



Fast and robust ellipse detection algorithm for head-mounted eye tracking systems

Ion Martinikorena¹ · Rafael Cabeza¹ · Arantxa Villanueva¹  · Iñaki Urtasun¹ · Andoni Larumbe¹

Received: 30 October 2017 / Revised: 16 February 2018 / Accepted: 20 February 2018 / Published online: 23 May 2018
© The Author(s) 2018

Abstract

In head-mounted eye tracking systems, the correct detection of pupil position is a key factor in estimating gaze direction. However, this is a challenging issue when the videos are recorded in real-world conditions, due to the many sources of noise and artifacts that exist in these scenarios, such as rapid changes in illumination, reflections, occlusions and an elliptical appearance of the pupil. Thus, it is an indispensable prerequisite that a pupil detection algorithm is robust in these challenging conditions. In this work, we present one pupil center detection method based on searching the maximum contribution point to the radial symmetry of the image. Additionally, two different center refinement steps were incorporated with the aim of adapting the algorithm to images with highly elliptical pupil appearances. The performance of the proposed algorithm is evaluated using a dataset consisting of 225,569 head-mounted annotated eye images from publicly available sources. The results are compared with the better algorithm found in the bibliography, with our algorithm being shown as superior.

Keywords Eye tracking · Head mounted · Pupil detection

1 Introduction

The first experiments using eye trackers began in early twentieth century [1]. At that time, gaining an understanding of eye movements was one of the main objectives of those evaluations [2]. Today, the technology has evolved, considerably widening the range of applications for which eye trackers can be employed. As the computational capacity of the existing equipment increases and as the price of the available technology decreases, more powerful and computationally expensive algorithms have been introduced for eye tracker devices. Thus, the range of applications using eye trackers has also become wider, including human–computer interaction and eye movement analysis.

Over the last few years, considerable efforts have been made to broaden the use of this technology to new application environments. Making this technology more robust and cheaper is key in order to apply this knowledge to conditions

that are not completely controlled, i.e., outside the laboratory, such as in outdoor environments in which illumination cannot generally be controlled. Using eye trackers for driving experiments is one of the clearest examples, i.e., rapid light variations occur in an uncontrolled fashion, and most of the existing algorithms fail. Other cases are those carried out by users wearing head-mounted eye trackers in alternative environments such as shopping areas, and with individuals engaging in sports, work and other everyday activities. Moreover, the use of head-mounted devices also produces elliptical-shaped pupils with high eccentricity compared to those obtained when remote eye trackers are used. These “wilder” frameworks produce undesirable image artifacts, such as reflections, occlusions, blurring, and cases in which the pupil is cut by contact lenses or glasses or by problems caused by an eye mask.

As far as is known, gaze estimation methods use the center of the pupil to estimate the Point or Regard (PoR) or the Line of Sight (LoS), depending on the kind of experiment that is being carried out. Consequently, an accurate detection of the pupil center is key in obtaining a reliable measurement of gaze. Eye tracking is considered the algorithm that is employed to analyze the image captured by the camera, while gaze estimation is used to refer to the procedure that is responsible for estimating gaze using the

✉ Arantxa Villanueva
avilla@unavarra.es

Rafael Cabeza
rcabeza@unavarra.es

¹ Electrical and Electronics Engineering Department, Public University of Navarre, Pamplona, Spain

results of the eye tracking stage [3]. The present proposal contributes to the area of eye tracking. More specifically, this paper presents a novel algorithm for detecting the pupil center in non-controlled environments in a more robust and accurate manner. The algorithm shows outstanding results compared to other methods that were previously published using state-of-the-art challenging eye tracking databases [4].

The accurate detection of the pupil center, together with the detection of corneal glint(s), has been studied since the very beginning of the technology, and several methods have been published [5]. However, the number of studies that have considered natural environments is reduced, wherein most of the methods that work in laboratory conditions fail. Recently, Fuhl et al. presented a paper [4] in which well-known algorithms, such as Starburst and some of the other more recent state-of-the-art algorithms, were evaluated. The Starburst algorithm [6] bases its method on an initial approximation of the pupil center, from which rays with varying angular resolution are calculated. This method is based on detecting the pupil and the contour points in the rays, assuming that a significant gray level variation is produced according to a threshold. Each one of the contour points calculated in the first pass of the algorithm is iterated, and new contour points are detected. For each iteration, an ellipse is fitted using the potential contour points, and its center is calculated until no significant variation is produced. The algorithm proposed by Świrski et al. [7] is devoted to solving the cases in which the angle between the eye and camera's optical axis is high, producing elliptical pupils with high eccentricity; thus, the assumption of circularity fails. The method proposed is based on using Haar features representing center-contour appearance. The result of the convolution with Haar features is used for a segmentation process of the image in which the threshold is calculated by employing a k -means algorithm. The detected region is considered the pupil, and an ellipse-fitting procedure is carried out using the edge points calculated by the Canny operator. The SET algorithm [8] uses a semiautomatic procedure. First, a threshold is manually selected to obtain a binary image in which the pupil is contained. For the blobs that are obtained, a signature value is calculated using the values of the x and y components of the contour points with respect to the center of the blob as a function of the angle. Both distributions are approximated by a sinusoidal function. The blob for which the aspect ratio between the sinusoidal functions is closer to one is selected as the pupil, i.e., the more circular shape. The PupilLabs algorithm was developed together with the open source code known as Pupil [9]. This algorithm presents high robustness in the presence of glints that overlap the pupil. As in the algorithm suggested by Świrski, Haar features are employed. Afterward, a Canny operator is used, and the edge points having darker gray values are selected. The resulting segments are analyzed using specific connectivity rules and curvature cri-

teria. Ellipse-fitting techniques are also applied in order to select the best candidate. The ExCuSe method is one of the most recent algorithms [10], and it uses different approaches based on the presence of glints. On the one hand, in cases when a glint is detected, the edge points are calculated using Canny. A thinning procedure is then applied to the calculated edges, and specific ad hoc rules are applied in order to select the segments that are potential candidates for being part of the pupil contour. As in the rest of the algorithms, an ellipse is calculated, and its center is estimated. On the other hand, when no glints are detected, the pupil is segmented using an automatic threshold calculated from the image information. Subsequently, angular integral projection function (AIPF) is employed. This transform obtains the center of different projection angles using the binarized pupil information. The projections are weighted by using the gray level. The data that are calculated as a result of the projections are used to estimate an approximate pupil center. This point is employed to crop the image, and the aforementioned edge-processing procedure is applied in order to refine the pupil contour detection. The ray-tracing algorithm proposed by Starburst is also applied. Finally, the ELSe algorithm [11] proposes the use of an edge-processing algorithm similar to the one used by ExCuSe. After the edge selection stage, an ellipse is fitted for all the sets of points that are potential candidates to be pupil contour points. If the ellipses do not match a specific area, shape and gray level criteria are rejected. For the rest of the ellipses, a goodness parameter is calculated using the gray level and the shape information. The best of them is selected to be the pupil ellipse, assuming that a goodness threshold is exceeded. In cases when no ellipse is detected, a convolution is performed using circular masks to obtain a probability map that is further post-processed to approximate the pupil center. Using a completely different perspective, we found that some works employ deep learning, i.e., convolutional neural networks (CNN), to estimate pupil center. CNN have been demonstrated to be the best solution for many artificial vision problems. Valuable efforts have been made in eye tracking for low-resolution systems, i.e., for images captured with a webcam [12] for which the results are far from the ones obtained by high-resolution systems. Regarding the topic under study in this paper, we found the recent work in which a CNN-based method was applied to high-resolution images obtained in the "wild" [13].

This paper presents a novel algorithm, the fast robust ellipse detection algorithm (FREDA) algorithm, that beats the existing algorithms in terms of robustness and accuracy. The proposed method is based on the fast radial symmetry transform (FRST) [14] which is based on calculating the point presenting the highest radial symmetry in the image that is assumed to be the pupil center. This method was tested using the same framework that was used for the five state-of-

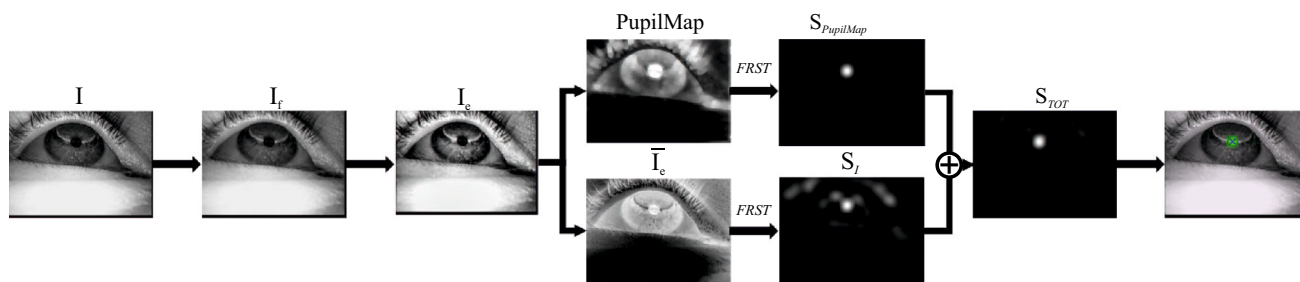


Fig. 1 Flowchart of FREDa algorithm

the-art methods mentioned before and showed outstanding results.

In Sect. 2, the algorithm is described in detail, as well as the two center refinement stages. In addition, the set of images used to evaluate the algorithms is presented in this section. Section 3 shows the performance of the algorithms in the presented datasets, as well as a comparison to the ELSe algorithm. Finally, in Sect. 4 the conclusions of the present paper are explained.

