Journal of Biomedical Optics

BiomedicalOptics.SPIEDigitalLibrary.org

Robust eye tracking based on multiple corneal reflections for clinical applications

Clara Mestre
Josselin Gautier
Jaume Pujol

SPIE.

Clara Mestre, Josselin Gautier, Jaume Pujol, “Robust eye tracking based on multiple corneal reflections for clinical applications,” *J. Biomed. Opt.* **23**(3), 035001 (2018), doi: 10.1117/1.JBO.23.3.035001.

Robust eye tracking based on multiple corneal reflections for clinical applications

Clara Mestre,* Josselin Gautier, and Jaume Pujol
Universitat Politècnica de Catalunya, Davalor Research Center (dRC), Terrassa, Spain

Abstract. A set of methods in terms of both image processing and gaze estimation for accurate eye tracking is proposed. The eye-tracker used in this study relies on the dark-pupil method with up to 12 corneal reflections and offers an unprecedented high resolution imaging of the pupil and the cornea. The potential benefits of a higher number of glints and their optimum arrangement are analyzed considering distinct light sources configurations with 12, 8, 6, 4, and 2 corneal reflections. Moreover, a normalization factor of the pupil-glint vector is proposed for each configuration. There is a tendency for increasing accuracy with the number of glints, especially vertically (0.47 deg for 12 glints configuration versus 0.65 deg for 2 glints configuration). Besides the number of corneal reflections, their arrangement seems to have a stronger effect. A configuration that minimizes the interference of the eyelids with the corneal reflections is desired. Finally, the normalization of the pupil-glint vectors improves the vertical eye tracking accuracy up to 43.2%. In addition, the normalization also limits the need for a higher number of light sources to achieve better spatial accuracy. *© 2018 Society of Photo-Optical Instrumentation Engineers (SPIE)* [DOI: 10.1117/1.JBO.23.3.035001]

Keywords: eye tracking; gaze estimation; interpolation; corneal reflection; pupil-glint vector; pupil fitting.

Paper 170661R received Oct. 10, 2017; accepted for publication Feb. 1, 2018; published online Mar. 2, 2018.

1 Introduction

Video oculography (VOG) has become the most popular eye tracking technique in the last few decades due to its performance, versatility, and low intrusiveness. Actually, some video-based systems represent an interesting alternative to the scleral search coil technique,[1] which is considered to be the gold standard in oculomotor research.[2]

Nowadays, most video-based commercial eye-trackers use the pupil-corneal reflection technique. It is based on the assessment of gaze position from the pupil-glint vectors, that is, the relative distance between the centers of the pupil and one or more corneal reflections. In the image, these reflections are called glints. The number of glints depends on the number of infrared (IR) light sources. The eye tracking process can be divided into two stages: the first one consists of processing the eye images in order to locate the center of the pupil and the glint, and the second estimates gaze position from the detected features in the images.

There are several methods for image processing and eye detection.[3] Some methods rely on the detection of eye features on the images. The pupil is more commonly used as an image feature than the limbus since it has a higher contrast and is less likely to be occluded by the eyelid. Most approaches address pupil detection by thresholding[4,5] or by gradient-based methods, e.g., the Canny edge detector.[6,7] Other approaches consider both methods and decide the one to use depending on the intensity histogram of the images.[8]

Once the pupil has been detected, most existing methods refine its position with ellipse fitting. While simple methods such as the direct least squares fitting of ellipses are highly affected by outliers (points which do not correspond to the pupil edge), there exist other approaches that are more robust to points not lying exactly on the pupil edge. On the one hand, voting-based methods, such as the Hough transform, are effective and exhaustive, although computationally expensive and limited to circular shapes, hence, near-frontal images. On the other hand, searching-based methods, such as the random sample consensus (RANSAC) paradigm,[9] are based on selecting the best of a set of possible candidate ellipses. The RANSAC method is effective in the presence of a relatively large and unknown amount of outliers. It consists in fitting iteratively an ellipse to a small data subset and finding the one with the most agreement within the complete set of candidate pupil edge points.

The application of RANSAC in eye tracking was first described by Li et al.[10] when they proposed the well-known Starburst algorithm. First, the corneal reflection is located through an adaptive threshold and removed by radial interpolation. Then, the pupil edge points are detected at the position along a limited number of rays where the gradient is above a fixed threshold. An ellipse is fitted to the edge points using the RANSAC paradigm. Finally, the result of the ellipse fitting is further optimized using a model-based approach. Despite being computationally costly, the pupil tracking based on Starburst algorithm is highly parallelizable and able to achieve up to 530 frames/s with high-resolution images using a general purpose graphics processing unit.[11]

Yuille et al.[12] proposed a more complex model based on deformable templates, which represents the eyelids with two parabolas and the iris with a circle. This method was extended by Lam and Yan.[13] The combination of both elliptical and complex eye models may quicken the localization and improve

*Address all correspondence to: Clara Mestre, E-mail: clara.mestre@upc.edu

1083-3668/2018/$25.00 © 2018 SPIE

the tracking accuracy.[14,15] Other methods, classified as appearance-based,[3] detect the eyes directly from their appearance in the images, either in the intensity or in a transformed domain. These methods require a large amount of eyes' data of different subjects under different face orientations and illuminations to be trained.

Gaze estimation is the following process, which infers gaze position from the information that has been previously extracted from the images. Gaze estimation methods are typically divided into two main groups: geometry-based and interpolation-based methods. The former methods estimate gaze position based on 3-D models of the eye. The parameters typically used for geometric modeling of the eye include cornea radii, angles between visual and optical axes, index of refraction of the different ocular media, iris radius, and the distance between the pupil and cornea centers. Most geometrical approaches require camera calibration and a geometric model external to the eye composed of light sources, camera and monitor position, and orientation.[3] There is a wide variety of possible setups, from one camera and a single light source[16,17] to multiple cameras and light sources,[18,19] including several other combinations.[17,20,21]

Interpolation-based methods describe the point of gaze as a generic polynomial function of image features (mapping function). As mentioned previously, the pupil and glint centers are commonly used as image features. Subject calibration is required to retrieve the unknown coefficients of the expression. Although the polynomial equation determines not only the accuracy of the system but also the required user calibration process, there are no standards regarding the best mapping function. Several studies analyzed the influence of the order and the number of terms of the polynomial equation on the performance of eye tracking systems.[22–26] Although extensive research has been done to determine the best mapping function, it is not clear whether the conclusions can be generalized to other VOG systems due to the distinct hardware and methodology used in the different studies.

