Augmenting the robustness of cross-ratio gaze tracking methods to head movement

Flávio Luiz Coutinho*
Computer Science Department
University of São Paulo

Carlos H. Morimoto†
Computer Science Department
University of São Paulo

Abstract

Remote gaze estimation using a single non-calibrated camera, simple user calibration or calibration free, and robust to head movements are very desirable features of eye tracking systems. Because cross-ratio (CR) is an invariant property of projective geometry, gaze estimation methods that rely on this property have the potential to provide these features, though most current implementations rely on a few simplifications that compromise the performance of the method. In this paper, the CR method for gaze tracking is revisited, and we introduce a new method that explicitly compensates head movements using a simple 3 parameter eye model. The method uses a single non-calibrated camera and requires a simple calibration procedure per user to estimate the eye parameters. We have conducted simulations and experiments with real users that show significant improvements over current state-of-the-art CR methods that do not explicitly compensate for head motion.

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

Keywords: eye gaze tracking, cross-ratio, head movement compensation

1 Introduction

Several methods have been developed to track eye movements as described in [Hansen and Ji 2010]. Because we are particularly interested in eye trackers for interactive applications, our focus is on camera based eye tracking devices that are non intrusive, remote, low cost, and easy to setup. Camera based eye trackers capture and process images of a person's eye. During image processing, relevant eye features are detected and tracked and used to compute the point of regard (PoR). Typical eye features used are the iris and pupil contours, eye corners, and corneal reflections generated by near infrared light sources (active illumination).

Remote eye tracking methods can be classified into two groups [Hansen and Ji 2010]: interpolation based methods and model based methods. Interpolation based methods map image features to gaze points. Model based methods estimate the 3D gaze direction and its intersection with the scene geometry is computed as the PoR. Interpolation based methods have simpler requirements than model based methods but head movement is more restricted in general. Model based methods, on the other hand, offers greater freedom of movement though they require more complex setups.

*email: flc@ime.usp.br
†email: hitoshi@ime.usp.br

ETRA 2012, Santa Barbara, CA, March 28 – 30, 2012.
© 2012 ACM 978-1-4503-1225-7/12/0003 $10.00

Model based methods use geometric models of the eye to estimate the line of sight in 3D [Shih and Liu 2003; Hennessey et al. 2006; Guestrin and Eizenman 2008; Model and Eizenman 2010]. Important elements of an eye model for gaze tracking are: the centers and radii of the eyeball and cornea, modeled as spheres; the position of the foveola, the central region of the fovea; the center of the pupil; the optical axis of the eye defined by the centers of the eyeball, cornea, and pupil; and the visual axis of the eye, defined by the line connecting the foveola and the point of regard, that also passes through the center of corneal curvature. The angle between the optical and visual axis is usually referred to as the κ angle.

Most model based methods rely on stereo cameras [Nagamatsu et al. 2008; Model and Eizenman 2010], although single camera solutions have also been suggested in [Guestrin and Eizenman 2006; Hennessey et al. 2006]. In both cases, the cameras need to be calibrated and the scene geometry must be known so that the PoR can be computed. Therefore freedom of movement is achieved by an increase in complexity of system setup.

In its simplest form, the cross-ratio (**CR**) technique [Yoo et al. 2002] for eye tracking would, in theory, allow freedom of head motion, while not requiring any kind of system or user calibration. Unfortunately, it has been shown by [Guestrin et al. 2008] that simplifying assumptions have too big of an impact on the accuracy of the basic **CR** technique. Many extensions have been suggested to improve its accuracy [Yoo and Chung 2005; Coutinho and Morimoto 2006; Kang et al. 2007; Coutinho and Morimoto 2010; Hansen et al. 2010], though these methods are still sensitive to head motion.

In this paper we revisit the basic **CR** technique and explicitly model two of the main simplification assumptions of the basic technique: the assumption that $\kappa = 0$, and that the pupil and corneal reflections are coplanar. A simple calibration procedure is required per user to compute the parameters of the eye model. This method requires a single non-calibrated camera, and we show by simulation and user experiments that our new technique significantly improves the robustness to head movements.

The next section describes the **CR** technique in detail and reviews recent literature about extensions of the method. Section 3 introduces the *Planarization of the CR Features* technique (**PL-CR**) that explicitly models κ and computes the intersection of the visual axis with a virtual plane Π_G. By bringing all relevant features onto Π_G, CR can be applied for gaze estimation. The results of simulation and user experiments are given and discussed in Section 4, and Section 5 concludes the paper.