2 Methods

The proposed approach uses the fast radial symmetry transform as the basis for detecting the pupil center. The FRST was also used in the algorithm presented by Skodras et al. [15] for remote eye tracking systems. Our contribution was directed at validating the use of the symmetry transform in head-mounted, gray-scale, high-resolution images. The algorithm bases the pupil center estimation on the detection of the highest radial symmetry point as the result of fast radial symmetry transform [14]. This transform detects circularly shaped zones in an image; thus, it is particularly appropriate for detecting the pupil center, assuming that the pupil’s appearance is typically circular. Nevertheless, in cases where the pupils appearance is more elliptical, this method tends to mark the center closer to the foci of the apparent ellipse. To avoid this problem, the FREDa I and FREDa II variations are presented, which incorporate an additional center refinement stage. The presented methods were developed using MATLAB.

The stages of the FREDa algorithm are summarized as follows (see Fig. 1): first, a preprocessing stage of the image is applied in order to adapt it to the subsequent processes. Then, the radial transform is computed both, on the negative of the preprocessed image, labeled \bar{I}_e , and on the created pupil-enhanced image, labeled PupilMap. The two contributions are summed, and the center is defined by taking the coordinates of the maximum point of the resulting transformation, defined as S_{TOT} .

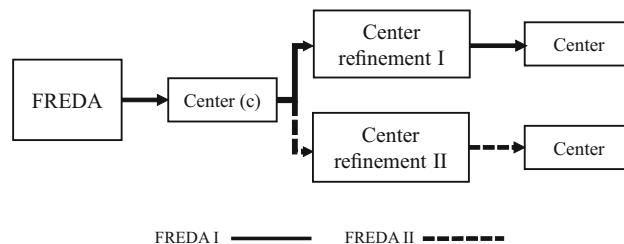


Fig. 2 Flowchart of the FREDa I and FREDa II algorithms

The further center-refining stages of the FREDa I and FREDa II take the center, c , given by FREDa on the source image as a starting point (see Fig. 2). The center is chosen as the seed point for the successive region growths with which it is intended to fit the pupil. Thus, the corrected pupil center is considered the center of the best fitted ellipse to the region most closely matching the pupil shape. The difference between the two algorithms lies in the way in which the similarity between the pupil and the growing region is determined.

2.1 FREDa

2.1.1 Image preprocessing

First, an image preprocessing stage is implemented in order to prepare it for the subsequent processes. This stage is compounded by two operations: a low-pass filter and an adaptive histogram equalization (see Fig. 1). Due to the calculation of the image gradient in the posterior radial symmetry transform, a low-pass filter is applied, resulting in the I_f image, in order to reduce the effect of noise on the border detection. A 5×5 Gaussian filter is used to implement the low-pass filter.

Adaptive image equalization is then performed, calculating the output image I_e , to increase the contrast between the pupil and the background, thus obtaining more defined pupil edges. This procedure equalizes the histogram by small patches of the image rather than the entire image. Assuming that the pupil size is approximately a 10th of the image’s width, a subdivision of 10 columns and 10 rows is selected

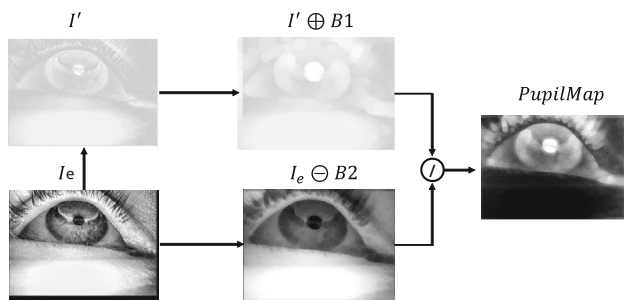


Fig. 3 PupilMap construction

to which the equalization is applied. The output histogram of each region approaches a uniform distribution. To eliminate block effects between adjacent regions, they are combined using bilinear interpolation. To prevent noise from increasing in uniform areas of the image, the contrast is limited to a threshold that is chosen empirically, having a value of 0.01.

2.1.2 Pupil enhancement: PupilMap

In this step, specific transformations are applied to the image in order to enhance the pupil region, thus facilitating the posterior identification and center estimation using the radial symmetry transform. The steps are based on the method proposed by Skodras et al. [15] for RGB images obtained with remote eye tracking systems. The process was modified to adapt it to gray-scale images, head-mounted and high-resolution images. Figure 3 depicts the steps involved in obtaining the enhanced pupil image. In summary, the process consists in dividing a bright pupil image by a dark pupil one, thus increasing the contrast between the pupil area and the rest of the image. As seen in the obtained enhanced image, the pupil is intended to be the brightest part of the image; thus, only the positive directions of the gradient are taken into account when applying the radial symmetry transform.

Equation (2) shows the operations to obtain the PupilMap, where \oplus and \ominus symbolize, the morphological dilation and the erosion, respectively.

$$\text{PupilMap} = \frac{I' \oplus B1}{I_e \ominus B2 + \varepsilon} \quad (1)$$

$B1$ and $B2$ are flat, circular structuring elements whose radii are $R_{B1} = \text{Imagewidth}/20$ and $R_{B2} = R_{B1}/2$, respectively. The use of circular structuring elements emphasizes round patterns in the image, thus increasing the radial symmetry of the pupil zone. We applied a parabolic gray-scale transformation G to the I_e input image ($G(I_e) = I'$). The G transformation brightens dark pixel areas that have gray levels below 0.2, approximating the negative transform, while the light parts, i.e., above 0.8, remain unchanged, approximating the identity transform. For normalized gray values

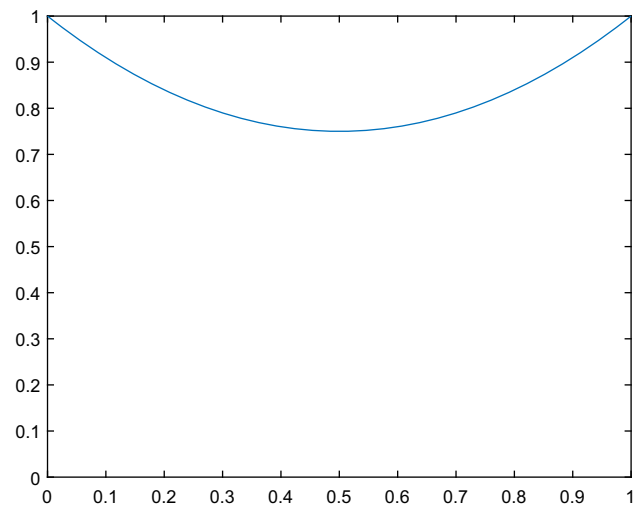


Fig. 4 G gray-scale transform (I')

between 0.2 and 0.8, the contrast is significantly reduced (see Fig. 4). Thus, when dividing the dilated transformed image of I' by the eroded input of I_e , the brighter areas in I_e will tend to cancel out while the pupil zone will be enhanced. To avoid dividing by zero, the factor ε is summed. This factor is defined as:

$$\varepsilon = \text{mean}(I_e \ominus B2) \quad (2)$$

2.1.3 Fast radial symmetry transform: FRST

As described previously, in our approach, pupil center estimation is based on the detection of the point presenting the highest radial symmetry in the image. The implemented method is a modification of the transform proposed by Loy et al. [14]. This radial symmetry transform is a highly efficient computational approach.

The transform is calculated for a set of radii, $n \in N$, where the values in N are selected empirically, considering the even numbers between 20 and 34 pixels for the application proposed. A discontinuous range of integers is selected in order to improve the computing speed. This reduction does not affect the accuracy of the center estimation.

First, the gradient of the image is calculated by a Sobel 3×3 operator. Only significant gradient values are considered. A threshold is empirically chosen as 5% of the maximum magnitude value of the gradient obtained in each image. Only gradient values greater than this threshold are considered, thereby reducing the number of pixels to be computed in the transform. Once the gradient values are calculated, the FRST is applied in order to detect the pupil center. Next, the FRST is summarized for clarity [14].

For each significant point, p , of the gradient, the affected pixel, p_{af} , is defined as the point located at a distance n from p and to which the gradient vector in p points at, as follows:

$$p_{af} = p + \text{round} \left(\frac{g(p)}{\|g(p)\|} n \right) \tag{3}$$

Notice that, as the positive values of the gradient are associated with directions from dark to bright regions, only pixels that are in bright zones will be affected. As the images to be transformed have the pupil zone brighter (see Fig. 1), this means that only bright zones will be detected by the transform, thus avoiding dark circular shapes, which are caused by reflections or other bright artifacts in the original image.

For each radius n , an orientation projection O_n and a magnitude projection image M_n are created using the affected pixels p_{af} in the following ways:

$$O_n(p_{af}) = O_n(p_{af}) + 1 \tag{4}$$

$$M_n(p_{af}) = M_n(p_{af}) + \|g(p)\| \tag{5}$$

$O_n(p_{af})$ represents the number of pixels voting for p_{af} while $M_n(p_{af})$ represents the contribution, in terms of magnitude, of the voting pixels for a radius of value n . The contribution of the radial symmetry of the radius n is obtained by combining both matrices and convolving them with a Gaussian smoothing mask, A_n , with a mean, μ , equal to $2n$ in the following way:

$$S_n = (M_n \cdot O_n^\alpha) * A_n \tag{6}$$

where α denotes the radial strictness parameter. The α parameter determines how strictly the radial symmetry must be for the transform to return a high interest value. High α values eliminate non-radially symmetric features, while choosing a low α value includes non-circular symmetry points of interest. The parameter is set, empirically, as $\alpha = 2$.

