposed dataset.

To obtain eye images, we manually cropped many thousands of eye regions from several NIR images at the first stage; the sample input images are presented in Fig. 1. At the second stage, we used the manually cropped images to train the eye detector based on the histogram of oriented gradients combined with the SVM classifier. This detector was used to automatically extract the eye regions. The example of eye region detection is shown in Fig. 2. After the detection step, we thoroughly checked each detected region and we removed the false positives. We created and checked 85,000 eye images of various people (37 different persons) captured in various lighting conditions and situations; the dataset contains images of different quality with different properties. In the following paragraphs, we will show the examples of eyes that can be found in the proposed dataset.



**Fig. 4.** Examples of eye reflections without glasses that are included in the proposed dataset.

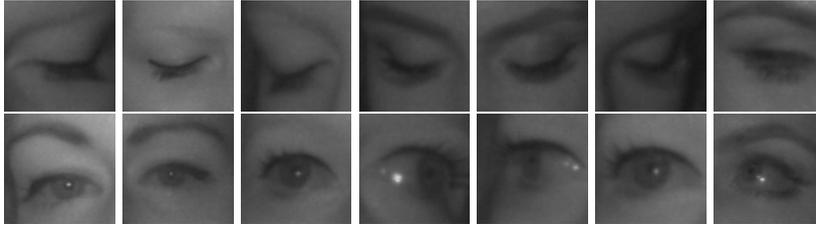
For example, the eyes of persons with eye glasses are shown in the first row in Fig. 3. The problems that often occur with the glasses are reflections. In the dataset, we focused on this problem and we provide lots of images with reflexions. Moreover, for each eye image, we provide three state of reflexions based on the size of reflexion areas in each image; no reflexions (the first row in Fig. 3), small reflexions (Fig. 3), and big reflexions (the third row in Fig. 3). In Fig. 4, the dataset examples show that the reflections can also occur without glasses (e.g. in sclera and pupil).

In general, a lot of women use synthetic eyelash and eyebrow, which can cause the problem for the recognizers of gaze direction and eye state. Therefore, in the dataset, we also provide the information about the gender of each person. The examples of open and closed eyes of woman with a tinted eyelash or eyebrow are shown in Fig. 5.

## 2.2 Annotation and Statistics of Dataset

In order to simplify the comparison of algorithms, the images are divided into several categories, which also makes the dataset suitable for training and testing classifiers. We annotated the following properties:

- subject ID; in the dataset, we collected the data of 37 different persons (33 men and 4 women)
- image ID; the dataset consists of 84,898 images
- gender [0 - man, 1 - woman]; the dataset contains the information about gender of the person in each image (man, woman)



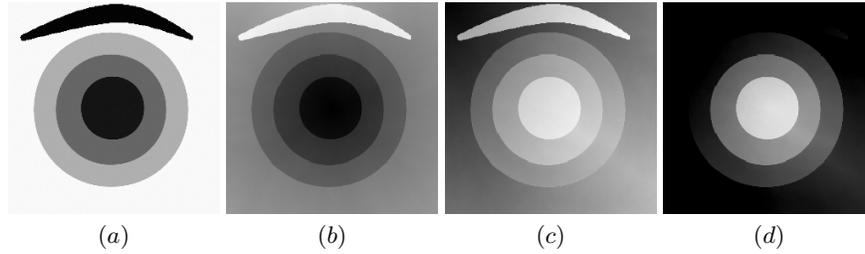
**Fig. 5.** Examples of women eyes with a tinted eyelash or eyebrow that are included in the proposed dataset.

- glasses [0 - no, 1 - yes]; the information if the eye image contains glasses is also provided for each image (with and without the glasses); 24,001 images with glasses
- eye state [0 - closed, 1 - open]; this property contains the information about two eye states (41,945 closed and 42,953 open);
- reflections [0 - none, 1 - small, 2 - big]; we annotated three reflection states based on the size of reflections (66,060 none, 6,129 small, and 12,709 big reflections)
- lighting conditions [0 - bad, 1 - good]; based on the amount of light during capturing the videos (53,630 bad, 31,268 good)
- sensor ID [01 - RealSense, 02 - IDS, 03 - Aptina]; at this moment, the dataset contains the images captured by three different sensors (Intel RealSense RS 300 sensor with  $640 \times 480$  resolution, IDS Imaging sensor with  $1280 \times 1024$  resolution, and Aptina sensor with  $752 \times 480$  resolution)
- aside from the previous properties, we also provide the annotation for approximately 15,000 pupil points (images)

In summary, the dataset contains eye images in low resolutions, images with reflections in the eyes, or with reflections on glasses that are caused by the IR illuminator placed in front of the person. Some eye images in which the head is not aiming at the camera are also included. All these types of images in this dataset make detection of eye parts more difficult than it would be in the previously mentioned eye datasets. The collection is available for the public [1].