2 Cross-ratio based eye tracking

A method for remote eye gaze tracking based on the cross-ratio invariant property of projective geometry was introduced by [Yoo et al. 2002]. The method uses 4 light sources arranged in a rectangular shape, placed on a surface of interest. Typically this surface is the computer screen and each light source is attached at a screen corner. When a person faces the screen, four corneal reflections are generated on the cornea surface. These reflections, together with the observed pupil center, are then used to compute the PoR. The

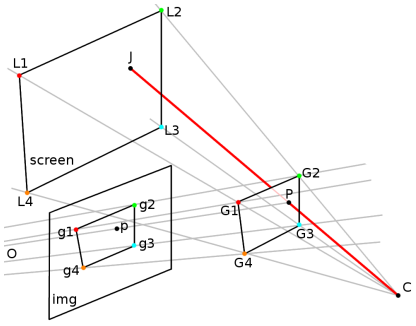

Figure 1: *Geometric setup used by the cross-ratio method for remote eye gaze tracking.*

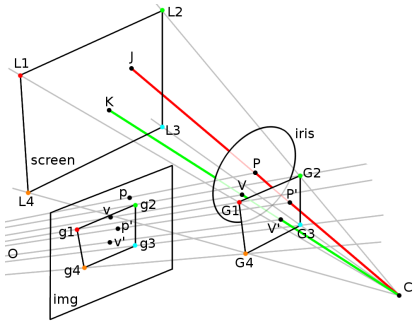

Figure 2: *Realistic geometric setup that should be considered for cross-ratio based eye gaze tracking.*

PoR is computed using cross-ratios, an invariant property of projective geometry. Figure 1 illustrates the geometric setup considered in this method, that shows the following elements: L_i, the light sources at the screen corners; G_i, the corneal reflections of L_i; g_i, the images of G_i; J, the point of regard; P, the pupil center; C, the center of curvature of the cornea; p, the image of P; O, the camera projection center.

In its basic form, the **CR** method for remote eye gaze tracking assumes g_i as projection of G_i, G_i as projection of L_i, and that each one of these groups (g_i, G_i, and L_i) is coplanar, defining the planes Π_g, Π_G and Π_L (note that Π_g is coincident with the image plane). Besides that, $p \in \Pi_g$ is the projection of $P \in \Pi_G$ and P is the projection of $J \in \Pi_L$.

Points in Π_g are the result of the composition of two projective transformations of Π_L, and therefore the composition is also a projective transformation. This way, being invariant to projective transformations, cross-ratios can be used to compute J.

Since the cross-ratio method is based on projective transformations between planes, these transformations can also be described by means of homographies. In this case, p can be expressed as: $p = \mathbf{H_2}(\mathbf{H_1}(J))$, where $\mathbf{H_1}$ is the homography that transforms points from Π_L to Π_G and $\mathbf{H_2}$ the one that transforms points from Π_G to Π_g. Homographies $\mathbf{H_1}$ and $\mathbf{H_2}$ can be combined into a single transformation $\mathbf{H}$ that directly transforms points from Π_L to points in Π_g. Matrix $\mathbf{H}$ can be estimated from the correspondence between points g_i and L_i, and J can be computed as: $J = \mathbf{H}^{-1}(p)$.

To facilitate the presentation and discussion of other gaze tracking methods based on cross-ratios we define the $\mathbf{CRf}$ function. The $\mathbf{CRf}$ function receives g_i, p, and the dimensions of the rectangle formed by L_i as inputs. It returns the point in Π_L that corresponds to p in Π_g (in other words, it returns the PoR). Since the dimensions of the rectangle formed by L_i is usually constant considering a typical gaze tracking scenario, we can drop the dimensions from the input arguments of the $\mathbf{CRf}$ function. Thus, we will define the following notation for this function (note that $\mathbf{CRf}$ can be computed using homographies and not necessarily using cross-ratios):

$$PoR = \mathbf{CRf}(g_i,\ p) \tag{1}$$

The method so far does not impose any restriction on the eye position and no previous parameter value needs to be used. It is, therefore, an elegant and simple solution that tolerates head movements and is calibration-free. Unfortunately, large gaze estimation errors are observed when this basic form of the **CR** method is used. [Guestrin et al. 2008] explain the large observed estimation error, identifying two major sources of errors which are, in fact, two simplifying assumptions that are not valid in practice. These assumptions are: first, P and G_i are coplanar; and second, $\overrightarrow{CP}$ is considered as the line of sight.

Figure 2 shows a more realistic geometric setup, that contains the following elements: L_i, the light sources placed at the screen corners; G_i, the corneal reflections of L_i; g_i, the images of G_i; C, the center of curvature of the cornea; P, the pupil center (coincident with iris center); J, the intersection between the optical axis and Π_L; P', the intersection of the optical axis with Π_G; V, intersection of the visual axis with iris disc; K, intersection of the visual axis with Π_L (the true PoR); V', intersection of the visual axis with Π_G; p, p', v, v': images of P, P', V and V'; and O, the camera center of projection.

Observing Figure 2 it is possible to notice what happens when p and g_i are directly used to compute the PoR using $\mathbf{CRf}$. First, p is the projection of P, a point that does not belong to Π_G. Consequently it is incorrect to assume that p and g_i are images of coplanar points. Besides that, the optical axis of the eye intersects the screen at J, which clearly does not correspond to the K (the true PoR). Next we describe new methods that were developed to deal with these sources of error.