For each radius n a smoothed voting map, S_n , is obtained, the values of which represent the contribution of each point to the local radial symmetry for a radius n . The final map is calculated by averaging all the voting maps as:

$$S(x, y) = \frac{1}{|N|} \sum_{n \in N} S_n(x, y) \tag{7}$$

In the algorithm presented in this paper, we propose to select the radius n for which S_n is the one that has a higher peak value, making S_n the final transformation and c the estimated pupil center.

$$S(x, y) = \max_{n \in N} \{S_n(x, y)\} \tag{8}$$

$$c = \max(S(x, y)) \tag{9}$$

2.2 FREDa I

As previously described, the FREDa I algorithm is an adaptation of the FREDa that aims to refine the estimated pupil

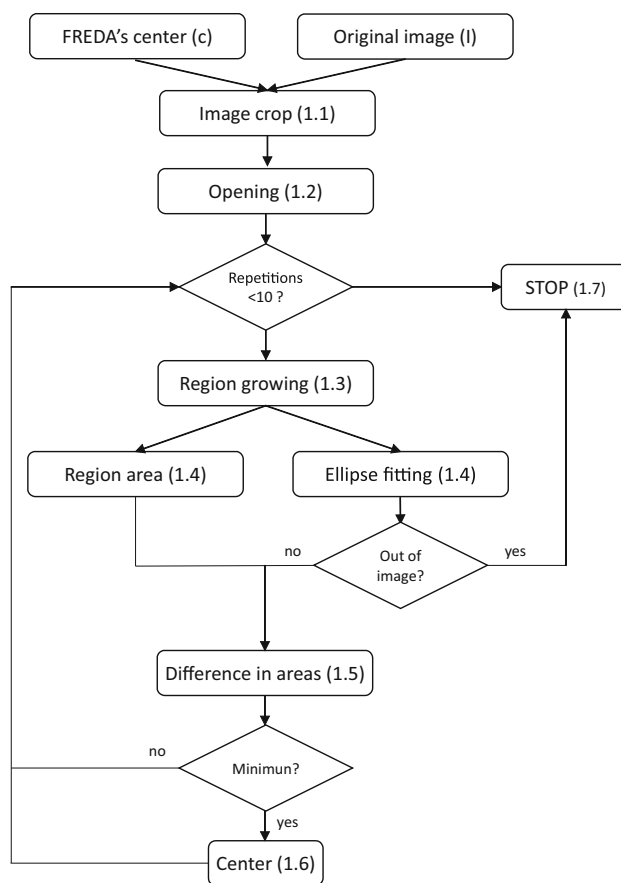


Fig. 5 FREDa I center refinement block diagram

center in pupil images with elliptical appearances. This refinement stage consists of various additional steps that are incorporated after the end of the FREDa. The flow diagram of this method is illustrated in Fig. 5.

First (step 1.1), the original image I is cropped, resulting in the image I_c , which focuses the region of interest on the center c obtained by the FREDa algorithm. The size of the rectangular cutout is chosen adaptively according to the radius n for which the maximum response is obtained in the FRST. Then, in order to cleanse the image of artifacts due to reflections or eyelashes, a morphological opening is applied (step 1.2). The structuring element used is a flat disk with a radius of $n/2$. Subsequently, an iterative procedure is carried out in which the best candidate for the pupil center is searched.

For each iteration i , a region growing operation (step 1.3) is performed as follows: starting from the seed point c , a region R is generated by appending a new pixel each time, whose intensity value difference with the mean of R is the minimum from all 8-connected neighbors. The growth is stopped when this intensity difference exceeds a threshold T_i . The initial value for T_i is 5 gray values, assuming 8-bit images and that in each iteration of the loop it is augmented

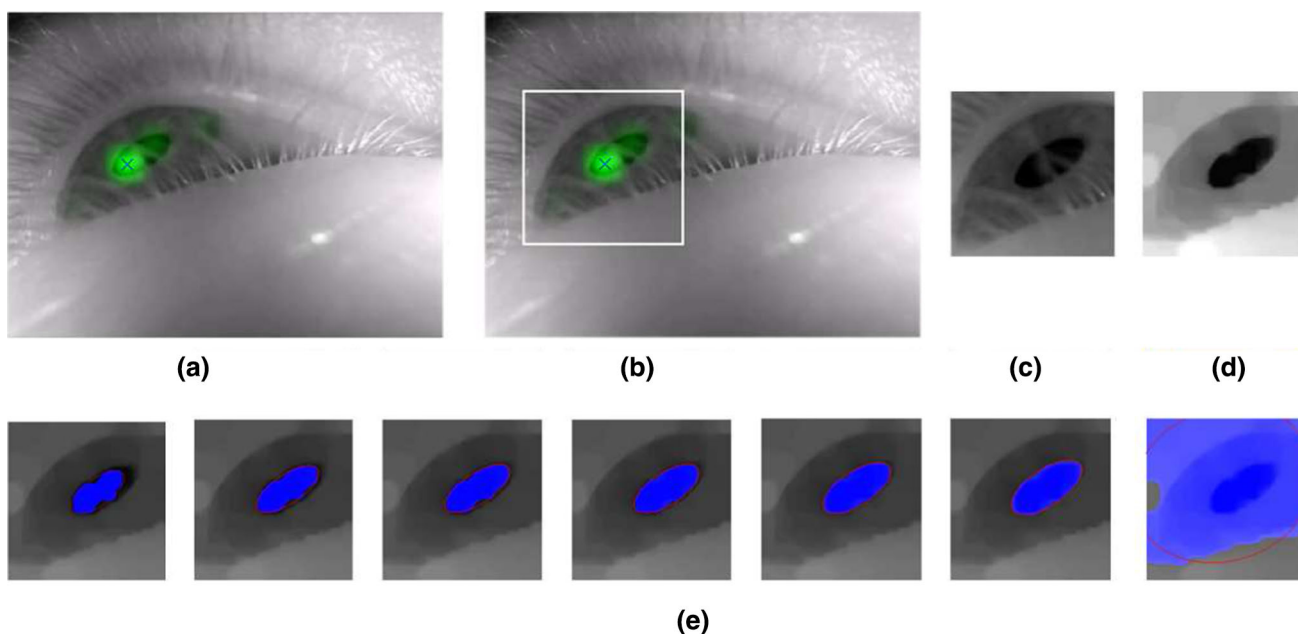


Fig. 6 Example of operation of the center refinement process of the FREDa I. **a** Pupil image with an elliptical appearance and the center incorrectly marked by the FRST, **b**, **c** cropping of the pupil area I_c (step

1.1). **d** Application of the morphological opening (step 1.2). **e** Successive iterations of the loop. The growing region (R) is presented in blue, and the fitted ellipse is in red (steps 1.3–1.7) (color figure online)

by a factor k as $T_{i+1} = T_i \cdot k$ where k is set as $k = 1, 3$. This increasing factor causes R to be larger in each iteration, thus enabling the finding of the most accurate approximation of the pupil area. Taking greater values of k leads to more rapid growth but may cause inaccurate approximations. In contrast, lower values permit more accurate pupil fitting, but more iterations will be needed. The selected value is a balanced choice between precision and rapid rising.

After region growing is completed, the resulting area is calculated and an ellipse, i.e., e , is fitted using the obtained region contour points (step 1.4). The fitted ellipse has the same normalized second central moment as the region. When the ellipse is calculated, it is checked if it gets out of I_c . In affirmative cases, the loop is interrupted assuming that R has increased out of the pupil (step 1.7) and the last saved center is considered as the new pupil center. Otherwise, (step 1.5), a normalized difference area parameter Δ is defined as:

$$\Delta = \frac{\text{Area}(e) - \text{Area}(R)}{\text{Area}(R)} \tag{10}$$

This parameter is used to evaluate the matching of the grown region to the pupil, assuming that, in a perfect adjustment, the fitted ellipse will perfectly match the region's contour, making the two areas equal. Therefore, if a minimum is obtained for the Δ parameter, the center of e is considered a better estimate for the pupil center, and its coordinates are saved (step 1.6). Finally, after a maximum of 10 iterations, the loop is finished, and the center that was saved last is consid-

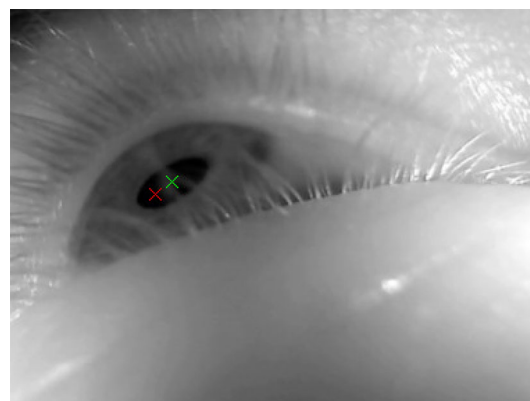


Fig. 7 Center obtained by the FREDa (red) and the center resulting from applying the refinement process of the FREDa I (green) (color figure online)

ered the new pupil center. This stopping criterion avoids for realizing unnecessary iterations, assuming that with a threshold $T_{10} = 5 \times 1.3^{10} \approx 70$ in the 10th repetition, the region R would contain all pixels belonging to the pupil area. A stop criterion based on the convergence of Δ has been tested with no satisfactory results. As the growing of R is not completely regular, the variation in Δ is not a monotonically decreasing function, thus preventing its use in estimating the stop condition. The described steps are graphically depicted in Fig. 6. Moreover, Fig. 7 shows the initially estimated center, as well as the one obtained after the refinement process.