### 3 Related Work

In the area of eye analysis and pupil localization, many methods were proposed in recent years. A head-mounted eye-tracking system (starburst) was presented in [11]. The starburst method combines feature-based and model-based approaches. In the first step, the corneal reflection is removed. In the next step, the pupil edge points are located using feature-based method. Finally, the detected edge points are used for ellipse fitting using RANSAC. Another pupil localization method that used RANSAC (for ellipse fitting) was proposed by Swirsky et al.

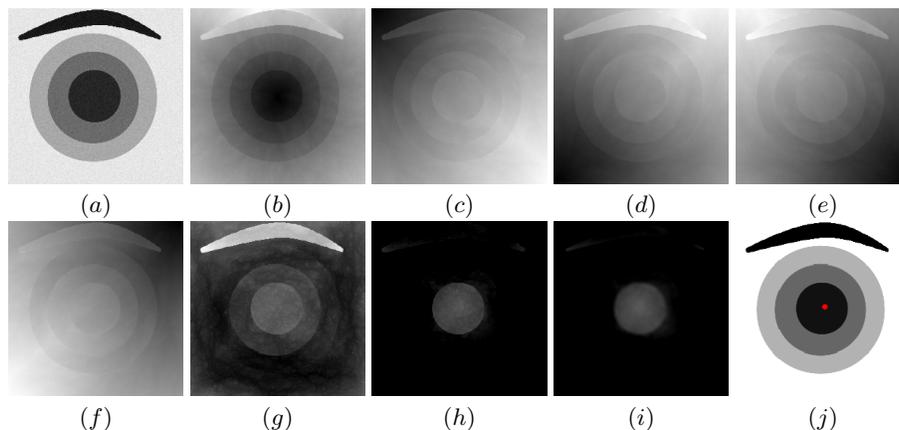


**Fig. 6.** An ideal eye model and the first steps of the proposed method for localization of the approximate (potential) regions of the eyeball. The input image (a). The visualization of the distance function from the centroid (b) and the left corner (c). The difference (d) (only non-zero distances are shown) between (c) and (b). The values of distance function are depicted by the level of brightness; in the input image, minimum amount of noise is added for more realistic visualization of distances.

in [16]. In this method, the approximate position of pupil is estimated using a Haar-like feature detector. In the next step, the  $k$ -means histogram segmentation with RANSAC is performed. An Exclusive Curve Selector (ExCuSe) algorithm for eye-tracking was proposed in [4]. The method uses the Canny edge detector to filter the edges that cannot correspond to the pupil, followed by the straight line removal technique. Finally, the ellipse is fitted using the direct least squares method. A low cost pupil detection method (known as SET ) was proposed in [8]. This method is based on the thresholding combined with segmentation. The pupil border is then extracted using the convex hull method and the best segment is selected. Another pupil detection method named Ellipse Selector (known as ElSe) was presented in [5]. The method is based on the edge filtering, ellipse evaluation, and pupil validation. The method for localization of iris center position was proposed in [6]. In this approach, the selected candidates of boundary points of iris are used to fit an ellipse using the RANSAC algorithm. Evaluation of the state-of-the-art pupil detection algorithms was presented in [3].

## 4 Proposed Method

The main idea of the proposed method for the localization of pupil center is based on the fact that the pupil center can be localized using the geodesic distances in the image. Since the approach also takes into account physiological properties of eyes, let us consider the following ideal case of eye model in Fig. 6 (a); we suppose that the position of eye region is obtained beforehand (e.g. using facial landmarks or classical eye or face detectors). In this model, the pupil is represented by the black circle area that is surrounded by a slightly brighter area of iris, and the iris area is surrounded by brighter sclera. The goal of the pupil localization methods is to find the black circular pupil area only. The location of the pupil center is crucial information for gaze direction recognition.



**Fig. 7.** A noisy eye model and the main steps of the proposed method. The input image (a). The visualization of the distance function from the centroid (b) and from the particular corners (c, d, e, f). The mean of all corner distances (g). The difference (h) between (g) and (b) (only non-zero distances are shown). The convolution step (i). The final position of pupil center (j). The values of distance function are depicted by the level of brightness.