2.1 CR with multiple alpha correction

[Yoo and Chung 2005] improved the **CR** method by correcting the gaze estimation error caused by the non-coplanarity of G_i and P. The PoR is computed in the following way:

$$PoR = \mathbf{CRf}(T_s(g_i,\ \alpha_i),\ p) \tag{2}$$

where T_s is a transformation defined by: $T_s(x,\ \alpha) = \alpha\,(x - g_0) + g_0$. In other words, T_s scales any image point x by α, relative to point g_0. This point is the image of the corneal reflection G_0, generated by a fifth light source that is placed near the camera's optical axis (note that when we refer to points g_i or G_i we are just considering the corneal reflections generated by the light sources attached to the screen corners). An important property of the G_0 corneal reflection, is that it belongs to line $\overline{OC}$, and as such, g_0 is the projection of C in the image plane.

The transformation of g_i by T_s is equivalent to perform scaling of G_i in space (relative to C), so that G_i and P become coplanar, and then projecting these transformed points into the image plane.

Each point g_i has its own scale factor α_i (because of this we will denote this method as the *cross-ratio with multiple alpha correction* method – **CR-M**α). These values are obtained by a calibration procedure where a person has to look at each L_i point. Each α_i is computed as: $\alpha_i = \|p_i - g_0\| / \|g_i - g_0\|$, where p_i corresponds to the projected pupil center p, when the person is gazing at L_i.

The idea behind this procedure lies in the fact that it is expected that p_i perfectly matches $T_s(g_i,\ \alpha_i)$ when the eye is gazing at L_i. A problem with this approach is that due to κ (the angle between the optical and visual axis of the eye, which is not taken into account by the method), p_i will not belong to the line $\overline{g_i g_0}$. This way, the calibrated α_i parameters may not be accurate enough to compensate the non-coplanarity of P and G_i.

2.2 CR with displacement vector correction

The *cross-ratio with displacement vector correction* (**CR-D**) method developed by [Coutinho and Morimoto 2006] is an extension **CR-M**α method [Yoo and Chung 2005], in which the error introduced due to the angle between the visual and optical axis are also compensated. For this method the PoR is computed by:

$$PoR = \mathbf{CRf}(g_i,\ T_s(p,\ \alpha)) + \vec{d} \tag{3}$$

The transformation of p using T_s can be thought of as a way to approximately compute p', the image of P', the point where the optical axis intersects the Π_G plane. Since P' and Π_G are coplanar, the first source of error of the basic **CR** method is compensated.

It is not enough, though, to correct the PoR estimation. As can be seen in Figure 2, the result of applying the function $\mathbf{CRf}$ to g_i and p' is the point J, displaced from the actual PoR K. To correctly compute K the displacement vector $\vec{d}$ must be added to J. The addition of $\vec{d}$ compensates the second source of error of the **CR** method, the displacement between J and K due to κ.

Parameters α and $\vec{d}$ are obtained by a calibration procedure where a person gazes at a set of on-screen target points. Let X be the set of n calibration points and $Y^{\alpha c}$ the set of estimated PoRs for a given αc (α candidate) without the addition of any displacement vector. Let $\Delta^{\alpha c} = \{x_i - y_i^{\alpha c} \mid x_i \in X,\ y_i^{\alpha c} \in Y^{\alpha c}\}$ be the set of displacement vectors given by the difference between calibration and estimated points. Based on the observation that for the optimum α value vectors in Δ^{α} should be approximately constant, the optimum α will be the αc value that minimizes the following summation:

$$\sum_{i=1}^{N} \parallel (x_i - y_i^{\alpha c}) - mean(\Delta^{\alpha c}) \parallel \tag{4}$$

After the α parameter is computed, $\vec{d}$ is taken as the mean vector of the Δ^{α} set.

2.3 CR with dynamic displacement vector correction

[Coutinho and Morimoto 2010] have further extended the **CR-D** method to dynamically correct the displacement vector, thus we will call it **CR-DD**. The goal of the **CR-DD** method is to improve gaze tracking accuracy under head movements, in particular depth movements of the head, the type of head movement that most affects the **CR-D** method.

If it is possible to measure the eye distance to the screen, it is possible to adjust $\vec{d}$ so that its length is adequate to the eye distance in a given moment, thus minimizing error. This solution is not ideal, since the length and orientation of $\vec{d}$ are functions of both eye distance and rotation, but it is possible to compensate a portion of the error introduced due to eye translations in z.

Consider $\vec{d_0}$ the reference displacement vector obtained by the calibration procedure of the **CR-D** method, which was executed at a reference distance z_0. As $\|\vec{d_i}\|$ is directly proportional to current distance z_i, a displacement vector $\vec{d_i}$ for an arbitrary distance z_i can be computed by:

$$\vec{d_i} = \left[\frac{z_i}{z_0}\right] \vec{d_0} \tag{5}$$

Therefore the displacement vector $\vec{d}$ can be adjusted according to the ratio z_i/z_0, which can be inferred by size variations of the quadrilateral formed by g_i.

