Computing and Informatics, Vol. 36, 2017, 257–282, doi: 10.4149/cai_2017_2_257

PRECISE NON-INTRUSIVE REAL-TIME GAZE TRACKING SYSTEM FOR EMBEDDED SETUPS

Antonio García-Dopico, Antonio Pérez
José Luis Pedraza, María Luisa Córdoba

Departamento de Arquitectura y Tecnología de Sistemas Informáticos
Facultad de Informática, Universidad Politécnica de Madrid
Campus de Montegancedo s/n
E-28660, Boadilla del Monte, Madrid, Spain
e-mail: {dopico, aperez, pedraza, mcordoba}@fi.upm.es

Abstract. This paper describes a non-intrusive real-time gaze detection system, characterized by a precise determination of a subject's pupil centre. A narrow field-of-view camera (NFV), focused on one of the subject's eyes follows the head movements in order to keep the pupil centred in the image. When a tracking error is observed, feedback provided by a second camera, in this case a wide field-of-view (WFV) camera, allows quick recovery of the tracking process. Illumination is provided by four infrared LED blocks synchronised with the electronic shutter of the eye camera. The characteristic shape of corneal glints produced by these illuminators allows optimizing the image processing algorithms for gaze detection developed for this system. The illumination power used in this system has been limited to well below maximum recommended levels. After an initial calibration procedure, the line of gaze is determined starting from the vector defined by the pupil centre and a valid glint. The glints are validated using the iris outline to avoid glint distortion produced by changes in the curvature on the ocular globe. In order to minimize measurement error in the pupil-glint vector, algorithms are proposed to determine the pupil centre at sub-pixel resolution. Although the paper describes a desk-mounted prototype, the final implementation is to be installed on board of a conventional car as an embedded system to determine the line of gaze of the driver.

Keywords: Pattern recognition, image processing, eye-tracking, gaze-detection, embedded systems

1 EYE DETECTION AND GAZE TRACKING SYSTEMS

Line of gaze determination has been the subject of a great number of studies for several years, with applications in very different fields including medical research [1], psychological analysis of human behaviour [2], behaviour characterisation of car drivers under fatigue situations [3, 4], and also as a computer interface, both for conventional computers and to assist people with a high degree of physical disability [5, 6].

Different approaches have been used, in many cases involving head-mounted devices [7, 8] or devices placed in front of the eyes [9] or face [10]. Some non-intrusive systems have been developed during the last decade [11, 12, 13, 14, 15]. Gaze tracking techniques have been used as viable technologies in several applications [16, 17].

The approach chosen by the authors for tracking a car driver's gaze, is just one of many possible applications of this technology.

The conditions and concentration needed for driving a car strongly limit driver movements. In practice, this means that a driver's head typically makes low-amplitude movements quickly returning to the initial position. Due to this fact, the first assumption is to consider a fixed head situation and a narrow field of view (NFV) camera pointing directly to one of the person's eyes. As head movements may occur at any time, the image of the eye may frequently be lost at random moments. To recover it, a second camera covering the whole face of the person locates the eye allowing the NFV camera to be reoriented.

One of the non-intrusive methods most frequently used for determining the gaze direction is based on the detection of a glint – corneal reflection – produced by the reflection of a light source that also illuminates the face. Useful surveys and the theoretical groundwork for estimating the point-of-gaze from the coordinates of the centres of the pupil and corneal reflections can be found in [18] and [19]. Proper illumination has to be used to obtain images that allow for optimal processing. With this purpose, the illumination source is usually based on infrared LEDs, both to avoid unwanted distractions and also to minimize the influence of external changes in ambient illumination [20, 21]. The image of the eye, including the glints, is captured by an infrared sensitive camera while filtering natural illumination as much as possible to guarantee stable light conditions. Other groups have also developed non-intrusive systems relying solely on ambient illumination and detecting the eyes in an image of the face [22].

Starting from a fixed head situation, gaze direction can be determined by means of the vector defined by the centre of the pupil and a fixed point in the eye (the above mentioned glint). Real-time image processing algorithms must be applied to the images gathered by the camera to determine the location of these points. The algorithms involved in the proposed method are described in [23]. Some other proposals are analysed in [24, 25]. When high resolution is needed in determining the gaze direction, subpixel computation of the pupil and glint centres can be used with the largest possible vector [26]. The image should contain only the iris, thus

requiring a camera with a very narrow field of view. In this situation, the camera has to track and follow the person's eye to compensate for even the slightest head movements while keeping the eye centred in the image.

A problem arises when, due to a sudden head movement, the eye slips out of the NFV camera field of view. In this case, the eye has to be quickly located again. To guarantee quick recovery, several authors propose the use of another camera with a wide field of view (WFV) that gathers images of the person's head in which the eye position can be located [27]. Depending on the application, the use of two cameras to obtain stereoscopic images can be appropriate, thus obtaining information to focus the NFV camera if needed [28, 29, 12].

Face analysis is used in different fields and the techniques employed are often based on detecting skin colour in the captured image [30, 31, 32], sometimes decomposing the face region into its wavelet domain subimages [33]. An increasing interest can be observed focusing on the development of techniques for automatic analysis and identification of faces, with application to the field of robotics, expression analysis, authentication or teleconferencing [34, 35]. Face analysis has been also used as a complementary technique to improve eye-trackers by means of a mutual collaboration between the eye-tracker and the face recognition system [36]. There are several applications of real-time gesture detection and recognition using sequences of video images [37] that characterise the relevant elements in a person's face: eyes, eyebrows, nose or mouth [38, 39].

In applications used for detecting the line of gaze of an automobile driver, non-intrusive systems are the most suitable as they do not affect the driving process, thus meeting all the conditions for usual behaviour during the experiment [40]. At the same time, ocular movements are captured with a video camera mounted on the car dashboard so that tracking and analysis can be carried out in real time [41].

The above mentioned techniques imply using some calibration procedure before starting the experiment. Calibration is performed by directing the subject's gaze towards a set of previously defined points. In this way, the system can extract the parameters needed to relate the pupil-glint vector with the actual gaze direction [42]. These parameters are applied later during the gaze determination process.

The possibility of the glint disappearing due to strong torsion of the eye or some head movements has to be considered. In addition, the eye's geometry has to be taken into account because the change in the ocular globe curvature – between the iris and the sclera – can cause strong distortions of the glint when it appears near the contour of the iris or outside of it. Figure 1 shows the geometry of the ocular globe, representing the curvature changes. Turning the head or the eye more than 50 degrees causes some glints to be severely distorted and makes the determination of their co-ordinates and geometrical characteristics difficult.

Some authors use two illumination sources to try to guarantee the existence of at least one glint [43]. Even in this case, there are situations in which glints disappear.

