

Detecting Eye Position and Gaze from a Single Camera and 2 Light Sources

Carlos H. Morimoto
Dpto de Ciência da Computação - IME/USP
Rua do Matão 1010, São Paulo, SP 05508, Brazil
hitoshi@ime.usp.br

Arnon Amir and Myron Flickner
IBM Almaden Research Center
650 Harry Rd, San Jose, CA 95120, USA
{arnon,flick}@almaden.ibm.com

Abstract

We introduce a new method for computing the 3D position of an eye and its gaze direction from a single camera and at least two near infra-red light sources. The method is based on the theory of spherical optical surfaces and uses the Gullstrand model of the eye to estimate the positions of the center of the cornea and the center of the pupil in 3D. The direction of gaze can then be computed from the vector connecting these two points. The point of regard can also be computed from the intersection of the direction of gaze with an object in the scene. We have simulated this model using ray traced images of the eye, and obtained very promising results. The major contribution of this new technique over current eye tracking technology is that the system does not require to be calibrated with the user before each user session, and it allows for free head motion.

1 Introduction

An eye gaze tracker (EGT) is a device that can compute the direction of gaze (DoG), i.e., the line of sight of the eye. If information about the scene is available, the point of regard (PoR) can be computed as the intersection of the DoG with a scene object. The scene can easily be constrained to a computer screen and the DoG and PoR information made available to any computer application, as another form of input device.

EGTs have helped psychologists, neurologists and ophthalmologists for several decades to study and build theories and models about eye movements and their behaviors. Several EGT methods have been employed in such studies, from physical probes, electric muscular activity, to magnetic sensors, and a wide range of optical and visual techniques. For a review of EGT methods see [1, 8]. There are quite a few commercial EGT devices on the market, which are mainly used by professionals as lab equipment for experiments and medical diagnosis. The use of EGTs to enhance human computer interaction (HCI) in general have

been suggested in many studies [4], and it has been proven to be useful for people with physical disabilities [3].

Some of the techniques for EGT are head mounted, requiring the user to wear some sort of helmet or glasses. Other techniques are remote in the sense that they do not require any device to be in physical contact with the user. In general, head mounted EGTs are more accurate and allow for free head motion, but they might cause some discomfort. Therefore remote EGTs are more appropriate for general HCI applications. Both head mounted and remote EGTs require an initial calibration procedure.

This paper describes a camera based EGT. Such EGTs track the limbus (the contour of the iris) or the pupil. It is also common to use an external light source to generate a reflection on the cornea (glint), which is used as a reference point. Pure limbus tracking techniques can use the edges of the eye socket as the reference. Assuming a spherical cornea, the glint position as viewed by the camera does not change with rotations but the pupil does. The glint and the pupil center define a 2D vector that can be used to compute the PoR using a direct mapping from the glint-pupil vector to scene or computer screen coordinates. The purpose of the calibration procedure is to measure a sufficient number of points to create this mapping. In general it requires the user to fixate her gaze at known scene coordinates in a particular order.

Although the calibration procedure might take only a few seconds, it has to be made before each user session. This simple model does not allow head motion, so that the system requires recalibration when the user changes his head position (typically, a remote EGT requires recalibration if the head moves just a few inches).

The need for frequent calibration and the requirement to keep the face considerably still are two of the major deficiencies in current remote EGTs. Elevated costs and limited accuracy (about 1 degree error) are other factors that has hindered the wide spread use of such devices. Contributions in order to make remote EGTs more robust and low cost are presented in [6]. Recently, Shih *et al.* [7] have shown that without information about the cornea and pupil

sizes, at least two cameras and two light sources are needed to recover eye position and gaze in 3D. The next section introduces the computational model used to build a lower cost alternative to estimate the eye position and gaze using a single camera, that does not require frequent user calibration and allows for free head motion. Section 3 presents results from simulations that show the accuracy and feasibility of the system, and Section 4 concludes the paper.