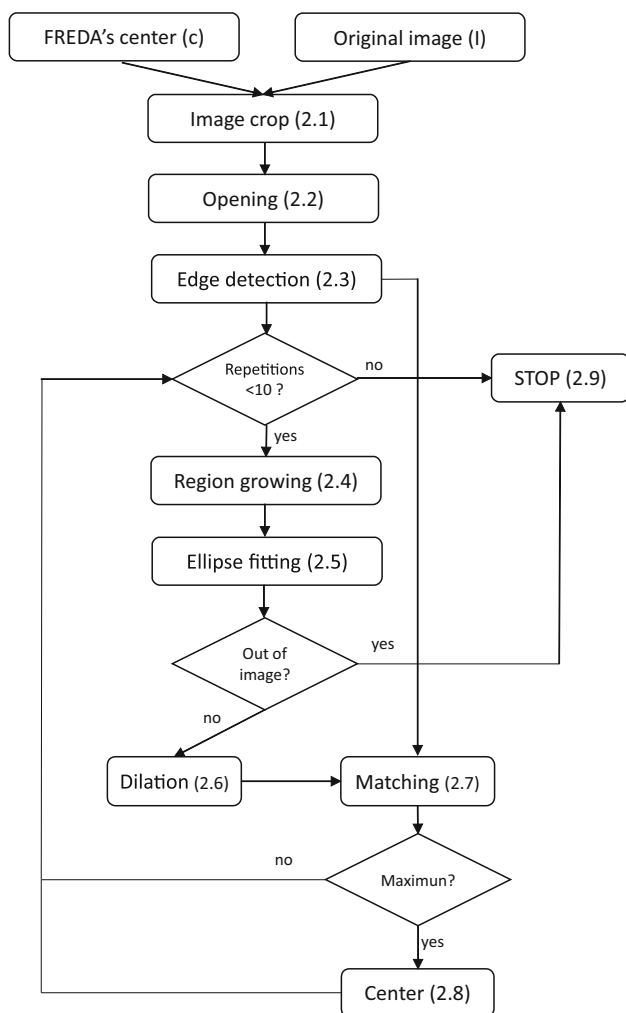


Fig. 8 FREDA II center refinement block diagram

2.3 FREDA II

In a similar manner to that of the FREDA I, the FREDA II algorithm is constructed by adding a center-refining stage to the FREDA as an additional alternative to improve the accuracy in the detection of the center in pupils with elliptical appearance. Figure 8 shows the flow diagram of the proposed method.

The first two steps, i.e., image cropping (step 2.1) and morphological opening (step 2.2), are identical to those of the FREDA I. The same structuring elements and parameters are used in both approaches. In step 2.3 a Canny edge filter is applied. The selected parameters are $\sigma = \sqrt{2}$ for the Gaussian filter and, $T_H = 0.3$ and $T_L = 0.02$ for the high and low thresholds, respectively. This edge image, labeled C , is used in a subsequent step to find the best center candidate. Once the opening is performed, an iterative procedure is carried out in which, after successive region growths, the best candidate for pupil center is determined. So, for each itera-

tion, a region growing operation is executed using the same configuration parameters as in FREDA I. After the process of region growth ends, i.e., R , an ellipse e is fitted to the region contour points, taking the one that has the second central moment of the region (step 2.5). It is verified that this ellipse does not get out from the image cutout, I_c . If this happens, the loop is terminated (step 2.9), and the last saved center is given as result. Otherwise, a binary image, E , is created applying a morphological dilation to the obtained ellipse with a squared 3×3 structuring element (step 2.6). Then, the number m of pixels in the intersection between E and the edge image, C , is obtained as $m = \|E \cap C\|$ (step 2.7). The previous dilation facilitates the matching between the two binary images. The result is compared with the previously stored value of m . If the value obtained is greater, the current value is saved, and the center of e is saved as the best estimation of the pupil center (step 2.8). The idea behind this method is to consider that, in a perfect adjustment of R to the pupil area, the fitted ellipse will obtain a maximum number of matching pixels with the edge image; in other words, m will reach its maximum value. Finally, if 10 iterations are completed, the ellipse center that was stored last is considered the corrected new pupil center (step 2.9). As in the FREDA I, it has been shown that in less than 10 iterations, the region R gets out from I_c or practically covers the pupil zone. The described steps are graphically represented by an example in Fig. 9, and both the center obtained by the FRST and the one obtained after the refinement process are shown in Fig. 10.

2.4 Evaluation images

For the evaluation of the algorithms, three collections of public databases containing eye images were used, totaling 225,569 images. Together with each collection of images the image coordinates of the pupil center are attached, which are used as references for the evaluation of the accuracy of the algorithms.

2.4.1 Tübingen collection

This collection of images was published by Fuhl et al. for the evaluation of the ExCuSe [10] and ELSe [4] algorithms. It consists of a total of 94,113 eye images of 384×288 pixels divided into 24 sets corresponding to 24 different subjects, of which the first 17 correspond to the publication of the ExCuSe algorithm, and the remaining 7 were presented with the ELSe algorithm. Sets I–IX were obtained in a road driving experiment [16] using the Dikablis eye tracking system (Ergoneers Inc., Manching, Germany), while sets X–XVII were recorded during an experiment that involved a search for products in a supermarket [17] using the same eye tracking device. Two images of each of the 24 sets are shown in Fig. 11.

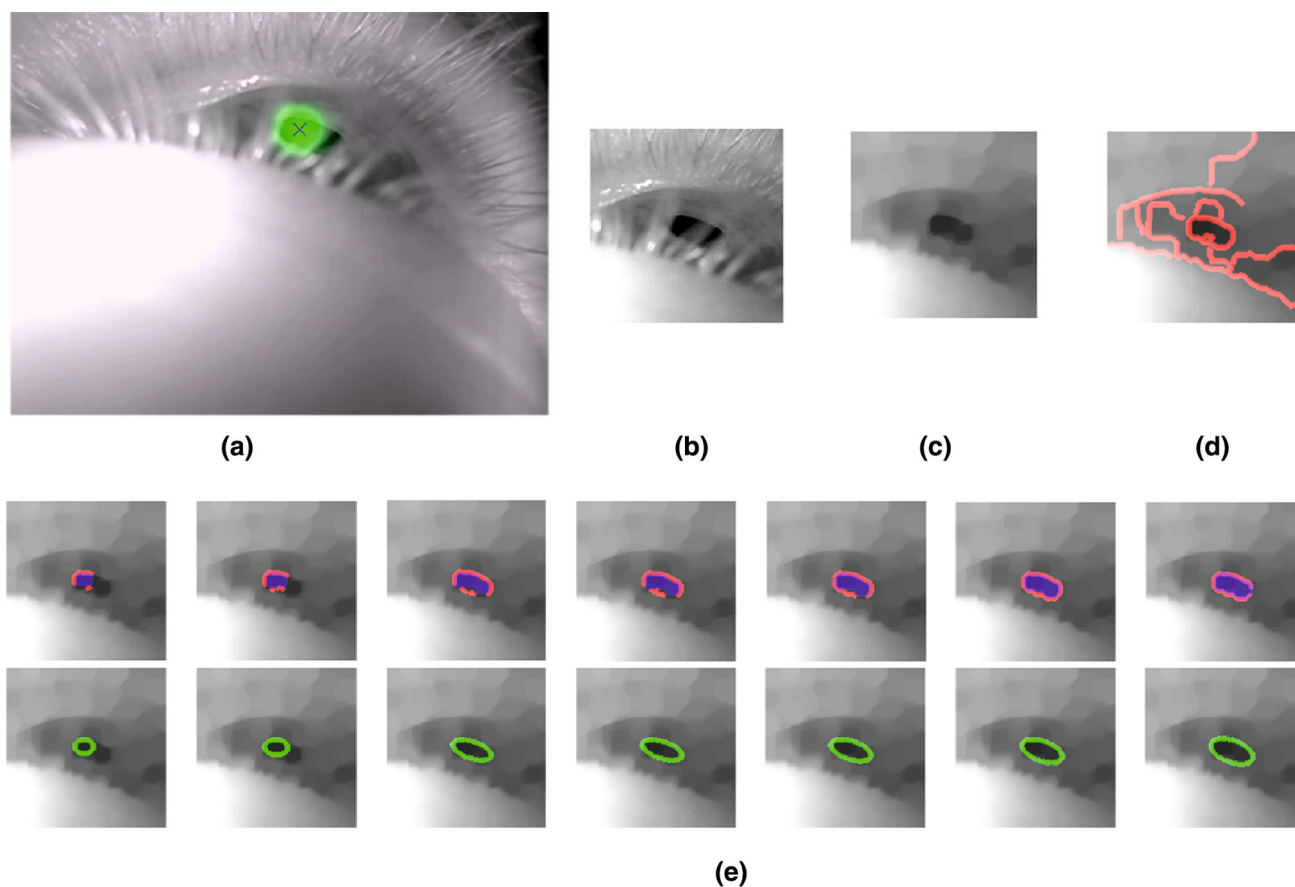


Fig. 9 Example of operation of the center refinement process of the FREDA II. **a** A pupil image with an elliptical appearance and the center being badly marked by the FRST, **b** image crop I_c (step 2.1), and **c** morphological opening (step 2.2). **d** Edge image C obtained with the

Canny filtering (step 2.3). **e** Successive iteration of the loop. The growing region R is presented in blue, matching pixels are in red, and the fitted ellipse e is in green (steps 2.4–2.9) (color figure online)

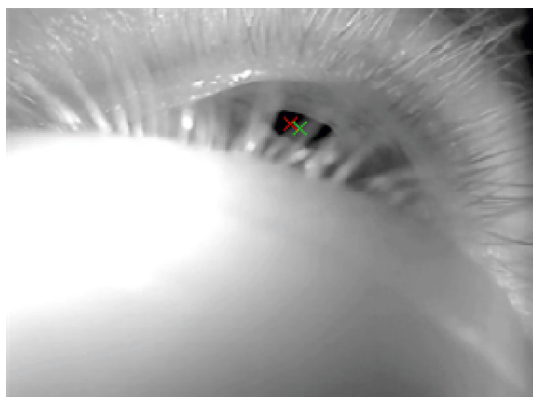


Fig. 10 Center obtained by the FREDA (red) and the center resulting from applying the refinement process of the FREDA II (green) (color figure online)

The sets of images between the XVIII and XXIV, corresponding to the study of ELSe algorithm, present numerous artifacts that greatly complicate the estimation of the cen-

ter of the pupil. The sets between the XVIII and XXII were recorded during the road driving experiment and are characterized by a high level of blur, reflections and a low contrast of the pupil. Sets XXIII and XXIV, however, were recorded from Asian subjects, for whom the main difficulty lies in pupil occlusions caused by eyelid and eyelash shadows. The marking of the images was done manually, and the error could be up to five pixels [4].

