Univ Access Inf Soc (2009) 8:241–257
DOI 10.1007/s10209-009-0149-0

LONG PAPER

A geometric approach to remote eye tracking

Arantxa Villanueva · Gintautas Daunys · Dan Witzner Hansen · Martin Böhme · Rafael Cabeza · André Meyer · Erhardt Barth

Published online: 5 March 2009

Abstract This paper presents a principled analysis of various combinations of image features to determine their suitability for remote eye tracking. It begins by reviewing the basic theory underlying the connection between eye image and gaze direction. Then a set of approaches is proposed based on different combinations of well-known features and their behaviour is evaluated, taking into account various additional criteria such as free head movement, and minimum hardware and calibration requirements. The paper proposes a final method based on multiple glints and the pupil centre; the method is evaluated experimentally. Future trends in eye tracking technology are also discussed.

Keywords Gaze estimation · Geometric modelling · Eye model · Eye tracking

A. Villanueva (✉) · R. Cabeza
Electrical and Electronics Engineering Department, Public University of Navarre, Navarre, Spain
e-mail: avilla@unavarra.es

R. Cabeza
e-mail: rcabeza@unavarra.es

G. Daunys
Department of Electronics, Siauliai University, Siauliai, Lithuania
e-mail: g.daunys@tf.su.lt

D. W. Hansen
Technical University of Denmark/IT University, Copenhagen, Denmark
e-mail: witzner@itu.dk

M. Böhme · A. Meyer · E. Barth
Institute for Neuro- and Bioinformatics, University of Lübeck, Lübeck, Germany
e-mail: boehme@inb.uni-luebeck.de

A. Meyer
e-mail: meyer@inb.uni-luebeck.de

E. Barth
e-mail: barth@inb.uni-luebeck.de

1 Introduction

An eye tracker is a system for analysing eye movements. As the eye scans the environment or focuses on particular objects in the scene, an eye tracker simultaneously localises the eye position and tracks its movement over time to determine the direction of gaze.

Recent advances in eye tracking technology and the availability of more accurate gaze trackers have joined the efforts of many researchers working in a broad spectrum of disciplines. Duchowski [9] gives a review of tracking technologies, their applications, and the human visual system. The interactive nature of some eye tracking applications offers, on the one hand, an alternative human–computer interaction technique for activities where hands can barely be employed and, on the other, a solution for disabled people who maintain eye movement control. Individuals with disabilities are often unable to perform certain everyday tasks independently. Assistive technologies such as eye tracking can help disabled people to maintain their independence in certain key areas [5, 6]. Eye tracking has been demonstrated to be a valuable means of interaction for various groups of people with severe disabilities, including those with cerebral palsy, motor neurone disease (MND), and amyotrophic lateral sclerosis (ALS); see [5] for a survey. Controlled eye movement is maintained in advanced stages of many of these conditions. This makes eye interaction an inestimable communication

tool. Furthermore, eye tracking can be an attractive alternative even for those individuals who are still able to use other communication tools, such as a head mouse or switches; users report that, depending on the application, eye tracking can be less fatiguing and even faster than alternative forms of interaction [5, 6].

The accuracy required for gaze-based interaction depends on the precise type of application; see [5] for an overview of various types of augmentative and alternative communication (AAC) applications. Text input is possible with as little as two fields on the screen, which are used to select characters hierarchically; this type of application requires only low accuracy (5° or worse). Full on-screen keyboards require higher accuracy, around 1–2°. Interaction with standard graphical user interfaces such as Windows requires higher accuracy still, around a tenth of a degree; since this level of accuracy cannot be achieved using an eye tracker, special methods such as zooming and fish-eye views need to be used [5].

Whereas only a few years ago the standard in eye tracking was for systems to be intrusive, i.e. they either required the user's head to be fixated or equipment to be mounted on the user's head, systems have now evolved to the point where the user is allowed much more freedom in head movements while maintaining good accuracy (1° or better). For example, Electro-oculography (EOG), as illustrated in Fig. 1c, was a popular method 40 years ago. This type of system measures the potential differences at specific points of the skin around the eye via electrodes. Movements of the eye inside its orbit cause signal variations. While EOG systems produce good results, intrusiveness and lack of handling head movements are among their limitations. Bite bars, chin rests (Fig. 1a) and head-mounted eye trackers (Fig. 1b) have previously been used since they, by construction, minimise head movements relative to the camera observing the user. As no reference for head position is needed, methods that rely on fixed head positions implicitly assume that a certain movement of observed feature points (e.g. Purkinje reflexes or the centre of the pupil) corresponds to a fixed amount of eye rotation. The results obtained with these kinds of systems seem to be satisfactory when it comes to accuracy. Despite the effort involved in constructing more comfortable head mounted systems [1], less intrusive techniques are obviously desirable. The ideal in this respect would be an eye tracker with a minimal degree of intrusiveness, allowing relatively free head movement while maintaining high accuracy.

The last few years have seen the development of so-called remote eye trackers, which do not require the user to wear helmets nor to be fixated. Instead, the systems employ strategies with one or several cameras and with possible use of external light sources emitting invisible light (infrared, IR) on the user. The light sources produce

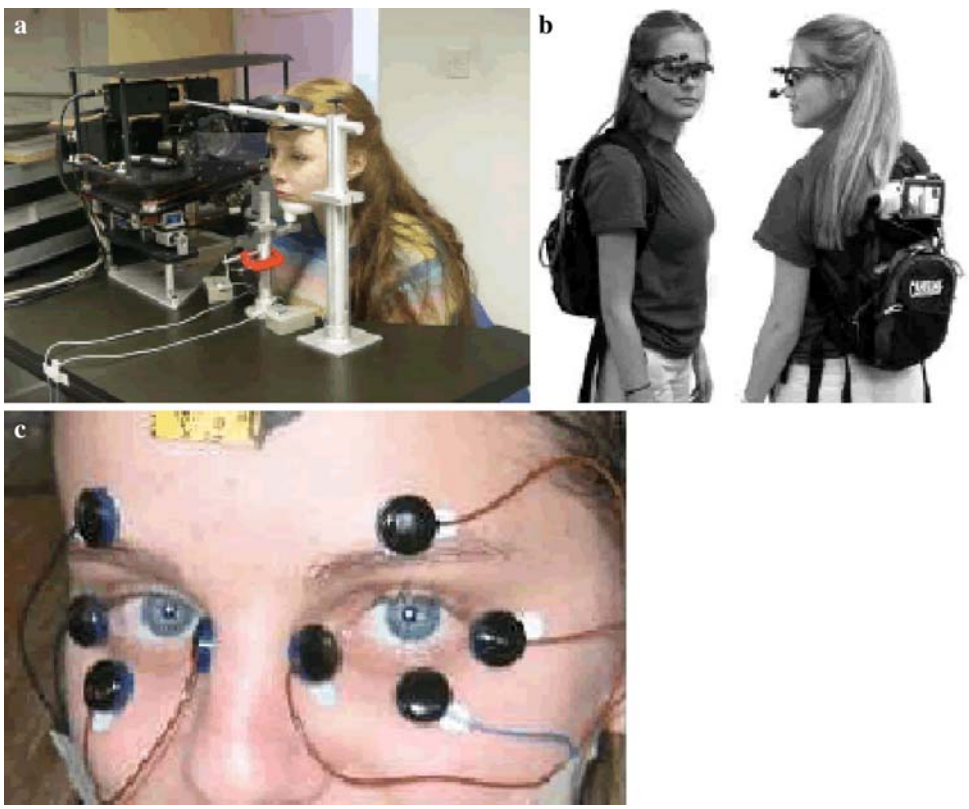

Fig. 1 Different types of eye tracking systems. **a** System with head fixation using a chin rest. **b** Head-mounted eye tracking system. **c** EOG system. Images courtesy of Fourward Technologies, Visual Perception Laboratory at the Rochester Institute of Technology, and Metrovision

stable reflections on the surface of the eye, which are observable in the images.

The first remote eye tracking systems that appeared in the literature used multiple cameras [3, 4, 25, 29, 34], usually in some kind of stereo setup. Many of these systems use a pan-and-tilt mechanism to keep a narrow-field-of-view camera aimed at the eye when the head moves.

A more recent trend in remote eye tracking has been to use a single fixed high-resolution camera with a wide field of view. The advantage of this is that the cost and size of the system is reduced, because no pan-and-tilt mechanism is required and only one camera is used. The first system of this kind was the commercial Tobii eye tracker [31]; it allows a head movement volume of 30 × 15 × 20 cm and a maximum head motion speed of 15 cm/s. Recently, several academic groups have built similar single-camera systems [11, 13, 20]. Guestrin and Eizenman's system [11] allows only small head movements, but it appears that their well-founded approach would allow larger head movements if the resolution of the camera was higher.

In this paper, a principled geometrical analysis of the eye tracking problem is conducted, focussing on eye trackers based on video (a.k.a. *video occulography*), with a special emphasis on eye trackers that extract features such as the reflections and centre of the pupil for gaze estimation. The underlying objective is to explore the geometry of the situation without assuming specific hardware and image analysis algorithms and to find the minimum set of image features necessary for remote eye tracking, in order to minimise hardware and software requirements, and thus make eye tracking accessible to as many potential users as possible. The paper also presents an implementation of a remote eye tracker based on the identified minimal set of features.

1.1 Eye tracker components

An eye tracker consists of several parts, and a general overview of these is provided in Fig. 2. A video-based eye tracker obtains its information from one or more cameras (ImageData).The first step of an eye tracker is to find the initial eye position (*Detection* component) in the images. The position is used for initializing the *Eye tracking* component, which, in turn, aims at following the eye overtime. Based on information obtained from the eye region and possibly head pose, the *Gaze estimation* component will then determine where the user is looking. This information is then used in the gaze-based application.