2 Computational Model

The optical theory of spherical surfaces for paraxial rays [5] are combined with the Gullstrand model of the eye [2] to recover the 3D centers of the cornea and the pupil. The direction of gaze is computed from these two points. The cornea is an almost spherical transparent membrane that covers the iris. The pupil is a circular aperture in the center of the iris that regulates the amount of light coming into the eye. It is placed between the cornea and the eye lens, and lies inside the aqueous humor. The cornea and the aqueous humor have different indexes of refraction, thus the light actually travels through at least 3 different mediums and is refracted at each medium boundary. The air-cornea boundary however is responsible for most of the refraction. We use the following data from the Gullstrand model: cornea radius of 7.7mm, cornea index of refraction of 1.376, and the pupil distance to the cornea center of 3.6mm.

2.1 Computing the center of the cornea

The reflection of external light sources will be used to estimate the center of the cornea, which will be modeled as a spherical convex mirror for this purpose. Figure 1 shows the image formation process for a convex mirror of radius r and centered at C . Without loss of generality, let the origin, O , be at the camera focal center, and define the principal axis of the cornea as the line OC . The vertex of the cornea V and its focal point F are placed on this line, V at the surface and F halfway between V and C .

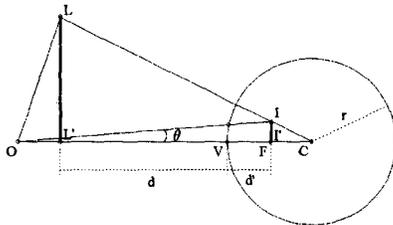


Figure 1. Image formation on a convex mirror.

Consider that the camera is calibrated (internal and exter-

nal parameters are known) and that a light source is located at a known position L . This light creates a virtual upright image I behind the cornea, and its position can be computed from the convex mirror formula for paraxial rays [5]:

$$\frac{1}{f} = \frac{2}{r} = \frac{1}{d} + \frac{1}{d'} \quad (1)$$

where the focal distance $f = |VF| = r/2$, $d = |VL|$, $d' = |VI|$, and L' and I' are the perpendicular projections of L and I respectively onto the principal axis of the mirror. For EGT applications it is reasonable to assume that the value of d is much greater than r , so that I' is formed at the focal point F of the mirror.

Now consider a second light source positioned on the optical axis of the camera, close to O . This light source generates a glint at V . Therefore, with this second on-axis light source it is possible to compute the direction of the line OV (note that O , V , F , and C are colinear), and the angle θ between the lines OV and OI (the direction OV and OI can be computed from the camera images of the glints, but not their magnitude). The lines OV and OI also define a plane which contains all the points and lines shown in Figure 1.

Now the center of the cornea can be computed as follows. Let the known coordinates of O and L be $O = (0, 0)$, and $L = (L_x, L_y)$. To estimate $C = (C_x, 0)$ note that C belongs to the line OV , and I is located at the intersection of the lines LC and OI . The equation of the line LC is defined as

$$y - L_y = (x - L_x) \frac{-L_y}{C_x - L_x} \quad (2)$$

Since I is formed halfway between V and C , it has the coordinates $(C_x - r/2, \tan \theta (C_x - r/2))$, which can be substituted in (2) so that

$$[\tan \theta (C_x - \frac{r}{2}) - L_y](C_x - L_x) = -L_y(C_x - \frac{r}{2} - L_x) \quad (3)$$

Rearranging the terms in (3), we have that the coordinate C_x is given by the smallest positive solution of the following second order polynomial:

$$C_x^2 - C_x(\frac{r}{2} + L_x) + \frac{r}{2}(L_x - \frac{L_y}{\tan \theta}) = 0 \quad (4)$$

Note that when multiple lights are available, a set of equations is obtained, which can be used to compute a more robust least square solution, and also used to compute the radius of the cornea.

2.2 Computing the 3D center of the pupil

To compute the center of the pupil, consider now the cornea as a concave spherical surface separating air (with index of refraction $i_{or} \approx 1$) from the interior of the eye (i_{or}

= 1.376). The refraction equation of spherical surfaces for paraxial rays is

$$\frac{n}{d} + \frac{n'}{d'} = \frac{n' - n}{r} \quad (5)$$

where n and n' are the indexes of refraction of the interior of the eye and air respectively, d and d' are the distances of the object and image to the vertex of the surface V , and r is the radius of curvature of the surface.