2.4.2 Świrski collection

This set of images was published by Świrski et al. [11] and contains 600 high-resolution images (640×480 pixels) corresponding to both eyes of two different subjects. The images were obtained through a low-cost head-mounted system with infrared illumination under laboratory conditions. Its main advantage is the good quality of the images in terms of the images being mainly devoid of reflections and the nice contrast of the pupil with respect to the rest of the eye. However, the main difficulty in detecting the pupil center lies in the

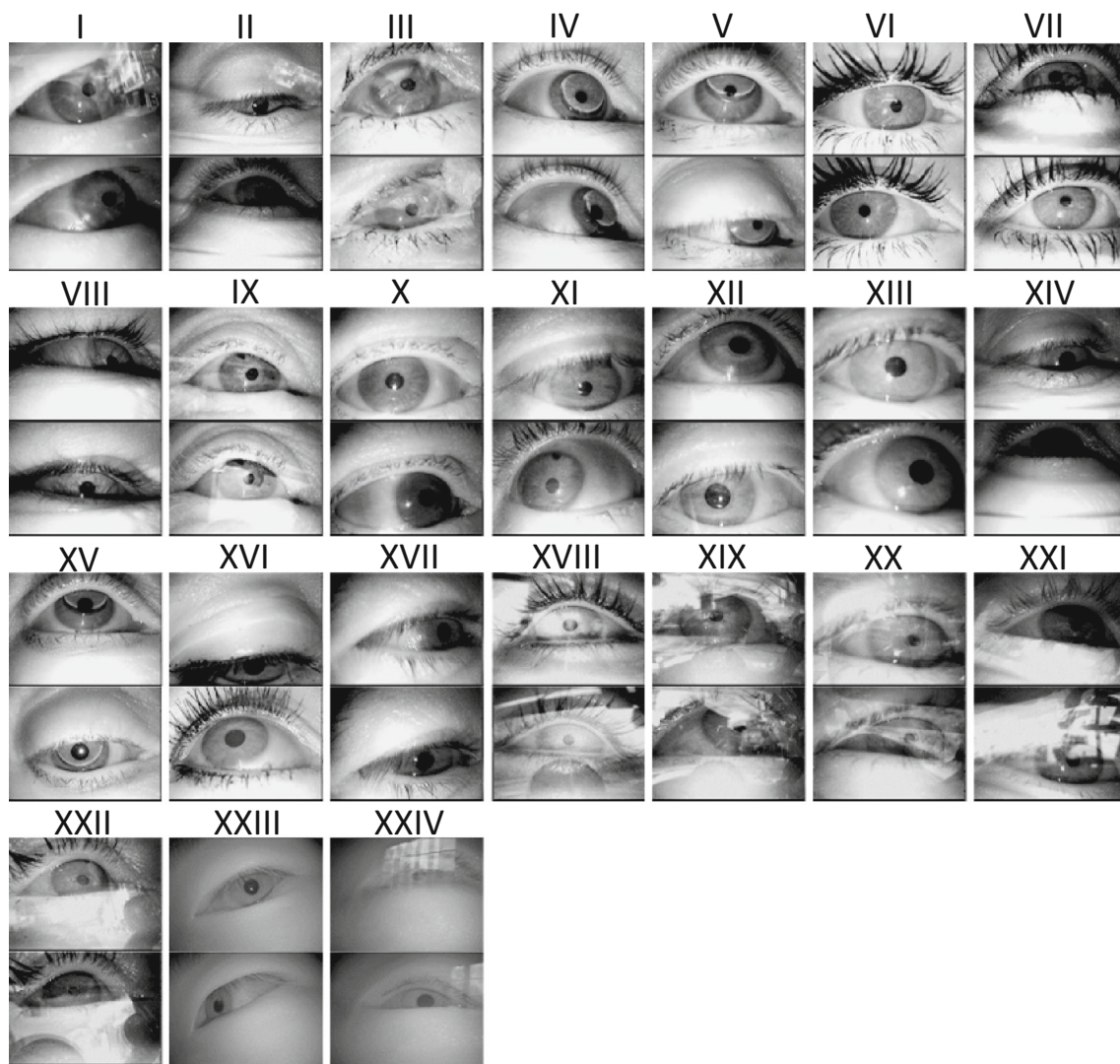


Fig. 11 Examples of each of the 24 sets of images from the Tübingen collection, showing the difficulties in determining the pupil center. Each pair of images corresponds to a subject of the study

eccentricity of the pupils due to the high degree of angulation of the camera with respect to the axis of sight. This fact also causes occlusions of the pupil by the eyelashes or eyelids in the image (see Fig. 12). Marking of the pupil center was performed by adjusting an ellipse with respect to at least 5 points manually placed over the pupil border resulting in a highly precise marking.

2.4.3 Labeled pupils in the wild (LPW) collection

The set of images called “Labeled Pupils in the Wild”, or LPW, published by Tonsen et al. [18] comes from a total of 66 high-quality videos from 22 different subjects. Each video contains approximately 2000 frames of 640×480 pixels, obtained at a frequency of 95 FPS, resulting in a total of 130,856 eye images. The collection covers a wide range

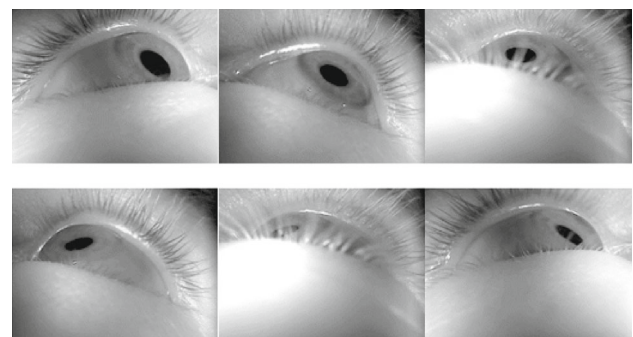


Fig. 12 Examples of images from the Świrski collection

of situations during both outdoor and indoor events. Each user was recorded in two indoor locations and one outdoor location. The change in the lighting conditions drastically

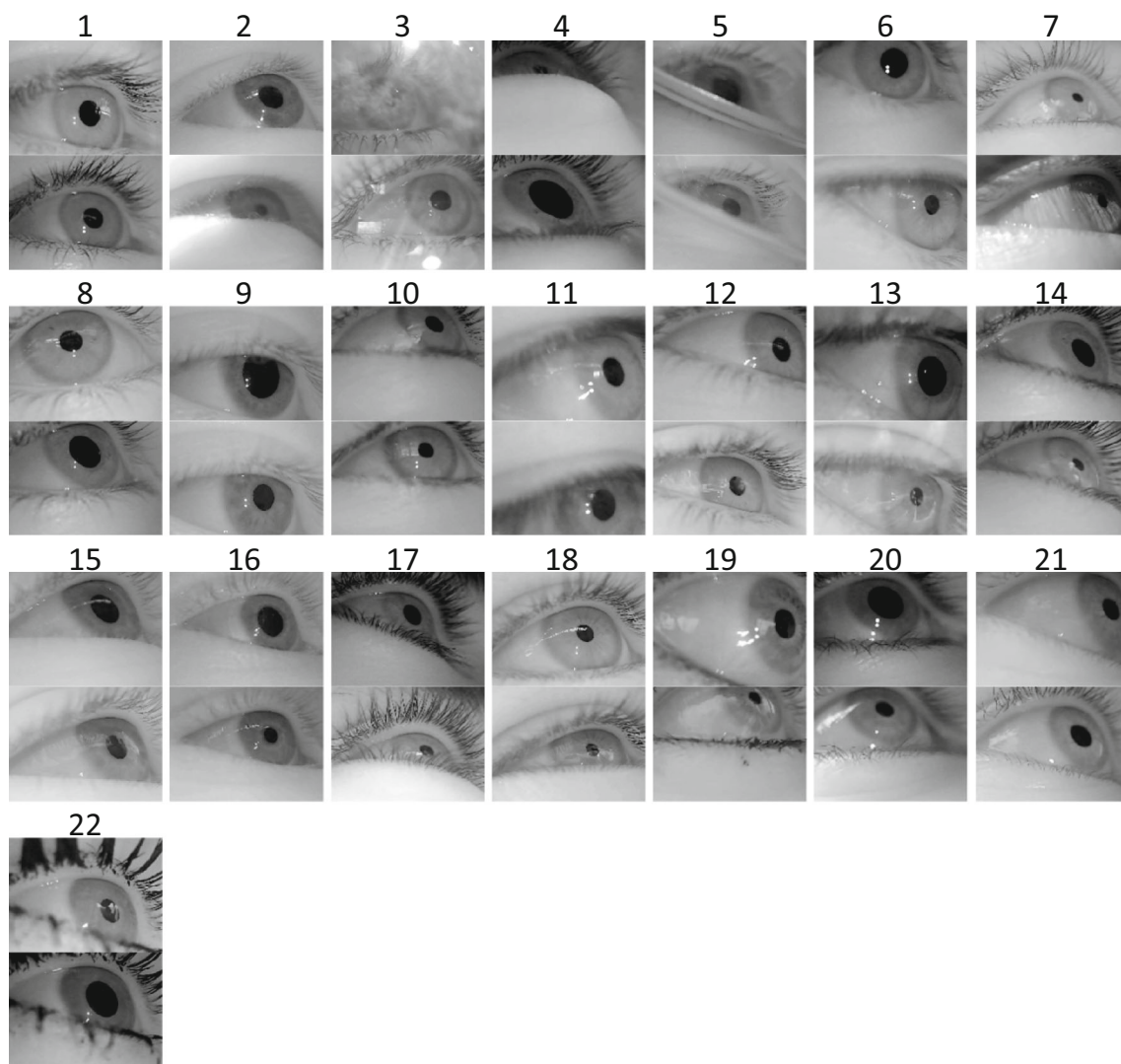


Fig. 13 Examples of images from the LPW collection. Each pair of images corresponds to a subject of study. The changing conditions of the recording environment are appreciated, both in overall illumination and in pupil dilation