In summary, a connection must be found between the captured images of the subject's eye and gaze direction. This paper proposes a study of various geometrical models for gaze estimation based on point features such as pupil centre and glints in a single-camera system. The objective

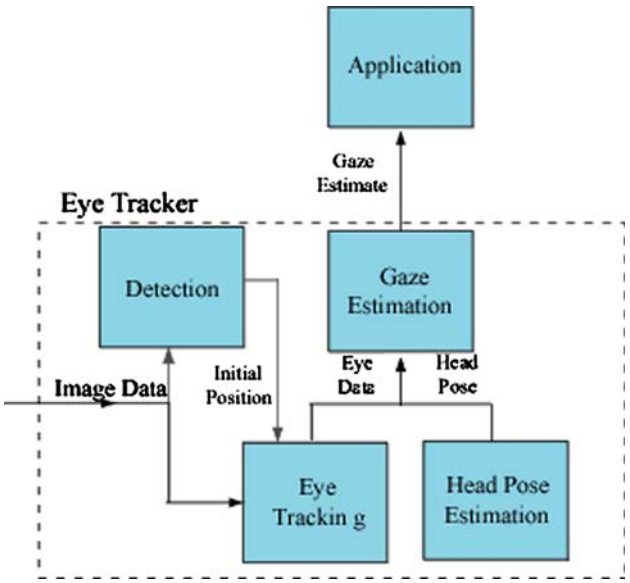

Fig. 2 Components of video-based eye trackers

is to review all possible feature combinations and to evaluate whether the resulting models can be used to estimate gaze and, if so, what their requirements are in terms of setup and calibration and whether the user is allowed to move the head.

Section 2 reviews basic theory about gaze estimation. In Sect. 3, the model for the eyeball is proposed. Based on purely geometrical principles, a set of models for gaze estimation are proposed and evaluated; for those readers who wish to skip the mathematical details, Sect. 3.5.5 summarises the properties of the various models. An implementation of the model that was found to be the most suitable for remote eye tracking is described in Sect. 4. The conclusions and discussion of the topic are presented in Sect. 5.

2 Fundamentals of gaze estimation

The usual components of video-oculography systems are illumination sources (often IR) and one or more cameras. Although the basic components are the same in most systems, this general setup admits multiple variations. The type of illumination sources and cameras, their quantity and location result in different properties of the image data, and consequently the algorithms employed change. Early VOG systems are discussed by Duchowski [9] and Young and Sheena [35].

For the purpose of gaze estimation, the point where the subject is looking needs to be inferred given the image data (i.e. from the tracker). The 3D direction of gaze is defined as line of sight (LoS) and the 2D point at which the user is looking (i.e. the intersection of 3D direction of gaze with a

2D surface) as point of regard (PoR). Both of them are referred to in this paper as *gaze*. In other words, a mapping described by parameters **c**:

$$\Phi_{\mathbf{c}} : \mathbb{R}^m \rightarrow \mathbb{R}^3$$

from an m-dimensional feature space to world coordinates is sought. For screen-based applications, the output domain of Φ is a subspace $\Omega \subseteq \mathbb{R}^2$. Clearly, if a 3D direction vector is required for the application, the camera system (either a single camera or a stereo rig) needs to be calibrated. To avoid confusion with camera calibration, the process of gathering data and finding the parameters **c** of the transformation Φ is called *gaze calibration*. The parameters vector **c** may contain parameters related to the system hardware, coefficients of a polynomial expression and human specific variables. Gaze calibration is usually performed by assuming that the user looks at N predefined points (target values) t_i on the screen, while relating these to calculated features of the eye x_i. From the set of tuples $D = \{(t_i, X_i)\}_{i=1}^{N}$ the mapping $\Phi_{\mathbf{c}}$ should be inferred. Calibration of human specific parameters in **c** is called *subject specific calibration*. Notice that some methods require additional knowledge of head pose estimation. The process of gaze calibration is shown in Fig. 3.

The input feature vectors for Φ may be the centre of the iris or higher-dimensional feature coordinates such as the image region of the eye. Different approaches can be proposed to establish the formal relationship between the image and gaze, i.e. $\Phi_{\mathbf{c}}$. If a general-purpose expression is proposed, such as quadratic or cubic expression, the set of parameters **c** will consist of unknown coefficients. On the other hand, if a geometrical model is used, **c** could involve parameters related to the tracker configuration, such as screen position or camera intrinsic parameters among others, and subject-specific variables such as corneal radius or head position.

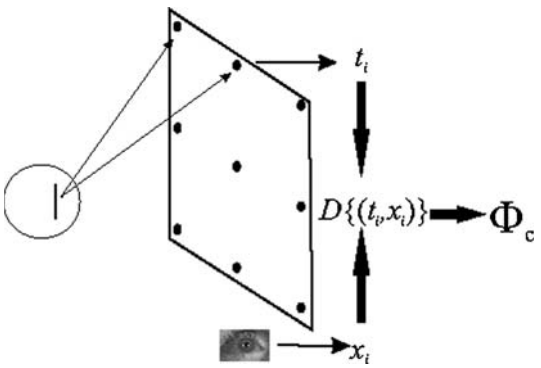

Fig. 3 The calibration consists of asking the subject to gaze at specific marks on the screen, allowing for inference of the mapping function $\Phi_{\mathbf{c}}$

Obviously, the more calibration points are used, the better are the chances to be able to infer the mapping from the image to the monitor. It would even be possible to sample the entire function space given a sufficient number of calibration points. However, a complex and time-consuming calibration is tedious and may cause discomfort for the user. As long as accuracy is maintained, it is preferable to use as few calibration points as possible. The ideal system would therefore be one that does not need calibration of either the user or of the parameters of the system geometry, that is head pose invariant, fast and has high accuracy. Such a system may be difficult to obtain, since several of these constraints are conflicting. To the authors' knowledge, all current eye trackers with high accuracy (around 1° or less) need calibration of either the geometry or the user.

2.1 Gaze estimation methods

Several conflicting issues have to be solved when trying to estimate gaze: accuracy, restricted head movement, robustness, and ease of calibration. Very good accuracies for gaze estimation may be obtained when the head is fixed or an excessive amount of calibration data is used. Other methods handle head movements, but require additional information, such as knowing the distance between the camera and the eye or using several cameras.

Methods for gaze estimation often use an eye model to make the predictions and can be divided into feature-based and appearance-based methods. There are two main strategies for making the prediction. One relies on geometric information and the other is based on statistical learning principles.

2.1.1 Feature-based gaze estimation

Feature-based methods use features such as contours, eye corners, and reflections for gaze determination. IR-based eye trackers primarily use feature-based methods, since the centre of the eye and the glint (reflection) are easily obtained [10, 14, 15, 21].

A classical feature-based approach uses the assumption that the vector between the corneal reflection and the pupil centre in the image only changes with eye movements but remains constant with minor head movement [9]. The Dual-Purkinje-Image techniques are accurate [7, 23], but require highly controlled light conditions in addition to specialised hardware. For a single camera and a single light source model, Morimoto et al. [21] proposed a geometric approach utilising two-second-order polynomials to represent the mapping of the glint-pupil vector to the screen coordinates. The method required the user's head to be in a relatively fixed position. Stereo cameras can be used for

locating the 3D position and 3D gaze direction of the eye [24, 29]. Multiple cameras can also be used by assigning one camera for eye tracking and one or more for head pose estimation [3, 4]. Several methods are able to compensate for significant head movements while maintaining a high accuracy [16, 17]. Unfortunately, however, high accuracy comes at the expense of modelling the geometry of light sources, camera and user. In turn, this means complicating the system with calibration procedures to obtain metric information or added cost due to several cameras or specialised hardware. Rather than using several cameras, one may opt to use several light sources.

2.1.2 Appearance-based gaze estimation

The feature extraction process may be prone to errors. The appearance-based methods do not explicitly extract features, but use all the image information as input. Therefore, the dimensionality of the input space is much higher than for feature-based methods. Traditional approaches are neural networks or manifold learning [26, 30] on cropped images. These approaches often need a large set of calibration points to be successful.

3 Eye model and algorithms for gaze estimation

The main objective of this study is to shed light on the connection between image data and gaze (PoR/LoS) based on geometrical modelling. This paper focusses on feature-based gaze estimation where the objective is to derive models that relate point features extracted in the image (i.e. centre of pupil, reflections) to gaze coordinates (PoR/LoS) on the screen through a model of the eye. This encompasses aspects related to the geometry of the system, as well as eye physiology.

The desired properties of a gaze estimation method are the following: it should allow free head movement, have minimum hardware requirements (minimum number of cameras and lighting sources), involve a minimum of calibration and require a minimal number of image features. The objective of this study is to analyse mathematical models based on point features in the image that describe the connection between the point an individual is looking at on the screen and the rest of the elements of the system, e.g. head location, eye dimensions, parameters of the camera, as well as the features extracted from the captured images of the eye.

To this end, the analysis starts with the simplest features of the image. The significant points in the image are the centre of the pupil and the glint or glints produced by one or more light source(s). In the following, the estimation of these features is described in two ways, an accurate approach and a simplified approach. Then a set of algorithms for gaze estimation is presented.

This section describes the eye model which forms the basis for the gaze estimation algorithms. The involved geometric entities are introduced, with a discussion of how these relate to each other in a 3D setting.

Entities defined in a 3D geometric setting and those defined in the image and other local coordinate systems are distinguished.