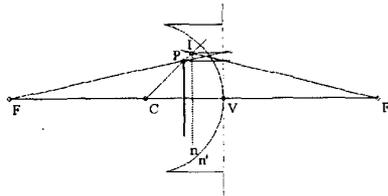


Figure 2. Refraction on concave surfaces.

The position of the image of the pupil is determined by (5). Using the values from the Gullstrand model the real pupil P is located 4.1mm behind the cornea ($d = 4.1$). Therefore the image I of the pupil is formed about 3.5mm behind the cornea, as shown in Figure 2. As a simplifying assumption, we assume that the locus of I is a sphere located 3.5mm inside the cornea.

The computation of the 3D position of the center of the pupil starts with the detection of the center of I' , the projection of I onto the camera plane. I' is then used to compute the pupil vector OI' . The 3D point I is defined by the intersection of the pupil vector and the inner sphere. We will consider that the 3D vector connecting C and I defines the direction of gaze.

In practice, the distance between the pupil and the cornea might need to be refined for each user through a procedure similar to the calibration procedure of current remote eye trackers. This "calibration" would need to be performed only once per user though, instead of once per session.

3 Simulation Results

Synthetic images of the eye of resolution 640x480 pixels were generated using ray tracing. A camera with vertical field of view 1.7 degrees was placed at the origin, as well as one light source (on-axis). To test the results for different positions of the off-axis light source, several lights were placed every 50mm to the right and above the camera, though not all of them are used to demonstrate the results. To better understand the behavior of the model presented in Section 2, the translation of the eye was limited to the principal axis of the camera, i.e., there are only depth changes of the eye (changes in d). The following examples show this parameter varying from 300 to 800mm.

To illustrate the accuracy of the spherical mirror model for the computation of the cornea center, the true 2D position of the glints in the camera image are computed and used to estimate the 3D cornea position. Figure 3 shows the estimation error of this model using the true glint coordinates for the light sources at 150 and 300mm from the camera. Since the center of the cornea is on the principal axis of the camera, the results are the same for the light sources at the right and top of the camera.

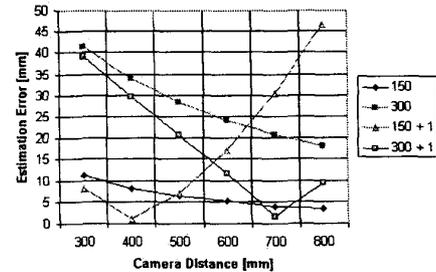


Figure 3. Results for true glint coordinates.

Because the model is valid for paraxial rays only, rays outside the paraxial region (narrow region around the principal axis, so that paraxial rays must be within this region, and be parallel or have small angles with the principal axis) do not come to focus at a common point. This phenomenon is known as spherical aberration. Therefore, it is expected that a light source closer to the camera will have smaller errors. Their glints will make very small angles with the principal axis, so that even a small glint detection deviation will cause a large estimation error. The estimation error will also depend on the depth of the cornea. The farther the cornea, the closer to paraxial will be the reflections relative to the camera, therefore smaller the error. These two effects can be clearly observed in Figure 3. The estimation error from the light sources at 150 and 300mm monotonically decrease with the camera distance, and the light at 300mm have consistently a larger estimation error than the light at 150mm. To demonstrate the influence of the errors in the detection of the glints, the lines labelled "150+1" and "300+1" in Figure 3 show the estimation error in the computation of the center of the cornea when 1 pixel deviation is added to the true position of the glints. Note that at some cases, this deviation actually makes the reflection closer to what is expected by the model, resulting in smaller estimation errors.

Figure 4 shows the estimation error for light sources at 50, 100, 150 and 300mm. The position of the glints in the ray traced images were manually extracted. The results for the lights at 150 and 300mm are very similar to the ones presented in Figure 3 using the true position of the glints. The light at 100mm generates even better results, but the

error from the light at 50mm increases with depth. This result is also expected since the glint generated by this light source makes a very small angle with the principal axis, i.e., the estimation error is mainly due to the error in detecting the correct glint position.