One of the biggest concerns about remote VOG systems is the tolerance to head movements. Although complete head pose invariance is difficult to achieve, the geometry-based methods seem to be more robust to head movements.[24] On the other hand, the accuracy of interpolation-based methods decreases as the user moves away from the calibration position, especially with movements in depth.[22] The normalization of the pupil-glint vectors with respect to the distance between two glints in the eye image seems to reduce the effect of head movements.[25,27] Other scaling factors of the pupil-glint vectors have also been proposed for systems consisting of four IR light-emitting diodes (LEDs).[28] They obtained comparable results to Cerrolaza et al.[25] and matched the performance of more complex geometrical-based methods that require system calibration.[28]

A surge toward developing multiple IR light sources eye tracking systems appeared during the last decades. Several approaches used two IR light sources, one placed near the camera optical axis (on-axis) and the other slightly off-axis in order to generate bright and dark pupil images, respectively.[29–31] This strategy allows to detect the pupil on the images relatively easily by differencing the bright and dark pupil images and thresholding. Yoo and Chung[32] proposed a technique using five IR LEDs and two cameras to estimate gaze position under large head motion. The wide field-of-view camera tracks the face continuously to properly position the other camera, which has a zoom lens to capture magnified images of the eye. One LED is placed on axis to produce bright pupil images and a glint on the cornea. The other four are placed on the corners of the monitor and produce four glints. Gaze position is estimated by computing the cross-ratio of a projective space. Coutinho and Morimoto[33] extended this method by considering the deviation between the visual and optical axes. Similarly, a method relying on homography normalization and using four IR LEDs was proposed by Hansen et al.[34] The offset between the optical and visual axes is modeled to a much higher degree than the cross-ratio based methods, hence achieving better accuracy. Although none of the methods are invariant to depth or in-plane head movements, this homography normalization-based method showed better performance.[34] These methods represent an alternative to the fully calibrated systems since only the light source's position information is needed.

Hennessey and Lawrence[28] described the drawbacks of using a single corneal reflection to compute the pupil-glint vector (e.g., distortion or deletion in large eye rotations). They proposed a technique to track a pattern of four corneal reflections and applied a second order interpolation equation to map the pupil-glint vector onto gaze position. In this method, an algorithm that compensates for translation, distortion, addition, and deletion of corneal reflections is applied and the pupil-glint vector is formed from the pupil center to the centroid of the corneal reflections pattern. Thus, the resulting vector is robust to loss, translation, and distortion of the glints. The proposed technique managed to estimate the point of regard in all head positions and eye rotations tested while up to 27% of the time the point of regard would have been lost if only one corneal reflection was used.[28]

The use of multiple IR light sources has also become common in recent portable commercial eye-trackers. For example, Tobii Pro Glasses 2 (Tobii, Falls Church, Virginia) is a wearable eye tracking system that embeds eight IR LEDs per eye. The Oculus Rift DK2 system (SensoMotoric Instruments, Berlin, Germany) is an eye-tracker embedded in a virtual reality head mounted display, which contains six IR LEDs per eye.

The eye-tracker used in this study consists of a multiple-corneal reflections dark-pupil system, which offers an unprecedented high resolution imaging of the pupil and the cornea (640×480 pixels images with a field-of-view of 16 mm at the pupil plane). It is embedded in the Eye and Vision Analyzer (EVA) system (Davalor Salud, Spain), which is a stereoscopic virtual reality instrument to perform the optometric tests related to objective and subjective refraction, binocular vision, and accommodation while patients are watching a 3-D video game. The vergence-accommodation conflict is avoided by adjusting the accommodative plane with the vergence plane through an electro-optical lens. The EVA system allows to perform both visual diagnosis and visual therapy. The eye-tracker synchronously records both right and left eye movements during all the optometric tests. The intrusiveness of the whole system needs to be restricted to the least possible degree due to its wide clinical application requirements. Hence, the head movements are only restricted with a forehead rest.

This paper presents new methods for accurate eye tracking with multiple corneal reflections using interpolation-based techniques. The advantages of a higher number of glints and their optimum arrangement are analyzed to provide new insights for the community. Moreover, a normalization of the pupil-glint vectors method is proposed to increase the eye tracking spatial accuracy.

2 Methodology

2.1 *Experimental Methodology*

The study was approved by the Ethics Committee of Hospital Mutua de Terrassa (Terrassa, Spain). It followed the tenets of the Declaration of Helsinki and all subjects gave informed written consent after receiving a written and verbal explanation of the nature of the study.

Eye images of 20 subjects [mean age ± standard deviation (SD) of 31.9 ± 9.5 years] with normal or corrected-to-normal visual acuity were taken with the two cameras embedded in each of the two optical modules comprised in the EVA system [Figs. 1(a) and 1(b)]. Each optical module consists of three subsystems: the autorefractometer, the vision, and the eye-tracker [Fig. 1(c)]. The refractive error of participants was measured with the autorefractometer subsystem based on a Hartmann–Shack wavefront sensor. The vision subsystem allows the patients to see the liquid crystal on a silicon 2048×1536 pixels resolution microdisplay with a field-of-view of 26 deg horizontally and 19.8 deg vertically. The spherical and cylindrical refractive errors are corrected with an electro-optical lens and the rotation of two cylindrical lenses, respectively, which are adjusted in order to avoid the need for wearing glasses. Spherical refractive errors ranging from −18 diopters (D) to +13 D and cylindrical errors up to 5 D can be compensated. Finally, the eye-tracker subsystem consists of a complementary metal-oxide-semiconductor sensor recording 640×480 pixel images with a spatial resolution of 0.0045 deg and a frame rate of 30 Hz. The illumination system consists of a ring of 12 IR LEDs (950 nm). Participants' heads are partially immobilized with a forehead rest [Fig. 1(d)].

Participants were asked to sit down, put their head on a forehead rest, and fixate monocularly a black cross, which subtended an angle of 0.2 deg, on a mid-gray background. This cross was displayed in a sequence of nine positions of a 3×3 grid during both the calibration and validation procedures in the same order. The stimulus was displayed for 1.3 s at each position and eye images were acquired starting 0.3 s after the onset of the stimulus [Fig. 1(e)].

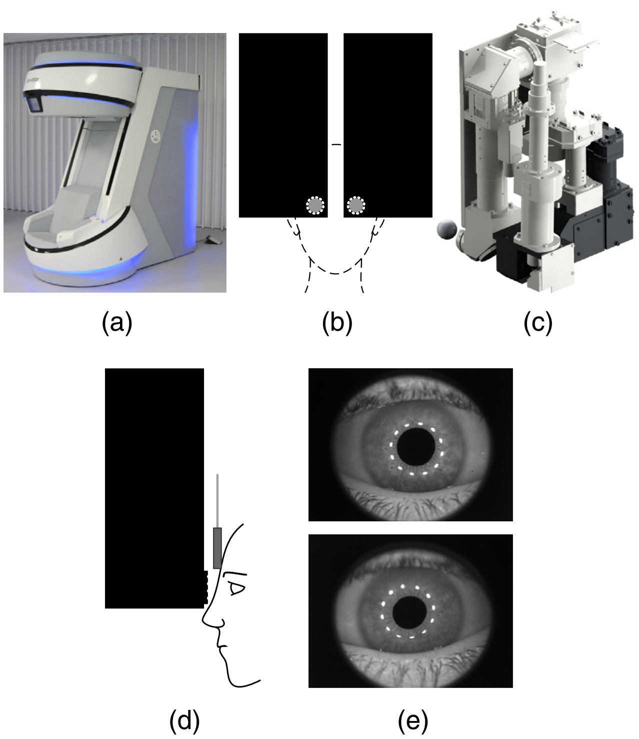


Fig. 1 Setup. (a) EVA system. Once the patient sits, the head of the machine goes down and adjusts its position according to the patient's height. (b) Frontal view of the optical modules. (c) Schematic representation of the optical module; the eye-tracker is shaded in dark gray while the vision and autorefractometer subsystems are represented in light gray. (d) Lateral view of the system. (e) Right (top) and left (bottom) eye images captured with the system.