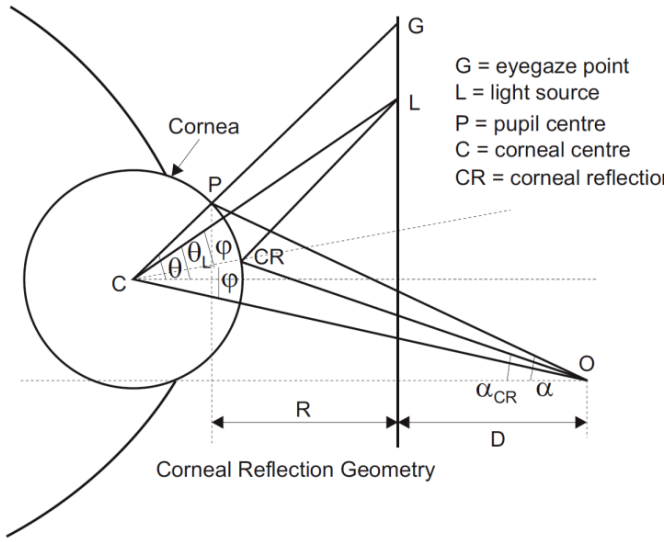


Figure 1. Ocular globe geometry

The basic idea of our work is somewhat similar to that described by Yoo [12] and Beymer [29], at least in the sense all three are using a Wide FOV camera (or camera-set) for detecting the subject's face and a Narrow FOV camera for identifying the pupil and glints from the user eye. Comparing with Yoo's paper, in our work we get the pupil center and the position of the glints with subpixel resolution. Moreover, we do not require two different, consecutively acquired images – with different illumination – for locating the pupil, as needed in the Yoo paper. In this way, it should be easier to make our system work with higher fps rate on updated hardware. In the case of the Beymer work, they propose a relatively complex system which needs two stereo cameras, a vertical setup for detecting the face and a horizontal setup for identifying and analyzing the eye. Their proposal includes a 3D model of the eye and they state that the results obtained are rather remarkable with a 22 sample experiment. Related to both [12] and [29], our system introduced an original infrared LED L shape which helps identifying the glints with great precision even when they are superimposed due to the specific position and orientation of the subject's eye at particular instants. In addition, our system is also distinguished because we use high luminosity LEDs activated during very short intervals with the purpose of obtaining very high contrast with any external illumination condition while preserving the maximum light pressure recommended for avoiding any eye harm. These characteristics are being motivated by the main application of our system as a tool to help investigating the behavior of car drivers under different environmental or psychological conditions.

2 SYSTEM OVERVIEW

A configuration with four illuminators has been chosen for the system to guarantee that every image will contain at least one glint without distortion. If there are more than one, they have to be identified. The specific shape of the illuminators allows optimization (simplification) of the algorithm used to identify the glint and the direction of eye rotation. Correct positioning of the illuminators with respect to the person's head ensures that at least one of the glints is valid in every frame. With this approach, the pupil has to be recognized in order to determine its centre. Also, the radius of the iris has to be determined, so as to consider only the glints found inside the iris area. Figure 2 represents a situation with glint distortion.

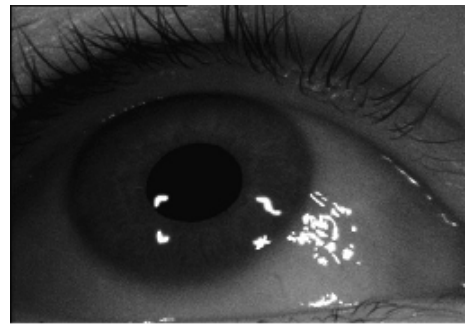


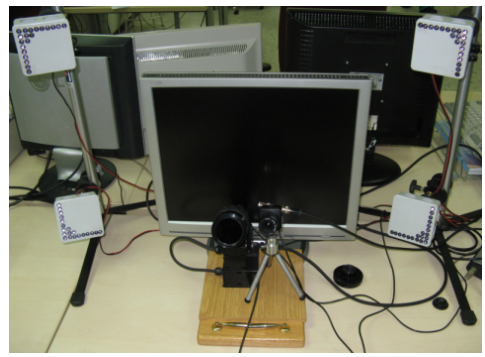
Figure 2. Example of glint distortion outside the iris zone

To minimize the influence of external illumination variations, the shutter of the NFV camera is set to allow just a small amount of light to enter (1/2 000 s) while suitable pulsating infrared illumination is used. Special care has been taken to limit the amount of illumination power applied, following opthalmologic recommendations in order to work well below maximum recommended levels [44, 45].

Every illuminator consists of 20 high intensity infrared LEDs with an aperture angle of 6 degrees (Figure 3 b)). Illuminators are located at an average of 75 cm from the subject's eyes to provide enough facial illumination and the generation of glints. This illumination is not continuous but synchronized with the electronic shutter of the NFV camera in such a way that very short 500-microseconds flashes are generated every 20 milliseconds.

Figure 3 a) shows a general view of the system including the four illuminators, the NFV video camera used to take images of the eye following the head movements (Sony XC-EI50) and the WFV camera for taking images of the face (Watek WAT-535-EX, Figure 3 c)). This second camera allows the determination of the eye position to reposition the NFV camera when the image of the eye has been lost.

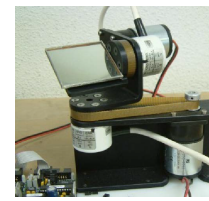
The NFV camera has a 1/4" infrared sensitive CCD with minimum illumination of 0.1 lux (F1.4). A 75 mm lens with a 2 × extender is used to obtain a total of 150 mm focal length. The F-stop has to be set over 5.6 to allow for sufficient depth



a) General view of the prototype

b) Illuminator

c) Cameras



d) Gimbal prototype

Figure 3. General view of the prototype implemented at the laboratory. Detail of illuminator, cameras and gimbal prototype.

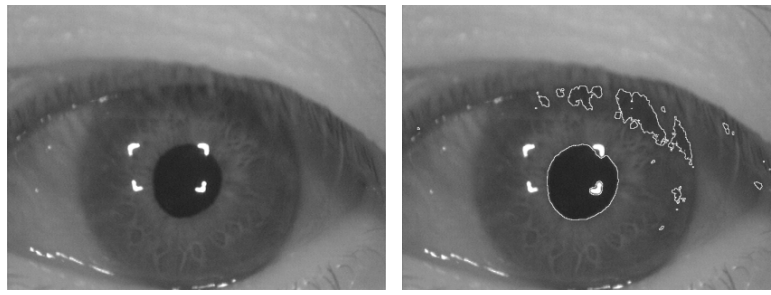
of field. Two different system prototypes have been implemented to follow the head movements (maintaining the eye centred in the image). The first one uses a pan-tilt unit to move the NFV camera and the second one has a fixed NFV camera and uses a mirror positioned in motorized gimbals to follow head movements. An alternative based on a gimbal mirror instead of a pan-tilt has also been taken into account 3 d). This mirror would be used in a real system instead of a lab prototype.

The WFV camera used to obtain an image of the face – to gather information about the absolute position of the eye – has a 1/3" infrared sensitive CCD with a minimum illumination of 0.5 lux (f1.4) and automatic electronic shutter.

3 DETERMINING THE PUPIL AND IRIS CENTRE AND RADIUS

The pupil-glint vector is the basis for determining gaze direction, hence the position of the pupil centre as well as the position of the glint have to be precisely determined. Due to the geometric features of the human eye, some of the glints may be invalid, especially those located outside the iris or near its contour. Therefore, the iris radius has to be determined to validate the glints.

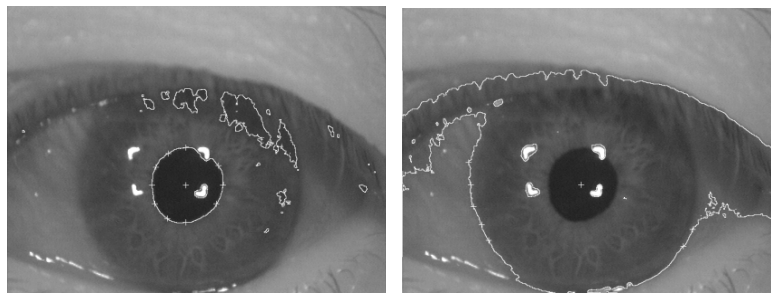
An important factor to be taken into account is the image resolution, since it is directly related to the final precision achieved in the point of gaze. The proposed



a) Original eye image b) Pupil contour

Figure 4. First estimation of the pupil contour

algorithm is able to determine the pupil centre at sub-pixel resolution, aiming at minimising measurement errors in the pupil-glint distance. It comprises the following phases (see Figure 4 and Figure 5).



a) Pupil contour and centre determination b) Iris radius determination

Figure 5. Final stages of estimating the pupil centre and iris contour and radius

3.1 Estimating the Pupil Centre

Because the pupil is nearly black, all the pixels with a grey level below a given threshold (close to black) are selected in the original image (Figure 4 a)). Then, a first approximation of the mass centre is computed by averaging the x co-ordinates and the y co-ordinates. The whole process is iterated, but using pixels in an area around the previously obtained pupil. Rejected points are considered remote stains or noise areas not relevant in the analysis. This iterative process is repeated until the resulting x and y co-ordinates are stabilized. This algorithm is summarized in Table 1.