Generally, an eye tracking system consists of one or more cameras $\{\mathbf{C}^i | i = 1,\ldots, N_c\}$ with corresponding projection matrices $\mathbf{P}^i$, a set of light sources $\{\mathbf{l}^j | i = 1,\ldots, N_l\}$ and a set of objects the viewer can gaze at, all defined in a common world coordinate system. The model discussed here only considers single camera setups (removing the superscript), a variable number of light sources (to be specified later) and a single gaze target object, a planar screen with origin $\mathbf{o}_{\text{mon}}$. Also, to reduce system complexity and ease camera calibration, it is assumed that the single camera has a fixed focus; this means that techniques such as depth-from-focus cannot be used to determine the 3D position of points in the image.

The geometrical analysis of the gaze tracking system uses three coordinate systems. The main coordinate system, i.e. the world coordinate system to which the rest of the elements are referenced, is the screen system $\mathbf{o}_{\text{mon}}$, a left-handed Cartesian coordinate system. By a convention that is common in computer graphics, the origin of the screen system lies at the top left corner of the screen, and its X axis runs from left to right while the Y axis runs from top to bottom. The Z axis is perpendicular to the monitor plane and points towards the user. The second coordinate system is the eye system; its origin is the centre of rotation $\mathbf{o}_{\text{e}}$ of the eye, and in the rest position of the eye, its axes are parallel to those of the screen system. A third system is used for the camera; the origin of this system is the projection centre $\mathbf{o}_{\text{cam}}$. XY plane of the system is image plane. Furthermore, coordinates in the image plane are referred to as image coordinates.

3.1 Eye model

The human eye is an organ of the visual system that is sensitive to light. Externally, the eye looks like a ball, so it is also called "eyeball" (Fig. 4). The cornea, which covers the front of the eyeball, is transparent to light. Inside the eye, light is focused by a lens onto the retina, a layer of light-sensitive photoreceptors at the rear of the eyeball. The human visual system is foveated, meaning that the distribution of photoreceptors in different parts of retina is unequal. Most photoreceptors are concentrated in a small region, the so-called fovea. This foveation is very important in gaze-based interaction, because the vision system

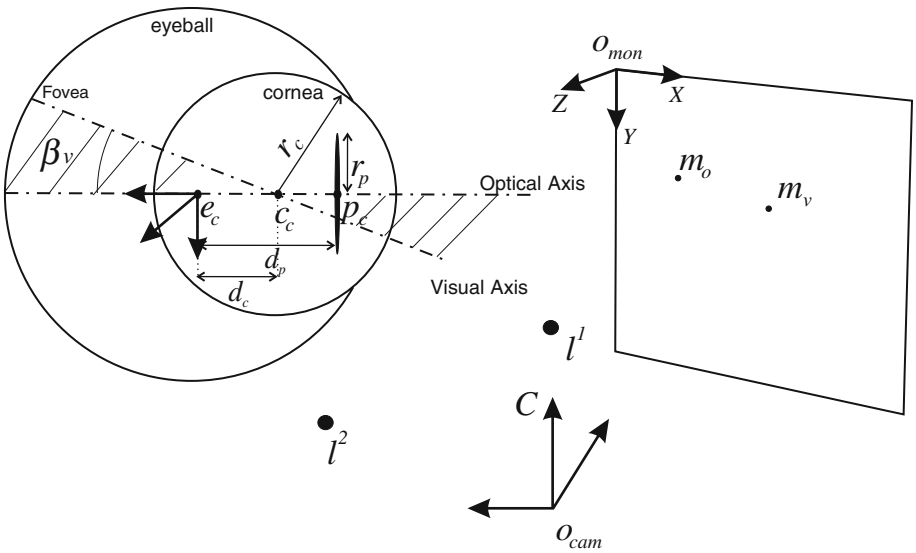

Fig. 4 Eyeball model

tends to orientate the eye so that important objects are projected onto the fovea. The fovea subtends a visual angle of about 1.2°; at a viewing distance of 50 cm, this corresponds to a circle with a diameter of 10.5 mm or, as a rule of thumb, the width of a thumbnail at arm's length.

The external geometry of the eyeball [33] is modelled as two spheres, the main eyeball sphere and the corneal sphere, with centres in different positions and radii of different sizes. The centre of the eyeball sphere is considered to be the centre of rotation of the eye, denoted by $\mathbf{e}_c$. The cornea is bounded by an inner and an outer spherical surface. The population average for the radius r_c of the external cornea surface is 7.8 mm; the centre of this external surface is denoted by $\mathbf{c}_c$. The internal surface has a radius of 6.5 mm, and its centre is offset slightly with respect to $\mathbf{c}_c$. This causes the thickness of the corneal surface to vary between the centre and the periphery. The refractive index for the cornea is about $n_c = 1{:}376$.

The pupil and iris are important elements of the eyeball. The iris is a pigmented circular muscle having a planar shape. The central aperture of the iris is the pupil. The anterior chamber is a cavity between the cornea and iris; it is filled with the aqueous humour, which has a refractive index $n_{ah} = 1{:}336$ very close to that of the cornea. Because the refractive indices of both media are very close, in the further analysis the internal boundary between aqueous humour and cornea is neglected. The combined effect of the anterior part of the eye up to the iris is approximated by modelling the cornea as a single spherical surface with radius r_c and refractive index $n = 1{:}3375$ (this approximation is taken from [11]).

The eye coordinate system has its origin at the centre of rotation of the eye, i.e. $\mathbf{o}_e = \mathbf{e}_c$. In this coordinate system, the cornea is thus centred at $\mathbf{c}_c^{(E)} = (0, 0, -d_c)^T$, where d_c is the distance between the corneal and eyeball centres. Consequently, $\mathbf{c}_c$ can be expressed in world coordinates as:

$$\mathbf{c}_c = \mathbf{e}_c - d_c\mathbf{k}', \tag{1}$$

where $\mathbf{k}'$denotes the unit vector of eye coordinate system Z axis in world coordinate system. Its orientation depends on eye rotation angles. The distance of the iris plane from $\mathbf{e}_c$ is denoted by d_p. The coordinates of the pupil centre in the eye coordinate system are thus $\mathbf{P}_c^{(E)} = (0, 0, -d_p)^T$. The position of the pupil centre in world coordinates is then:

$$\mathbf{P}_c = \mathbf{e}_c - d_p\mathbf{k}' \tag{2}$$

The pupil has a variable radius r_p.

The fovea centre has coordinates f_c. In the adult eye, the fovea typically lies about 4–5° temporally and 1.5° below the point where the optical axis intersects the retina. In the eye model adopted in this paper, the visual axis (LoS) is assumed to be the line connecting the fovea centre to the corneal centre, whereas the optical axis is considered to lie on the symmetry axis of the eye. As a consequence of the fovea position, there exists an angular offset between the visual and optical axes of the eye. This angular offset is modelled by means of a horizontal angle β_v and a vertical angle, α_v. The optical and visual axes intersect the screen in the points $\mathbf{m}_o$ and $\mathbf{m}_v$, respectively.

The eye model used in the analysis is summarised in Fig. 4.

3.2 Mathematical model for the image coordinates of the pupil centre

Eye rotations are described in the monitor coordinate system. Rotation about the X axis is described by an angle θ. Rotation about the Y axis is described by an angle ϕ.

Positive rotation is counterclockwise. Rotation matrices can be used to compute the coordinates of points in the eye after eye rotation. The rotation about the X axis by an angle θ is described by the matrix $\mathbf{R}_x$. The rotation about the Y axis by an angle ϕ is described by the matrix $\mathbf{R}_y$. Assuming that the rotation about the X axis is carried out first, followed by the rotation about the Y axis, the following rotation matrix is obtained:

$$\mathbf{R} = \mathbf{R}_y \times \mathbf{R}_x. \tag{3}$$

The individual entries of the matrix are as follows:

$$\mathbf{R}(\phi,\theta) = \begin{pmatrix} \cos\phi & -\sin\phi\sin\theta & \sin\phi\cos\theta \\ 0 & \cos\theta & \sin\theta \\ -\sin\phi & -\cos\phi\sin\theta & \cos\phi\cos\theta \end{pmatrix}. \tag{4}$$

Consequently, any point defined in the eyeball coordinate system can be transformed to screen coordinates using the following homogeneous transformation matrix:

$$\mathbf{T}(\phi,\theta) = \begin{pmatrix} \mathbf{R}(\phi,\theta) & \mathbf{e}_c \\ 0 & 1 \end{pmatrix}. \tag{5}$$

The third column of matrix (4) represents the direction of $\mathbf{k}'$, which points in the opposite direction of the optical axis of the eye. The optical axis of the eye can be expressed as a function of the eye centre and the point where the optical axis intersects the screen, i.e. $\mathbf{m}_o - \mathbf{e}_c = (-\Delta X, -\Delta Y, Z)$, where ΔX, ΔY denote the changes in the X and Y coordinates during rotation. These coordinate changes have a negative sign because rotation by positive angles causes the coordinates to decrease. Consequently, the rotation angles can be expressed as a function of the screen point $\mathbf{m}_o$:

$$\phi = \arctan\left(-\frac{\Delta X}{Z}\right); \tag{6}$$

$$\theta = \arctan\left(-\cos\phi\frac{\Delta Y}{Z}\right). \tag{7}$$

However, the PoR to be estimated is the point $\mathbf{m}_v$ where the visual axis intersects the screen, and not $\mathbf{m}_o$ where the optical axis intersects it. It is assumed that the orientation of the visual axis can be obtained from the optical axis by a rotation, described by the angles ϕ_V and θ_V. Such a rotation can be described by a matrix in the shape of equation (4) with parameters ϕ_V and θ_V, yielding the matrix $\mathbf{R}_V$ (ϕ_V and θ_V). From the rest position of the eye, where the optical axis is aligned with the Z axis of the screen system, the visual axis can be rotated into the Z axis of the screen system by applying the inverse matrix $\mathbf{R}_V^{-1}$. From such an initial orientation, a rotation by angles ϕ and θ will produce PoR shifts of ΔX and ΔY. The final orientation of the vector $\mathbf{k}'$ is obtained by the following equation:

$$\mathbf{k}' = \mathbf{R} \times \mathbf{R}_V^{-1} \times \mathbf{k}, \tag{8}$$

where the vector $\mathbf{k}$ is the orientation of the optical axis in the rest position of the eye. Now, the orientation $\mathbf{k}'$ obtained above can be inserted into (1) and (2) to find $\mathbf{c}_c$ and $\mathbf{p}_c$.To obtain an eye image, it is necessary to project 3D eye points onto the image plane. To do this, the orientation of the camera with respect to the screen needs to be known. The camera orientation is described by angles ϕ_C and θ_C, obtaining again a rotation matrix $\mathbf{R}_C$ in the form of equation (4) with parameters ϕ_C and θ_C. The projection of a point in screen coordinates onto the image plane of the camera can be calculated using the homogeneous transformation

$$\mathbf{W} = \mathbf{P} \times \begin{pmatrix} \mathbf{R}_C & \mathbf{o}_{cam} \\ 0 & 1 \end{pmatrix}, \tag{9}$$

where $\mathbf{P}$ is the projection matrix of the camera.

