Geometry Issues of a Gaze Tracking System

Arantxa Villanueva, Juan J. Cerrolaza, and Rafael Cabeza

Public University of Navarre,
Campus Arrosadia sn., 31006 Pamplona, Spain
{avilla, rcabeza}@unavarra.es

Abstract. One of the most confusing aspects that one meets as he introduces himself into gaze tracking is the variety, in terms of hardware equipment, of available systems providing solutions to the same matter, i.e. determining subject's gaze. Calibration permits adjusting trackers based on different hardware and image features to the subject. The negative aspect of calibration is that it permits the system to work properly but at the expense of a lack of control over the intrinsic behavior of the tracker. The objective of this work is to overcome this obstacle to explore more deeply the elements of a tracker from a purely geometrical point of view. Alternative models based on image features are evaluated. As result of this study a model has been constructed based on minimal calibration using one camera and multiple lighting with acceptable accuracy level.

Keywords: gaze-tracking, video-oculography, calibration, PoR, LoS.

1 Introduction

Among existing tracking technologies, systems incorporating video-oculography (VOG) use a camera or a number of cameras and try to determine the movement of the eye using the information obtained after studying the captured images. Normally they include infrared lighting in order to produce specific effects in the obtained images. The non intrusive nature of the trackers employing videooculography renders it an attractive technique. Video-oculographic eye tracking techniques can be divided into two main groups. Methods that measure the eye movements inside its orbit and methods that calculate the gaze direction, i.e. Line of Sight (LoS). For the methods of the second type if the position of the gazed area, such as a screen, is known the Point of Regard (PoR) is determined as the gazed point. The number of commercial systems and experimental trackers devoted to calculating PoR that incorporate VOG is really large and they admit multiple variations in terms of hardware, i.e. number of cameras and number of lighting sources. However, their basis seems to be the same; the image of the eye captured by the camera will change when the eye rotates or translates in 3D space. In fact most of them work in an acceptable way and can be used in real applications. However, from another point of view it seems essential to explore more deeply into the geometry and behavior of these systems. This will help to explain and

C. Stephanidis (Ed.): Universal Access in HCI, Part II, HCII 2007, LNCS 4555, pp. 1006–1015, 2007.

correct most of the problems arisen during tracking sessions as the subject moves his head or the tracking is lost. The arisen matter is evident: “what is the connection between the image in the camera and the observed point?” This study tries to analyze the mathematical connection between the image and the observed point in depth. Interesting works have been presented devoted to mathematical modelling of gaze tracking systems. The most relevant are the ones presented by Beymer&Flickner [1], Ohno&Mukawa [2] and Shih&Liu [3] and more recently the results by Guestrin&Eizenman [4] and Hennessey et al. [5]. All of them deal with single camera or stereo trackers that use some kind of infrared lighting. The work by Shih&Liu [3], from our point of view, presents high mathematical rigor.

We would like to emphasize the theoretical essence of the present paper. Although experimental results are included we consider that our main contribution is the construction of a geometrical basis and models as result of our work. The link between the image and the observed point is found deducing the minimal requirements of a system in terms of number of cameras or illuminators. Narrowly related with the hardware necessities, the most interesting features of the image for gaze determination can be identified. As a consequence of this finding a model for a gaze tracking system fully based on identifiable and recognizable variables is constructed. In the next section a more proper introduction to the studied matter is carried out. Sections 3 and 4 are devoted to the analysis of alternative models and to the construction of the final model for a gaze tracking system. Section 5 presents the experimental results and finally the conclusions obtained are exposed in section 6.