2.2 *Image Processing*

Eye images acquired during the experimental procedure were processed offline with an implementation of the algorithm in MATLAB (R2015b; MathWorks, Natick, Massachusetts). For simplicity, only data from the right eye were analyzed. The Starburst algorithm was extended in order to fit the characteristics of illumination sources, resolution of our eye images, and improve the accuracy of the original algorithm.

The location of the corneal reflections process was adapted to the content of the images used in this study, which have up to 12 glints. Then, an ellipse was fitted to the centroid of each glint using a direct least squares fitting method.[35] Instead of removing the glints by radial interpolation as in original Starburst, they were simply masked in order to avoid their interference in the pupil contour detection. Although the iterative process to detect the edge pupil points was essentially maintained from the original Starburst, the feature points were redefined as the positions along each ray where the gradient is maximum. That way, the feature points are located more precisely on the pupil edge, which otherwise would tend to underestimate the pupil border and its size.

One of the main challenges of pupil tracking is the detection of the pupil when it is partially occluded by dropped eyelids or downward eyelashes. In order to overcome this issue, our second proposal is an eyelid detector based on the visibility of the upper glints. When the complete ring of 12 glints was visible, the rays were traced from the estimated center along 360 deg, as was done originally. However, when some glint was missing, no rays were traced in that direction. Thus, pupil edge points were not located erroneously on the eyelid or eyelashes.

Once the edge pupil points were detected, the RANSAC method was applied to find the best fitted ellipse. A subset of six points instead of five, as originally suggested,[9] was chosen randomly but ensuring that they were equally distributed around all the regions of the pupil. Although these contributions produce a low improvement on accuracy, its main benefit is in terms of computing efficiency (Fig. 2). In addition, geometrical constraints on the maximum and minimum radius and eccentricity of the fitted ellipse were added based on anatomical parameters of the pupil.[36]

Since images were processed offline, computation time was not critical. The prototype version of this implementation written in MATLAB and run by a processor Intel i5-4200M CPU at 2.50 Hz with 8 GB of RAM operates at ∼1.11 frames/s. Around 70% of the algorithm's runtime is needed for the localization and masking of the 12 corneal reflections while the other 30% is needed for the pupil edge detection and ellipse fitting on the pupil. The results presented in this paper were obtained with this implementation. However, we have worked on a faster

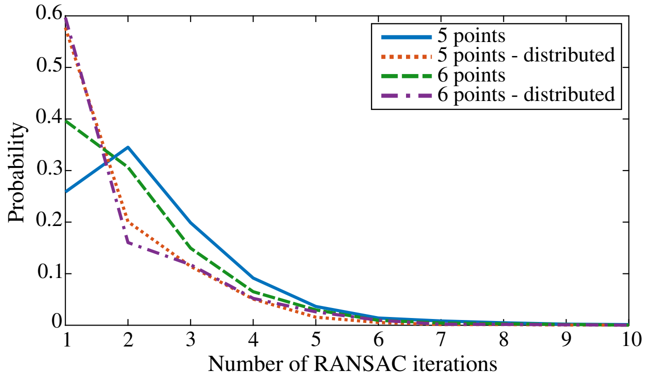


Fig. 2 Histogram of the number of RANSAC iterations with the original algorithm (5 points, solid), choosing 5 points and distributing them spatially (5 points—distributed, dotted), considering a subset of 6 points without constraints about distribution (6 points, dashed) and considering 6 points and distributing them spatially (6 points—distributed, dotted-dashed).

implementation of the algorithm using Nvidia compute unified device architecture (CUDA) parallelism in order to reduce the computation time and run the algorithm in real time. The hardware used to test this implementation consisted of a GPU Nvidia Quadro K5200 with 2304 CUDA cores and 8 GB of memory. The computation time could be reduced to around 2 ms/image.

The first objective of this study was to analyze the performance of the eye-tracker with different configurations of light sources. The tested configurations (Fig. 3) were chosen to study the optimal number of glints, the putative benefits of higher number of glints, and their optimum arrangement considering the possible interference of the eyelids. In a preliminary study, the different configurations were tested switching off the corresponding LEDs. After confirming by visual inspection that similar levels of image luminance could be obtained by retaining only two light sources and increasing their illumination power, we decided to acquire all eye images with the 12 LEDs switched on in order to simplify the experimental procedure. Then, the corresponding glints were removed from the eye images using radial interpolation assuming that their intensity profile follows a symmetric bivariate Gaussian distribution.

2.3 *Gaze Estimation*

Before estimating the gaze position with an interpolation-based method, the data obtained from the eye images were filtered using a trimmed mean to select the 50% of the 30 images available at each point. The trimmed or truncated mean reduces the effects of outliers on the calculated average by removing a certain percentage of the largest and smallest values before computing it.

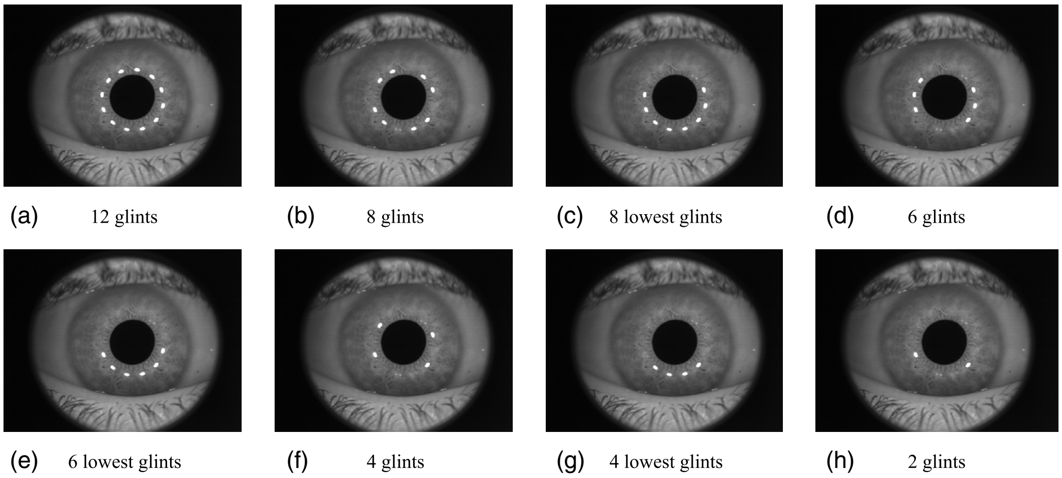
A second order polynomial Eq. (1) was used to map the tracked image features onto gaze position:

$$\begin{pmatrix} PoR_x \\ PoR_y \end{pmatrix} = C \cdot \begin{pmatrix} 1 \\ \vartheta_x \\ \vartheta_y \\ \vartheta_x^2 \\ \vartheta_y^2 \\ \vartheta_x \vartheta_y \end{pmatrix}, \tag{1}$$

where PoR_x and PoR_y are the horizontal and vertical coordinates, respectively, of the point of regard, C is the coefficient matrix determined during calibration, and ϑ_x and ϑ_y are the horizontal and vertical components, respectively, of the pupil-glint vector.

During the calibration procedure, Eq. (1) was used to calculate the polynomial coefficients in the matrix C assuming that PoR_x and PoR_y are the horizontal and vertical coordinates, respectively, of the stimulus, and computing the pupil-glint vector of the images captured during this procedure. During the validation procedure, the coordinates of the point of regard could be computed from the pupil-glint vector of the validation images and the known C matrix.

As mentioned previously, robustness against head movements is one of the main challenges of current video-based eye-trackers. The system used in this study exceeded the minimum hardware required (i.e., two light sources) to normalize the pupil-glint vectors, which was shown to improve overall spatial accuracy in interpolation-based eye tracking methods.[25,28] The normalization proposed by Sesma-Sanchez et al.[27] based on the interglint distance was adapted to the content of the images of each tested configuration.


(a) 12 glints (b) 8 glints (c) 8 lowest glints (d) 6 glints

(e) 6 lowest glints (f) 4 glints (g) 4 lowest glints (h) 2 glints

Fig. 3 Eye images with the tested light sources configurations. (a) 12 glints, (b) 8 glints, (c) 8 lowest glints, (d) 6 glints, (e) 6 lowest glints, (f) 4 glints, (g) 4 lowest glints, and (h) 2 glints.

When normalization was not applied, the components of the pupil-glint vector of the eye images captured during both calibration and validation procedures were computed as follows:

$$\vartheta_x = p_x - g_x, \quad \vartheta_y = p_y - g_y, \tag{2}$$

where p_x and p_y are the image coordinates of the pupil center. In the configurations with 12 and 8 glints, g_x and g_y are the image coordinates of the center of the ellipse fitted on the glints centroids, whereas in the configurations with 6, 4, and 2 glints, g_x and g_y are the image coordinates of the mean glints position. This difference among configurations is due to the fact that at least five points are required to fit an ellipse.

In the configurations with 12 and 8 glints, when normalization was applied, the horizontal and vertical components of the pupil-glint vector of the eye images captured during both calibration and validation procedures were defined as follows:

$$\vartheta_x = \frac{p_x - g_x}{r}, \quad \vartheta_y = \frac{p_y(1-k) - g_y}{r}, \tag{3}$$

where r is the major radius of the glints ellipse and k is a vertical weighting factor to attribute a higher weight to the glints in order to compensate for the higher uncertainty in the pupil detection, especially vertically. The same value of k was used for both calibration and validation procedures. Its optimum value for each configuration of light sources was determined empirically as the one that optimizes the accuracy of the eye-tracker averaged for all participants.

In the configurations with six or less glints, the normalized pupil-glint vectors of the eye images captured during both calibration and validation procedures were computed as follows:

$$\vartheta_x = \frac{p_x - g_x}{D}, \quad \vartheta_y = \frac{p_y(1-k) - g_y}{D}, \tag{4}$$

where g_x and g_y become the image coordinates of the mean glints position and D is the mean Euclidean distance between opposite glints.

Two normalization methods were proposed for the different configurations due to the limitation of the minimum number of points required to fit an ellipse. Although an ellipse could be fitted on 6 glints, preliminary results showed no robust results even when the normalization was not applied in those configurations. Therefore, in the configurations 6 glints and 6 lowest glints, the pupil-glint vectors were computed as in the configurations 4 glints, 4 lowest glints, and 2 glints, considering the mean glints position and the Euclidean distance between them. As a result, a unique normalization method was applied on each light source configuration.

The gaze estimation algorithm was also written in MATLAB. Approximately 30 ms were needed to compute the coefficients of the second order polynomial equation using the pupil-glint vectors extracted from the eye images captured during the calibration procedure, and 10 ms were needed to interpolate and compute the point of regard during the validation procedure.

2.4 *Evaluation*

The eye-tracker performance was evaluated by analyzing the horizontal and vertical accuracies. They were defined as the horizontal and vertical angular distances between the interpolated points of regard on the image plane using the eye images registered during the validation procedure and the real target positions that subjects were fixating on. The data reported in this paper correspond to the average horizontal and vertical accuracies obtained by averaging all the horizontal and vertical distances, respectively, over the 3×3 grid.

The determination of the optimum value of the factor k of Eqs. (3) and (4) requires to evaluate the eye-tracker's accuracy. Hence, it was determined from images obtained during the validation procedure.

2.5 *Statistical Analysis*

Statistical analysis was performed using IBM SPSS Statistics 23 (IBM; Armonk, New York). Nonparametric statistics were used after checking that most variables did not follow a normal distribution by applying the Shapiro–Wilk test and comparing the skewness and kurtosis statistics to the standard error.

Friedman tests were performed along both horizontal and vertical directions to compare the accuracy of the eight configurations. Significance was set at $p < 0.05$. When significance was obtained, post-hoc comparisons of configurations were made by Wilcoxon signed-rank tests with a Bonferroni adjustment given by the number of possible pairwise configuration comparisons, with significance $p < 0.05/28$. The same tests were also used to compare the accuracy of the eight configurations when the pupil-glint vectors were normalized.

Spearman's correlations were applied to identify associations between the differences in accuracy for certain pairs of configurations and other features of each configuration, such as the percentage of images in which the eye was detected or the percentage of images in which some glints were occluded.

Finally, the Wilcoxon signed-rank test was performed to compare the horizontal and vertical accuracies for each configuration without applying the normalization of the pupil-glint vectors and normalizing them.

