A depth compensation method for cross-ratio based eye tracking

Flavio L. Coutinho Carlos H. Morimoto *
Computer Science Department
University of São Paulo
{flc,hitoshi}@ime.usp.br

Abstract

Traditional cross-ratio methods (TCR) project a light pattern and use invariant properties of projective geometry to estimate the gaze position. Advantages of the TCR methods include robustness to large head movements and in general requires just a one time per user calibration. However, the accuracy of TCR methods decay significantly for head movements along the camera optical axis, mainly due to the angular difference between the optical and visual axis of the eye. In this paper we propose a depth compensation cross-ratio (DCR) method that improves the accuracy of TCR methods for large head depth variations. Our solution compensates the angular offset using a 2D onscreen vector computed from a simple calibration procedure. The length of the 2D vector, which varies with head distance, is adjusted by a scale factor that is estimated from relative size variations of the corneal reflection pattern. The proposed DCR solution was compared to a TCR method using synthetic and real data from 2 users. An average improvement of 40% was observed with synthetic data, and 8% with the real data.

CR Categories: I.2.10 [Vision and Scene Understanding]: Modeling and recovery of physical attributes—Eye Gaze Tracking I.3.6 [Methodology and Techniques]: Interaction techniques—Gaze Interaction

Keywords: single camera eye gaze tracking, free head motion, depth estimation, depth compensation

1 Introduction

The cross-ratio (CR) method for remote eye gaze tracking presented in [Yoo et al. 2002] is an elegant solution to the limitations of the pupil-corneal reflection (PCR) technique. In the PCR technique, gaze is estimated using a mapping from the pupil position (relative to corneal reflections) to screen coordinates, and this map is obtained by a position dependent calibration procedure. This method does not tolerate larger head motions and require frequent system recalibration. By projecting a rectangular light pattern on the cornea surface, the CR method estimates the point of regard (POR) using invariant properties of projective geometry. This solution avoids the use of a mapping that is position dependent, allowing larger head movements. An additional advantage of the CR method is that hardware requirements are as simple as the PCR: just a single non calibrated camera with at least four additional light sources attached to the screen corners of known dimensions are used. Therefore, it avoids complex hardware setups used by techniques that compute the 3D line of sight [Beymer and Flickner 2003; Shih and Liu 2003].

The original method proposed in [Yoo et al. 2002] has low accuracy though, since it is based on a simple geometrical model. Solutions were latter proposed considering more accurate models [Yoo and Chung 2005; Coutinho and Morimoto 2006], but still lacking performance under larger movements, specially movements in depth. In this paper we propose a depth compensation cross-ratio (DCR) method, an extension to these traditional cross-ratio (TCR) methods that presents better performance in the presence of large head movements.

The following section presents the CR method in more details and identifies the limitations that motivated this work. In section 3 our proposed solution is introduced and an evaluation of it using synthetic data is carried in section 4. Experimental results are shown in section 5 and section 6 concludes the paper.

2 Overview of the CR method

The idea of the CR method [Yoo et al. 2002] is to use four light sources attached to the screen corners to generate four corneal reflections. These reflections are considered as projections of the light sources and the pupil center P is considered to be the projection of the POR, with the cornea center E as the center of projection. When corneal reflections and the pupil are captured by the camera, another projection takes place. Since the composition of two projections is also a projection, the cross-ratio property can be applied to do a reverse projection of the pupil on the image plane into the POR on the screen plane.

2.1 Limitations

Besides the great potential, large estimation errors are observed by direct application of the basic CR method described above. The work in [Guestrin et al. 2008] presents a deep investigation for the reasons for POR estimation error, formally identifying two simplifying assumptions used by the basic CR method which are not true in practice. The first assumption is that the pupil center and corneal reflection lies on the same plane, whereas pupil location is defined by its distance to the cornea center and also eye rotation. The second assumption is that the $\overrightarrow{EP}$ vector is the line of sight but, in fact, it corresponds the optical axis of the eye, which is deviated from the visual axis (the actual line of sight).

Attempts to improve accuracy were made in [Yoo and Chung 2005; Coutinho and Morimoto 2006]. Both works try to compensate these assumptions by adjusting image points prior to POR calculation. In [Yoo and Chung 2005] this adjustment is done by scaling of the corneal reflections. Each corneal reflection is scaled by an individual scale parameter obtained via calibration. The purpose of this scaling is to bring the pupil center and the corneal reflections to a common plane. No compensation due to the angular deviation of the visual axis from the optical axis is performed, however. The work in [Coutinho and Morimoto 2006] extends this solution by compensating this angular deviation with the addition of a 2D offset to estimated POR. Scaling of the corneal reflections are also

*The authors would like to thank Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) for their financial support.