3.2.1 Accurate approach

The pupil, as imaged by the camera, is viewed through the cornea and aqueous humour. As both media have a refractive index that is noticeably different from that of air, the image of the pupil is affected by refraction. As mentioned in Sect. 3.1, the combined effect of the cornea and aqueous humour can be modelled using an effective index of refraction of 1.3375.

To account for these effects, the 3D propagation of light rays is modelled. The surface of the cornea is modelled as a refractive surface using the law of refraction, which states the following:

- The incident ray, the refracted ray and the normal at the point of refraction are in the same plane.
- The angle of incidence α_1 and the angle of refraction α_2 satisfy Snell's law $n_1 \sin\alpha_1 = n_2 \sin\alpha_2$.

Refraction significantly complicates the calculations since it cannot be described exclusively by means of linear operations. For each point on the pupil contour, an outgoing ray needs to be determined that is refracted at the cornea in such a way that it reaches the camera.

The range of possible initial ray directions can be restricted to a plane using the first of the two conditions stated above. Within this plane, the outgoing ray that is refracted directly onto the camera's centre of projection is then determined. In the same manner, the effect of refraction on the pupil centre can be estimated, and its projection onto the image plane calculated.

In order to obtain the pupil image, each point belonging to the pupil circumference [27] in the eyeball reference system, i.e. $(r_p \cos\vartheta, r_p \sin\vartheta, -d_p)$, $\vartheta = 0\ldots2\Pi$, can be projected onto the camera by means of refraction and

projection algorithms. Then, the pupil image can be treated as an ellipse, and its centre can be calculated geometrically as the centre of this ellipse.

3.2.2 Simplified approach

In addition to the full eye model presented above, a simplified model has also been elaborated. This model starts by eliminating corneal refraction. Consequently, any point belonging to the pupil circumference [27] in the eyeball reference system, i.e. $(r_p \cos\vartheta, r_p \sin\vartheta, -d_p)$, $\vartheta = 0\ldots2\Pi$, can be projected on the camera using the matrix $\mathbf{W}$.

Projecting the pupil circumference in this way yields an ellipse in the image plane. The centre of this ellipse is taken to be the centre of the pupil $\mathbf{P}'_{c_{im}}$. The position of this feature in the image depends not only on the point on the screen the pupil is directed at, i.e. $\mathbf{m}_o$, but also on the location of the user's head, the parameters of the eye, the location of the screen with regard to the camera projection centre, and the intrinsic parameters of the image acquisition system.

The expressions obtained for $\mathbf{P}'_{c_{im}}$ in the most general case are involved. In order to reduce the already existing complexity of the model, it is proposed that the centre of the projected pupil can be approximated by the projection of the centre of the pupil, i.e. projection of the point $\mathbf{P}_c$ in the eyeball coordinate system onto the image plane. This point is denoted by $\mathbf{P}_{c_{im}}$. One might assume that both points, i.e. the centre of the projected pupil $\mathbf{P}'_{c_{im}}$ and projection of the centre of the pupil $\mathbf{P}_{c_{im}}$ are the same point; however, because of the foreshortening that occurs during perspective projection, this is not true (In the case of an orthographic projection, the two points would actually coincide.).

This approximation—identifying the centre of the pupil in the image with the point $\mathbf{P}_{c_{im}}$—is popular in the eye tracking literature and can be justified from several points of view. Both points exhibit similar behaviour and the same symmetrical properties, and both behave similarly when the alternative parameters change their values. Also, because the size of the pupil disc is small compared to the distance between the pupil and the camera, the perspective projection can be approximated well by an orthographic projection (this approximation is also referred to as a *weak perspective* model). The projection of the pupil centre, i.e. $\mathbf{P}_{c_{im}}$, is easily calculated as the projection of the 3D point onto the image plane. Therefore,

$$\mathbf{P}_{c_{im}} = \mathbf{W} \times \mathbf{P}_c. \tag{10}$$

One of the most important characteristics to strive for in a model should be simplicity. The advantage of working with the expression $\mathbf{P}_{c_{im}}$ for the centre of the pupil is that the resulting model is much simpler than the equations deduced for $\mathbf{P}'_{c_{im}}$. In addition, the calibration process that precedes the tracking session can be assumed to reduce the errors caused by this approximation via a fine adjustment of the parameters of the model [32].

3.3 Mathematical model for corneal reflection

Light sources $\mathbf{l}_i$ illuminating the user's eye may produce reflections on the cornea. These reflections, as well as their projections onto the camera image plane, are referred to as glints. The camera image coordinates of the glint produced by light source $\mathbf{l}^i$ are denoted by $\mathbf{g}'^i_{im}$.

The corneal surface is assumed to be a perfect mirror, and thus the light is reflected in only one direction. This eliminates the need for deriving the location of the glint as the centre of gravity of various pixels (In fact, this model is not quite correct since part of the light is refracted on the corneal surface and thus changes the observed glint.). The location of the glint depends on the location of the cornea centre, the corneal radius, the relative orientation of the eyeball with respect to the camera, the light source location, and the camera parameters.

3.3.1 Accurate approach

The simplest case for computing the position of the corneal reflection is when the light source is located on the optical axis of the camera, as in bright-pupil systems [10, 15, 22]. A bright-pupil system uses on-axis illumination, which causes the pupil to appear as a bright disc in the image due to the light reflected off the retina. A dark-pupil system uses off-axis illumination, making the pupil appear as a dark, almost black disc (see also [9]).

To calculate the position of the virtual image, it is assumed that the incident rays are parallel. The continuations of the reflected rays in the opposite direction intersect in one point; this point is the virtual image of the light source. Let us consider a coordinate system xyz with the origin at the corneal centre and oriented such that the xyz plane contains the incident and reflected rays (In general, this coordinate system will be different from the eyeball coordinate system introduced earlier.). Let (y, z) be the point where the incident ray intersects the cornea (see Fig. 5), then the angle of incidence á can be expressed as

$$\sin\alpha = \frac{y}{r_c}. \tag{11}$$

The z coordinate can be expressed as

$$z = r_c \cos\alpha. \tag{12}$$

Because the continuation of the reflected ray intersects with the y axis at an angle of 2α, the virtual image is shifted

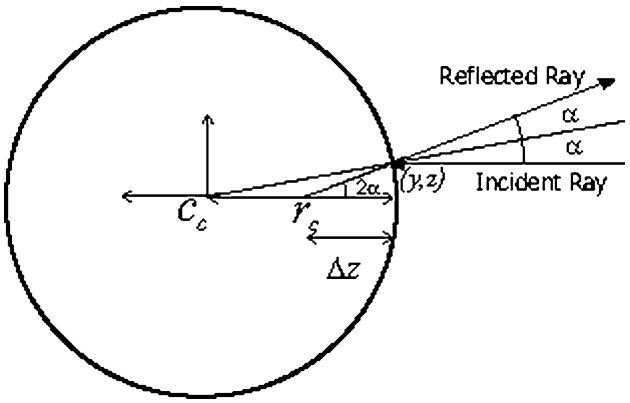

Fig. 5 Virtual image of corneal reflection

towards the centre of corneal curvature by a distance Δz, which can be approximated as follows:

$$\Delta z \simeq \frac{y}{\tan 2\alpha}. \tag{13}$$

From (11), the following is obtained

$$y = r_c \sin \alpha. \tag{14}$$

Substitution into (13) yields

$$\Delta z = \frac{r_c \sin \alpha}{\tan 2\alpha}. \tag{15}$$

If α is small, then

$$\begin{cases} \sin \alpha \approx \alpha \\ \cos \alpha \approx 1 \\ \tan 2\alpha \approx 2\alpha \end{cases}. \tag{16}$$

By substituting (16) into (15), the following is obtained

$$\Delta z \approx \frac{r_c}{2}. \tag{17}$$

As a consequence, the virtual image of the corneal reflection is at the midpoint between the corneal curvature centre and the closest point to the light source on the surface of the cornea. This result can be applied to those configurations for which α is small, i.e. close to the coaxial location of the light source. If the illumination is not coaxial, the approximation proposed for the glint is not valid. For this case, an alternative algorithm is used to determine the position of the glint in the image.

Given a light source with coordinates $\mathbf{l}^i$, a ray needs to be found that strikes the cornea and is reflected to reach the camera in its centre of projection. The path of the ray is governed by the law of reflection; to apply it, however, the point on the cornea surface where the reflection takes place needs to be known. To find this point, first equation (17), the solution for the coaxial case, is used to obtain a good approximation for the point of reflection. Starting from this point, a numerical minimisation technique is used to minimise the error that occurs in the law of reflection and thereby find the precise point where the reflection takes place. For this purpose, a non-linear Levenberg–Marquardt algorithm is used [19].