3 Results

Before analyzing the differences in terms of accuracy for the different tested configurations, the intrinsic repeatability of the algorithm is described. It is defined as the within-subject standard deviation of the accuracy in both horizontal and vertical directions. It might be thought as a descriptor of the random component of measurement error and is due to randomness in the selection of the initial subset of points for the RANSAC ellipse fitting. For the original 12 glints configuration, the within-subject standard deviation horizontally was 0.027 deg [95% confidence interval (CI), 0.026 deg to 0.027 deg], whereas vertically, it was 0.034 deg (95% CI, 0.033 deg to 0.035 deg).

Table 1 gives the descriptive statistics of horizontal and vertical accuracies for each configuration.

The Friedman test showed significant differences in both horizontal [$\chi^2(7) = 18.27,\ p = 0.011$] and vertical [$\chi^2(7) = 20.50,\ p = 0.005$] accuracies for the different configurations. The post-hoc test performed along each direction showed statistically significant differences horizontally between the configurations 8 lowest glints and 4 glints ($p = 0.001$). Any pairwise comparison vertically showed significant differences.

Although they were not statistically significant, the differences in accuracy between the configurations with the same number of light sources (i.e., 8 glints and 8 lowest glints, 6 glints and 6 lowest glints, and 4 glints and 4 lowest glints) are especially remarkable and might be justified by the fact that in some cases, the upper glints were occluded by the eyelid

Table 1 Median [interquartile range (IQR)] of the horizontal and vertical accuracies in degrees for different configurations.

Configuration	Horizontal accuracy (deg)	Vertical accuracy (deg)
12 glints	0.41 (0.19 to 0.49)	0.47 (0.40 to 0.70)
8 glints	0.42 (0.23 to 0.53)	0.53 (0.40 to 0.70)
8 lowest glints	0.38 (0.17 to 0.52)	0.54 (0.31 to 0.70)
6 glints	0.48 (0.24 to 0.55)	0.74 (0.47 to 0.84)
6 lowest glints	0.43 (0.18 to 0.55)	0.60 (0.48 to 0.87)
4 glints	0.51 (0.26 to 0.58)	0.69 (0.46 to 0.87)
4 lowest glints	0.39 (0.21 to 0.61)	0.57 (0.48 to 0.81)
2 glints	0.44 (0.21 to 0.57)	0.65 (0.45 to 0.87)

(Fig. 4). There was a moderate, positive, and significant correlation between the difference in vertical accuracy between the configurations 8 glints and 8 lowest glints and the percentage of images in which the glints 8, 9, and 11 -, and only these-, were occluded ($r_s = 0.46$, $p = 0.042$). There was no significant correlation horizontally.

In the four configurations with 6 and 4 glints, all the considered glints had to be visible, otherwise the eye could not be detected on that frame. Hence, the improvement in accuracy when the lowest glints were considered cannot be justified directly by the occlusion of some upper glints. Alternatively, it can be explained by the improvement in robustness defined as the percentage of images in which the eye is detected. There was a moderate, positive, and significant correlation between the difference in vertical accuracy between the configurations 6 glints and 6 lowest glints and the difference in robustness between both configurations ($r_s = 0.65$, $p = 0.003$). For the configurations 4 glints and 4 lowest glints, there was also a moderate, positive, and significant correlation ($r_s = 0.50$, $p = 0.029$). There was no correlation horizontally for the configurations 6 glints and 6 lowest glints nor for 4 glints and 4 lowest glints.

Table 2 shows the descriptive statistics of horizontal and vertical accuracies for each configuration when the normalization of the pupil-glint vectors was applied according to Eqs. (3) and (4).

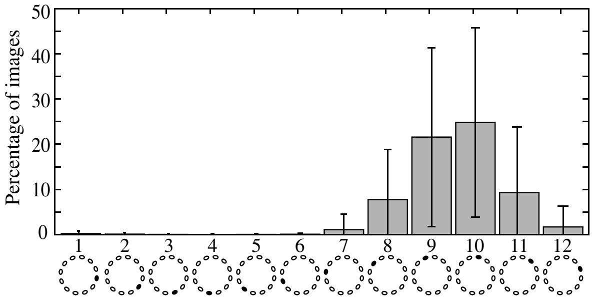


Fig. 4 Percentage of images averaged for all participants in which each glint is occluded. The corresponding occluded glints are represented in black in the schemes below the bars. Error bars show ±1 SD.

Table 2 Median (IQR) of the horizontal and vertical accuracies in degrees for different configurations when the pupil-glint vectors were normalized.

Configuration	Horizontal accuracy (deg)	Vertical accuracy (deg)
12 glints	0.44 (0.19 to 0.51)	0.39 (0.22 to 0.66)
8 glints	0.41 (0.20 to 0.52)	0.38 (0.29 to 0.73)
8 lowest glints	0.39 (0.16 to 0.54)	0.46 (0.32 to 0.64)
6 glints	0.46 (0.24 to 0.55)	0.42 (0.25 to 0.89)
6 lowest glints	0.45 (0.18 to 0.54)	0.39 (0.24 to 0.69)
4 glints	0.47 (0.24 to 0.60)	0.41 (0.28 to 0.93)
4 lowest glints	0.45 (0.22 to 0.63)	0.44 (0.27 to 0.72)
2 glints	0.43 (0.22 to 0.57)	0.37 (0.26 to 0.74)

The Friedman test showed significant differences in horizontal accuracy for the different configurations [$\chi^2(7) = 16.75$, $p = 0.019$]. The post-hoc test showed statistically significant differences between the configurations 8 lowest glints and 4 glints ($p = 0.001$). There were no significant differences in vertical accuracy for the different configurations [$\chi^2(7) = 6.04$, $p = 0.535$].

There were no statistically significant differences (Wilcoxon signed-rank test) in any configuration between the horizontal accuracy when the pupil-glint vectors were not normalized and when the normalization was applied. Moreover, in most configurations, the differences were lower than the horizontal within-subject standard deviation. However, the normalization significantly improved the vertical accuracy and the differences were above the vertical within-subject standard deviation in all configurations. The relative improvement of the median vertical accuracy due to the normalization of the pupil-glint vectors ranged from 43.2% for 6 glints to 14.8% for 8 lowest glints (Fig. 5).

4 Discussion

4.1 *Light Sources Configurations*

The median horizontal accuracy of the eye-tracker used in this study is systematically better than the median vertical accuracy in all configurations. A similar tendency was found by Cerrolaza et al.[24] in their study of polynomial mapping functions to optimize the calibration process of interpolation-based systems. Since there is no clearly preferred direction in the dispersion of gaze in tasks of sustained fixation,[37] it is hypothesized that this difference is due to the interference of the eyelid and eyelashes in the detection of the upper pupil region. The higher uncertainty in this region also implies a poorer repeatability of the algorithm vertically than horizontally.