Copyright © 2010 by the Association for Computing Machinery, Inc.
Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions Dept, ACM Inc., fax +1 (212) 869-0481 or e-mail permissions@acm.org.
ETRA 2010, Austin, TX, March 22 – 24, 2010.
© 2010 ACM 978-1-60558-994-7/10/0003 $10.00

performed, but a single parameter is used to scale all corneal reflections instead. A different calibration procedure was proposed to estimate both the scaling factor and the 2D offset.

This later solution achieves an accuracy of about 1^o of visual angle and tolerates head movements better than the PCR method. It also has the advantage of just requiring a one time per user calibration. However it was also shown that this solution were not able to compensate for larger user motions, specially depth motions when the user gets closer or farther to the screen.

3 Depth compensation method

First, it is important to understand why changes in depth increases estimation error. Figure 1 illustrates the origin of POR estimation error if no axis offset compensation is applied. Equations (1) and (2) describe the Δ_x and Δ_y lengths of the offset vector connecting uncorrected estimated POR to actual POR. α and β are the pan and tilt angles of the optical axis relative to the visual axis of the eye. γ and δ are the pan and tilt angles of the visual axis relative to the z axis. d is the distance of the eye to the screen along the z axis. We can see by the equations that when the eye translates in the plane parallel to the screen the offset size remains the same, but when the user translates on the direction perpendicular to the screen the offset size changes linearly. When the eye rotates the offset size also changes, but a significant contribution comes from the depth translation.

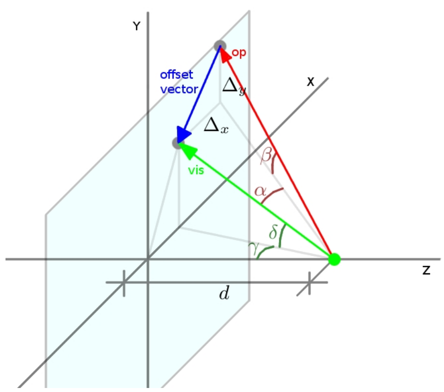

Figure 1: *source of POR estimation error due to angular offset between visual and optical axis of the eye.*

$$\Delta_x = [tan(\alpha + \gamma) - tan(\gamma)]\, d \quad (1)$$

$$\Delta_y = \left[\frac{tan(\beta + \delta)}{cos(\alpha + \gamma)} - \frac{tan(\delta)}{cos(\gamma)} \right] d \quad (2)$$

By measuring the translation on the z axis, the 2D offset vector obtained in calibration (performed at distance d_{ref}) can be scaled to its adequate size at current distance d. This solution is not ideal, since the orientation and length of the offset vector is a function of both eye distance and rotation, but it compensates the portion of the the introduced error associated with the translation in the z axis. The corrected offset vector can thus be computed by:

$$\overrightarrow{offset} = \left[\frac{d}{d_{ref}} \right] \overrightarrow{offset_{ref}} \quad (3)$$

3.1 Estimating depth change

To be able to scale the offset as described by (3), both the reference distance d_{ref} and the current distance d values are needed. Alternatively, by observing scale changes of the light pattern reflected on the cornea surface it is possible to measure the ratio d/d_{ref} without knowing absolute values of d and d_{ref}.

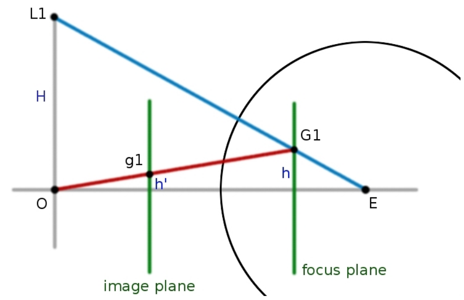

Figure 2: *corneal reflection formation for a single light source.*

Figure 2 illustrates the image formation of corneal reflection from a single light source. Let L_1 be the position of light source, O the camera's center of projection, E the cornea center and d the distance from O to E. Consider the camera with a focal length of f_{cam}, and the cornea as a spherical mirror with focal length f_{cor}. Consider also that corneal reflection G_1 is formed at the focus plane and is projected to point g_1 in the image plane. The heights h and h' (from G_1 and g_1, respectively, to camera axis) are given by:

$$h = \frac{f_{cor}H}{d}, h' = \frac{f_{cam}h}{d - f_{cor}} \quad (4)$$

Substituting equation for h in equation for h' we have:

$$h' = \frac{f_{cam}f_{cor}H}{(d - f_{cor})d} \rightarrow h' \sim \frac{f_{cam}f_{cor}H}{d^2} \quad (5)$$