Pupil center
1. Histogram
2. Threshold, close to black
3. Average coordinates of black pixels (centre)
4. Average coordinates of black pixels close to the previously obtained centre
5. Repeat (4) until the resulting coordinates are stable

Table 1. Algorithm for pupil centre computation

3.2 Searching for the Pupil Contour

Once a good estimate of the pupil centre has been obtained, a procedure to search for its contour is applied. This step is intended to check that previous searching has been successful and also serves for error rejection. To save computation time the image is clipped to include only a small area around the iris. By filtering the image, a noise reduction is performed and all pixels with a grey level higher than the pupil threshold are inverted. Then a border detection algorithm based on the Laplacian operator is applied (see Figure 4 b)). This algorithm is summarized in Table 2.

Pupil contour
1. Image clipping
2. Noise reduction
3. Threshold
4. Border detection using a Laplacian filter

Table 2. Algorithm to determine pupil contour

3.3 Determining the Pupil Centre and Radius

The first estimate of the pupil centre is insufficient due to its lack of accuracy and the possibility of finding a false pupil, propagating errors to the rest of the system. To confirm that the actual pupil has been discriminated, its contour must have an elliptical shape. In this case the validation is performed, whereas if this shape is not confirmed, an error is returned signalling the rest of the system that the pupil has not been found.

To confirm the pupil and to accurately compute its centre, a new algorithm has been developed. It consists of a searching procedure that starts from the previously estimated centre, applied to n diagonals crossing the pupil contour at specific points. If the contour is reached and the condition of equidistance to the centre is fulfilled for all intersecting points, then the position of the centre is guaranteed, and a first estimate of the pupil radius is obtained. During this process, wrong points must

be suppressed by considering their distance to the centre. The radius must be recomputed for each rejected point. This algorithm is summarized in Table 3.

Pupil centre and pupil radius
1. Drawing diagonals from the pupil centre previously obtained up to the already computed pupil contour
2. If the diagonal does not cross the pupil contour at an expected distance, the diagonal is rejected
3. If several diagonals are rejected, the pupil centre is also rejected and an error is returned to avoid tracking an erroneous pupil
4. Too long or too short diagonals are rejected
5. The pupil radius as the average of diagonal lengths is calculated
6. The pupil centre is recalculated as the intersection of the diagonals. If several values are found, voting is used to discard the wrong values.

Table 3. Algorithm for subpixel pupil centre and pupil radius computation

The same diagonals are used to compute a more accurate estimate of the centre of the pupil, obtaining sub-pixel resolution (see Figure 5 a)). First, both very long and very short diagonals are rejected. The centres of the remaining diagonals are used to compute the centre of the pupil. A new centre with higher accuracy is determined. If different centres are found, wrong values are discarded by voting.

3.4 Searching for the Iris Contour

The goal of this phase is to determine which glints are located inside the iris. The process is similar to the one used for the pupil. The starting image is a clipped image. First, the image is filtered, noise reduction is performed and then some pixels with high grey level values are inverted. A binary image is obtained with black pupil, iris and eyelashes, while most of the sclera is white. A Laplacian algorithm is applied on this image for edge detection. Finally, a binarization process renders the iris outline in white on a black background. This algorithm is summarized in Table 4.

Iris contour (same algorithm as pupil contour)
1. Image clipping
2. Noise reduction
3. Threshold
4. Border detection using a Laplacian filter

Table 4. Algorithm for iris contour determination

3.5 Determining the Iris Radius

Using the pupil centre as a first estimate of the iris centre, a process similar to the one used for determining the pupil radius is applied. A search for the iris contour is made following radial lines. Because all of the intersecting points must be situated at a similar distance from the centre, wrong points are rejected, and the radius of the iris is computed with the remaining points (Figure 5 b)). This algorithm is summarized in Table 5.

Iris radius (similar to pupil radius algorithm)
1. Diagonals are drawn from the previously calculated pupil centre to the iris contour
2. If the diagonal does not cross the iris contour at an expected distance, the diagonal is rejected
3. Diagonals that are too long or too short are rejected
4. The iris radius is calculated as the average of the diagonal lengths

Table 5. Algorithm for iris radius

4 GLINT DETECTION

The detection of adjacent sub-regions with common features (blobs), formed by glints or illumination patterns is a technique currently used with binary images modified by morphological processing (erode, dilate, open, etc.). Identifying sub-regions with similar shapes requires the computation of moments, since central moments are independent of the blob position in the image, normalised moments are independent from the area, and invariant moments are independent from rotation, shifting or change of scale. To calculate the central moments it is necessary to calculate each glint centroid $(\overline{x}, \overline{y})$ as follows. Let $Im[x, y]$ be the pixel value at (x, y) and m_{pq} the raw image moment pq. Then:

$$\overline{x} = \frac{m_{10}}{m_{00}}, \qquad \overline{y} = \frac{m_{10}}{m_{00}} \tag{1}$$

where

$$m_{pq} = \sum_p \sum_q x^p y^q Im[x, y]. \tag{2}$$

The invariant moments used for classifying the glints found are the first three moments defined by the Hu set of invariant moments [46]:

$$\phi_1 = \eta_{20} + \eta_{02}, \tag{3}$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2, \tag{4}$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \tag{5}$$

where

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{1+\frac{p+q}{2}}} \tag{6}$$

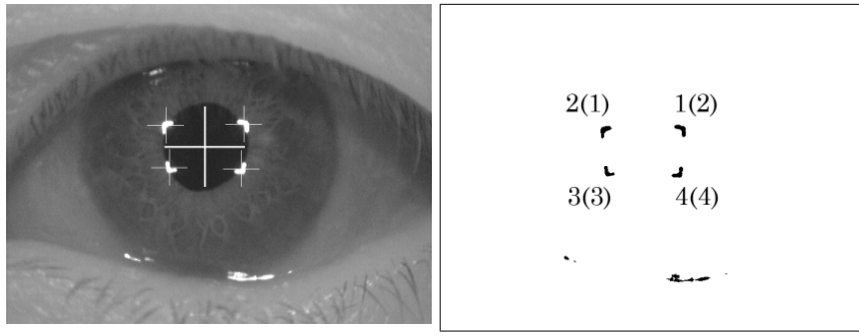
and

$$\mu_{pq} = \sum_p \sum_q (x - \overline{x})^p (y - \overline{y})^q Im[x, y]. \tag{7}$$

An algorithm that uses these techniques to detect the glints on the eye image has been implemented, with the aim of locating their position and classifying valid glints (Figures 6, 7, 8 and 9). For this task, the characteristic parameters of each blob are calculated.

The procedure carried out by the algorithm is divided into two phases:

1. A blob analysis of an image similar to Figures 6, 7, 8 and 9 is performed to isolate all the glints fulfilling certain restrictions.
2. Identification of valid glints among all of the isolated ones.