In the first step of the proposed method, the approximate (potential) regions of the pupil are localized. In general, the pupil can be located arbitrarily in the eye region, however (for simplicity), we consider that the position of pupil is close to the center of the eye region in our theoretical model. Therefore, suppose a point (centroid) that is placed in the center of this eye region. Let us compute the geodesic distance function from the centroid to all other points inside the image. The visualization of the distance function values is shown in Fig. 6 (b). In general, the geodesic distance between two points computes the shortest curve that connects both points along the image manifold. Therefore, the values of distance function are low inside the eyeball area; especially in the pupil and iris (Fig. 6 (b)). This step is also important for the removal of eyebrow. It can be observed that the values of distance function are high in the area of eyebrow and, therefore, this region is not expected to be a potential area of pupil. The potential location of pupil is in the areas with low distances.

In the next step, the area of pupil is detected more reliably. We finalize the removal of the eyebrow and sclera in the ideal case. For this purpose, suppose that the geodesic distance is calculated from an arbitrary point that is placed outside the eyeball area; say that in the left corner of the eye region (Fig. 6 (c)). In this case, it is visible that the distance values are high inside the pupil and iris regions. Let us calculate the difference between the distance function values computed from the left corner (Fig. 6 (c)) and the centroid (Fig. 6 (b)). In Fig. 6 (d), the result of difference is shown; only non-zero distance values are shown. It can be seen that the important areas of eyeball (iris and pupil) are correctly

indicated by the value of difference without the sclera and other unwanted eye parts (Fig. 6 (d)).

It is clear that the situation is more complicated in the images with noise (Fig. 7 (a)). One arbitrary point that is located in the left corner will not be enough to equally cover the whole image. Therefore, we suggest to calculate the geodesic distance from each image corner separately (Fig. 7 (c-f)) and use the mean of four distance matrices (Fig. 7 (g)) for the difference with the distance function values from the centroid (Fig. 7 (b)). Again, it can be seen that the eyebrow and sclera are removed using this difference step. In Fig. 7 (h), it can be observed that the mentioned difference gives the highest values in the pupil area in this case.

In the final step, the pupil center is localized by the convolution operation applied to image in Fig. 7 (h). Since we consider that a real center of pupil is the location with the maximum distance value in the previously calculated difference matrix (Fig. 7 (h)), we suggest the use of the Gaussian kernel. The center of the pupil is then determined as the location with the maximal value after the convolution step (Fig. 7 (j)). In Fig. 7 (i), the result of convolution operation is shown.