Using the approximation for h' (as f_{cor} is much smaller than d) and considering two different distances d_{ref} and d with respective heights h'_{ref} and h' we have that the ratio d/d_{ref} can be computed as follows:

$$\frac{d}{d_{ref}} = \sqrt{\frac{h'_{ref}}{h'}} \quad (6)$$

This way, by measuring size variations of the pattern formed by all four corneal reflections in the image plane, it is possible to estimate relative displacements of the cornea from an initial reference position.

For the DCR method the same calibration procedure of [Coutinho and Morimoto 2006] is used, but in addition to the scale factor and the offset vector parameters, a reference pattern size $size_{ref}$ is also computed (relative to calibration distance d_{ref}, which does not need to be known). The pattern size measure is taken as the sum of the diagonal lengths of the quadrilateral formed by the four corneal reflections. Since the calibration procedure involves the user looking at different points on the screen, the average pattern size for all calibration points is taken as $size_{ref}$. During eye tracking, the ratio d/d_{ref} is computed using (6) based on current observed pattern size and $size_{ref}$. Using (3), this ratio is then used to calculate the corrected 2D offset vector that will be applied to the estimated POR.

4 Evaluation using synthetic data

To evaluate the performance of the proposed DCR method, simulations using synthetic data were carried out. Images from the LeGrand eye model [Wyszecki and Stiles 1982] generated by ray tracing were used to simulate the cornea surface, the pupil and corneal reflections. In this model the cornea is a 0.8 cm radius sphere and the pupil is a 0.2 cm radius disc located 0.45 cm from the cornea center. A unified index of refraction of 1.336 is used for the cornea and aqueous humor. The visual axis, with an angular offset of 5.5^o from the optical axis, is used to direct the eye model to the observed points. The screen is a 17 inch monitor, which corresponds to a rectangle of size 34 cm × 27 cm. The camera was positioned 2 cm bellow the middle point of the bottom edge of the screen and the eye was placed along the axis perpendicular to the screen passing through its center.

The eye was positioned at distances from 30 cm to 90 cm, at intervals of 10 cm, from the screen. At each position the visual axis of the eye was directed to the center of each cell in a 7 × 7 grid covering the screen. The distance of 60 cm was used as reference to calibrate the system using points arranged in a 3 × 3 grid. The 9 calibration points consists of the the screen center plus the corners and the middle points of each edge of the external rectangle of the 7 × 7 grid.

The gaze error is defined as the angular difference between the estimated POR to the actual POR. In addition to the DCR method, we also simulated the static offset CR method [Coutinho and Morimoto 2006] for comparison. Average gaze errors are presented in table 1. Observe that at the 60 cm calibration distance the results were the same for both methods and as the distance from the calibration position increases the difference of the gaze error becomes more noticeable. At 30 cm, the average angular error was reduced by 53% when using the DCR method, while at 90 cm a reduction of 63% was observed. The average reduction for all distances was 43.24%, showing that our depth estimation method can significantly improve POR estimation accuracy.

Depth estimation errors using the method proposed in section 3 are shown in Figure 3. Since $d_{ref} = 60$ cm, the distance d was estimated using (6) for each eye position. Errors were computed by subtracting the estimated distance from the actual value. For all eye distances, the depth estimation error observed was below 8%.

distance	static offset	DCR - dynamic offset
30 cm	3.82 ± 1.09	1.78 ± 0.65
40 cm	2.12 ± 0.65	1.00 ± 0.41
50 cm	0.95 ± 0.45	0.66 ± 0.25
60 cm	0.51 ± 0.16	0.51 ± 0.16
70 cm	0.85 ± 0.22	0.48 ± 0.20
80 cm	1.33 ± 0.20	0.54 ± 0.23
90 cm	1.73 ± 0.18	0.64 ± 0.22

Table 1: *simulation results. Average errors and standard deviation in degrees.*

4.1 Inaccuracies in feature position estimation

To observe how errors in feature position estimation affects the DCR method, a condition that can occur when using a real gaze tracker setup, the same simulations described earlier were repeated adding perturbations to the positions of the four corneal reflections created by the screen light sources. Perturbations were uniformly distributed in the range of $[-1, 1]$ pixel, and was added independently to the x and y coordinates of each corneal reflection in the image. No perturbation was added for the calibration points. This way the calibrated parameters were the same that were used in the previous simulation.