2 Eye Tracking System Models

Regardless of the hardware used the goal of any tracker is to establish a connection between the features provided by the technology, i.e. image analysis results, and gaze. Quadratic or cubic expressions based on unknown coefficients that are deduced by means of the well-known calibration procedure are used to this end. Calibration consists in asking the subject to gaze at a 3x3/4x4 grid of known markers on the screen. This permits systems with fully different hardware and image features to work acceptably, but prevents researchers from determining the minimal system requirements and geometrical properties. The objective of this work is to overcome this obstacle and to build a geometrical model based on physical parameters for the gaze tracking system. The procedure selected in order to accomplish the work is to analyze separately the alternative features that can be extracted from the image. We define a model as a connection between the fixated point and alternative features extracted from the image, expressed as a function of hardware and subject parameters describing the gaze tracking system setup. The models proposed are based on single point features such as the center of the pupil and the glint, based on shapes such as the pupil ellipse or based on combinations of points and shapes. First a geometrical analysis is conducted for the models in which projective relations among the elements of the system are studied from a purely geometrical point of view. In this step corneal refraction is obviated and the determination of the gaze direction in a free head movement scenario is pursued for the models. Secondly corneal refraction is considered for those models that satisfied the previous step introducing additional limitations to the model. Figure 1 sketches the analysis for each model.

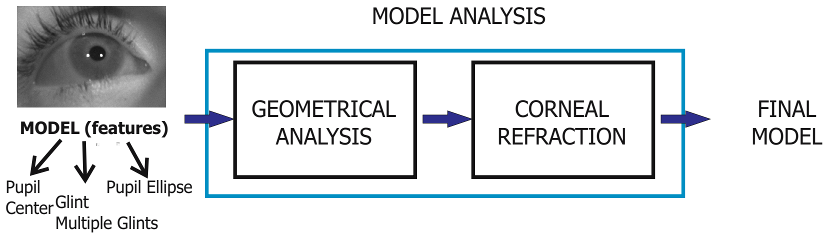

Fig. 1. Each model based on alternative image features is analyzed geometrically and from the point of view of corneal refraction

3 Geometrical Analysis

From a geometrical point of view the objective should be to find a model that can determine the 3D position of the gaze in a free head movement environment. The gaze line can be easily approximated by the visual axis. On the other hand the optical axis of the eyeball can be approximated as its symmetry axis. Optical and visual axes of the eye present an angular offset nasally whose value is about 5±1º. Once the 3D position of the optical axis is calculated with respect to a reference system, the visual axis can be calculated supposing that the horizontal angular offset between both axes is known and applying the corresponding torsion described in [6] in order to achieve an agreement with the eyeball orientation described by Listing's and Donder's laws. The optical axis of the eye contains three principal points of the eyeball since it is approximated as its symmetry axis, i.e. *A*, eyeball center, *C*, corneal center and *E*, pupil center. Therefore the geometrical matter is reduced to determine two points of the optical axis. Figure 2 represents a simplified description of the problem. Then the visual axis can be easily derived from it as it has been explained before. To follow the alternative models are introduced. Much effort has been made in order to simplify the most cumbersome mathematical aspects and to pay more attention in the obtained conclusions. More details can be found in [7].

Center of the Pupil and Corneal Reflection. First the models resulting from each feature separately are studied. Regarding to the center of the pupil if affine projection is assumed, the center of the pupil in the image can be approximated to the projection of point *E*. Regardless of the exact mathematical expressions relating the involved variables, the selected image feature gives just information about one principal point of the optical axis, i.e. *E*.

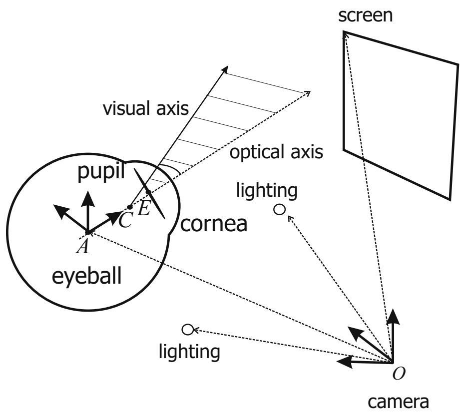

Fig. 2. Principal points of the system. The optical axis contains the center of the eyeball *A*, corneal center *C* and pupil center *E*.