## 5 Experiments

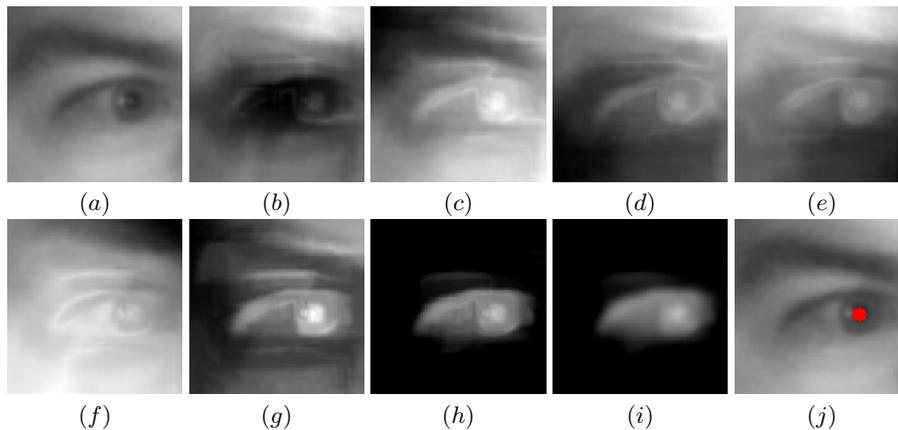


**Fig. 8.** Examples of BioID eye images.

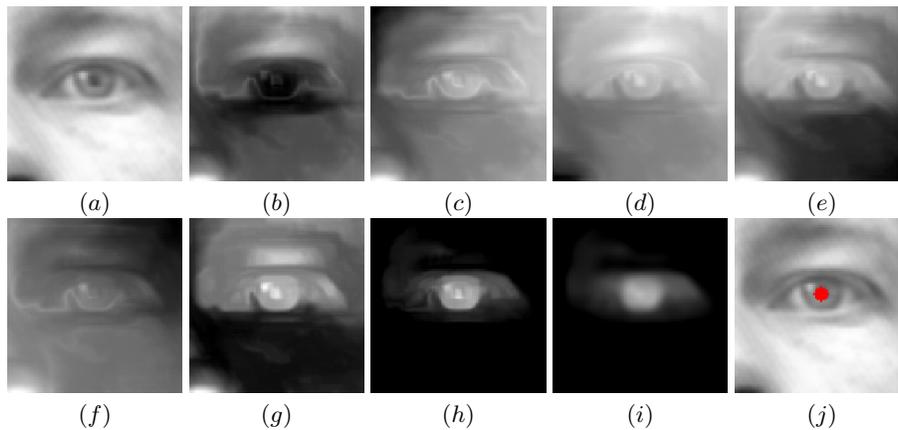
To evaluate the results of the presented method, we used two datasets; BioID [9] and the proposed dataset [1]. The BioID dataset contains 1521 gray level images with the resolution of 384x286 pixels in different indoor environments. We extracted the eye images based on the eye corner positions provided by the authors of the dataset. In Fig. 8, examples of BioID eye images that are used for experiments are shown. It is important to note that the eye images from BioID dataset were purposely extracted with the eyebrow to test the methods in complicated conditions. The size of each extracted eye image (from both datasets) is  $100 \times 100$  pixels in the following experiments.

In Fig. 9, the detection process of the proposed method can be seen in an image taken from the proposed dataset. In Fig. 10, the detection process of the proposed method can be seen in the image taken from the BioID dataset.

For comparison with the state-of-the-art methods, we used two renowned methods: ElSe and ExCuSe. For ElSe, we used the setting for remotely acquired



**Fig. 9.** The main steps of the proposed method in the image taken from the proposed dataset. The input image (a). The visualization of the distance function from the centroid (b) and from the particular corners (c, d, e, f). The mean of all corner distances (g). The difference (h) between (g) and (b) (only non-zero distances are shown). The convolution step (i). The final position of pupil center (j). The values of distance function are depicted by the level of brightness.

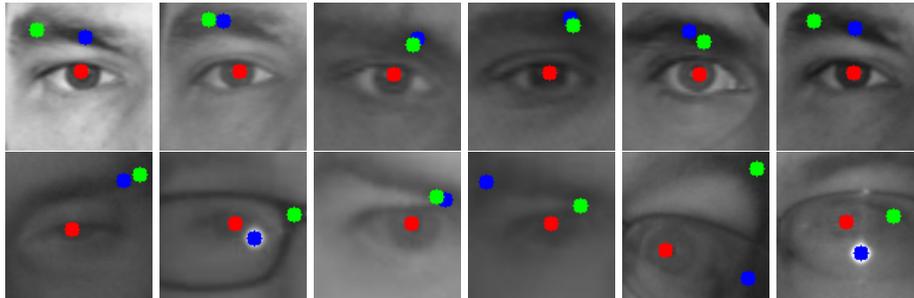


**Fig. 10.** The main steps of the proposed method in the image taken from the BioID dataset. The particular steps are described in Fig. 9.

images published by the authors of the algorithm in [3]. In Table 1, the detection results of methods are shown. The provided (absolute) error is calculated as the Euclidean distance between the ground truth of pupil center and the center provided by the particular detection method. From the results, it can be seen that the proposed detector based on the geodesic distance is the best performing ap-

**Table 1.** A comparison of average absolute errors.

	Error BioID (pixels)	Error Prop. Dataset (pixels)
Proposed method	5.5	6.2
ElSe	10.5	7.5
ExCuSe	11.0	15.7

**Fig. 11.** Examples of comparison of the methods for pupil center localization (the first row: BioID dataset, the second row: proposed dataset); proposed method - red, ElSe - green, ExCuSe - blue.

proach. Especially on the BioID dataset, the proposed method has a very small error 5.5 pixels. The achieved results reflect that the proposed method is developed to work also in the images with eyebrows; in contrast to the other tested methods. In the case of the proposed dataset, our method also achieved a lower error (6.2 pixels) than the state-of-the-art methods. However, it is important to note that ElSe achieved relatively good results (7.5 pixels) on the proposed dataset. Examples of images in which our method works better compared to other tested methods are shown in Fig. 11. In this figure, it is visible that the typical cause of errors is the presence of eyebrow.

As was mentioned in the previous section, the final step of the presented method is based on convolution with the Gaussian kernel. In our experiments, the kernel size was set to  $7 \times 7$  pixels, and the standard deviation was 3. We note that the average time needed for processing one eye region on an Intel core i3 processor (3.7 GHz) took approximately 9.0 milliseconds.

## 6 Conclusion

We believe that the contribution of the paper is twofold. We present a new method for localization of the pupil center. We offer the big dataset [1] used for evaluating the new method (together with the BioID dataset [10]).