affected the eye aperture, which exhibited a wide range of pupil sizes. An added difficulty is the high pupil eccentricity exhibited by certain images. All images were manually labeled. Figure 13 shows two example images from each user.

3 Results

We compared the precision of the pupil center estimation of each of the three proposed algorithms, namely FREDa, FREDa I and FREDa II, in the previously described datasets. For a performance evaluation comparison of the current approaches, the ELSe algorithm [18] was chosen as the reference method based on the analysis of state-of-the-art algorithms presented by Fuhl et al. [4], wherein the same image sets were used for testing. The detection error was

measured as the Euclidean distance between the center estimated by the algorithm and the labeled center. To normalize error rates among the images with different sizes, those from LPW and Świrski (640×480 pixel) were previously down-sampled to Tübingen's resolution (384×288 pixel).

Figure 14 shows the performance of the four algorithms obtained for the entire dataset, with the one gathering the three datasets together, as the detection rates for different pixel error values. The detection rate was defined as the number of correctly detected pupil centers up to a specific error distance normalized by the total amount of images. As can be observed, the FREDa I is superior to the other algorithms, up to a precision of 2 pixel error, which was closely followed by the FREDa II algorithm.

In Fig. 15 the detection rates of the four algorithms are shown, divided according to the collections. There were

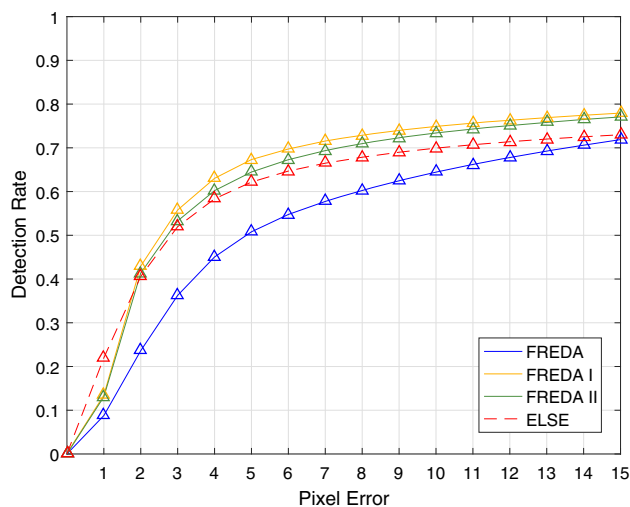


Fig. 14 Detection rates of the four algorithms for the entire dataset

notable differences in the results obtained in each collection for a specific algorithm. The FRED A was the most precise one for the Tübingen images, but its performance decayed drastically when it was evaluated in the Świrski and LPW datasets. In the same way, FRED A II was shown to be the best suited algorithm for the LPW and Świrski sets. In contrast, the results of the FRED A I and the ELSe algorithms were more balanced among the datasets, with the FRED A I being superior to the ELSe for the Tübingen and LPW collections, which, in practice, suppose the total of the images.

The observed variations in the error rates are caused by the elliptical appearance of the pupils from Świrski and LPW images. The Tübingen images are characterized by numerous challenging artifacts, but their appearance is almost circular. In this scenario, the use of radial information is a very robust and precise method for center detection, as can be seen in Fig. 15a. Nevertheless, the loss of performance shown for the Świrski and LPW collections, especially in the first case, demonstrates the necessity of a center-refining method to improve the accuracy for those types of images. Of the two presented approaches, the FRED A II was more precise than the FRED A I (Fig. 15b, c).

Table 1 shows the percentages of correctly determined pupil centers by each algorithm for each subset of the three collections. Because the error in the labeling of the images of Tübingen can be up to 5 pixels [11], a center was considered correctly estimated if the error was less than or equal to 5 pixels. The highest percentage obtained in each subset is marked in bold. According to the previous graphics, the FRED A algorithm was the most robust for the challenging images, being superior on 12 of the 24 subsets of the Tübingen collection. This result is clearly shown in Table 2, where the percentages of successfully determined centers by each algorithm are shown for the total of the three collections. The

FRED A obtained 67.17% of the corrected pupil centers compared to the 65.50% reached by the FRED A I, the 60.60% by ELSe and the 49.78% obtained by FRED A II.

In contrast, FRED A II was superior to its competitors for 14 of the 22 subsets of the LPW collection and for the Świrski images. Therefore, it can be argued that it is the most precise in high-quality images and in the presence of pupils with elliptical appearances. In addition, as can be seen in Table 2, the FRED A II obtained 76.84 and 86.83% of correctly estimated centers in LPW and Świrski collections, respectively, in contrast to the 65.86% and the 81.17% reached by its closest competitor, i.e., ELSe.

Regarding the FRED A I, it did not stand out for its performance in any of the collections since, as is shown in Table 2, it was inferior to the FRED A in the Tübingen collection and to the FRED A II in the LPW and Świrski collections. Nevertheless, as seen in Table 3, in which the number of correctly determined pupil centers is depicted for the entire dataset, the FRED A I was the most precise algorithm, reaching 69.73% of the hits, followed by the FRED A II with 67.29% of the hits.

Taking into account individual sets, it was observed that there were great differences in the rate of success among them. While the success rate exceeded 90% in numerous sets, the low rate observed particularly in sets XVIII and XIX of the Tübingen collection, as well as in sets 4 and 5 of the LPW collection, is remarkable. In the first two, the ELSe algorithm obtained the best results, with only 52.99 and 35.41% of correctly estimated centers and was closely followed by the FRED A, which obtained 50.63 and 33.47% of the hits. As shown in Fig. 16, in which three examples of each of the two sets are shown, pupil occlusions due to reflections, eyelids or blurring of the image, caused the detection of the center to be particularly difficult in these two cases.

With respect to the images of users 4 and 5 in the LPW collection, the success rates of the FRED A II were 52.20 and 21.02%, respectively, and 34.25 and 31.28% with ELSe. It can be seen in Fig. 17 that in user 4 the pupil may become inappreciable, as it was occluded by the eyelid and even partially cut by the image border. In case 5, however, it was the spectacle frames the subject wore, responsible for totally or partially concealing the pupil. The effect of the lens is also noticeable in the blurring of the image.

The algorithm presents several parameters that need to be tuned to enable the method to work. Some of the parameters are highly dependent on the working conditions and are not easily standardized, e.g., the values of the radius when calculating the FRST should be in accordance with the average size of the pupil in the camera, while others, such as those involved in the preprocessing stage are more difficult to select. To measure the robustness of the FRED A in terms of the specific values of the parameters, slight changes of $\pm 10\%$ were made to the size of the filters and to the lim-