At calibration distance z_0, besides computation of the α scale factor and the displacement vector $\vec{d_0}$, the reference size $size_0$ of the quadrilateral formed by g_i is also computed. The size of quadrilateral was taken as the sum of its diagonal lengths. After calibration of the α, $\vec{d_0}$ and $size_0$ parameters, gaze estimation is performed in the following way:

$$PoR = \mathbf{CRf}(g_i,\ T_s(p,\ \alpha)) + \sqrt{\frac{size_0}{size_i}}\ \vec{d_0} \tag{6}$$

2.4 Homography based methods

Another approach to compensate sources of errors for the basic **CR** method is to use a *homography* transformation to map the estimated gaze points (affected by both sources of errors) into the expected gaze points. This idea is presented by [Kang et al. 2007] and [Hansen et al. 2010]. In both cases the homographies used to correct the estimated gaze points are obtained by a calibration procedure, where a person has to gaze at some calibration target points.

In [Kang et al. 2007], the point of regard is computed by:

$$PoR = \mathbf{H_{LL}}(\mathbf{CRf}(g_i,\ p)) \tag{7}$$

where $\mathbf{H_{LL}}$ is a homography that transforms the estimated (incorrect) points in Π_L to expected (corrected) points in Π_L. Notice that no prior processing of the points passed as input to the $\mathbf{CRf}$ function is performed.

An advantage of the homography mapping is that there is no need for the extra light source responsible for generating corneal reflection G_0. The homography mapping can also be thought of as a generalization of the transformations realized by the **CR-M**α (scale) and **CR-D** (scale and translation) methods, being able to correct perspective distortions.

In the homography method (**HOM**) presented in [Hansen et al. 2010] the PoR is computed by:

$$PoR = \mathbf{H_{NL}}(\mathbf{CRf_N}(g_i,\ p)) \tag{8}$$

The function $\mathbf{CRf_N}$ is a variation of the $\mathbf{CRf}$ function in which the returned point is computed relative to a unitary square (normalized space), instead of being relative to the rectangle formed by L_i. The homography $\mathbf{H_{NL}}$ then transforms the estimated gaze points in normalized space to expected gaze points in screen space (Π_L).

The use of a normalized space adds another advantage to the homography method: the dimension of the rectangle formed by L_i does not need to be known. When the normalized space is not used and dimensions of L_i needs to be known, conversions between metric unit (physical size of the rectangle) and pixel unit must take place, during which eventual offsets between the L_i rectangle and useful screen area must also be taken into account. This way, the use of the normalized space facilitates implementation, by dissociation of the Π_L plane from the plane over which we want to track a person's gaze.

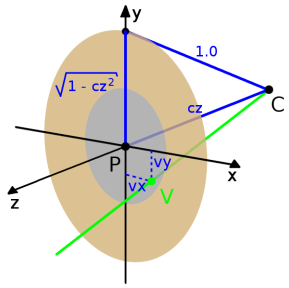

Figure 3: *Normalized eye model*

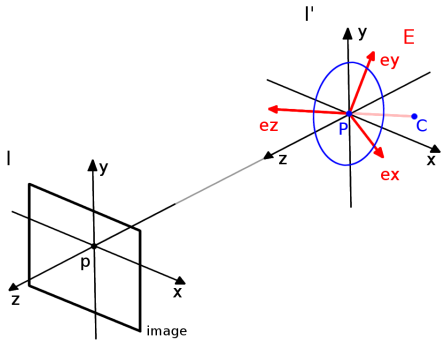

Figure 4: *Relevant coordinate systems for the **PL-CR** method.*

3 Planarization of the CR Features (PL-CR)

Recall Figure 2 that illustrates a more realistic geometric setup for the cross-ratio based methods for remote eye gaze tracking. It is straightforward to see that if v' can be computed, and v' and g_i are used to compute K using the basic cross-ratio principle then all sources of error regarding the geometric setup are eliminated. Remember that v' is the image of V', the intersection between the visual axis and Π_G. Therefore the use of V' satisfies the two simplifying assumptions assumed by the basic cross-ratio method: V' and G_i are coplanar; and V' is a point that belongs to the line of sight. Since the method brings the relevant features to a plane, we call it the *Planarization of the Cross-ratio Features* (**PL-CR**) method. For the **PL-CR** method the PoR is computed in the following way:

$$PoR = \mathbf{CRf}(g_i,\, v') \tag{9}$$

Since v' is the projection of V', which is defined by the intersection of $\overline{CV}$ with Π_G, we just have to estimate $\overline{CV}$ and Π_G. The **PL-CR** method assumes G_i to be coplanar, however we will be actually computing an approximation of Π_G which minimizes the distances from G_i to the computed plane. An orthographic camera model is assumed for the estimation of $\overline{CV}$ and Π_G.

3.1 Eye model

To estimate the visual axis in 3D space and compute its intersection with Π_G, we consider the eye model that is shown in Figure 3. This model considers the following orthonormal coordinate system: origin at the pupil/iris center P, plane xy coincident with the iris plane, with the y axis pointing in the upward direction, x in the horizontal direction and z perpendicular to the iris (corresponding to the optical axis of the eye).