3.3.2 *Simplified approach*

In the same manner as was done for the accurate approach, two possibilities are reviewed for the simplified approach: coaxial and non-coaxial location of the lighting with respect to the camera.

First, a coaxial location for the LED is selected. In this case, because the corneal centre is close to the position of the glint in the image, it is reasonable to approximate the glint by the projection of the corneal centre $\mathbf{c}_c = (0, 0, -d_c)$ onto the image, called $\mathbf{g}^0_{\text{im}}$:

$$\mathbf{g}^0_{\text{im}} = \mathbf{W} \times \mathbf{c}_c. \tag{18}$$

The same reasoning that was used for the pupil centre approximation can be used to justify the approximation for the glint. Therefore, the model proposed assumes that $\mathbf{g}'^0_{\text{im}} \simeq \mathbf{g}^0_{\text{im}}$. If the coaxial location for the LEDs cannot be assumed, the approximation proposed for the glint is not valid, i.e. the glint cannot be approximated by the projection of the corneal centre. So new analytical expressions must be found for the glint in the image. Assuming that the corneal surface is a specular reflector, the law of reflection says that the illumination source, the incident and reflected rays and the normal vector on the surface of reflection at the point of incidence are coplanar, as shown in Fig. 6 [28].

It is straightforward to deduce that the centre of the cornea is contained in the same plane, since the surface normal, which lies within the plane, crosses $\mathbf{c}_c$.

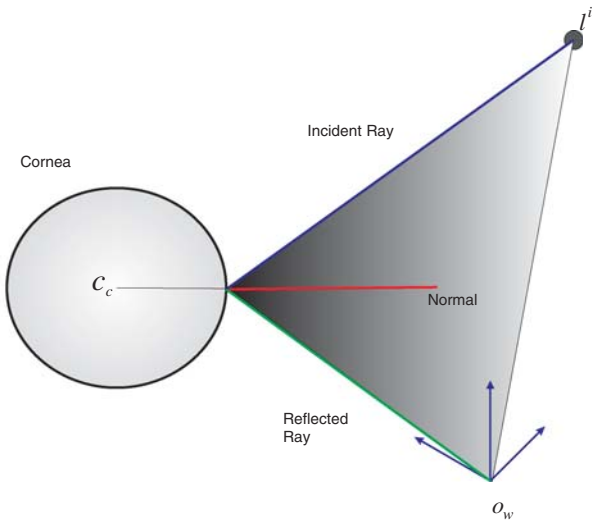

Fig. 6 Incident ray, reflected ray and normal at incidence point are coplanar

For this case, the position of the glint in the image can be deduced using the law of reflection. Every ray produced by the LED that hits the corneal surface is reflected. However, there is only a single ray, and consequently a single point on the cornea, where the reflection will produce the glint in the image; this will be the reflected ray containing the projection centre of the camera $\mathbf{o}_{\text{cam}}$. If one knows the position of the LED, which does not underlie any constraints, and the cornea ($\mathbf{c}_c$ and r_c), the desired glint position in the image can be determined.

3.4 Gaze detection algorithms based on the simplified model

The final objective is to derive an algorithm that estimates the user's gaze direction, i.e. the visual axis of the eye. This can be derived from the optical axis if the angular offset between both axes is known, as explained above. In the eyeball system, the Z axis is the optical axis of the eye. If the orientation of this coordinate system is known, the visual axis can be estimated as a function of the optical axis and the eyeball parameters. Consequently, the objective for any gaze estimation algorithm is first to determine the optical axis of the eye. The optical axis is a 3D line containing the centre of the eyeball $\mathbf{e}_c$, the corneal centre $\mathbf{c}_c$, and the pupil centre $\mathbf{p}_c$ and thus knowing either two of these, the optical axis can be determined as shown in Fig. 4.

At this point it is necessary to make clear two assumptions made in this simplified approach, which result from properties of the model established in previous sections:

- Refraction at the cornea is not taken into account for the algorithms based on the simplified approach. Once a feature or set of features is selected, the corresponding algorithm will be constructed and then, by means of geometrical relations, a set of conclusions will be extracted. Since refraction will surely modify the results obtained and will add new limitations to the model, a distinction needs to be made between the results and limitations of an algorithm that are due to the geometrical and projective relations involved, and those that are due to refraction. In other words, if an algorithm is proposed that is limited in some way or is underdetermined from a geometrical point of view, there is no point in introducing refraction into the algorithm, since further limitations are bound to arise. Refraction will be included in those algorithms that satisfy the geometrical analysis.
- The hardware of the tracker, lighting, screen and camera are assumed to be calibrated, and their positions with respect to the origin are assumed to be known.

3.5 Proposed algorithms

At first sight, based on the features obtained, five algorithms can be proposed:

- Algorithm based on the centre of the pupil.
- Algorithm based on the glint.
- Algorithm based on multiple glints.
- Algorithm that combines the centre of the pupil and the glint (PCCR).
- Algorithm that combines the centre of the pupil and multiple glints.

Figure 7 depicts these algorithms.

A more careful analysis leads to the conclusion that the first two algorithms can be regarded as equivalent. In other words, the first algorithm is based on the location of the pupil in 3D space, whereas the glint position is a consequence of the position of the cornea in 3D space. There is a virtual line connecting three fundamental points of the eyeball, i.e. the eyeball centre $\mathbf{e}_c$, corneal centre $\mathbf{c}_c$ and pupil centre $\mathbf{p}_c$, which represents the optical axis of the eye. From a geometrical point of view, it makes no difference whether one determines the 3D location of the corneal centre or that of the pupil centre because they both lie on the same 3D line, the optical axis of the eye. In short, the features represent the projections of two points that move jointly on the same 3D line. Therefore, the results obtained

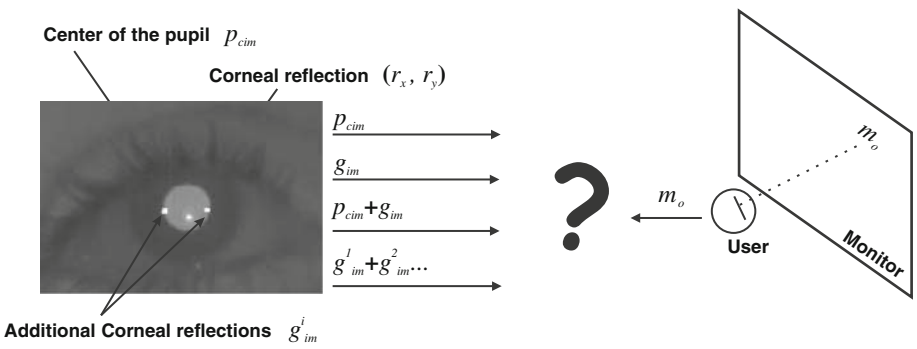

Fig. 7 Proposed image features and algorithms

from the analysis of one of the algorithms can straightforwardly be carried over to the other.

From an implementation point of view, there are of course differences between the two points, the corneal centre has the disadvantage that it is harder to measure because it cannot be observed directly, whereas the pupil has the disadvantage that changes in pupil size can cause the pupil centre to shift. However, these effects are not considered in the proposed geometrical derivations.

Therefore, the list of algorithms can be reduced to the following:

- Algorithm based on the centre of the pupil or the glint.
- Algorithm based on multiple glints.
- Algorithm that combines the centre of the pupil and the glint (PCCR).
- Algorithm that combines the centre of the pupil and multiple glints.

The objective is to analyse each algorithm in order to determine its limitations with respect to head movements, its hardware requirements, and the set of parameters that need to be calibrated. The number of parameters (unknowns) of the model is closely connected to the number of calibration points needed. The main focus of this paper is not on calibration procedures for gaze tracking systems, but on geometrical models for determining the LoS or PoR. However, it is assumed that a subject-specific calibration is carried out for each algorithm to determine the necessary unknowns. Consequently, the parameters that have to be calibrated for each algorithm are named.

For those readers who wish to skip the technical details, the main properties of the various algorithms are summarised in Sect. 3.5.5.

3.5.1 *Algorithm based on the centre of the pupil or the glint*

The image features considered for this algorithm are either the pupil centre or the glint, i.e. $x = \{\mathbf{P}_{c_{im}}\}$ or $x = \{\mathbf{g}^0_{im}\}$. Of the two, the pupil centre will be investigated in detail. From (10), the formal relationship between $\mathbf{P}_c$ and $\mathbf{P}_{c_{im}}$ is obtained. This equation permits one to estimate the position of the pupil centre, provided that the model parameters contained in $\mathbf{W}$, i.e. $\mathbf{e}_c$, and d_p are known.

The most outstanding characteristic is that the algorithm based on the centre of the pupil depends on the eyeball centre position $\mathbf{e}_c$. In other words, it does not allow for free head movement, and in its most general configuration it is an algorithm based on six unknown parameters ($\mathbf{e}_c$, d_p, β_v, and α_v) since the hardware setup is already calibrated as explained above.

This algorithm would permit several fixed positions for the eyeball $\mathbf{e}_c$. If a centred $\mathbf{e}_c$ position with respect to the camera is assumed, the number of model parameters is reduced since just the distance of this point to the camera is unknown.

Similar results would have been obtained for the glint algorithm. The analysis can be extended further to reduce the number of parameters used, but the head movement constraint cannot be removed; this makes the centre of the pupil (or, alternatively, the glint) unsuitable as a single feature on which to base a tracking algorithm. The model for the general setup permits the equations to be adjusted for different head locations (by changing $\mathbf{e}_c$), but once the system is calibrated for a certain head position, the algorithm is only guaranteed to perform well if the same head position is maintained, i.e. if the conditions that were used during calibration are not changed.