Although there are no statistically significant differences of vertical accuracy with the distinct tested configurations, there is a tendency for increasing accuracy with the number of glints, especially in the vertical direction. There is a significant negative correlation between the number of glints and the median vertical accuracy (Table 1) of the best configurations (i.e., 12 glints, 8 lowest glints, 6 lowest glints, 4 lowest glints, 2 glints; $r_s = -0.90$, $p = 0.037$). The correlation is weaker and not

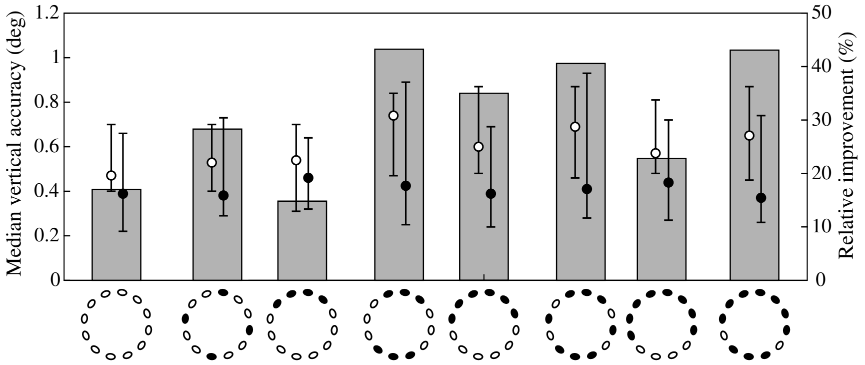


Fig. 5 Median vertical accuracy without normalizing the pupil-glint vectors (empty circles) and applying the normalization (solid circles). Each configuration of light sources is represented in the schemes below the bars, where active and inactive LEDs are represented in white and black, respectively. Error bars show the interquartile range. Bars correspond to the relative improvement of vertical accuracy due to the normalization.

significant in the horizontal direction ($r_s = -0.50$, $p = 0.391$). The between-subjects variability of the accuracies is rather similar in all configurations and along both directions.

The arrangement of the light sources seems to have a stronger effect than the number of glints itself. To our knowledge, this is the first study that addresses the question of the best positioning of the IR LEDs to optimize the accuracy of a VOG system. Figure 4 confirms the intuitive thought that the upper glints are the most likely to be occluded by the eyelid. The fact that the lower eyelid hardly ever interferes with the glints justifies our choice of considering the lowest corneal reflections in the configurations 8 lowest glints, 6 lowest glints, and 4 lowest glints. However, one should bear in mind the specific eye-tracker setup used in this study with the cameras placed in front of the eyes. The results regarding the optimum arrangement of light sources might not be applicable to other systems in which the cameras are located in other positions.

In the configurations in which an ellipse is fitted on the glints (12 glints, 8 glints, and 8 lowest glints), the total number of active glints was not always visible. As can be seen in Fig. 6, there was a considerable percentage of images in which some glints were occluded, especially in the 12 glints configuration [Fig. 6(a)]. However, since a dataset of at least five points is required to fit an ellipse, at least five corneal reflections needed to be visible so as to track the eye in each frame. The main difference between the configurations 8 glints and 8 lowest glints was the number of glints available to fit the ellipse, which was not always 8 due to eyelid occlusion. The mean ± SD percentage of images in which all eight corneal reflections were visible with the 8 glints configuration was 76.9% ± 19.7% [Fig. 6(b)], whereas with the 8 lowest glints configuration, they were all visible in 95.6% ± 5.5% of the images [$p = 0.001$; Fig. 6(c)]. Therefore, the improvement in accuracy when the lowest glints were considered might be explained by a more robust ellipse fitting with a larger dataset of points.

In the configurations with six or less corneal reflections, the average glints position was used to compute the pupil-glint vectors. In these configurations, the number of glints must be the same in all frames. Otherwise, the components of the pupil-glint vectors would be modified regardless of eye movements, which in turn would lead to an incorrect measurement of eye position. Hence, the advantage of the configuration 6 lowest glints over the 6 glints is reflected in the robustness of the system (i.e., the percentage of frames in which the eye is detected).

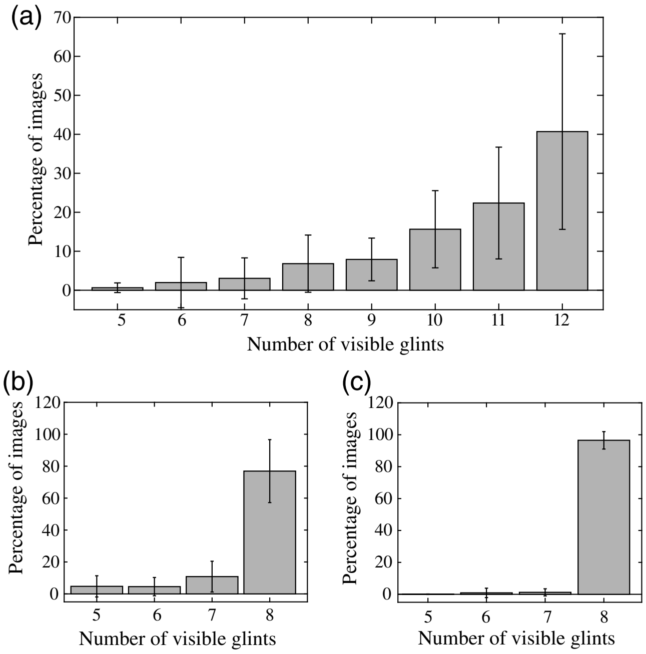


Fig. 6 Percentage of images averaged for all participants in which there were five or more visible glints for the configurations (a) 12 glints, (b) 8 glints, and (c) 8 lowest glints. Error bars show ±1 SD.

The mean ± SD robustness of 6 glints configuration was 89.9% ± 11.4%, whereas with the 6 lowest glints configuration, it was 96.6% ± 3.4% ($p = 0.001$). Similarly, the mean ± SD robustness of 4 glints configuration was 89.9% ± 11.5%, whereas with the 4 lowest glints configuration, it was 96.9% ± 2.6% ($p = 0.001$).

The improvement of accuracy in the configurations in which more data is available (higher robustness) suggests that an increase in sampling frequency might lead to a better performance of the eye-tracker not only regarding temporal measurements but also in terms of spatial accuracy.

4.2 *Normalization of the Pupil-Glint Vectors*

Our results suggest that the normalization of the pupil-glint vectors is an effective method to improve the accuracy of VOG systems.

Previous studies working with eye-trackers with two or more IR light sources tested the tolerance to head movements in depth applying different types of normalization.[24,27,28] The head movements in other directions (parallel to the display) were not tested since they were shown to be considerably less problematic in VOG systems.[22] To do so, they acquired eye images locating the subjects' head in three different positions separated by 5 cm. The experimental procedure of our study did not include testing in different locations of the head due to the shadow depth of field of the eye-tracker's cameras. Nevertheless, the head of the patients was not fully immobilized.