Relevant points for this model are C (the center of corneal curvature) and V (the point where the visual axis intersects the iris). C belongs to the z axis and its coordinates are given by $(0, 0, -c_z)$. V belongs to the xy plane and has coordinates $(v_x, v_y, 0)$. This model has, therefore, 3 parameters (v_x, v_y and c_z) that are estimated by a calibration procedure. Similar to other gaze tracking methods, the calibration procedure consists of finding values for v_x, v_y and c_z that minimize the gaze estimation error for a set of calibration points. Since the model parameters are independent on eye location, the calibration procedure needs to be done just once per person.

This model is a normalized model where the cornea radius has a value of 1.0. This way, the iris radius is given by $\sqrt{1 - {c_z}^2}$. The use of a normalized eye eliminates the need to know absolute values of the eye structures. What is important, in this case, are the ratios between model elements.

3.2 Coordinate systems

Besides the normalized eye model, it is also important to define 3 orthonormal coordinate systems, shown in Figure 4: the image coordinate system $\mathbf{I}$ (represented by the $\mathbf{F_I}$ matrix), the translated image coordinate system $\mathbf{I}'$ (represented by the $\mathbf{F'_I}$ matrix) and the eye coordinate system $\mathbf{E}$ (represented by the $\mathbf{F_E}$ matrix).

Coordinate system $\mathbf{I}$ has its xy plane coincident with the image plane, z axis perpendicular to xy, origin in p, and units given in pixels. Coordinate system $\mathbf{I}'$ has x, y and z axis equal to those from $\mathbf{I}$, with origin in P.

Since an orthographic camera model is being used, the projection of a given point in the image plane is equivalent to the projection of the corresponding point in the plane xy of $\mathbf{I}'$. The distance between the origins of $\mathbf{I}$ and $\mathbf{I}'$ is unknown and can have an arbitrary value. We will assume $\mathbf{I}'$ to be our reference coordinate system. Estimation of the visual axis and the plane Π_G will take place relative to this reference system. This way, any point that does not have an explicit indication of a coordinate system is assumed to be relative to $\mathbf{I}'$.

Coordinate system $\mathbf{E}$ is also centered in P with its orthonormal axes defined by:

$$\vec{e_z} = \vec{n}/\|\vec{n}\| \tag{10}$$

$$\vec{e_x} = (\vec{up} \times \vec{e_z})/\|\vec{up} \times \vec{e_z}\| \tag{11}$$

$$\vec{e_y} = \vec{e_z} \times \vec{e_x} \tag{12}$$

where $\vec{n}$ is the normal to the iris (it represents the optical axis of the eye) and the $\vec{up}$ vector is a reference to the world vertical direction. Without this reference, there would be infinite possibilities for the $\vec{e_x}$ and $\vec{e_y}$ vectors of the $\mathbf{E}$ coordinate system, and consequently infinite possibilities for the V_E point when transformed to the reference coordinate system $\mathbf{I}'$.

3.3 Visual axis estimation

Estimation of the visual axis consists of finding coordinates of C and V in the reference coordinate system $\mathbf{I}'$. C and V can be computed by:

$$C = s\, \mathbf{F_E}\, C_E \tag{13}$$

$$V = s\, \mathbf{F_E}\, V_E \tag{14}$$

where $C_E = (0, 0, -c_z)$, $V_E = (v_x, v_y, 0)$ and s is a scale factor given by:

$$s = \frac{r_t}{\sqrt{1 - {c_z}^2}} \tag{15}$$

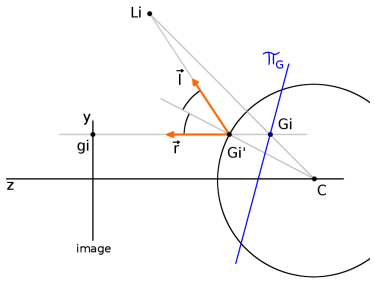

Figure 5: *Corneal reflection formation, assuming orthographic projection. Also depicted in the figure point G_i and plane Π_G.*

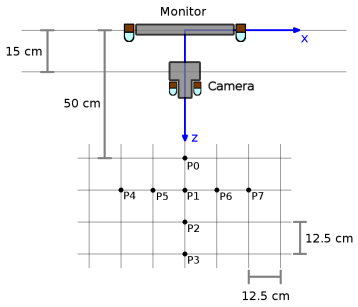

Figure 6: *Capture layout setup used for collecting simulated and real user data.*

that has the role of scaling the normalized eye model so that its dimensions match the dimensions of the eye in the image at a given instant t, with r_t being the iris radius (in pixels) at t. Because the iris can have an elliptical shape in the image, its radius is defined by the length of its major axis.

The $\vec{up}$ vector used to define the $\mathbf{E}$ coordinate system is a reference to the real world vertical direction. The $(0, 1, 0)$ vector in the $\mathbf{I}'$ coordinate system may not correspond to the real world vertical direction if the camera is pointed upwards, downwards or is rotated around its optical axis. Assuming that the screen plane is parallel to the world vertical direction, the $\vec{up}$ vector can be inferred by the positions of the L_i light sources by:

$$\vec{up} = \frac{(L_1 - L_4) + (L_2 - L_3)}{\|(L_1 - L_4) + (L_2 - L_3)\|} \tag{16}$$

3.4 Iris normal estimation

Consider the points G_0 and C and their projections on the image g_0 and c. Assuming the cornea as a spherical surface O, G_0 and C are collinear and c coincides with g_0 in the image.