3.5.2 *Multiple glints*

In this section, the approximation that was derived for the case of a coaxially located LED is not considered, since more than one illuminator is used and the approximation $\mathbf{g}'^i_{im} \simeq \mathbf{g}^i_{im}$ is thus not valid. The set of image features considered is $x = \{\mathbf{g}^1_{im} \ldots \mathbf{g}^n_{im}\}$.

From Sect. 3.3, it is known that for each illumination source the incident and reflected rays and the normal vector on the reflection surface at the point of incidence are coplanar. This configuration is thoroughly analysed by Shih and Liu [28].

As mentioned before, it is straightforward to deduce that the centre of the cornea is contained in the same plane, since the normal, which is contained in the plane, crosses $\mathbf{c}_c$. The matter at hand is now to study if any new improvement can be achieved by increasing the number of LEDs. Each illuminator yields a plane containing the incident and reflected rays, the corneal centre $\mathbf{c}_c$, and the camera projection centre $\mathbf{o}_{cam}$. It is known that the corneal centre and the projection centre of the camera are contained in all these planes and consequently in their intersection.

If two or more LEDs are used, with known positions, this intersection is thus a 3D line containing $\mathbf{c}_c$. If the corneal radius was known, one could determine the 3D position of the corneal centre. However, this result does not add anything new, from a point of view of performance, since $\mathbf{e}_c$ or $\mathbf{P}_c$ would need to be known additionally in order to determine the optical axis of the eye and consequently the fixated point.

Independently of the chosen point, a solution with six parameters is obtained, as in the previous section; the parameters are $\mathbf{e}_c$ (or $\mathbf{P}_c$), r_c, β_v. and α_v. Adding more glints has thus not changed the general properties of the algorithm: it performs acceptably in a fixed head condition after calibration.

3.5.3 Pupil centre + glint (PCCR)

The algorithm presented next uses both the centre of the pupil and the glint as features to determine the fixation point [22, 35], i.e. $x = \{\mathbf{g}^0_{\text{im}}, \mathbf{P}_{c_{\text{im}}}\}$. The illumination is assumed to be coaxial. It is commonly assumed that the difference vector between these two features in the image remains constant as the head moves. This is not true and there are several reasons for this. One important reason is that the glint moves as the head moves, as well as when the gaze changes. Even with a fixed head position, a change of gaze rotates the eyeball and this will force a translation of the corneal sphere and thus the movement of the glint in the image. Furthermore, minor changes of glint position may be due to the fact that the eyeball is not entirely spherical. The influence of small head movements (less than 1 cm) on the difference vector is indeed minimal, and the technique is used successfully in many eye trackers where the camera is fixed relative to the eye to compensate for small amounts of slippage. However, larger head movements (tens of centimetres), which are commonly encountered in remote eye trackers, cause significant changes in the difference vector and the technique is therefore no longer valid.

Nevertheless, this conclusion still does not disqualify this combination of features. The topic of discussion is to check if this two-feature combination, not necessarily in the form of a difference vector, can solve the head constraint. Assuming that the eyeball position $\mathbf{e}_c$ is unknown, the objective will be to check the possibility of determining the points $\mathbf{p}_c$ and $\mathbf{c}_c$ to deduce the 3D orientation of the optical axis. The analysis regarding head movements can be simplified by using the approximations described in Sects. 3.2.2 and 3.3.2, however, this should still apply for the more accurate models.

So far, under the approximated model the glint, $\mathbf{g}^0_{\text{im}}$, is the projection of corneal centre $\mathbf{c}_c$ onto the image plane. Furthermore, the projection of the centre of the pupil $\mathbf{P}_c$ results in the point $\mathbf{P}_{c_{\text{im}}}$ in the image. The observed positions of these features in the image are back projected into space on two lines as illustrated in Fig. 8. One line, r_m, joins the origin of the camera, $\mathbf{o}_{\text{cam}}$, and the pupil centre, $\mathbf{P}_{c_{\text{im}}}$. The other line, r_r, intersects $\mathbf{o}_{\text{cam}}$ and the glint $\mathbf{g}^0_{\text{im}}$. By construction, $\mathbf{p}_c$ and $\mathbf{c}_c$ lie on lines r_m and r_r. Even though the distance, d_{cp}, between them is known, there is an infinite number of points on the lines so that they are at distance d_{cp} apart. Therefore, if the position of the eyeball, $\mathbf{e}_c$, is unknown, it is not possible to determine the optical axis invariantly of head pose. If $\mathbf{e}_c$ is known and either the pupil centre or the glint is known, then the problem is similar to the one just described.

The parameters required are $\mathbf{e}_c$, d_p, β_v, and α_v, i.e. the same ones as for the algorithm that used only the pupil centre or the glint individually.

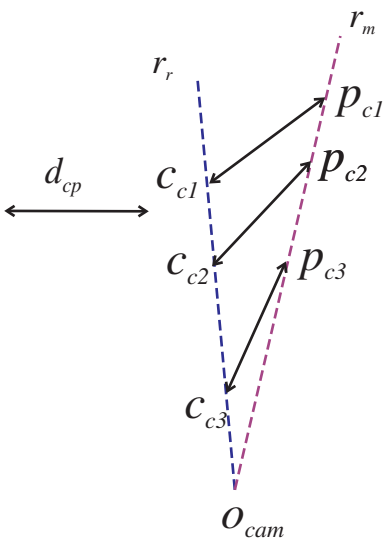

Fig. 8 Combining the pupil centre and glint does not provide a unique solution for the optical axis of the eye

3.5.4 Pupil centre + multiple glints

This section analyses the case where the pupil centre and at least two glints are used as image features, i.e. $\mathrm{x} = \{\mathbf{g}^1_{\text{im}} \ldots \mathbf{g}^n_{\text{im}}, \mathbf{P}_{c_{\text{im}}}\}$. As discussed in Sect. 3.5.2, the position of the corneal centre $\mathbf{c}_c$ in space can be determined if glints from at least two LEDs are used and the radius of corneal curvature r_c is known.

This exact knowledge of the position of $\mathbf{c}_c$ allows to break the ambiguity that existed when only one glint was used (Sect. 3.5.3). Again, the centre of the pupil $\mathbf{P}_{c_{\text{im}}}$ is back-projected into space, obtaining a ray that contains the pupil centre $\mathbf{p}_c$. Assuming that the distance d_{cp} between the corneal centre $\mathbf{c}_c$ and the pupil centre $\mathbf{p}_c$ is known, a second constraint is obtained for the position of $\mathbf{p}_c$: It must lie on a sphere with radius d_{cp} around $\mathbf{c}_c$.

This sphere is intersected with the ray obtained by back-projecting $\mathbf{P}_{c_{\text{im}}}$. In the general case, two possible solutions are obtained for the position of $\mathbf{p}_c$. Assuming that the gaze vector points roughly towards the camera rather than away from it, the position for $\mathbf{p}_c$ that is closer to the camera is chosen.

By determining the positions of $\mathbf{c}_c$ and $\mathbf{p}_c$, the optical axis of the eye has been found. This can now be corrected by the optical and visual axes angular offset to obtain the visual axis. Typically, this offset is determined using a user-dependent calibration procedure, which can also be used to correct for any residual errors remaining in the model. Unlike the methods discussed so far, this method does not require the head to remain fixed during use.

Note that this approach allows refraction of the pupil image at the corneal surface to be modelled easily: because the position of the cornea and its radius are known, the ray

obtained by back-projecting $\mathbf{P}_{c_{im}}$ can be refracted at the corneal surface, and this refracted ray is then used to compute the position of the pupil centre $\mathbf{p}_c$.

As remarked above, the model requires the radius of corneal curvature r_c and the distance d_{cp} between corneal centre and pupil centre to be known. The easiest approach is to use population averages for these parameters, and experience confirms that this is sufficient for achieving satisfactory performance. Section 4 discusses accuracy measurements performed on an implementation of this algorithm.

3.5.5 Summary

The previous sections have shown that to determine the direction of gaze while allowing head movement, it is necessary to know the position of the pupil centre and at least two glints in the image. The other combinations of features can only be used when the head is fixed.

After deriving this result geometrically in the preceding sections, a more intuitive understanding of the results will now be presented. The first case to be considered is that only the pupil centre is known. It is obvious that the pupil centre moves when the gaze direction changes. However, because movements of the whole head also cause the pupil centre to move, head movements are indistinguishable from eye movements; hence, the pupil centre can only be used for eye tracking if the head is fixed. A similar argument applies when a single glint is used.

Even though the either the pupil centre or the glint alone are theoretically sufficient for eye tracking in a fixed-head scenario, a combination of both of these features ("pupil centre, corneal reflex", "PCCR") is usually used in practice. This is because small head movements of a centimetre or two, which invariably occur even if a chin rest or some other fixation device is used, cause only a small change in the relative position of the pupil centre and glint, whereas eye movements cause a large change in this relative position. The PCCR method is thus far less sensitive to small head movements than the theoretically equivalent approach of using the pupil centre or glint alone.

However, the PCCR method cannot compensate for large head movements of tens of centimetres, especially when the distance between the head and the camera changes. To see why this is so, let us consider a scenario where the gaze direction remains the same while the head moves closer to the camera; as a result, the eye becomes larger in the camera image and, hence, the pupil centre and glint move further apart in the image. Compare this to what happens when the eye rotates away from the camera: again, the pupil centre and glint move further apart. Hence, the PCCR method cannot distinguish between a head movement and an eye movement.