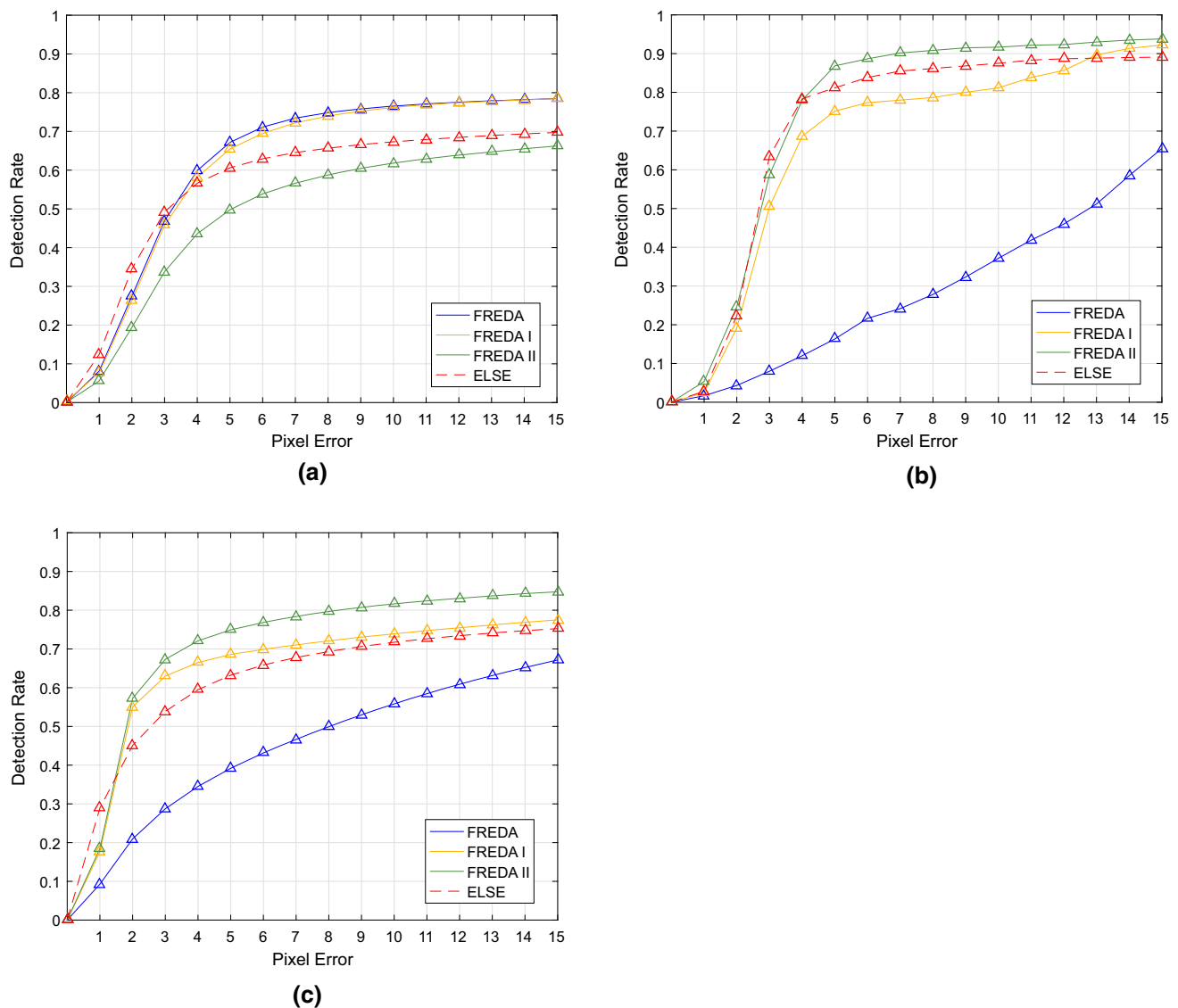


Fig. 15 Detection rates of the four algorithms separated according to the collections. **a** Tübingen, **b** Świrski, **c** LPW

its of the contrast stretching transform. The overall result did not change, and the conclusions are still valid. The large number of images involved compensated for possible biases, and on average, the result remained the same. Moreover, the datasets involved presented different types of images, and the parameters values were valid across datasets demonstrating the robustness of the method and the lack of sensitivity to the involved parameters. In fact, the FRST was the only stage that presented problems regarding the elliptical pupils in the Świrski collection and that did not depend on any parameter.

3.1 Computing time

The average processing times per image obtained for each algorithm were 44 ms for the FREDA, 63 ms for the FREDA I and 68 ms for the FREDA II. This result showed a 43%

increase in the computing time when using the first center refinement and a 53% increase when the second refinement was used. Regarding the FREDA, from our measurements, it can be deduced that half of the time is used in the preprocessing stage, while the other half is used when computing the FRST. The computation time of the refinement stages can be determined from the differences of the FREDAs I and II with respect to the computation time of the FREDA. As shown in Table 3, in contrast, the increase in the detection was 27% (and increase from 54.76 to 69.73%) for the FREDA I and 23% (an increase from 54.76 to 67.29%) for the FREDA II. Since the algorithms were implemented in MATLAB the computation times were not directly comparable to the time that was obtained with ELSe since a version compiled in C++ language was used. The average processing time per image observed for ELSe was 8 ms. A commercial

Table 1 Percentage of correctly detected centers of the four algorithms for each set of images

	FREDA	FREDA I	FREDA II	ELSe
<i>Tübingen</i>				
I	89.38	85.70	69.58	85.95
II	74.45	75.64	52.87	63.76
III	84.31	80.30	55.05	65.31
IV	92.54	92.92	88.14	83.05
V	96.06	95.60	93.30	84.73
VI	88.59	87.95	86.34	77.27
VII	78.22	76.89	72.47	59.61
VIII	71.58	72.22	51.43	67.30
IX	91.09	89.33	68.63	86.72
X	93.57	87.50	88.33	78.93
XI	88.55	72.37	72.37	75.27
XII	73.47	86.26	86.83	79.01
XIII	73.31	74.54	67.62	73.73
XIV	91.68	93.39	61.62	84.22
XV	77.41	73.00	62.53	57.30
XVI	89.03	86.48	75.26	59.95
XVII	94.77	96.64	75.75	89.18
XVIII	50.63	48.70	24.09	52.99
XIX	33.47	32.17	25.14	35.41
XX	61.68	59.45	37.94	68.79
XXI	67.96	67.81	60.93	41.32
XXII	62.38	60.44	30.01	56.61
XXIII	91.19	98.11	97.33	93.40
XXIV	43.60	51.72	47.55	51.20
<i>LPW</i>				
1	63.70	97.45	95.88	88.23
2	74.45	93.58	93.15	50.15
3	55.72	62.32	61.00	50.20
4	18.25	35.53	52.20	34.25
5	14.03	19.72	21.02	31.28
6	47.41	84.85	91.65	63.09
7	56.48	85.30	92.05	70.07
8	54.43	89.27	90.62	84.17
9	32.48	66.52	70.05	70.90
10	27.72	67.90	78.85	65.80
11	20.28	53.43	65.07	56.18
12	49.25	85.13	86.68	88.93
13	28.44	55.94	59.66	52.31
14	31.33	55.52	65.60	74.95
15	15.47	60.18	72.22	65.93
16	66.86	85.73	89.53	87.47
17	21.83	72.82	86.48	67.05
18	58.12	88.63	92.12	83.33
19	32.03	52.98	62.33	41.13

Table 1 continued

	FREDA	FREDA I	FREDA II	ELSe
20	53.83	86.18	94.35	23.05
21	24.57	78.27	90.53	62.10
22	15.12	30.52	35.63	78.58
<i>Świrski</i>	21.67	75.17	86.83	81.17

Table 2 Percentage of correctly detected centers of the four algorithms for each collection of images

	FREDA	FREDA I	FREDA II	ELSe
Tübingen	67.17	65.50	49.78	60.60
LPW	43.15	69.88	76.84	65.86
Świrski	21.67	75.17	86.83	81.17

Table 3 Percentage of correctly detected centers of the four algorithms for each entire dataset

	FREDA	FREDA I	FREDA II	ELSe
Total	54.76	69.73	67.29	64.69

version of the FREDA has been implemented for which the computing times are below 10 ms per frame. Consequently, the estimated times for FREDA I and FREDA II would be approximately 14 and 15 ms, respectively. The contribution of the preprocessing stage to the final result was also measured. It was observed that it was specifically significant for higher error values up to 5 pixels for which the accuracy was improved about 2–5% meaning that it contributes to improve the robustness of the method when facing “wild” images. In contrast, the improvement is negligible for lower error values. The refinement stages facilitate the processing of more elliptical images. Hence, knowing the influence of each one of the steps and depending on the working environment, alternative work flows can be selected with varying computation times. Moreover, no video sequences were considered in the paper. It is easy to deduce that in a real scenario the image to be processed can be cropped according to the guess obtained from the previous frame, thus reducing computation times.

3.2 Comparison to CNN

CNN have merged as an effective solution for solving several artificial vision problems, such as object detection or scene recognition. The work by Fuhl et al. [13] presents a comparison among several methods based on CNN using an extended version of the Tübingen dataset that was employed in this paper. They train the network using a random set of images consisting of 50% of the images obtained from the

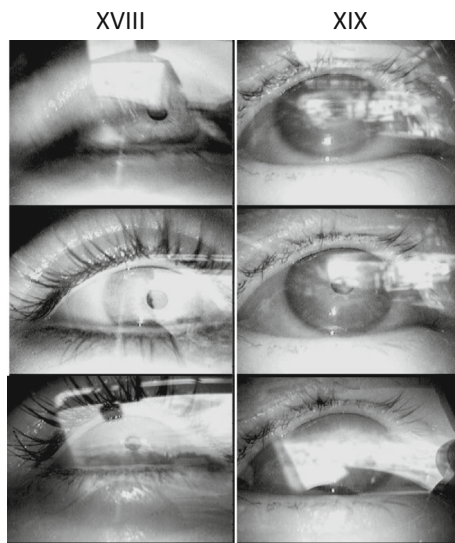


Fig. 16 Examples of images from sets XVIII and XIX of Tübingen collection

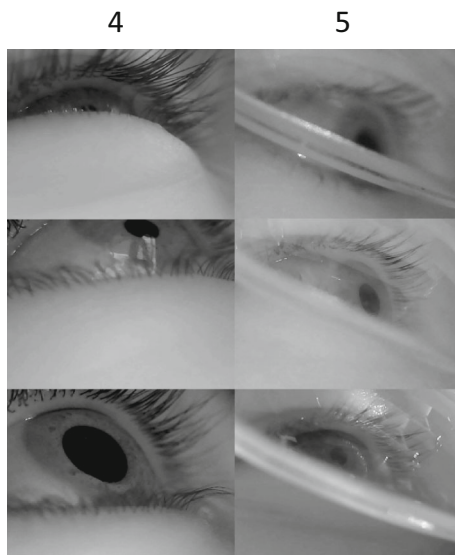


Fig. 17 Examples of images from sets 4 and 5 of the LPW collection

alternative datasets forming the database. The test is carried out in the other half of the images.