The iris normal can be computed directly using $\vec{n} = P - C$, and the projection of $\vec{n}$ in the image will be $\vec{m} = p - g_0$. Assuming an orthographic camera model, then $n_x = m_x$ and $n_y = m_y$. These values can be directly extracted from the image points p and g_0. The one missing value is the n_z coordinate of $\vec{n}$, whose module is given by: $|\vec{n}| = \sqrt{{n_x}^2 + {n_y}^2 + {n_z}^2}$. Using the scale factor s previously introduced, it is also known that: $|\vec{n}| = s\,c_z$. Combining these two equations we have: $n_z = \sqrt{s^2\,{c_z}^2 - {n_x}^2 - {n_y}^2}$.

3.5 Plane Π_G estimation

We need to estimate a 3D plane where G_i can be assumed as projections of L_i, with C being the projection center. This implies that G_i belongs to the line segments $\overline{L_iC}$. Also, if G'_i is the point on the spherical cornea surface where the specular reflection due to L_i occurs, and G'_i is projected to the image as g_i, then g_i, G'_i and G_i are collinear points. This way, two lines that contains G_i are defined and G_i can be computed by their intersection (see Figure 5).

The line defined by g_i and G'_i is simple to be described considering the orthographic camera assumption. By this hypothesis, coordinates x and y of g_i, G'_i and G_i are the same, and the vector representing the reflected light ray is given by $\vec{r} = (0, 0, 1)$. The first line is then defined by: $R_i : g_i - a_i\,\vec{r}$, with x and y coordinates of g_i being extracted directly from the image. For the z coordinate any arbitrary positive value bigger than the cornea radius can be used to ensure that g_i is in front of the eye.

The second line is given by C and L_i, but just C is known. In order to define this second line, L_i must be computed as well. L_i can be expressed by the following equation: $L_i = G'_i + b_i\,\vec{l_i}$, where: $\vec{l_i}$ is the vector that corresponds to the light ray that reaches G'_i (see Figure 5); G'_i can be computed by the intersection of line R_i with the cornea surface (a sphere of radius s, centered in C); and finally, $\vec{l_i}$ can be obtained by reflecting $\vec{r}$ at G'_i.

If each equation for L_i is taken individually, it is not possible to compute L_i because b_i remains unknown at each equation. However, from knowledge of the distances between the L_i points (i.e., knowledge of the dimensions of the rectangle formed by L_i), an overdetermined system of six equations with four unknowns (b_1, b_2, b_3, and b_4) can be defined and solved.

Once L_i is computed, we are able to define $\overline{L_iC}$ and compute its intersection with R_i, thus obtaining G_i. With G_i, the plane Π_G can finally be estimated.

3.6 Estimation of v'

After computation of C, V and Π_G, estimation of v' is straightforward. First V' is computed as the intersection of $\overline{CV}$ with Π_G. Next we project V' to the image plane. Since an orthographic camera model is used $v' = (V'_x, V'_y, 0)$.

4 Experimental design and results

We compare the performance of the planarization method (**PL-CR**) with the homography normalization method **HOM**, the cross-ratio with displacement vector **CR-D** (which is an extension of the **CR-M**α), and also with the the cross-ratio with dynamic displacement vector **CR-DD** (which also explicitly compensates for some head motion).

For the simulation experiment, ray-traced images using a virtual eye based on LeGrand's eye model were generated using the layout setup displayed in Figure 6.

Two head movements are investigated, translation parallel to the monitor screen (or lateral translation), and depth translation perpendicular to the center of the screen. Therefore, the layout shown in Figure 6 correspond to a T-shape configuration with 8 positions. Depth translations are numbered 0 to 3, being P_0 the closest position and P_3 the farthest. Lateral translations were numbered from left to right, being P_4 the left most position, and P_7 the right most. Each position is distant 12.5cm from each other (within its set). Position P_1 was used to calibrate the systems, and the same calibration was used to estimate the gaze when the eye was moved to the other positions. For the simulations, the visual axis of the eye

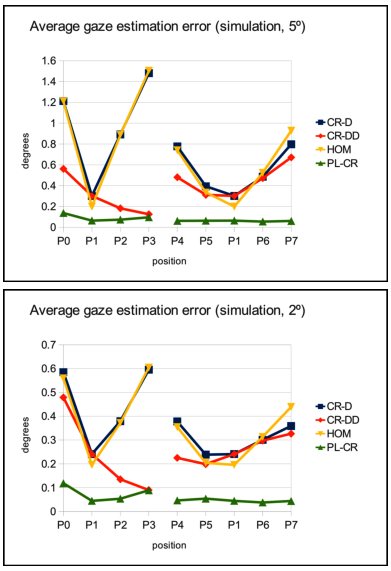

Figure 7: *Average gaze estimation error for $\kappa = 5^o$ (top) and $\kappa = 2^o$ (bottom).*