On the other hand the corneal reflection or glint in the image is a consequence of the reflection caused by the illumination source on the corneal surface and in the best case can provide information about the corneal center position *C*. These two features separately provide information about one principal point of the optical axis, i.e. *E* or *C* respectively. Therefore the optical axis can hardly be solved. An acceptable behavior of the model could be expected in a fixed head situation (fixed *A*). However, free head movement is pursued and consequently these models are rejected.

It is commonly assumed that the distance between these two features in the image compensates for possible head movements during a tracking session but from the point of view of this work this approximation is not valid and there exists a dependence between this vector value and the head position. In order to simplify the analysis let us propose a rough approximation of both features. Assuming affine projection one could back-project the center of the pupil from the image generating the line r_m. Considering a coaxial location of the led with respect to the camera we define r_r as the back-projection of the glint into 3D space. Close approximations of points *E* and *C* are contained in r_m and r_r respectively. Knowing the distance between *E* and *C* does not solve the indetermination, since more than one combination of points in r_m and r_r can be found at the desired distance and therefore there is no unique solution. Therefore once again the 3D optical axis position is not determined (see figure 3).

Multiple Glints. Following the law of reflection each illuminator will result in a plane containing the incident ray, the reflected ray and the normal at the point of incidence. Consequently the corneal center *C* as well as the projection center of the camera *O* will be contained in the plane. It is deduced that if more than one illuminator are used, *C* contained in the normal, and *O* contained in the reflected ray, will be contained in the resulting intersection line of the planes [3]. Once the corneal 3D line is deduced knowing the corneal radius by means of a calibration process one could deduce the corneal center 3D position. Consequently it is stated that two illumination sources, i.e. two glints, are enough in order to determine the 3D position of the corneal center. However it does not add anything new, from a purely geometrical point of view, since A and E remain unknown and the optical axis is not determined.

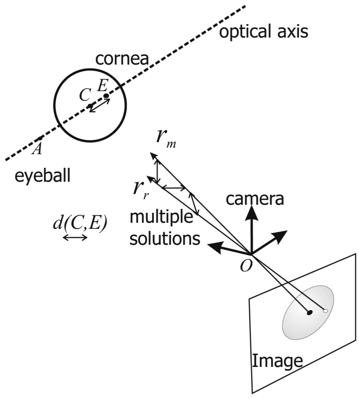

Fig. 3. The combination of pupil center and glint position does not allow determining the optical axis of the eye since multiple combinations can be found in r_r and r_m having the same distance

Pupil Center + Multiple Glints. From preceding sections we know that employing two glints leads to the 3D determination of the corneal center C. This breaks the indetermination arisen for the model based on the center of the pupil and one glint. In other words once C is found, the center of the pupil can be easily found knowing r_m and the distance between both points. The advantage with respect to the single glint model is that the introduction of a second illuminator resolves the position of the corneal center. No approximation has been considered for the estimation of C whereas affine projection is assumed for the center of the pupil.

The Projected Pupil Ellipse. It is already known that the projection of the pupil results in a shape that can be approximated by an ellipse [1] [2]. The back-projection of the pupil from the image into 3D space would be a cone, i.e. back projection cone, and it could be assured that there is at least one plane that intersects the cone in a circular section containing the pupil. The theory about conics claims that parallel intersections of a quadric result in equivalent conic sections. Considering the back projection cone as a quadric, it is clear that if we find a plane with a circular section for the specific quadric, i.e. back projection cone, infinite pupils of different sizes could be defined employing intersecting parallel planes. The direction of the planes that result in circular sections is not a trivial task and is derived from [8]. The absolute conic that using homogenous coordinates (x^0, x^1, x^2, x^3) is defined as $x^0 = 0$;

$$\sum_{i=1}^{3}(x^i)^2 = 0 \tag{1}$$

lying on the plane at infinity is the place of all the cyclic sections. The intersection of a quadric with the absolute conic is a circumference. The mathematical solution of this intersection finds out the direction of the parallel planes or sets of parallel planes that intersect the quadric with a circular shape. For the case under analysis two possible solutions resulting in circular sections of the cone are found. Two groups of an infinite number of planes can be calculated, each of them intersecting the back-projection cone in a circular shape of different sizes and containing a suitable solution for the gaze tracking matter. Consequently no solution can be provided for the optical axis since the size of the 3D pupil is supposed to be unknown and variable during a tracking session and consequently not valid to select one of the intersections.