Now let us consider the situation where multiple glints are used. In this case, the spatial position of the centre of the corneal sphere in space can be determined, assuming that the corneal radius (i.e. the size of the cornea) is known. The distance of the cornea from the camera can be determined because the glints will lie closer together if the cornea is further away, and vice versa. To determine the distance accurately, it is necessary to know the size of the cornea, because a larger cornea that is further away will look the same as a smaller cornea that is closer to the camera. However, because the size of the cornea does not vary too much across the population, satisfactory results can be obtained by using a population average for this value.

Because the centre of the corneal sphere moves when the gaze direction changes, it can be used to track the gaze. Again, however, this only works if the head is fixed, because head movements also change the position of the cornea and are thus indistinguishable from eye movements.

Finally, let us consider the case where both multiple glints and the pupil centre are known. As above, the glints can be used to determine the centre of the corneal sphere. This position gives one point on the optical axis of the eye; the pupil centre is a second point on the optical axis. Because two points unambiguously determine the orientation of a line in 3D space, the orientation of the optical axis and, hence, the direction of gaze, can be determined.

4 Implementation and results

A system has been implemented [20] that uses the "pupil centre + multiple glints" approach (see Sect. 3.5.4). The hardware setup for this system is shown in Fig. 9. It consists of a high-resolution camera (1,280 × 1,024 pixels) and two infrared LEDs mounted to either side of the camera. The system is mounted below an LCD monitor with a display area of 36 × 28 cm.

The camera uses a fixed-focus lens that has a focal length of 16 mm, which provides a field of view of about 60 × 50 cm at a distance of 50 cm from the camera. The lens is fitted with a filter that lets only near-infrared light pass.

The LEDs have a peak wavelength of 870 nm; they provide general illumination and generate glints on the surface of the cornea. These glints are used to find the eye in the camera image and determine the location of the corneal centre in space, as described in Sect. 3.5.2.

The gaze estimation algorithm (Sect. 3.5.4) requires the radius of corneal curvature r_c and the distance d_{cp} between corneal centre and pupil centre to be known; population averages are used for these values. To correct for errors resulting from individual variations in these parameters as

Fig. 9 Remote eye tracker system setup. The eye tracking hardware consists of a single high-resolution camera below the display and two infrared LEDs to either side

well as other residual errors, a bilinear correction function is used that is calibrated using a 5-point calibration pattern.

The image processing component, which extracts the position of the pupil and the glints from the camera image, is based on the Starburst algorithm [18], which was reimplemented and modified to fit the needs of the remote eye tracking setting. The general outline of the algorithm is as follows: a difference of Gaussians is applied to find the glints; these show up as maxima in the filtered image and are segmented using an adaptive threshold. Various heuristics are applied to eliminate false positives (for example, valid glints always turn up in pairs). The centre of a glint is found by computing the centroid of the segmented region. Searching for the darkest pixel in the vicinity of the glints yields an initial guess for the pupil centre. Rays are then shot outwards from this centre, and pupil contour points are found by searching for derivative maxima on the rays. Secondary rays are shot from the detected pupil contour points and additional contour points are detected on these rays. Finally, an ellipse is fitted through the contour points. Figure 10 shows a graphical outline of the algorithm.

The eye tracking software was implemented in C++ and runs under the Windows operating system on an Intel Pentium 4 PC with 3 GHz and 1 GB of RAM. The tracker is designed to be used with the head at a distance of 60 cm from the screen and allows head movements in a volume of about 20 × 20 × 20 cm around this point. It achieves an accuracy of around 1.5°; it is hoped that this can be improved further by fine-tuning the image processing. The tracker currently runs at 15 Hz; this limit is imposed by the frame rate of the camera. The experiments conducted show that the software should be able to run at 50 Hz or more before it becomes limited by the processing power of the PC.

Table 1 shows the result of accuracy measurements performed on the eye tracker with four test subjects. All four subjects were Caucasians, and none of the subjects wore glasses or contact lenses. Each subject calibrated the eye tracker with their head in the centre of the working range, and eye tracking accuracy was then measured in several different head positions, displaced by 10 cm from the calibration position in different directions; a chin rest was provided to position the subject's head accurately, but, of course, no such chin rest is necessary in actual use.

To measure the accuracy of the eye tracker, a rectangular grid of nine gaze targets was displayed on the screen, and subjects were asked to fixate each target in turn. The pattern of gaze targets had a width of 27 cm and a height of 22 cm; five of the gaze targets were identical to the calibration points. For each gaze target, 40 gaze samples were collected; the acquisition was started by pressing a key. No temporal filtering or averaging was applied to the gaze samples.

Table 1 shows the root mean square (RMS) error for each head position, averaged over all gaze targets and test subjects. A systematic trend or difference in the error depending on the test subject or gaze target was not observed.

The lowest error, around 1°, was obtained in the head position used for calibration. Movement of the head parallel to the monitor plane increased the error by about 0.2–0.4°; moving the head towards the front and back of the working range increased the error more strongly, by about 1–1.5°. This is mainly because the image of the eye is slightly defocused at the front and back of the working range, so that the positions of the CRs and the pupil cannot be measured as accurately.

The average error over all head positions was 1.57°; other systems report accuracies of 0.5–1° (e.g. [31]), but note that most reported results use temporal filtering or averaging on the gaze samples, which was not used here.

5 Discussion and conclusion

The main objective of this paper has been to explore and clarify the fundamental geometrical aspects of gaze tracking systems. The main contribution of the study is a mathematical analysis of alternative gaze tracking models. The results obtained can help to improve the geometry of gaze tracking systems and contribute to further studies. Some of the most commonly employed eye image features, such as pupil centre and glint, were selected to construct

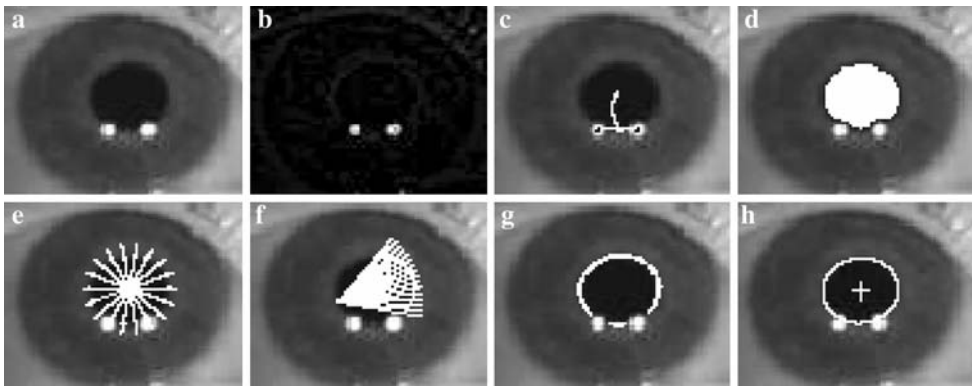

Fig. 10 Outline of the pupil and glint location extraction algorithm: **a** eye region from input image, **b** extract glints using difference of Gaussians, **c** find approximate pupil centre as darkest pixel in vicinity of glints, **d** initial segmentation of pupil using adaptive threshold, **e** find contour points on rays shot from centre of pupil, **f** shoot secondary rays to find more contour points, **g** extracted contour points, **h** *t* ellipse to extracted points

Table 1 Accuracy measurements performed on the remote eye tracker for four test subjects

Head position	(*x*, *y*, *z*)	RMS gaze error (degrees)
Centre (calib. pos.)	(0, 0, 0)	1.01
Left	(−10, 0, 0)	1.33
Top	(0, 10, 0)	1.23
Top left	(−10, 10, 0)	1.44
Front	(0, 0, −10)	2.54
Back	(0, 0, 10)	1.88
Overall		1.57

Root mean square (RMS) gaze error was measured in different head positions; head position coordinates (in centimetres) are given relative to the centre of the working range, which was at a distance of 58 cm from the screen

gaze tracking models. Each model was analysed from a purely geometrical point of view, in order to evaluate its tolerance for head movement, hardware necessities and calibration requirement.

These issues are important for users, since they determine how many calibration points are needed, whether they need sit still and what the minimal size of objects is that can be selected on the screen. These issues are thus directly related to users' acceptance of an eye tracking system.

Of the different approaches that were analysed, the approach based on a single camera and multiple glints was shown to allow head movement and require minimal calibration; these results were confirmed on tests of an implementation of this approach.

Despite the advances in remote eye tracking systems in recent years, there are still quite a number of areas in which improvements have to be made if these systems are to see widespread use in human–computer interfaces, including, but not limited to, augmentative and alternative communication (AAC) applications. Some of these areas concern the theoretical foundation of eye tracking; others are of a more technical nature. Some of the areas in which it is believed that worthwhile advances can be made in the coming years are:

- *Tolerance towards glasses* Systems that use infrared illumination often do not work well for users who wear glasses because of reflections on the surface of the glasses. The existing systems can usually be made to work with glass wearers to a certain extent, but only for some head orientations where no interfering reflections occur. For other head orientations, the reflections can obscure the user's eyes completely, making eye tracking impossible. One way of dealing with this problem might be to use more than two infrared illuminators. At any given time, the system would use two of the illuminators. If the system detected that the user's pupils were being obscured by reflections, it would switch to a different set of illuminators at a different angle relative to the user and the camera. In this way, the reflections should shift off the eyes or even be eliminated entirely.
  To achieve high accuracy in the presence of glasses, the eye model may have to be augmented with a model of the glasses to account for their effect on the image of the eye. However, preliminary tests indicate that the accuracy is still tolerable even if the effect of the glasses is not modelled.
- *Suitability for outdoor use* Existing eye trackers are mainly designed for indoor use and have trouble coping with outdoor settings or even with sunlight coming through a window. For many applications, such as wheelchair-mounted AAC solutions, this is a severe limitation.