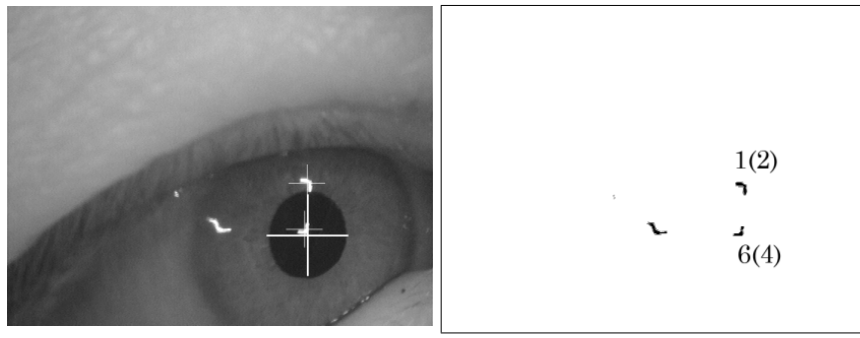


a) Original eye image b) Binarized image

Blob	Area	$\overline{x}$	$\overline{y}$	Iris	Size	Glint
1	194	401.9	236.6	yes	yes	2
2	192	273.7	238.2	yes	yes	1
3	146	279.2	307.8	yes	yes	3
4	170	397.2	309.9	yes	yes	4
5	37	209.4	451.9	no	no	–
6	6	223.5	457.5	no	no	–
7	4	476.8	478.2	no	no	–
8	227	398.0	487.0	no	yes	–
9	105	437.8	487.0	no	yes	–

Figure 6. Example of glint detection with 4 glints detected

Blob analysis is carried out as follows. First of all, the image is loaded and binarized with a certain threshold. An array is reserved and initialised to hold the information related to each blob. The image is scanned to search for a white pixel.



a) Original eye image b) Binarized image

Blob	Area	$\overline{x}$	$\overline{y}$	Iris	Size	Glint
1	169	503.2	329.8	yes	yes	2
2	3	288.0	344.3	no	no	–
3	3	289.0	348.3	no	no	–
4	1	346.0	388.0	no	no	–
5	188	358.7	400.8	no	yes	–
6	116	499.1	405.9	yes	yes	4
7	2	369.5	404.0	yes	no	–

Figure 7. Example of glint detection with 2 glints detected

This first pixel establishes the origin of a new blob initiating a propagation process that scans all neighbouring pixels, increasing the size of the blob. If the maximum number of pixels is exceeded, an error message is returned. This procedure tags any pixels already scanned to avoid repetition.

Each pixel in a blob is scanned using connectivity 8 by means of recursive calls. Labelling of each pixel is performed in an auxiliary labelling image, and the central moments and limits of the blob are computed in a cumulative way while new pixels are located. When the process has finished, the gravity centre for each blob is computed.

The detection of glints by the proposed algorithm is fairly reliable, although blobs corresponding to shiny patches of skin or different parts of the eye are sporadically detected. Therefore, these undesired blobs must be rejected before starting the identification of the searched glints. Two techniques are used for this purpose: first, rejecting blobs by size (too big or too small) and second, rejecting all blobs located outside the iris. The radius of the iris was previously computed as described in Subsection 3.5. This algorithm is summarized in Table 6.

Once a purge of erroneous glints has been performed, the identification phase is carried out working with the previously computed gravity centres. An index from one to four is assigned to each glint, starting from the top left corner. In this way, each of the glints has a fixed relative position with regard to the others. This is an important feature for glint identification.

Glint detection
1. Threshold, close to white
2. An untagged white pixel is searched to set the origin of a new blob
(a) All the neighbours are scanned recursively adding detected white pixels to the blob. All the scanned pixels are tagged
(b) Step 2 (a) is repeated until there are no more neighbours to scan
3. Step 2 is repeated until there are no more untagged pixels in the image
4. Blobs that are too big or too small are removed
5. Blobs that are outside the iris contour are removed
6. The glints are identified starting form the top left corner, from 1 to 4

Table 6. Algorithm for glint detection

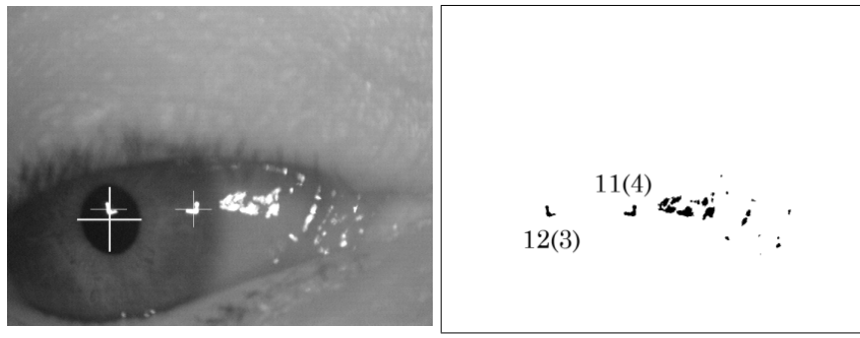
The algorithm described includes the computation of the absolute co-ordinates of the centre of each glint. These co-ordinates, together with those for the pupil centre allow the pupil-glint vector to be determined. In a number of cases, when more than one glint can be identified, a vector for each of them is obtained, allowing to use redundant information to determine the point of gaze.

5 CALIBRATION METHOD

There are two main methods for mapping eye image features to point-of-gaze co-ordinates [47]: geometry-based models – considering eye, camera and illuminator characteristics and position, and general interpolation-based methods. Cerrolaza et al. conclude stating that a lower error in the calibration position can be observed in polynomial-based methods. They even state that using higher order polynomials and complete mathematical expressions do not enhance accuracy. These authors establish that using only one corneal reflection is not enough unless the system avoid or strongly limit subject's head motion.

On the other hand, some methods have been proposed to avoid calibration by using four glints, as the cross-ratio (CR) based eye tracking described in [48]. However, this eye-tracker does not provide good estimations because of the geometry model simplifications it assumes. A recent proposal found in [49] improves CR eye-tracker precision, but it needs to do a calibration step for each person and it also requires to find all four glints in every image, a limitation shared with the original CR eye-tracker.

The mapping method employed in our system is originally based on the system geometry, but it has been simplified so as to be finally based on a quadratic pupil-reflection expression evaluated by polynomial regression. In order to allow some degree of head movement, a maximum of four glints is used to determine the point of gaze. Even considering up to four pupil-glint vectors, the calibration process is identical as if only one glint would be used because for each calibration point the



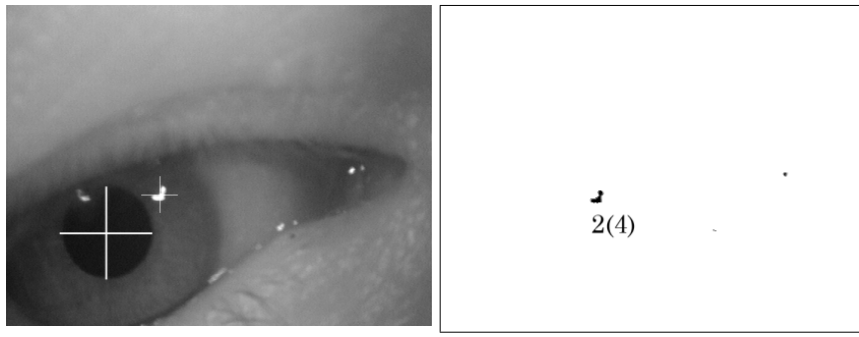
a) Original eye image b) Binarized image

Blob	Area	$\overline{x}$	$\overline{y}$	Iris	Size	Glint
1	8	467.8	311.8	no	no	–
2	15	477.4	340.1	no	no	–
3	24	453.5	342.2	no	no	–
4	146	391.2	352.0	no	yes	–
5	114	440.6	355.3	no	yes	–
6	366	376.2	363.3	no	yes	–
7	18	452.6	352.4	no	no	–
8	16	521.1	352.8	no	no	–
9	47	427.4	355.5	no	no	–
10	12	417.5	354.0	no	no	–
11	187	317.5	372.1	yes	yes	4
12	129	178.7	372.5	yes	yes	3
13	306	408.6	372.6	no	yes	–
14	259	447.1	373.6	no	yes	–
15	67	513.2	374.2	no	yes	–
16	26	581.7	372.0	no	no	–
17	82	506.4	388.0	no	yes	–
18	7	531.4	411.9	no	no	–
19	31	477.6	416.3	no	no	–
20	12	563.2	415.7	no	no	–
21	42	563.4	428.5	no	no	–
22	2	533.5	443.0	no	no	–

Figure 8. Example of glint detection with 2 glints detected and many blobs

system calculates the pupil center and the coordinates of the four corneal reflections. The system determines the coordinates of such points by means of the algorithms described in previous Sections 3 and 4.