In Fig. 18, we find a comparison of the best results they obtained over the 50% of the testing images and our results from the use of the whole database with the FREDAs and FREDAs I. It would not be fair to include the results over the training images in the comparison. From the figure, it can be deduced that our results are slightly better for any error value. Except for the pixel errors over 12 pixels for which the CNN obtains a somewhat better rate but still comparable. In the case of the CNN, 79% of the pupil centers were estimated within an error of 15 pixels, while this value decreased to 78% in the case of our approach. Regarding the gaze estimation

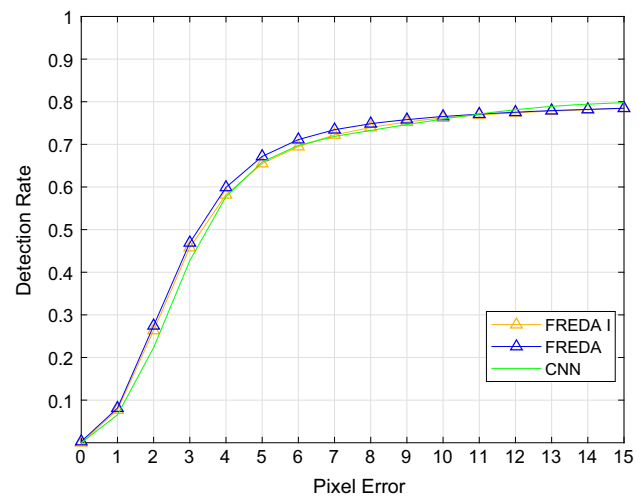


Fig. 18 Comparison among FREDAs, FREDAs I and the best results obtained by the CNN [13]

error, these type of errors do not allow a reliable estimation of gaze, i.e., these images would have to be rejected or an additional refinement stage would be required in order to obtain a more accurate estimation.

If we take into account that our method was tested over a larger number of images, the improvement is still remarkable. Moreover, in order to carry out a reliable comparison, the CNN should be trained only using the half of the datasets to be tested over the rest of the datasets, i.e., over completely unknown samples, or over entirely new databases, such as that of the LPW, as it has been made in the present paper. From our results in Fig. 15, it can be observed how the results can vary among different databases. Considering the training requirements of the CNN and their computational load and looking at the results obtained from our automatic procedure, it can be concluded that our method is superior in terms of accuracy and is fully comparable regarding robustness. Regardless of the undoubted potential of deep learning techniques and the valuable efforts made to apply them to eye tracking [13], the variability in the data of the topic under study and the labeling difficulties of the eye tracking images have prevented CNN from obtaining satisfying results to date.

4 Conclusions

A new algorithm, the FREDAs, with two additional center-refining steps (FREDAs I and FREDAs II) has been developed for eye center detection in head-mounted systems, based on the calculation of the radial symmetry of the pupil. The FREDAs algorithm is publicly available.¹ After evaluating

¹ <http://gi4e.unavarra.es/code/>.

their performance on a large set of images obtained under a wide variety of conditions, the FRED A I showed greater precision in the detection of the pupil center, surpassing the ELSe algorithm, which has been used as a reference among the published algorithms to date. In addition, it showed better results than other works using completely different perspectives, such as CNN.

A fast radial symmetry transform was chosen as the basis for the pupil center estimation in order to develop a robust method for difficult images that have been taken in real scenarios. Although it was shown to be an effective method for circularly appearing pupils, there was a lack of precision when the pupils possessed an elliptical shape. Thus, two approaches with additional center refinement steps (FRED A I and FRED A II) were developed to solve this inconvenience and the results showed that the FRED A II was the best suited for elliptical pupil images. However, its precision decayed in response to challenging images where the pupil is not well defined due to strong reflections, blurring, partial occlusions by eyelids or eyelashes, etc. In these cases, the center refinement stage of the FRED A I was more reliable, reaching higher detection rates than ELSe. Therefore, it can be concluded that the FRED A I algorithm is a robust and efficient approach for eye tracking systems, as it is able to obtain a high rate of detection in a great number of challenging situations that are common in those systems.

Acknowledgements We would like to acknowledge the Spanish Ministry of Economy, Industry and Competitiveness for their support under Contract TIN2014-52897-R in the framework of the National Plan of I+D+i in the framework of the National Plan of I+D+i

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Duchowski, A.T.: *Eye Tracking Methodology: Theory and Practice*. Springer, New York Inc, Secaucus (2007)
2. Huey, E.B.: *The Psychology and Pedagogy of Reading, with a Review of the History of Reading and Writing and of Methods, Texts, and Hygiene in Reading*, p. 16. Macmillan, New York (1908)
3. Majaranta, P., Aoki, H., Donegan, M., Hansen, D.W., Hansen, J.P.: *Gaze Interaction and Applications of Eye Tracking: Advances in Assistive Technologies*, 1st edn. Information Science Reference—Imprint of: IGI Publishing, Hershey (2011)
4. Fuhl, W., Tonsen, M., Bulling, A., Kasneci, E.: Pupil detection for head-mounted eye tracking in the wild: an evaluation of the state of the art. *Mach. Vis. Appl.* **27**(8), 1275–1288 (2016)
5. Hansen, D.W., Ji, Q.: In the eye of the beholder: a survey of models for eyes and gaze. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**, 478–500 (2010)

6. Li, D., Winfield, D., Parkhurst, D.J.: Starburst: a hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches. In: *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)—Workshops*, vol. 3, CVPR '05 (Washington, DC, USA). IEEE Computer Society, p. 79 (2005)
7. 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, ETRA '12* (New York, NY, USA). ACM, pp. 173–176 (2012)
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)
9. Kassner, M., Patera, W., Bulling, A.: Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction. *CoRR*, vol. abs/1405.0006 (2014)
10. Fuhl, W., Kübler, T., Sippel, K., Rosenstiel, W., Kasneci, E.: ExCuSe: Robust Pupil Detection in Real-World Scenarios, pp. 39–51. Springer, Cham (2015)
11. Fuhl, W., Santini, T.C., Kübler, T.C., Kasneci, E.: Else: ellipse selection for robust pupil detection in real-world environments. *CoRR*, vol. abs/1511.06575 (2015)
12. Krafka, K., Khosla, A., Kellnhofer, P., Kannan, H., Bhandarkar, S., Matusik, W., Torralba, A.: Eye tracking for everyone. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016)
13. Fuhl, W., Santini, T., Kasneci, G., Kasneci, E.: Pupilnet: convolutional neural networks for robust pupil detection. preprint, vol. abs/1601.04902 (2016)
14. Loy, G., Zelinsky, A.: Fast radial symmetry for detecting points of interest. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**, 959–973 (2003)
15. Skodras, E., Fakotakis, N.: Precise localization of eye centers in low resolution color images. *Image Vis. Comput.* **36**, 51–60 (2015)
16. Kasneci, E., Sippel, K., Aehling, K., Heister, M., Rosenstiel, W., Schiefer, U., Papageorgiou, E.: Driving with binocular visual field loss? A study on a supervised on-road parcours with simultaneous eye and head tracking. *PLoS ONE* **9**, 1–13, 02 (2014)
17. Sippel, K., Kasneci, E., Aehling, K., Heister, M., Rosenstiel, W., Schiefer, U., Papageorgiou, E.: Binocular glaucomatous visual field loss and its impact on visual exploration—a supermarket study. *PLoS ONE* **9**, 1–7, 08 (2014)
18. Tonsen, M., Zhang, X., Sugano, Y., Bulling, A.: Labeled pupils in the wild: a dataset for studying pupil detection in unconstrained environments. *CoRR*, vol. abs/1511.05768 (2015)

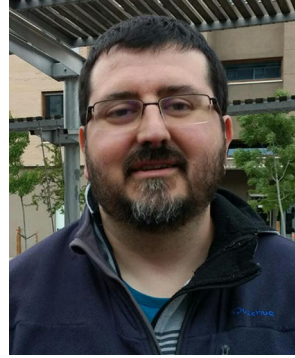


Ion Martinikorena was born in Spain in 1986. He received the B.E. and M.E. degrees from Public University of Navarre, Spain, in 2013 and 2017, respectively. He was a member of the GI4E (Gaze Interaction for Everybody) group during 2016 and 2017. His main areas of research interest are computer vision, machine learning, deep learning and gaze estimation.



Rafael Cabeza was born in Soria, Spain, in 1967. He received the M.Sc. degree in Physics in 1990 from the University of Zaragoza, Zaragoza, Spain, and the Ph.D. degree (with honors) in Telecommunications Engineering in 1996 from the Public University of Navarre, Pamplona, Spain. Since 1999, he has been an Associate Professor in the Department of Electrical and Electronics Engineering at Public University of Navarre, where, from 2000 to 2010, he was the Head of the Sig-

nal Processing, Microelectronic, and Instrumentation Research Group. His current research interests include image processing and computer vision.



Iñaki Urtasun was born in a small town in Navarre in 1978. He joined the IT service of the Public University of Navarre in 2005. While he was working, he received the B.E. (Telecommunications). He is interested in digital image processing, deep learning and classic arcade games.



Arantxa Villanueva was born in Pamplona, Spain, in 1974. She received the M.Sc. and the Ph.D. degrees in Telecommunications Engineering from the Public University of Navarre in 1998 and 2005, respectively. She is currently Associate Professor in the Electrical and Electronical Engineering Department at the Public University of Navarre and works as researcher in the GI4E (Gaze Interaction for Everybody) group. Her main research interests are in

the area of image processing and gaze estimation.



Andoni Larumbe was born in Spain in 1993. He received the B.E. (Telecommunications Engineering) and M.E. (Biomedical Engineering) degrees from Public University of Navarre, Spain, in 2015 and 2017, respectively. He joined Public University of Navarre in 2016. Since 2017, he is a PhD student in the GI4E (Gaze Interaction for Everybody) research group. His main areas of research interest are computer vision, machine learning, deep learning and gaze estimation.