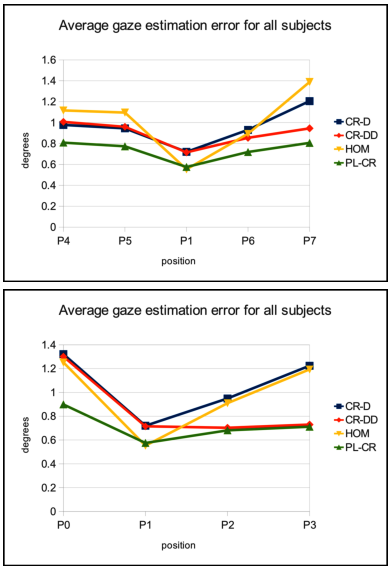

Figure 8: *Average gaze estimation computed using all 9 subjects.*

was pointed to the center points of a 7×7 grid, evenly spaced on a 17" monitor, i.e., 49 gaze points were collected at each position P_i, $i \in [0, 7]$. Light sources L_i were positioned at each screen corner. A similar setup was also used for the user experiments.

4.1 Simulation experiment results

Figure 7 shows the average estimation error in degrees for the **CR** methods **CR-D**, **CR-DD**, **HOM**, and **PL-CR**. Each graph presents the average gaze estimation error for all methods at each position. The graph's vertical axis corresponds to the visual angle error in degrees. The horizontal axis corresponds to each head position. Notice that we repeat position P_1 in both graphs so that we have two continuous ordered set of positions. This facilitates observation of translation effects in each axis individually. One set represents translations in the z axis and comprises positions P_0, P_1, P_2 and P_3. The other represents translations in the x axis and are formed by P_4, P_5, P_1, P_6 and P_7.

As expected, the head motion compensation (HMC) methods (**CR-DD** and **PL-CR**), perform better, i.e., have smaller average gaze estimation error, than non-HMC methods (**CR-D** and **HOM**) as the eye moves away from the calibration position for both simulation conditions ($\kappa = 5^o$ and $\kappa = 2^o$). The major observed difference between the two conditions is that for the smaller κ, the overall error is also smaller for all methods. For $\kappa = 5^o$, the maximum error observed for all methods and positions is 1.5^o of visual angle, while for $\kappa = 2^o$, a maximum error of 0.6^o is observed. This indicates that for subjects with smaller κ improvement for the HMC methods will be less noticeable than for subjects with larger κ values.

Between non-HMC methods it is possible to note they are more affected by depth translations (along z) than by lateral translations (along x). This is related to how κ affects gaze estimation results.

Between the HMC methods, the **PL-CR** method achieves better gaze estimation accuracy (maximum error of 0.13^o considering both conditions) than the **CR-DD** method (maximum error of 0.67^o for $\kappa = 5^o$ and 0.47^o for $\kappa = 2^o$). Results for the **PL-CR** method are also more stable across all positions when compared to the **CR-DD** method.

This difference is due to the different approaches of each method. Although the **CR-DD** method compensates head movement, the compensation applied is incomplete, as eye rotation is not taken into account. This effect can be observed by the smaller improvement of the **CR-DD** method for translations in x when compared to translations in z. Also note that even for translations in z the improvement is more significant at position P_3, where distance variation from the screen is maximum and the amount of eye rotation needed to scan through the entire screen is minimum. At this position the estimation error for the **CR-DD** method and **PL-CR** method are identical.

Our **PL-CR** method for estimation of V' (intersection of visual axis with corneal reflection plane Π_G) compensates all aspects of eye movements (position and rotation). This explains the smaller variation of gaze estimation error through all head positions.

4.2 User experiment results

A group of 9 subjects participated in the user experiment. They were all male, aged 25 to 45 years old. A chin rest fixed to a tripod was used during data collection. The monitor screen and camera was fixed, and the chin rest was moved to each location shown in Figure 6 for data collection. A software was used for the data acquisition. The software was responsible for displaying a circular target at each of the 49 test points on screen and storing the image of the subject's eye. Starting from the top left point among the 49 test points, the target was displayed in a left to right and top to bottom sequence. At each test position the target stayed for about 1.3 seconds (equivalent to 40 video frames). During this time 20 images of the subject's eye were stored. Also, during this interval, the size of the circular target varied from an initial radius of 20 pixels to a final radius of 5 pixels to serve as visual stimulus. Since multiple samples were used for each test point, the gaze estimation error for a given target point was computed as the average gaze estimation error for all samples for that target.

Figure 8 shows the average estimation error in degrees for the cross-ratio based methods **CR-D**, **CR-DD**, **HOM**, and **PL-CR** considering all subjects that participate in the user experiment. As in the simulations, we also observe that the HMC methods exhibit better performance than the non-HMC methods, as the head moves away from the calibration position P_1. This performance improvement is

clear in 7 out of the 9 participants. For one participant with a small κ angle of less than 0.5^o the performance improvement were not as noticeable.