The Pupil Ellipse + Glint. To simplify the analysis once again we deduce a 3D line by means of the back-projection of the glint in the image which is supposed to contain an approximation of C, i.e. r_r. Each possible plane intersecting the back-projection cone in a circumference determines a pupil center E and an optical axis that is calculated as the 3D line perpendicular to the pupil plane that crosses E (see figure 4). The solution is easily deduced if the distance between the center of the pupil E and the corneal center C is known. The pupil plane for which the optical axis meets the r_r line at the known distance from E will be selected as solution to the tracking problem. In addition the intersection between the optical axis and the r_r line will be the corneal center C. Therefore the introduction of the glint permits the selection of one of the planes for each one of the two possible orientations (see The Projected Pupil Ellipse). However a more rigorous mathematical analysis leads to conclude that just one solution is possible as the second one requires the assumption that the center of the cornea, C, remains closer to the camera than the center of the pupil E and this situation is not feasible.

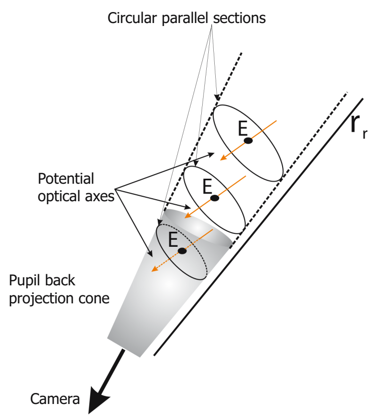

Fig. 4. Each plane intersects the cone in a circle resulting in an optical axis crossing its center E perpendicularly. The correct optical axis is the one that intersects r_r at the right distance, i.e. *d(C, E).*

The Pupil Ellipse + Multiple Glints. It is already known that the combination of two glints and the center of the pupil provides a solution to the tracking problem. Therefore at least the same result is expected if the pupil ellipse is considered since it contains the value of the center. The most outstanding difference amongst models based on the pupil ellipse with one or multiple glints is the fact that employing the information provided exclusively by the glints, the corneal center can be accurately determined. Consequently the data about the distance between both centers, i.e. pupil and corneal centers can be ignored. The known point C must be located in one of the optical axes calculated from the circular sections and crossing perpendicularly the corresponding center E. The optical axis among the possible ones that contains the previously calculated C will be selected as solution for the tracking problem.

4 Refraction Analysis

Since any ray of light coming from the back part of the eye suffers refraction and consequently a deviation in its direction when it crosses the corneal surface the obtained pupil image must be considered as the projection of a virtual pupil. The opposite path could be studied and any point of the pupil contour in the image can be back-projected from the image and refracted in the cornea. Starting from the pupil image in the camera the cornea, i.e. its center and radius, must be known to apply refraction. Consequently the model based on the pupil ellipse and the glint fails this analysis since it does not accomplish a previous determination of the corneal center. Contrary to this model the one based on the pupil center and two glints makes a prior computation of the corneal center, however, it can no longer be assumed it is the center of the real pupil the one contained in r_m but the center of the virtual pupil. One could expect that E will be contained in a 3D line obtained as a consequence of the refraction of r_m when crossing the cornea. Although this approximation can be considered as valid in some cases it is geometrically not true, since refraction through a spherical surface is not a linear transformation. The model based on two glints and the shape of the pupil provides the most accurate solution to the matter. The model deduces the value of C employing exclusively the two glints of the image. Then each pupil contour back-projected ray from the image is calculated and its refraction estimated into the cornea. The center of the pupil should be a point located at a known distance from C that represents the center of a circle whose perimeter is fully contained in the refracted lines and is perpendicular to the line connecting pupil and corneal centers as shown in figure 5.