To determine the point of gaze, its coordinates (x_G, y_G) and the components of the pupil-glint vector (d_x, d_y) have to be correlated. Their relationship depends on the eye's geometry and on the relative position of the eye with respect to the glint and the point of gaze. An analytical study of the problem has shown that a quadratic



a) Original eye image b) Binarized image

Blob	Area	$\overline{x}$	$\overline{y}$	Iris	Size	Glint
1	25	578.3	309.3	no	no	–
2	201	263.8	349.1	yes	yes	4
3	6	459.5	403.5	no	no	–

Figure 9. Example with one glint detected

function of (d_x, d_y) provides a good approximation for (x_G, y_G). Analytically, the relationship can be expressed as follows:

$$\frac{x_G - a}{\sqrt{x_G^2 - 2ax_G + b^2}} = K_1(d_x) + K_2, \tag{8}$$

$$\frac{y_G - c}{\sqrt{y_G^2 - 2ay_G + d^2}} = K_3(d_y) + K_4 \tag{9}$$

where a, b, c, d, K_1, K_2, K_3 and K_4 are constants that depend on the geometry of the problem.

In this quadratic function, the co-ordinates of x_G and y_G can be considered independent or related as follows:

$$x_G = C_0 + C_1(d_x) + C_2(d_y) + C_3(d_x)(d_y) + C_4(d_x)^2 + C_5(d_y)^2, \tag{10}$$

$$y_G = K_0 + K_1(d_x) + K_2(d_y) + K_3(d_x)(d_y) + K_4(d_x)^2 + K_5(d_y)^2. \tag{11}$$

If enough pairs $[(d_x, d_y), (x_G, y_G)]$ are available, the values for C_i and K_i can be determined by means of a polynomial regression, using a function that generates the regression matrix R and the coefficients:

$$C = (R^t R)^{-1} R^t X, \tag{12}$$

$$K = (R^t R)^{-1} R^t Y. \tag{13}$$

As the eye and head parameters differ for different users, a set of parameters have to be determined in each case. Therefore a calibration procedure has been established to obtain the necessary pairs (d_x, d_y), (x_G, y_G) allowing the parameters to be determined. The calibration procedure is as follows:

A number of points is defined in the screen. The user has to position the mouse pointer on each of these points and while fixing his/her gaze on the selected point press the mouse button. The system will identify the position of the point on the screen, as well as the pupil centre and the glint positions in the eye image.

The calibration process presents two critical aspects:

1. A very accurate determination of the position of the centre of the pupil is needed.
2. Fixing the subject's gaze on a point requires some practice. This is important for correct calibration of the system, as subjects tend to find it difficult to focus on exactly the same point, which causes dispersion of measurements.

a) b) c)

Figure 10. Location of eye positions from the face image: 10 a) original image, 10 b) face contour detection-and-clipping and eye location, and 10 c) final result

6 FACE AND EYE DETECTION

Face detection is used to locate the position of the eye whenever its image eludes the NFV camera due to a sudden head movement. Although this kind of movement is not very frequent, it can cause the system to loose track of the eye.

A fixed WFV camera is used to gather images of the face to give information about the absolute position of the eye. This camera is also infrared sensitive, gathering monochrome images that allow the eye position to be determined.

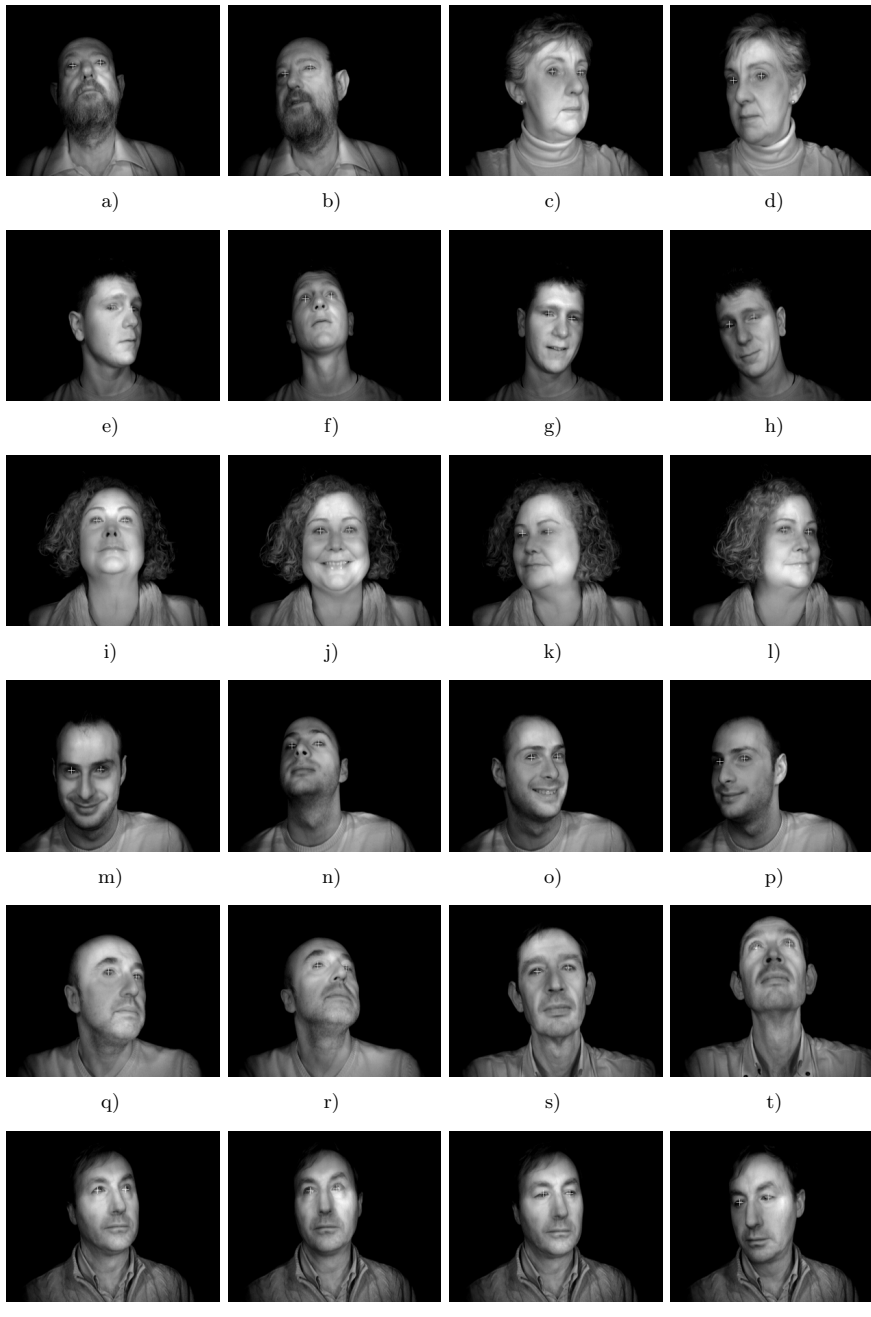


Figure 11. Face processing results

The first problem is the elimination of unwanted reflections due to ambient light. In order to minimize the influence of external illumination variations, which are unavoidable in a conventional driving session, the NFV camera electronic shutter, working at a speed of 1/2000 s, has been synchronized with the illuminators working in pulsed mode (500 microseconds illumination flashes per frame). The system is suitable for working outdoors because the pulsed infrared illumination can be much higher than the ambient light (even with the sun shining above), maintaining the average illumination power in a level which avoids damaging the eye, following ophthalmologic recommendations and working well below dangerous levels.

The algorithm developed to determine the face contour, (Figure 10), applies an image filter and blob detection to the original image (Figure 10 a)) so as to isolate the face area. The resulting image contains only the face contour allowing the eyes and other features such as mouth and eyebrows to be distinguished. Applying anthropometric considerations, the image is clipped in such a way that only an area around the eyes is included (Figure 10 b)). Finally, blob detection is again applied to locate the eyes (Figure 10 c)) and results are validated using anthropometric distances. The algorithm is summarized in Table 7.

Face and eye detection
1. Noise reduction
2. Blob detection to localize the face
3. Image clipping around the eyes
4. Blob detection to localize the eyes
5. Rejection of erroneous blobs using anthropometric distances

Table 7. Algorithm for face and eye detection

The described process returns the eye position information and it is applied whenever the eye leaves the image field captured by the NFV camera. This process has been integrated with the pupil-glint vector determination process, providing an efficient system to track and determine the point of gaze of a user that is especially robust for usual head movements.