For most systems, the main factor that causes problems in outdoor settings is their reliance on active infrared illumination, which can be drowned out by strong sunlight. Development of algorithms that do not depend on glints, or that can switch to an alternative mode of operation when the glints are not found, would result in better outdoor performance.

- *Ease of setup/use* Remote eye tracking systems are typically based on a physical model of the eye, the eye tracking system (camera and illuminators), and the monitor. Because of this, they require the spatial relationship between the camera, the illuminators, and the monitor plane to be known. These measurements are usually obtained by hand, a process that is time-consuming, error-prone, and difficult for an end user to carry out. Beymer and Flickner [3] calibrated the orientation of the monitor plane automatically using a mirror to reflect the image of a draughtboard pattern taped to the monitor back into the camera. It is planned to implement a similar automatic calibration in the system reported in this paper.
- *Sensor technology* High-accuracy eye trackers need high-quality images in high resolution. For this reason, the remote eye trackers that exist today typically use high-resolution industrial cameras with relatively high-grade lenses. This makes the systems quite expensive, even before labour costs for assembly are taken into account. For example, the camera and lens used in the eye tracker described here have a combined price of around 1,000 USD. This puts the system out of reach of many potential users.

  However, high accuracy is not needed for all types of applications. If reduced accuracy is acceptable, an obvious idea for reducing costs is to use cheap, off-the-shelf hardware such as web cameras. Due to the lower quality of the images and the unknown geometry of the camera and possible light sources, image analysis and gaze estimation are more difficult (see e.g. Hansen and Pece [12]). If the cameras use visible light, this may make eye tracking more difficult, since certain invariants that are typically exploited in the active illumination case no longer apply; it may, however, be possible to find new invariants that apply also under passive illumination.

  If past trends are anything to go by, the resolution and image quality of webcams can be expected to increase in the coming years, which will make the idea of using this type of camera for eye tracking ever more attractive. Also, recent developments in the area of alternative image sensors, such as 3D time-of-flight (TOF) cameras [8], seem to hold promise for eye tracking.

Robust, affordable eye trackers would have a broad range of potential applications. They would of course be invaluable for AAC applications, but beyond that, eye tracking has the potential to become a new general-purpose interaction medium. Eye tracking may change the way people interact with technology and the way visual information is communicated. Current work on gaze guidance [2] has the goal of augmenting a video or visual display with a recommendation of how to view the information, of what is to be seen.

The advances currently being made in eye tracking hardware and software may finally help make widespread low-cost eye tracking a reality.

Acknowledgments Our research has received funding from the European Commission within the Network of Excellence COGAIN (contract no. IST-2003-511598) of the 6th Framework Programme. Additionally, research at the University of Lübeck has received funding from the European Commission within the project GazeCom (contract no. IST-C-033816) of the 6th FP. All views expressed herein are those of the authors alone; the European Community is not liable for any use made of the information.

References

1. Babcock, J.S., Pelz, J.B.: Building a lightweight eyetracking headgear. In: Proceedings of the Eye Tracking Research and Applications symposium, pp. 109–114. ACM Press (2004)
2. Barth, E., Dorr, M., Böhme, M., Gegenfurtner, K.R., Martinetz, T.: Guiding the mind's eye: improving communication and vision by external control of the scanpath. In: Rogowitz, B.E., Pappas, T.N., Daly, S.J. (eds.) Human Vision and Electronic Imaging, Proceeddings of SPIE, 2006, vol. 6057. Invited Contribution for a Special Session on Eye Movements, Visual Search, and Attention: a Tribute to Larry Stark
3. Beymer, D., Flickner, M.: Eye gaze tracking using an active stereo head. In: Proceedings of Computer Vision and Pattern Recognition (CVPR), vol. 2, pp. 451–458 (2003)
4. Brolly, X.L.C., Mulligan, J.B.: Implicit calibration of a remote gaze tracker. In: Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW 2004), vol. 8, p. 134 (2004)
5. COGAIN: D3.1 user requirements report, with observations of difficulties users are experiencing. Technical report, European Union Network of Excellence COGAIN (contract no. IST-2003-511598) of the 6th Framework Programme, 2005. http://www.cogain.org/results/reports/COGAIN-D3.1.pdf (2005)
6. COGAIN: D3.2 report on features of the different systems and development needs. Technical report, European Union Network of Excellence COGAIN (contract no. IST2 003-511598) of the 6th Framework Programme, 2006. http://www.cogain.org/results/reports/COGAIN-D3.2.pdf (2006)
7. Crane, H., Steele, C.: Accurate three-dimensional eye tracker. J. Opt. Soc. Am. **17**(5), 691–705 (1978)
8. CSEM: SwissRanger SR-3000, CSEM SA, Zurich, Switzerland. http://www.swissranger.ch (2006)
9. Duchowski, A.T.: Eye tracking methodology, theory and practice. Springer, London (2003)
10. Ebisawa, Y.: Unconstrained pupil detection technique using two light sources and the image difference method. In: Visualization and Intelligent Design in Engineering, pp. 79–89 (1989)
11. Guestrin, E.D., Eizenman, M.: General theory of remote gaze estimation using the pupil center and corneal reflections. IEEE Trans. Biomed. Eng. **53**(6), 1124–1133 (2006)

12. Hansen, D.W., Pece, A.E.C.: Eye tracking in the wild. Comput. Vis. Image. Underst. **98**(1), 155–181 (2005)
13. Hennessey, C., Noureddin, B., Lawrence, P.: A single camera eye-gaze tracking system with free head motion. In: Proceedings of Eye Tracking Research and Applications (ETRA), pp. 87–94 (2006)
14. Hutchinson, T.E.: Human-computer interaction using eye-gaze input. IEEE Trans. Syst. Man Cybern. **19**(6) (1989)
15. Ji, Q., Yang, X.: Real time visual cues extraction for monitoring driver vigilance. In: ICVS: Second International Workshop on Computer Vision Systems, Vancouver, Canada (2001)
16. Karmali, F., Shelhamer, M.: Compensating for camera translation in video eye movement recordings by tracking a landmark selected automatically by a genetic algorithm. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Proceedings, pp. 5298–5301 (2006)
17. Kolakowski, S.M., Pelz, J.B.: Compensating for eye tracker camera movement. In: ETRA 2006: Proceedings of the 2006 symposium on Eye Tracking Research and Applications, pp. 79–85 (2006)
18. 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 IEEE Vision for Human-Computer Interaction Workshop at CVPR, pp. 1–8 (2005)
19. Marquardt, D.: An algorithm for least-squares estimation of nonlinear parameters. SIAM J. Appl. Math. **11**, 431–441 (1963)
20. Meyer, A., Böhme, M., Martinetz, T., Barth, E.: A single-camera remote eye tracker. In: Perception and Interactive Technologies. Lecture Notes in Artificial Intelligence, vol. 4021, pp. 208–211. Springer, New York (2006)
21. Morimoto, C.H., Koons, D., Amir, A., Flickner, M.: Pupil detection and tracking using multiple light sources. IVC **18**(4), 331–335 (2000)
22. Morimoto, C.H., Mimica, M.R.M.: Eye gaze tracking techniques for interactive applications. Comput. Vis. Image. Underst. **98**(1), 4–24 (2005)
23. Müller, P., Cavegn, D., d'Ydewalle, G., Groner, R.: A comparison of a new limbus tracker, corneal reflection technique, purkinje eye tracking and electro-oculography. In: d'Ydewalle, G., Rensbergen, J.V., (eds.) Perception and Cognition, pp. 393–401. Elsevier, Amsterdam (1993)
24. Newman, R., Matsumoto, Y., Rougeaux, S., Zelinsky, A.: Real-time stereo tracking for head pose and gaze estimation. In: International Conference on Automatic Face and Gesture Recognition, pp. 122–128 (2000)
25. Ohno, T., Mukawa, N.: A free-head, simple calibration, gaze tracking system that enables gaze-based interaction. In: Eye Tracking Research and Applications (ETRA), pp. 115–122 (2004)
26. Pomerleau, D.A., Baluja, S.: Non-intrusive gaze tracking using artificial neural networks. In: CMU-CS-TR (1994)
27. Rabbetts, R.B.: Bennett and Rabbetts clinical visual optics, 3rd edn. Butterworth-Heinemann, Elsevier, Edinburgh (1998)
28. Shih, S.W., Liu, J.: A novel approach to 3-D gaze tracking using stereo cameras. IEEE Trans. Syst. Man Cybern. B. **34**(1), 234–245 (2004)
29. Shih, S.-W., Wu, Y.-T., Liu, J.: A calibration-free gaze tracking technique. In: Proceedings of the 15th International Conference on Pattern Recognition, pp. 201–204 (2000)
30. Tan, K.-H., Kriegman, D.J., Ahuja, N.: Appearance-based eye gaze estimation. In: IEEE Workshop on Applications of Computer Vision, pp. 191–195 (2002)
31. Tobii: Tobii 1750 eye tracker, Tobii Technology AB, Stockholm, Sweden. http://www.tobii.se (2002)
32. Villanueva, A., Cabeza, R., Porta, S.: Eye tracking: pupil orientation geometrical modeling. Image Vis. Comput. **24**(7), 663–679 (2006)
33. Wyszecki, G., Stiles, W.S.: Color science: concepts and methods, quantitative data and formulae, 2nd edn. Wiley, New York (1982)
34. Yoo, D.H., Chung, M.J.: A novel non-instrusive eye gaze estimation using cross-ratio under large head motion. Comput. Vis. Image Underst. **98**, 25–51 (2005)
35. Young, L., Sheena, D.: Methods & designs: survey of eye movement recording methods. Behav. Res. Methods Instrum. **7**(5), 397–429 (1975)