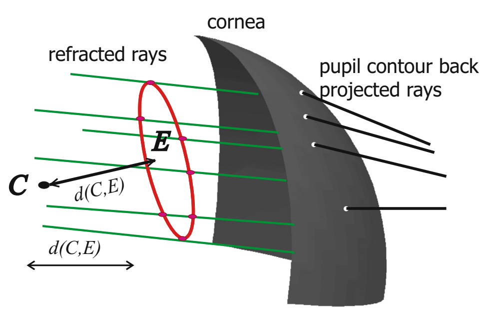

Fig. 5. The pupil center E is deduced from the refracted rays of the pupil contour from the image. It represents the center of a circumference that contains all the refracted rays and is connected to C perpendicularly at the correct distance $d(C, E)$.

5 Experimental Results

From the prior analysis, the model based on two glints and the shape of the pupil appears as the only potential model for the gaze tracking system. When a real eye tracking system is considered, real hardware must be considered. That means that certain intrinsic tolerances and noise of the elements composing the eye tracker need to be introduced. The reduced size of the glint in the image at certain user camera distances introduces certain indetermination in the position of the corneal reflection and consequently in the corneal center computation. Based on the experimental indetermination value of this feature a formal study was carried out that showed the sensitivity of the corneal center estimation with respect to glint position indetermination. In order to reduce the sensitivity the solution adopted by this study, is to increase the number of illumination sources in order to obtain an average value for the point *C*. If more than two illuminators are employed, alternative pairs can be used to estimate the pursued point. An average of the obtained values by means of the possible combinations reduces the sensitivity of the model to glint indetermination. Consequently a model based on the pupil ellipse and multiple lighting is proposed. Ten users were selected for the test. The working distance was selected in the range 400-500 mm from the camera. Figure 6a shows the selected fixation marks uniformly distributed in the gazing area whose position is known with respect to the camera. The position in mm for each point is shown. Ten consecutive images were acquired and grabbed for each fixation. The images have been captured with a calibrated Hamamatsu C5999 camera and digitalized by a Matrox Meteor card with a resolution of 640x480 (RS-170). The LEDs used for lighting have a spectrum centered at 850 nm. Four LEDs were selected to produce the needed multiple glints. They were located in the lower part of the camera and its positions with respect to the camera were measured. The images present a dark pupil and four bright glints (see figure 6b). The next step was to process each image separately to extract the glints coordinates and the ellipse of the pupil. As said in the introduction, this paper deals with geometrical modelling of gaze tracking systems. From this point of view aspects such as image processing algorithms used and additional experimental details are obviated to focus the reader in the results obtained for the proposed geometrical model.

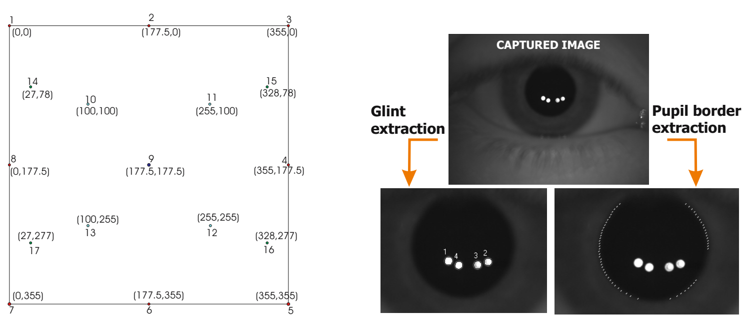

Fig. 6. a) Test points. b) From each image the glints positions and the pupil contour are extracted.