Based on the experiment results, we can conclude that the proposed method based on the geodesic distance achieved very promising detection score. It is worth mentioning that various other distances can be used as well. We leave the experiments with various types of distance for future work.

## References

1. Dataset, P.: Proposed dataset. url is not included due to the double-blind review process
2. Ferhat, O., Vilarino, F., Sanchez, F.J.: A cheap portable eye-tracker solution for common setups. *Journal of Eye Movement Research* 7(3) (2014)
3. Fuhl, W., Geisler, D., Santini, T., Rosenstiel, W., Kasneci, E.: 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*. pp. 1716–1725. *UbiComp '16*, ACM, New York, NY, USA (2016), <http://doi.acm.org/10.1145/2968219.2968340>
4. Fuhl, W., Kübler, T., Sippel, K., Rosenstiel, W., Kasneci, E.: Excuse: Robust pupil detection in real-world scenarios. In: Azzopardi, G., Petkov, N. (eds.) *Computer Analysis of Images and Patterns*. pp. 39–51. Springer International Publishing, Cham (2015)
5. Fuhl, W., Santini, T.C., Kübler, T.C., Kasneci, E.: Else: Ellipse selection for robust pupil detection in real-world environments. *CoRR* abs/1511.06575 (2015), <http://arxiv.org/abs/1511.06575>
6. George, A., Routray, A.: Fast and accurate algorithm for eye localization for gaze tracking in low resolution images. *CoRR* abs/1605.05272 (2016), <http://arxiv.org/abs/1605.05272>
7. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. *Tech. Rep. 07-49*, University of Massachusetts, Amherst (October 2007)
8. Javadi, A.H., Hakimi, Z., Barati, M., Walsh, V., Tcheang, L.: Set: a pupil detection method using sinusoidal approximation. *Frontiers in Neuroengineering* 8, 4 (2015), <https://www.frontiersin.org/article/10.3389/fneng.2015.00004>
9. Jesorsky, O., Kirchberg, K.J., Frischholz, R.: Robust face detection using the hausdorff distance. In: *Proceedings of the Third International Conference on Audio- and Video-Based Biometric Person Authentication*. pp. 90–95. *AVBPA '01*, Springer-Verlag, London, UK, UK (2001), <http://dl.acm.org/citation.cfm?id=646073.677460>
10. Jesorsky, O., Kirchberg, K.J., Frischholz, R.W.: Robust face detection using the hausdorff distance. In: Bigun, J., Smeraldi, F. (eds.) *Audio- and Video-Based Biometric Person Authentication*. pp. 90–95. Springer Berlin Heidelberg, Berlin, Heidelberg (2001)
11. Li, D., Winfield, D., Parkhurst, D.J.: Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*. pp. 79–79 (June 2005)
12. McMurrough, C.D., Metsis, V., Rich, J., Makedon, F.: An eye tracking dataset for point of gaze detection. In: *Proceedings of the Symposium on Eye Tracking Research and Applications*. pp. 305–308. *ETRA '12*, ACM, New York, NY, USA (2012)

13. Pan, G., Sun, L., Wu, Z., Lao, S.: Eyeblink-based anti-spoofing in face recognition from a generic webcam. In: 2007 IEEE 11th International Conference on Computer Vision. pp. 1–8 (Oct 2007)
14. Sequeira, A., Chen, L., Wild, P., Ferryman, J., Alonso-Fernandez, F., Raja, K.B., Raghavendra, R., Busch, C., Bigun, J.: Cross-eyed - cross-spectral iris/periorcular recognition database and competition. In: 2016 International Conference of the Biometrics Special Interest Group (BIOSIG). pp. 1–5 (Sept 2016)
15. Song, F., Tan, X., Liu, X., Chen, S.: Eyes closeness detection from still images with multi-scale histograms of principal oriented gradients. *Pattern Recognition* 47(9), 2825 – 2838 (2014)
16. Świrski, L., Bulling, A., Dodgson, N.: Robust real-time pupil tracking in highly off-axis images. In: *Proceedings of the Symposium on Eye Tracking Research and Applications*. pp. 173–176. ETRA '12, ACM, New York, NY, USA (2012), <http://doi.acm.org/10.1145/2168556.2168585>
17. Villanueva, A., Ponz, V., Sesma-Sanchez, L., Ariz, M., Porta, S., Cabeza, R.: Hybrid method based on topography for robust detection of iris center and eye corners. *ACM Trans. Multimedia Comput. Commun. Appl.* 9(4), 25:1–25:20 (Aug 2013), <http://doi.acm.org/10.1145/2501643.2501647>