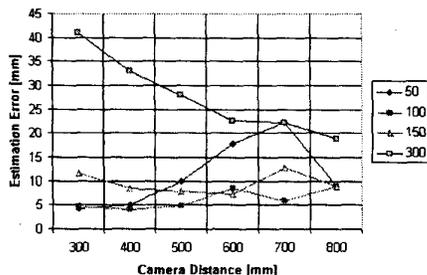


Figure 4. Results for ray tracing.

Figure 5 shows the gaze estimation error relative to the true line of sight in degrees. The simulation data was generated by fixating the gaze at the point (100 200 0), which is coplanar with the camera and light sources. The position of the glints in the ray traced images generated by the lights at 50, 100, 150, and 300mm, and the center of the pupil, were again manually extracted. The estimation errors are considerably larger when the cornea is closer to the camera, again because of spherical aberrations in the computation of the center of the pupil. For regular user-to-monitor distances (500-600mm), the average gaze estimation error is about 5 degrees. This value was also obtained for other arbitrary cornea positions (non-coplanar with the principal axis). Observe that the gaze estimation error is about the same for the lights at 50, 100, 150mm, although the error in the computation of the cornea center are different, and the light at 300mm actually gives the best results, even though it has the largest position error.

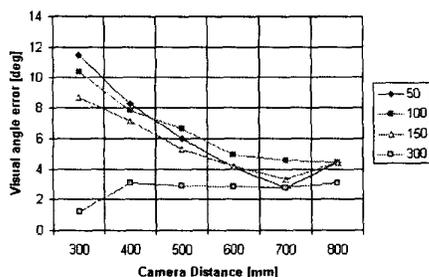


Figure 5. Results for gaze estimation.

There is obviously a tradeoff between how close to the camera the light source has to be placed, so that a dynamic selection mechanism could be used to determine which light

sources should be considered in the minimization of the estimation errors.

4 Conclusion

This paper presents a new remote eye gaze tracking (EGT) technique that estimates the 3D positions of the cornea and the pupil using one single camera and at least two light sources. More lights could be used to make the system more robust. The vectors defined by two light sources and information about the scene are used to compute the direction of gaze and the point of regard. The model is computationally very efficient and therefore appropriate for real time implementation. The technique has two main contributions over current state of the art EGTs: its ability to handle free head motion and the elimination of the need for user session calibrations. Simulations show that the accuracy of the system still have to be improved. Future work includes extensions of the model to handle non-paraxial rays, the use of more lights to increase accuracy and robustness, and a real-time prototype implementation.

Acknowledgements

We thank the Fundação ao Amparo à Pesquisa do Estado de São Paulo (FAPESP) and the Conselho Nacional de Pesquisa e Desenvolvimento (CNPq) for their financial support.

References

- [1] H. Collewijn. Eye movement recording. In R. Carpenter and J. Robson, editors, *Vision Research A Practical Guide to Laboratory Methods*, chapter 9, pages 245–285. Oxford University Press, 1998.
- [2] H. Emsley. *Visual Optics*. Butterworths, London, 5th ed., 1977.
- [3] T. Hutchinson, K. W. Jr., K. Reichert, and L. Frey. Human-computer interaction using eye-gaze input. *IEEE Trans. on Systems, Man, and Cybernetics*, 19:1527–1533, Nov/Dec 1989.
- [4] R. Jacob. The use of eye movements in human-computer interaction techniques: What you look at is what you get. *ACM Trans. on Information Systems*, 9(3):152–169, April 1991.
- [5] F. Jenkins and H. White. *Fundamentals of Optics*. McGraw-Hill, New York, NY, 4 edition, 1976.
- [6] C. Morimoto, D. Koons, A. Amir, and M. Flickner. Pupil detection and tracking using multiple light sources. *Image and Vision Computing*, 18(4):331–336, March 2000.
- [7] S. Shih, Y. Wu, and J. Liu. A calibration free gaze tracking technique. In *Proc. Int. Conf. on Pattern Recognition*, pages 201–204, Barcelona, Spain, September 2000.
- [8] L. Young and D. Sheena. Methods & designs: Survey of eye movement recording methods. *Behavioral Research Methods & Instrumentation*, 7(5):397–429, 1975.