7 RESULTS

The gaze-tracking system is able to work in real time even with a simple experimental platform based on an intel Core2 Q8200 CPU running at 2.33 GHz and with 4 GB of RAM. This platform can process 25 frames per second for 720×576 pixel images. It permits gaze tracking for a gaze angle range over ± 45 degrees. Table 8 shows the approximate intervals spent on processing the algorithm stages.

As the gaze direction is estimated using the vector linking the pupil centre to one of the glints generated by the illuminators, the accuracy of gaze tracking directly depends on the detection accuracy of both the pupil centre and the glints under

Available time between frames	40 ms
Eye tracking	2 ms
Eye position detection	1 ms
Point of gaze detection:	
Pupil and Iris determination	14 ms
Glint detection	3 ms
Total processing time per frame	20 ms

Table 8. Time spent on main algorithm stages

different conditions. For instance, the use of contact lenses or conventional glasses may have some impact on the results.

Figure 11 shows that both user eyes are correctly detected with the described prototype from the images taken by the WFV camera, allowing quick recovery from tracking errors of the NFV camera due to sudden movements of the user. Using the WFV images the position of both eyes is obtained and the NFV tracking can be reset and continue tracking the proper eye.

In the figure several faces are shown where the eyes are properly detected even though the subjects are looking to different directions, either because they move their head or their eyes. Some of the figures show the subject with half closed eyes or with glints from teeth or from ear rings.

User's facial expression does not affect the detection, the individual can be serious, normal, smiling or even laughing. Moreover, the use of infrared light removes background objects reducing the complexity of the scene and allowing easy processing of the images.

Figures 12 and 13 show the results obtained by processing eye images from several individuals while they are looking at different objects from a distance. The robustness of the algorithm for detecting both the pupil centre and a sufficient number of led reflections is shown by the wide range of different positions and orientations of the pupil in each image.

In Figures 12 and 13 each row shows multiple shots of an eye from the same subject. The first column in both Figures shows the cases in which the four led reflections are located and correctly identified. In all these images the pupil is quite well centered in its image frame, the reflections are inside the iris radius and there are no appreciable distortions.

The remaining eye-images of each individual show situations with some additional difficulties. In all these cases, the individual is looking up, down or sideways. Furthermore, the pupil is not properly centered in the image, which may be due to a head movement or to a sudden eye movement (*saccadic eye movement*) which can not be corrected in real time by the positioning system of the camera. In almost all of the referred cases, at least one reflex is located and correctly identified. Thus, all four reflections are located in Figures 12 a)–12 d)–12 e)–12 f)–12 h)–12 i)–12 k)–12 m) and 13 a)–13 b)–13 c)–13 e)–13 f)–13 i)–13 m)–13 o), three of the four reflections are identified in Figures 12 g)–12 j)–12 n)–12 o) and 13 g)–13 j)–13 n)–13 p), two reflec-

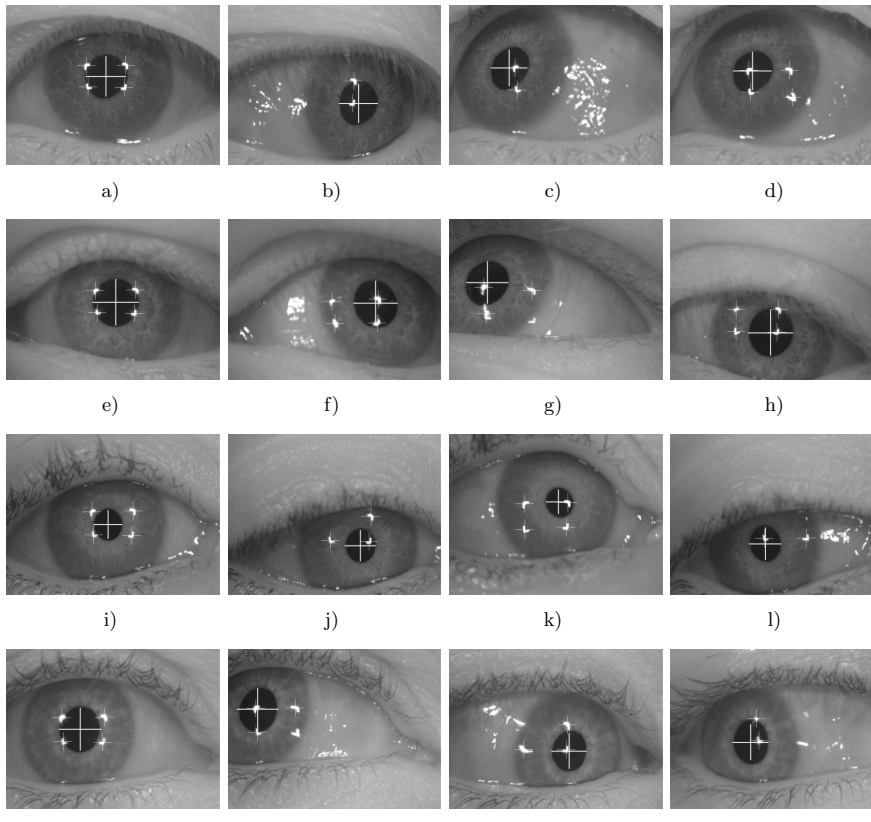


Figure 12. Eye processing results 1

tions in Figures 12 b)–12 c)–12 l)–12 p) and 13 d)–13 k), and finally, a single reflection is identified in Figure 13 l). Furthermore, no reflections can be identified in the image corresponding to Figure 13 h), even though several reflection blobs have been located. However, none of them conforms to the expected shape of the standard reflection types. This is largely due to the partial occlusion of the iris and pupil by the eye lashes, causing distortion on the reflections. In all other cases at least one reflection is detected, and this is sufficient to identify the point of gaze, since the reflection type has been correctly identified, allowing the required formula to be applied.

Finally, the robustness of the identification system may be even better justified by the fact that some of the sample images used correspond to individuals wearing contact lenses, such as the cases shown in Figures 12 e)–12 f)–12 g)–12 h)–12 m)–12 n)–12 o)–12 p) and 13 e)–13 f)–13 g)–13 h).

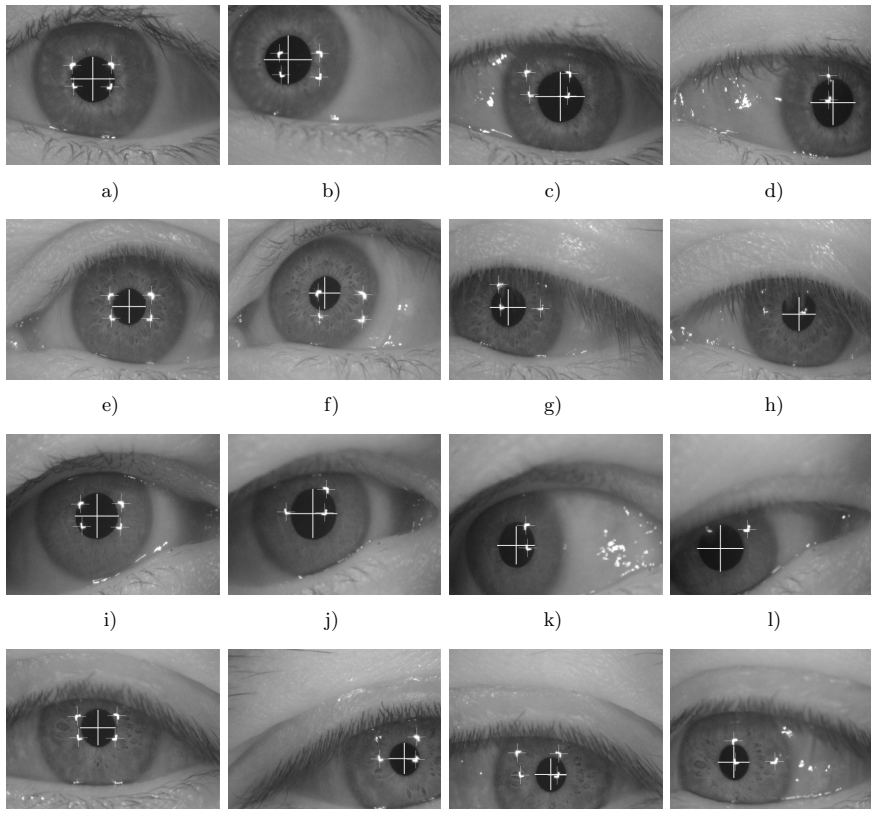


Figure 13. Eye processing results 2

8 CONCLUSIONS

A non-intrusive eye-gaze detection system prototype has been developed. The system works in real time processing 25 frames per second for 720×576 pixel images taken by the eye camera. It allows gaze tracking for a range of gaze angles over ± 45 degrees.