Once the hardware is defined and in order to apply the constructed model based on the ellipse and glints positions, some individual subject characteristics need to be calculated such as corneal radius, angular offset between optical and visual axes and the distance between corneal and pupil centers. The constructed model based on multiple glints and pupil shape permits, theoretically, determining this data by means of a single calibration mark. In the practice and in order to increase the confidence on the obtained values, three fixations were selected for each subject to estimate a mean value for these parameters. The aim of table 1 is to show a quantitative evaluation of the model competence for two, three and four LEDs. For each subject the average error for the 17 fixation marks was calculated in visual degrees since this is the most significative measurement of the model performance. It is clear that the model with four LEDs presents the lowest errors. In average the model with two LEDs presents an error of 1.08º the model with three LEDs 0.89º and the model with four 0.75º. Therefore it can be said that, in average, the models with three and four LEDs render acceptable accuracy values (<1º). As expected an increase in the number of illumination sources results in an improvement of the system tracking capacity.

Table 1. Error quantification (degree) of the final model using 2, 3 and 4 LEDs for ten users

Subject	1	2	3	4	5	6	7	8	9	10
2LED	1.47	0.85	1.46	0.90	0.92	0.97	1.24	0.78	1.19	1.06
3LED	1.06	0.80	1.35	0.58	0.75	0.78	1.20	0.79	0.74	0.86
4LED	1.04	0.76	1.01	0.62	0.72	0.71	0.62	0.65	0.59	0.80

6 Conclusions

The intrinsic connection between the captured image from the eye and gaze has been explored. A model for a video-oculographic eye tracking system has been constructed. A model is understood as a geometrical connection between subject's gaze and the variables describing the elements of the system together with the data extracted from the image. The objective was not to find the most robust system but to find out the minimal features of the image that are necessary in order to solve the gaze tracking problem in an acceptable way. It has been demonstrated that the model based on the pupil ellipse and multiple glints allows for a competent tracking and matches the pursued requirements, i.e. permits free head movement, has minimal calibration requirement and presents an accuracy in the range of the already existing systems with longer calibrations and more restrictions for the head. Theoretically one point of calibration is enough to adjust the model. In addition the minimal hardware needed by the system is also determined, i.e. one camera and multiple infrared lighting. Geometrical modelling of gaze tracking systems will provide a theoretical basis to find out image-gaze connections that tolerate head movement, explain and correct the loss of accuracy during a tracking session and allow for simpler and shorter calibration procedures.

Acknowledgments

COGAIN (511598): Communication by Gaze Interaction. Network of Excellence FP6-2003-IST-2.

References

1. Beymer, D., Flickner, M.: Eye gaze tracking using an active stereo head. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. vol. 2, Wisconsin, pp. 451–458 (2003)
2. Ohno, T., Mukawa, N.: A free head, simple calibration, gaze tracking system that enables gaze-based interaction. In: Proceedings of the Eye Tracking Research and Applications Symposium, San Antonio (TX), pp. 115–122 (2004)
3. Shih, S.W., Liu, J.: A novel approach to 3-D gaze tracking using stereo cameras. IEEE Transactions Systems Man. and Cybernetics Part -B 34(1), 234–245 (2004)
4. Guestrin, E., Eizenman, M.: General theory of remote gaze estimation using pupil center and corneal reflections. IEEE Transactions on Biomedical Engineering 53(6), 1124–1133 (2006)
5. Hennessey, C., Noureddin, B., Lawrence, P.: A single camera eye-gaze tracking system with free head motion. In: ETRA ’06: Proceedings of the 2006 symposium on Eye tracking research and applications, pp. 87–94. ACM Press, New York (2006)
6. Fry, G., Treleaven, C., Walsh, R., Higgins, E., Radde, C.: Definition and measurement of torsion. In: American Journal of Optometry and Archives of American Academy of Optometry. vol. 24., Philadelphia, pp. 329–334 (1947)
7. Villanueva, A.: Mathematical Models for Video Oculography. Thesis, Public University of Navarra, Pamplona, SPAIN (2005)
8. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press, Cambridge (2004) ISBN: 0521540518