On the one hand, it is hypothesized that the improvement in accuracy shown in all configurations when the pupil-glint vectors were normalized might be partially due to the compensation of small, although not quantified, head movements allowed by the forehead rest, since previous works obtained satisfactory results applying similar normalization methods with this purpose.[24,27] On the other hand, as seen in Fig. 5, the relative improvement in accuracy due to normalization was different for each configuration. This implies that the normalization of pupil-glint vectors might have further effects besides the compensation of head movements and might be due to the k factor, whose optimum value varies among configurations.

The main improvement when the pupil-glint vectors were normalized was in terms of vertical accuracy. Actually, the differences in horizontal direction were neither statistically significant nor relevant, since in most configurations they were below the within-subject standard deviation, which means that they might be simply due to the intrinsic variability of the algorithm. The stronger effect of normalization vertically than horizontally might be justified by the fact that most of the coefficients of the mapping function have a higher value in the polynomial equation for determining the PoR_y than in the equation for the PoR_x. Thus, when the normalization of the pupil-glint vectors was applied, the change in the computed gaze positions was more prominent in the vertical direction than horizontally.

As discussed previously, horizontal accuracy was below 0.5 deg and already better horizontally than vertically when the pupil-glint vectors were not normalized. This suggests that the weaker effect of normalization horizontally might be also explained by the fact that horizontal accuracy might be limited by lack of exactitude in stages prior to gaze estimation, such as the image acquisition and processing or the dispersion of gaze itself due to fixational eye movements. The pronounced improvement vertically contributed to reduce the difference in performance between both directions and equalize the horizontal and vertical accuracies.

Since normalization had a small effect in the horizontal direction, the between-subjects variability of horizontal accuracy was similar than when the normalization was not applied. However, the interquartile range of vertical accuracy was considerably wider with normalization, except for the configurations 8 lowest glints, 6 lowest glints, and 2 glints in which it was rather similar. As shown by the error bars in Fig. 5, the distribution of vertical accuracy tends to become more asymmetric when normalization was applied. This means that in most participants, normalization improved vertical accuracy although in some subjects with poorer accuracy, the effect was weaker.

Several eye tracking methods published previously were evaluated using the Euclidean distance between the estimated point of gaze and the true eye position instead of considering separately the horizontal and vertical directions. The accuracy of the eye-tracker used in this study was also computed as the Euclidean distance for the purpose of comparing it with existing methods. Its median value was 0.6 deg for the 12 glints configuration and normalizing the pupil-glint vectors, which is better than the average accuracy of 1 deg of visual angle shown by the original Starburst algorithm.[10] Other interpolation-based methods using one glint obtained an accuracy around 0.8 deg.[26,38] Cerrolaza et al.[24] obtained a considerably better accuracy with two IR LEDs, a second order interpolation equation, the interglint distance to normalize the pupil-glint vectors, and with the patients' head stabilized using a chin rest (0.2 deg horizontally and 0.3 deg vertically).

Comparable values of accuracy were obtained with geometry-based and head pose invariant models.[19] Several systems consisting of more than two IR LEDs rely on the cross-ratio[33,39,40] of a projective space or homography normalization.[34] Although these methods allow head movement, their optimum accuracy values, which are below 0.5 deg, were shown with the head stabilized with a chin rest.[34,39]

To conclude, different lightening configurations for on-axis eye tracking have been proposed and studied. In particular, the interference of the corneal reflections with the upper eyelid has been emphasized. One should take into account that the high variability in the anatomical shape of eyelids leads to high variability of the results. Then, the configuration with the best performance might be different depending on factors such as the ethnicity or the age of the eye-tracker's users. The proposed normalization of the pupil-glint vectors seems to be an effective method to improve the accuracy of VOG systems. It also counteracts the tendency for increasing accuracy with the number of glints. Therefore, if they are properly positioned, our normalization proposal allows to be independent from the need for higher number of light sources.

Disclosures

The authors have commercial relationship with Davalor Salud (financial support).

Acknowledgments

This research was supported by the Spanish Ministry of Economy and Competitiveness under the grant DPI2014-56850-R, the European Union, and Davalor Salud. C.M. would like to thank the Generalitat de Catalunya for a PhD studentship grant. The authors would like to thank Arantxa Villanueva and Rafael Cabeza for the insightful discussions and Juan Mompeán for his help with the implementation of the algorithm using CUDA.

References

1. J. N. van der Geest and M. A. Frens, "Recording eye movements with video-oculography and scleral search coils: a direct comparison of two methods," *J. Neurosci. Methods* **114**, 185–195 (2002).
2. H. Collewijn, "Eye movement recording," in *Vision Research: A Practical Guide to Laboratory Methods*, R. H. S. Carpenter and J. G. Robson, Eds., pp. 245–285, Oxford University Press, Oxford, United Kingdom (1998).
3. D. W. Hansen and Q. Ji, "In the eye of the beholder: a survey of models for eyes and gaze," *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(3), 478–500 (2010).
4. S. Goñi et al., "Robust algorithm for pupil-glint vector detection in a video-oculography eyetracking system," in *Proc. 17th Int. Conf. Pattern Recognition*, IEEE, Cambridge (2004).