The sub-pixel determination of the pupil centre enhances accuracy and the use of four specially shaped infrared illuminators allows continuous eye gaze determination, even with high eye or head torsion. Synchronization between the illuminators and the camera shutter avoids interference by ambient light sources while maintaining illumination power below dangerous levels that could harm the eye.

The absolute eye position, obtained by processing the image of a wider field of view camera, is used for quick recovery from occasional tracking errors produced by sudden head movements or very slow blinking. The prototype has been tested

for various individuals and lighting conditions and it presents a robust system with high potential for use in different applications.

REFERENCES

[1] Grice, S. J.—Halit, H.—Farroni, T.—Baron-Cohen, S.—Bolton, P.—Johnson, M. H.: Neural Correlates of Eye-Gaze Detection in Young Children with Autism. Cortex, Vol. 41, 2005, No. 3, p. 342, doi: 10.1016/S0010-9452(08)70271-5.

[2] Bengler, K.—Bernasch, J. H.—Löwenau, J. P.: Comparison of Eye Movement Behaviour During Negotiation of Curves on a Test-Site and in a Driving Simulator. Applied Psychology, University Regensburg, 1994.

[3] Ji, Q.—Yang, X.: Real-Time Eye, Gaze, and Face Pose Tracking for Monitoring Driver Vigilance. Real-Time Imaging, Vol. 8, 2002, No. 5, p. 357, doi: 10.1006/rtim.2002.0279.

[4] D'Orazio, T.—Leo, M.—Distante, A.: Eye Detection in Face Images for a Driver Vigilance System. Intelligent Vehicles Symposium, 2004 IEEE, 2004, pp. 95–98, doi: 10.1109/IVS.2004.1336362.

[5] Kim, K.—Ramakrishna, R. S.: Vision-Based Eye-Gaze Tracking for Human Computer Interface. Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Vol. 2, 1999, pp. 324–329.

[6] Sesin, A.—Adjouadi, M.—Ayala, M.—Barreto, A.—Rishe, N.: A Real-Time Vision Based Human Computer Interface as an Assistive Technology for Persons with Motor Disability. WSEAS Transactions on Computer Research, 2007, Vol. 2, pp. 115–121.

[7] Reulen, J. P.—Marcus, J. T.—Koops, D.—de Vries, F. R.—Tiesinga, G.—Boshuizen, K.—Bos, J. E.: Precise Recording of Eye Movement: The IRIS Technique. Part 1. Medical and Biological Engineering and Computing, Vol. 26, 1988, No. 1, pp. 20–26, doi: 10.1007/bf02441823.

[8] Parkhurst, D. B. D.: Open-Source Software for Real-Time Visible-Spectrum Eye Tracking. The 2nd Conference on Communication by Gaze Interaction (COGAIN 2006), 2006, pp. 18–20.

[9] Robinson, D. A.: A Method of Measuring Eye Movemnent Using a Scieral Search Coil in a Magnetic Field. IRE Transactions on Bio-Medical Electronics, Vol. 10, 1963, No. 4, pp. 137–145, doi: 10.1109/TBMEL.1963.4322822.

[10] Kaufman, A.—Bandopadhay, A.—Shaviv, B.: An Eye Tracking Computer User Interface. Proceedings of the IEEE 1993 Symposium on Research Frontiers in Virtual Reality, 1993, pp. 120–121, doi: 10.1109/VRAIS.1993.378254.

[11] Morimoto, C. H.—Koons, D.—Amir, A.—Flickner, M.: Pupil Detection and Tracking Using Multiple Light Sources. Image and Vision Computing, Vol. 18, 2000, No. 4, p. 331, doi: 10.1016/S0262-8856(99)00053-0.

[12] Yoo, D. H.—Chung, M. J.: A Novel Non-Intrusive Eye Gaze Estimation Using Cross-Ratio Under Large Head Motion. Computer Vision and Image Understanding, Vol. 98, 2005, No. 1, pp. 25–51, doi: 10.1016/j.cviu.2004.07.011.

[13] Noureddin, B.—Lawrence, P. D.—Man, C. F.: A Non-Contact Device for Tracking Gaze in a Human Computer Interface. Computer Vision and Image Understanding, Vol. 98, 2005, No. 1, pp. 52–82, doi: 10.1016/j.cviu.2004.07.005.

[14] Hansen, D. W.—Hansen, J. P.: Eye Typing with Common Cameras. Proceedings of the 2006 Symposium on Eye Tracking Research & Applications (ETRA '06), ACM, New York, NY, USA, 2006, pp. 55–55, doi: 10.1145/1117309.1117340.

[15] Droege, D.—Geier, T.—Paulus, D.: Improved Low Cost Gaze Tracker. Proceedings of the 3rd COGAIN Conference, 2007, pp. 37–40.

[16] Talmi, K.—Liu, J.: Eye and Gaze Tracking for Visually Controlled Interactive Stereoscopic Displays. Signal Processing: Image Communication, Vol. 14, 1999, No. 10, pp. 799–810.

[17] Parkhurst, D.—Niebur, E.: A Feasibility Test for Perceptually Adaptive Level of Detail Rendering on Desktop Systems. Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization (APGV '04), ACM, New York, NY, USA, 2004, pp. 49–56, doi: 10.1145/1012551.1012561.

[18] Morimoto, C. H.—Mimica, M. R. M.: Eye Gaze Tracking Techniques for Interactive Applications. Computer Vision and Image Understanding, Vol. 98, 2005, No. 1, pp. 4–24, doi: 10.1016/j.cviu.2004.07.010.

[19] Guestrin, E.—Eizenman, M.: General Theory of Remote Gaze Estimation Using the Pupil Center and Corneal Reflections. IEEE Transactions on Biomedical Engineering, Vol. 53, 2006, No. 6, pp. 1124–1133, doi: 10.1109/TBME.2005.863952.

[20] Mulligan, J.: Image Processing for Improved Eye Tracking Accuracy. Behavior Research Methods, Instruments, & Computers, Vol. 29, 1997, No. 1, pp. 54–65, doi: 10.3758/bf03200567.

[21] Smith, P.—Shah, M.—da Vitoria Lobo, M.: Monitoring Head/Eye Motion for Driver Alertness with One Camera. Proceedings of 15th International Conference on Pattern Recognition, 2000, Vol. 4, pp. 636–642, doi: 10.1109/ICPR.2000.902999.

[22] Stiefelhagen, R.—Yang, J.—Waibel, A.: Tracking Eyes and Monitoring Eye Gaze. Proceedings of the Workshop on Perceptual User Interfaces (PUI '97), 1997, pp. 98–100.

[23] Pérez, A.—Córdoba, M. L.—García, A.—Méndez, R.—Muñoz, M.—Pedraza, J. L.—Sánchez, F.: A Precise Eye-Gaze Detection and Tracking System. 11th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG03), 2003, pp. 105–108.