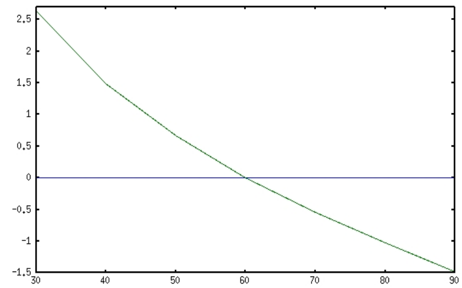

Figure 3: *depth estimation error (in cm) for the eye at distances from 30 cm to 90 cm relative to the computer screen.*

Average errors are presented in table 2. At 30 cm distance the use of the DCR method reduced the average error by 52%, similar to what was observed in the first simulation. This is explained by the fact that when the eye is closer to the screen, the corneal reflection pattern in the images appears with maximum size and the perturbation is not so significant compared to farther distances. At 90 cm a 29% reduction was observed. Average reduction for all distances was of 27.5%, a smaller value than what was observed at the first simulation, but even under perturbation the DCR method was able to improve accuracy of POR estimation. Note that as the eye moves away from the calibration position, the average error for the dynamic offset method grows at a slower rate than for the static method.

distance	static offset	DCR - dynamic offset
30 cm	3.79 ± 1.23	1.81 ± 0.77
40 cm	2.29 ± 0.78	1.19 ± 0.55
50 cm	1.21 ± 0.64	1.00 ± 0.50
60 cm	1.08 ± 0.52	1.08 ± 0.53
70 cm	1.30 ± 0.53	1.11 ± 0.52
80 cm	1.60 ± 0.73	1.10 ± 0.58
90 cm	2.15 ± 0.95	1.52 ± 0.60

Table 2: *results obtained when corneal reflections were perturbed. Average errors and standard deviation in degrees.*

Table 3 shows depth estimation results based on the reference distance of 60 cm. It is possible to see that depth estimation is quite robust to noise. It should be noted, however, that as distance increases the pattern formed by the corneal reflections get smaller in the images, making the perturbation in the $[-1, 1]$ pixel range more significant. This is observed by the increase of the standard deviation.

distance	average estimated distance
30 cm	32.6356 ± 0.0702
40 cm	41.4809 ± 0.1545
50 cm	50.7194 ± 0.2817
60 cm	59.9514 ± 0.4202
70 cm	69.4644 ± 0.6870
80 cm	79.0274 ± 1.1479
90 cm	88.7901 ± 1.4968

Table 3: *average depth estimation results and standard deviation (in cm) for inaccurate corneal reflection detection.*

5 Experimental results

To observe the real performance of the proposed method, a experiment was conducted for two subjects. Both were already familiar with gaze tracking devices and the experiment was similar to the simulations. An important difference from the simulation setup is that a fixed focus camera was used. This way, the user-camera distance remained constant (about 50 cm) when testing different user-screen distances. In fact it was the screen that was moved during experiments while the user and camera remained at a fixed position. A chin rest was used to keep users at a fixed position and the screen was placed at distances of 50, 60, 70 and 80 cm from the chin rest. Markers over a table were used to place the screen and the chin rest at the specified positions. Note that the actual user-screen distance is not the same as the chin rest-screen as the head usually rests about 2-3 cm closer to the screen. However as head position is fixed, there will always be a 10 cm translation between each screen position. The dimension of the used screen is 34 cm $\times$ 27 cm.

At each screen position the users looked at 49 tests points (center of each cell of a 7×7 grid covering the screen) and a image (640×480 pixels) was captured for each of them. Calibration was performed for the distance of 60 cm, and the 9 calibration points were extracted from the 49 test points in the same way that was done in the simulations. Due to the fact that a fixed focus camera was used, keeping the user-camera distance constant, the ratio d/d_{ref} had to be computed differently:

$$\frac{d}{d_{ref}} = \frac{size_{ref}}{size} \tag{7}$$

Average angular estimation errors are shown in table 4. When DCR method was used, an average reduction of 7.29% in the average error was observed for the first subject, while for the second subject the reduction was of 8.69%. Depth compensation improves accuracy of POR estimation compared to when no depth compensation is performed. However, observed improvement in accuracy is not as expressive as was observed in simulations, even in the simulation where inaccurate feature detection was considered.