5. A.-H. Javadi et al., "SET: a pupil detection method using sinusoidal approximation," *Front. Neuroeng.* **8**, 4 (2015).
6. L. Świrski, A. Bulling, and N. Dodgson, "Robust real-time pupil tracking in highly off-axis images," in *Proc. of the Symp. on Eye Tracking Research and Applications (ETRA '12)*, pp. 173–176, ACM Press, Santa Barbara (2012).
7. W. Fuhl et al., "Else: ellipse selection for robust pupil detection in real-world environments," in *Proc. of the Ninth Biennial ACM Symp. on Eye Tracking Research & Applications (ETRA '16)*, pp. 123–130, ACM Press, Charlseton (2016).
8. W. Fuhl et al., "ExCuSe: robust pupil detection in real-world scenarios," in *Computer Analysis of Images and Patterns*, G. Azzopardi and N. Petkov, Eds., pp. 39–51, Springer, New York (2015).
9. M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM* **24**(6), 381–395 (1981).
10. D. Li, D. Winfield, and D. J. Parkhurst, "Starburst: a hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches," in *Proc. Vision for Human-Computer Interaction Workshop, IEEE Computer Vision and Pattern Recognition Conf.* (2005).
11. J. Mompean et al., "GPU-accelerated high-speed eye pupil tracking system," in *27th Int. Symp. Computer Architecture and High Performance Computing (SBAC-PAD)*, pp. 17–24, Florianópolis, Brazil (2015).
12. A. L. Yuille, P. W. Hallinan, and D. S. Cohen, "Feature extraction from faces using deformable templates," *Int. J. Comput. Vision* **8**(2), 99–111 (1992).
13. K. M. Lam and H. Yan, "Locating and extracting the eye in human face images," *Pattern Recognit.* **29**, 771–779 (1996).
14. G. Chow and X. Li, "Towards a system for automatic facial feature detection," *Pattern Recognit.* **26**(12), 1739–1755 (1993).
15. J.-Y. Deng and F. Lai, "Region-based template deformation and masking for eye-feature extraction and description," *Pattern Recognit.* **30**(3), 403–419 (1997).
16. E. D. Guestrin and M. Eizenman, "General theory of remote gaze estimation using the pupil center and corneal reflections," *IEEE Trans. Biomed. Eng.* **53**(6), 1124–1133 (2006).
17. A. Villanueva and R. Cabeza, "Models for gaze tracking systems," *J. Image Video Process.* **2007**(4), 2 (2007).
18. T. Ohno and N. Mukawa, "A free-head, simple calibration, gaze tracking system that enables gaze-based interaction," in *Proc. of the Eye Tracking Research and Applications Symp. on Eye Tracking Research and Applications (ETRA '04)*, pp. 115–122, ACM Press, San Antonio (2004).
19. D. Beymer and M. Flickner, "Eye gaze tracking using an active stereo head," in *Proc.IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, Vol. 2, IEEE Computer Society (2003).
20. J. J. Kang et al., "Investigation of the cross-ratios method for point-of-gaze estimation," *IEEE Trans. Biomed. Eng.* **55**(9), 2293–2302 (2008).
21. C. Hennessey, B. Noureddin, and P. Lawrence, "A single camera eye-gaze tracking system with free head motion," in *Proc. of the 2006 Symp. on Eye Tracking Research and Applications (ETRA '06)*, pp. 87–94, ACM Press, San Diego (2006).
22. C. H. Morimoto and M. R. M. Mimica, "Eye gaze tracking techniques for interactive applications," *Comput. Vision Image Understanding* **98**(1), 4–24 (2005).
23. K. P. White, T. E. Hutchinson, and J. M. Carley, "Spatially dynamic calibration of an eye-tracking system," *IEEE Trans. Syst. Man. Cybern.* **23**(4), 1162–1168 (1993).
24. J. J. Cerrolaza, A. Villanueva, and R. Cabeza, "Study of polynomial mapping functions in video-oculography eye trackers," *ACM Trans. Comput. Interact.* **19**(2), 1–25 (2012).
25. J. J. Cerrolaza, A. Villanueva, and R. Cabeza, "Taxonomic study of polynomial regressions applied to the calibration of video-oculographic systems," in *Proc. Symp. Eye Tracking Research and Applications*, pp. 259–266 (2008).
26. X. L. C. Brolly and J. B. Mulligan, "Implicit calibration of a remote gaze tracker," in *Conf. Computer Vision and Pattern Recognition Workshop*, p. 134 (2004).
27. L. Sesma-Sanchez, A. Villanueva, and R. Cabeza, "Gaze estimation interpolation methods based on binocular data," *IEEE Trans. Biomed. Eng.* **59**(8), 2235–2243 (2012).
28. C. A. Hennessey and P. D. Lawrence, "Improving the accuracy and reliability of remote system-calibration-free eye-gaze tracking," *IEEE Trans. Biomed. Eng.* **56**(7), 1891–1900 (2009).
29. A. Tomono, M. Iida, and Y. Kobayashi, "A TV camera system which extracts feature points for non-contact eye movement detection," *Proc. SPIE* **1194**, 2–12 (1990).
30. Y. Ebisawa and S.-I. Satoh, "Effectiveness of pupil area detection technique using two light sources and image difference method," in *Proc. of the 15th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, pp. 1268–1269, IEEE, San Diego (1993).
31. C. H. Morimoto et al., "Pupil detection and tracking using multiple light sources," *Image Vision Comput.* **18**, 331–335 (2000).
32. D. H. Yoo and M. J. Chung, "A novel non-intrusive eye gaze estimation using cross-ratio under large head motion," *Comput. Vision Image Understanding* **98**, 25–51 (2005).
33. F. Coutinho and C. Morimoto, "Free head motion eye gaze tracking using a single camera and multiple light sources," in *19th Brazilian Symp. on Computer Graphics and Image Processing*, pp. 171–178, IEEE, Manaus (2006).
34. D. W. Hansen, J. S. Agustin, and A. Villanueva, "Homography normalization for robust gaze estimation in uncalibrated setups," in *Proc. of the Symp. on Eye-Tracking Research and Applications (ETRA '10)*, pp. 13–20, ACM Press, Austin (2010).
35. A. Fitzgibbon, M. Pilu, and R. B. Fisher, "Direct least square fitting of ellipses," *IEEE Trans. Pattern Anal. Mach. Intell.* **21**(5), 476–480 (1999).
36. D. Atchison and G. Smith, "The pupil," in *Optics of the Human Eye*, 1st ed., pp. 21–29, Butterworth-Heinemann, Edinburgh (2000).
37. C. Cherici et al., "Precision of sustained fixation in trained and untrained observers," *J. Vision* **12**(6), 31 (2012).
38. P. Blignaut, "Mapping the pupil-glint vector to gaze coordinates in a simple video-based eye tracker," *J. Eye Mov. Res.* **7**(1), 1–11 (2014).
39. J.-B. Huang et al., "Towards accurate and robust cross-ratio based gaze trackers through learning from simulation," in *Proc. Symp. Eye Tracking Research and Applications*, pp. 75–82 (2014).
40. Z. Zhang and Q. Cai, "Improving cross-ratio-based eye tracking techniques by leveraging the binocular fixation constraint," in *Proc. Symp. Eye Tracking Research and Applications*, pp. 267–270 (2014).

Clara Mestre received her degree in optics and optometry from the Universitat Politècnica de Catalunya (UPC) in 2014 and her MSc in photonics from UPC, Universitat Autònoma de Barcelona, Universitat de Barcelona, and Institut de Ciències Fotòniques in 2015. She is currently pursuing her PhD in optical engineering in the Davalor Research Center (dRC), UPC. Her research interests are in eye tracking and ocular movements, especially in the relationship between the eye's motility and binocular vision.

Josselin Gautier received his MS degree in electronic and informatics engineering from the University of Nantes, in 2007 and his PhD in computer science from the University of Rennes 1, in 2012. From 2013 to 2015, he was a postdoctoral research fellow with the VHS Laboratory from Anglia Ruskin Cambridge University. Thereafter, he established as postdoctoral researcher with the dRC, UPC, Spain. His interests cover oculomotricity, eye tracking, visual attention, and their clinical applications.

Jaume Pujol received his BS in physics from the Universitat Autònoma de Barcelona (1981) and his PhD from the same university (1990). He directed the Optics and Optometry Department (1994 to 2000) and Center for Sensors, Instruments, and Systems Development (CD6) (1997-2009) among others, and was the main entrepreneur of VISIOMETRICS S.L. He is the director of the dRC, president of the Davalor clinical council, and scientific director of the CD6. His research is focused on visual biophotonics.