[24] Li, D.—Winfield, D.—Parkhurst, D.: Starburst: A Hybrid Algorithm for Video-Based Eye Tracking Combining Feature-Based and Model-Based Approaches. IEEE Computer Society Conference on Computer Vision and Pattern Recognition – Workshops (CVPR Workshops), 2005, pp. 79–79, doi: 10.1109/CVPR.2005.531.

[25] Daunys, G.—Ramanauskas, N.: The Accuracy of Eye Tracking Using Image Processing. Proceedings of the Third Nordic Conference on Human-Computer Interaction (NordiCHI '04), ACM, New York, NY, USA, 2004, pp. 377–380, doi: 10.1145/1028014.1028074.

[26] Zhu, J.—Yang, J.: Subpixel Eye Gaze Tracking. Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition (FGR '02), IEEE Computer Society, Washington, DC, USA, 2002, pp. 131–136.

[27] Singh, S.—Papanikolopoulos, N.: Advances in Vision-Based Detection of Driver Fatigue. Proceedings of the ITS America Ninth Annual Meeting, 1999.

[28] Tang, B.—Ait-Boudaoud, D.—Matuszewski, B.—Shark, L.: An Efficient Feature Based Matching Algorithm for Stereo Images. Geometric Modeling and Imaging – New Trends, 2006, pp. 195–202, doi: 10.1109/GMAI.2006.6.

[29] Beymer, D.—Flickner, M.: Eye Gaze Tracking Using an Active Stereo Head. Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '03), 2003, Vol. 2, pp. II–451–8, doi: 10.1109/CVPR.2003.1211502.

[30] Bala, L.-P.—Talmi, K.—Liu, J.: Automatic Detection and Tracking of Faces and Facial Features in Video Sequences. ITG FACHBERICHT, 1997, pp. 251–256.

[31] Bakić, V.—Stockman, G.: Menu Selection by Facial Aspect. Proceedings of Vision Interface, Vol. 99, 1999.

[32] Bagci, A. M.—Ansari, R.—Khokhar, A.—Cetin, A. E.: Eye Tracking Using Markov Models. Proceedings of the 17th International Conference on Pattern Recognition (ICPR 2004), IEEE, 2004, Vol. 3, pp. 818–821, doi: 10.1109/ICPR.2004.1334654.

[33] Turkan, M.—Pardas, M.—Cetin, A. E.: Edge Projections for Eye Localization. Optical Engineering, Vol. 47, 2008, No. 4, Art. No. 047007.

[34] Matsumoto, Y.—Zelinsky, A.: An Algorithm for Real-Time Stereo Vision Implementation of Head Pose and Gaze Direction Measurement. Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000, pp. 499–504, doi: 10.1109/AFGR.2000.840680.

[35] Huang, W.—Mariani, R.: Face Detection and Precise Eyes Location. Proceedings of the 15th International Conference on Pattern Recognition, 2000, Vol. 4, pp. 722–727, doi: 10.1109/ICPR.2000.903019.

[36] Valenti, R.—Sebe, N.—Gevers, T.: Combining Head Pose and Eye Location Information for Gaze Estimation. IEEE Transactions on Image Processing, Vol. 21, 2012, No. 2, pp. 802–815, doi: 10.1109/TIP.2011.2162740.

[37] Heinzmann, J.—Zelinsky, A.: Robust Real-Time Face Tracking and Gesture Recognition. Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI '97), 1997, pp. 1525–1530.

[38] Betke, M.—Mullally, W. J.—Magee, J. J.: Active Detection of Eye Scleras in Real Time. IEEE CVPR Workshop on Human Modeling, Analysis and Synthesis, 2000, pp. 1–8.

[39] McKenna, S.—Gong, S.—Würtz, R.—Tanner, J.—Banin, D.: Tracking Facial Feature Points with Gabor Wavelets and Shape Models. In: Bigün, J., Chollet, G., Borgefors, G. (Eds.): Audio- and Video-Based Biometric Person Authentication. Springer Berlin, Heidelberg, Lecture Notes in Computer Science, Vol. 1206, 1997, pp. 35–42, doi: 10.1007/BFb0015977.

[40] Pappu, R.—Beardsley, P.: A Qualitative Approach to Classifying Gaze Direction. Proceedings of the 3rd International Conference on Face & Gesture Recognition (FG '98), IEEE Computer Society, Washington, DC, USA, 1998, pp. 160–165.

[41] Bebis, G.—Fujimura, K.: An Eigenspace Approach to Eye-Gaze Estimation. ISCA 13th International Conference on Parallel and Distributed Computing Systems, 2000, pp. 604–609.

[42] Ramanauskas, N.—Daunys, G.—Dervinis, D.: Investigation of Calibration Techniques in Video Based Eye Tracking System. In: Miesenberger, K., Klaus, J., Zagler, W., Karshmer, A. (Eds.): Computers Helping People with Special Needs (ICCHP '08). Springer, Berlin, Heidelberg, Lecture Notes in Computer Science, Vol. 5105, 2008, pp. 1208–1215.

[43] Ebisawa, Y.: Improved Video-Based Eye-Gaze Detection Method. IEEE Transactions on Instrumentation and Measurement, Vol. 47, 1998, No. 4, pp. 948–955, doi: 10.1109/19.744648.

[44] International Commission on Non-Ionizing Radiation Protection: Guidelines on Limits of Exposure to Broad-Band Incoherent Optical Radiation (0.38 to 3 μM). Health Physics, Vol. 73, 1997, No. 3, pp. 539–554.

[45] International Commission on Non-Ionizing Radiation Protection: ICNIRP Statement on Light-Emitting Diodes (LEDS) and Laser Diodes: Implications for Hazard Assessment. Health Physics, Vol. 78, 2000, pp. 744–752, doi: 10.1097/00004032-200006000-00020.

[46] Hu, M. K.: Visual Pattern Recognition by Moment Invariants. IRE Transactions on Information Theory, Vol. 8, 1962, No. 2, pp. 179–187, doi: 10.1109/TIT.1962.1057692.

[47] Cerrolaza, J. J.—Villanueva, A.—Cabeza, R.: Study of Polynomial Mapping Functions in Video-Oculography Eye Trackers. ACM Transactions on Computer-Human Interaction (TOCHI), Vol. 19, 2012, No. 2, Art. No. 10, doi: 10.1145/2240156.2240158.

[48] Yoo, D. H.—Kim, J. H.—Lee, B. R.—Chung, M. J.: Non-Contact Eye Gaze Tracking System by Mapping of Corneal Reflections. Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition, 2002, pp. 101–106, doi: 10.1109/AFGR.2002.1004139.

[49] Coutinho, F.—Morimoto, C.: Improving Head Movement Tolerance of Cross-Ratio Based Eye Trackers. International Journal of Computer Vision, Vol. 101, 2013, No. 3, pp. 459–481, doi: 10.1007/s11263-012-0541-8.

Antonio García-Dopico received his M.Sc. degree in computer engineering and the Ph.D. degree in computer science from the Universidad Politécnica de Madrid (UPM), Spain, in 1993 and 2001, respectively. He is currently Associate Professor in the Department of Computer Systems Architecture and Technology at UPM. His research interests include parallel and distributed computer systems.

Antonio Pérez received his M.Sc. degree in telecomunication engineering and the Ph.D. in computer science from the Universidad Politécnica de Madrid (UPM), Spain, in 1979 and 1982, respectively. He is currently Full Professor in the Department of Computer Systems Architecture and Technology at UPM. His research interests include computer architecture, fault tolerant computers, and microprocessor systems design.

José Luis Pedraza is Associate Professor in the Department of Computer System Architecture and Technology at the Universidad Politécnica de Madrid (UPM) in Spain. He received his M.Sc. degree in electronics in 1981 from the University of Salamanca (Spain) and his Ph.D. in 1987 from the same university. His research interests include real time, image processing and embedded systems.

María Luisa Córdoba received her M.Sc. degree in computer engineering and the Ph.D. degree in computer science from the Universidad Politécnica de Madrid (UPM), Spain, in 1986 and 1996, respectively. She is currently Associate Professor in the Department of Computer Systems Architecture and Technology at UPM. Her research interests include computer architecture and high performance computing.