Some reasons can be pointed for the reduced performance that was observed using a real setup. Inaccuracy in feature detection is one of them as was verified in section 4.1 and may be larger than the $[-1, 1]$ interval considered for simulations. Another reason is that cornea surface of subjects may not fit the spherical model used in simulations, specially at the region closer to the limbus. There is also the possibility of subject dependent variations of the cornea surface. All these reasons end up affecting the pattern of corneal reflections in the images from which are extracted the scale measure used to estimate depth variation. The method for depth variation calculation is also subject to some error as it is an approximation that considers a condition of paraxial rays that are not meet in practice. The orientation of the screen relative camera and the eye may also affect this approximation. Finally, we noticed that actual orientation of the 2D onscreen offset vector can also change for different eye rotations and positions but, despite the fact that the offset size is adjusted dynamically, its orientation remains fixed.

distance	static offset	DCR - dynamic offset
first subject		
50 cm	0.95 ± 0.54	0.97 ± 0.47
60 cm	0.96 ± 0.43	0.96 ± 0.41
70 cm	0.80 ± 0.31	0.62 ± 0.30
80 cm	1.14 ± 0.43	1.04 ± 0.56
second subject		
50 cm	1.33 ± 0.55	1.16 ± 0.53
60 cm	0.68 ± 0.44	0.67 ± 0.43
70 cm	1.19 ± 1.07	1.12 ± 1.06
80 cm	1.30 ± 1.10	1.11 ± 0.70

Table 4: *average gaze estimation error and standard deviation (in degrees) for both subjects.*

6 Conclusion

This paper has presented a solution to the problem shared by current cross-ratio (CR) remote eye tracking methods regarding accuracy decay with changes in eye position along the optical axis of the camera. [Coutinho and Morimoto 2006] have presented a solution that compensates the angular offset between the optical and visual axis of the eye with a fixed 2D offset vector applied directly on the onscreen gaze estimates. As the length of the 2D offset vector varies with head motion, particularly with depth, the use of a fixed vector does not suffice to accurately estimate the POR in the presence of large depth changes. Our proposed method dynamically adjusts the vector by a scale factor derived from relative depth change estimation. The main contribution of our method to current CR techniques is an 8% accuracy improvement using real user data and a 40% improvement using synthetic data. Our method is very simple to compute and does not require any extra hardware. A relevant discovery from our experimental data is that eye roll (rotation along the z axis), which was not considered in our simulation model, may be a significant source of error. Future work includes further investigation for a solution that dynamically adjusts not only the length of the 2D offset vector but also its orientation.

References

BEYMER, D., AND FLICKNER, M. 2003. Eye gaze tracking using an active stereo head. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, 451–458.

COUTINHO, F., AND MORIMOTO, C. 2006. Free head motion eye gaze tracking using a single camera and multiple light sources. In *Computer Graphics and Image Processing, 2006. SIBGRAPI '06. 19th Brazilian Symposium on*, 171–178.

GUESTRIN, E. D., EIZENMAN, M., KANG, J. J., AND EIZENMAN, E. 2008. Analysis of subject-dependent point-of-gaze estimation bias in the cross-ratios method. In *ETRA '08: Proceedings of the 2008 symposium on Eye tracking research & applications*, ACM, New York, NY, USA, 237–244.

SHIH, S., AND LIU, J. 2003. A novel approach to 3d gaze tracking using stereo cameras. *IEEE Transactions on systems, man, and cybernetics - PART B* (Jan), 1–12.

WYSZECKI, G., AND STILES, W. S. 1982. *Color Science: Concepts and Methods, Quantitative Data and Formulae*. John Wiley & Sons.

YOO, D. H., AND CHUNG, M. J. 2005. A novel non-intrusive eye gaze estimation using cross-ratio under large head motion. *Comput. Vis. Image Underst. 98*, 1, 25–51.

YOO, D. H., LEE, B. R., AND CHUNG, M. J. 2002. Non-contact eye gaze tracking system by mapping of corneal reflections. In *FGR '02: Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, IEEE Computer Society, Washington, DC, USA, 101.