Figures 9 and 10 show the error distribution at each position as a heat map for one of the 7 participants whose results demonstrate a significant improvement of the **PL-CR** method over the others. Axes in each heat map represent the indices of the target fixation points. From calibration, the participant have $\kappa = 2.6^o$. Hot colors (red) correspond to large errors, and cold colors (blue and green) correspond to small errors. A dark blue region correspond to eye tracking failure, where gaze was not estimated.

Figure 9 shows the error distribution of the 4 methods as the eye translates perpendicular to the screen, at the positions P_0, P_1, P_2, and P_3. Each row of Figure 9 corresponds to a position (top rows are closer to the screen), and each column shows the performance of one of the methods **CR-D**, **HOM**, **CR-DD**, and **PL-CR**. Observe that, at the calibration position (P_1) all methods have reasonable performance (small errors). At $P0$, the top row presents large errors for all methods. This can be explained because the participants are too close to the monitor, requiring larger eye rotations to gaze at the top of the screen, relative to the camera, that was placed near the bottom of the screen. At these conditions, the corneal reflections might disappear or become unstable because they are reflected from the sclera. For positions P_2 and P_3, farther from the screen and away from the calibration position, observe that the proposed method **PL-CR** outperforms all other methods.

Figure 10 shows the error distribution of the 4 methods as the eye translates parallel to the screen. Each column of Figure 10 corresponds, from left to right, to positions P_4, P_5, P_1, P_6, and P_7 (the results for P_1 are repeated for convenience, since they correspond to the calibration position). From top to bottom, the rows of Figure 10 corresponds, respectively, to the methods **CR-D**, **HOM**, **CR-DD**, and **PL-CR**. Observe again that as the head moves away from the calibration position, the error increases more significantly for all methods other than the **PL-CR**.

5 Conclusion

In this paper we introduced a novel eye gaze tracking technique that explicitly compensates the two main simplification assumptions used in basic cross-ratio methods [Yoo et al. 2002]: firstly that the κ angle between the visual and optical axis is zero, and secondly that the pupil is coplanar with corneal reflexions [Guestrin et al. 2008]. A 3 parameter eye model is used to represent κ and a virtual reflection plane Π_G. The model parameters are computed using a simple calibration procedure per user. The new method computes the intersection of the visual axis with Π_G so that, because all feature points are now coplanar, cross-ratio becomes more reliable. We have shown results using simulated data that demonstrate the robustness of the new method to head translation and rotation, and also compare its performance with other cross-ratio based methods. These results were also validated using gaze data collected from 9 volunteers. The results from simulated and real user data also indicate that the performance improvement of the new head motion compensation method is more significant for eyes with large κ values.

Acknowledgements

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.

References

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.

COUTINHO, F. L., AND MORIMOTO, C. H. 2010. A depth compensation method for cross-ratio based eye tracking. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ACM, New York, NY, USA, ETRA '10, 137–140.

GUESTRIN, E., AND EIZENMAN, M. 2006. General theory of remote gaze estimation using the pupil center and corneal reflections. *Biomedical Engineering, IEEE Transactions on 53*, 6 (june), 1124 –1133.

GUESTRIN, E. D., AND EIZENMAN, M. 2008. Remote point-of-gaze estimation requiring a single-point calibration for applications with infants. In *ETRA '08: Proceedings of the 2008 symposium on Eye tracking research & applications*, ACM, New York, NY, USA, 267–274.

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 *Proceedings of the 2008 symposium on Eye tracking research & applications*, ACM, New York, NY, USA, ETRA '08, 237–244.

HANSEN, D. W., AND JI, Q. 2010. In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence 32*, 478–500.

HANSEN, D. W., AGUSTIN, J. S., AND VILLANUEVA, A. 2010. Homography normalization for robust gaze estimation in uncalibrated setups. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ACM, New York, NY, USA, ETRA '10, 13–20.

HENNESSEY, C., NOUREDDIN, B., AND LAWRENCE, P. 2006. A single camera eye-gaze tracking system with free head motion. In *Proc. of the ETRA 2006*, 87–94.

KANG, J. J., GUESTRIN, E. D., MACLEAN, W. J., AND EIZENMAN, M. 2007. Simplifying the cross-ratios method of point-of-gaze estimation. In *30th Canadian Medical and Biological Engineering Conference (CMBEC30)*.

MODEL, D., AND EIZENMAN, M. 2010. User-calibration-free remote gaze estimation system. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ACM, New York, NY, USA, ETRA '10, 29–36.

NAGAMATSU, T., KAMAHARA, J., IKO, T., AND TANAKA, N. 2008. One-point calibration gaze tracking based on eyeball kinematics using stereo cameras. In *ETRA'08*, 95–98.

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.

YOO, D. H., AND CHUNG, M. J. 2005. A novel non-intrusive eye gaze estimation using cross-ratio under large head motion. *Computer Vision and Image Understanding 98*, 1, 25 – 51. Special Issue on Eye Detection and Tracking.

YOO, D. H., KIM, J. H., LEE, B. R., AND CHUNG, M. J. 2002. Non-contact eye gaze tracking system by mapping of corneal reflections. In *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, 94 –99.

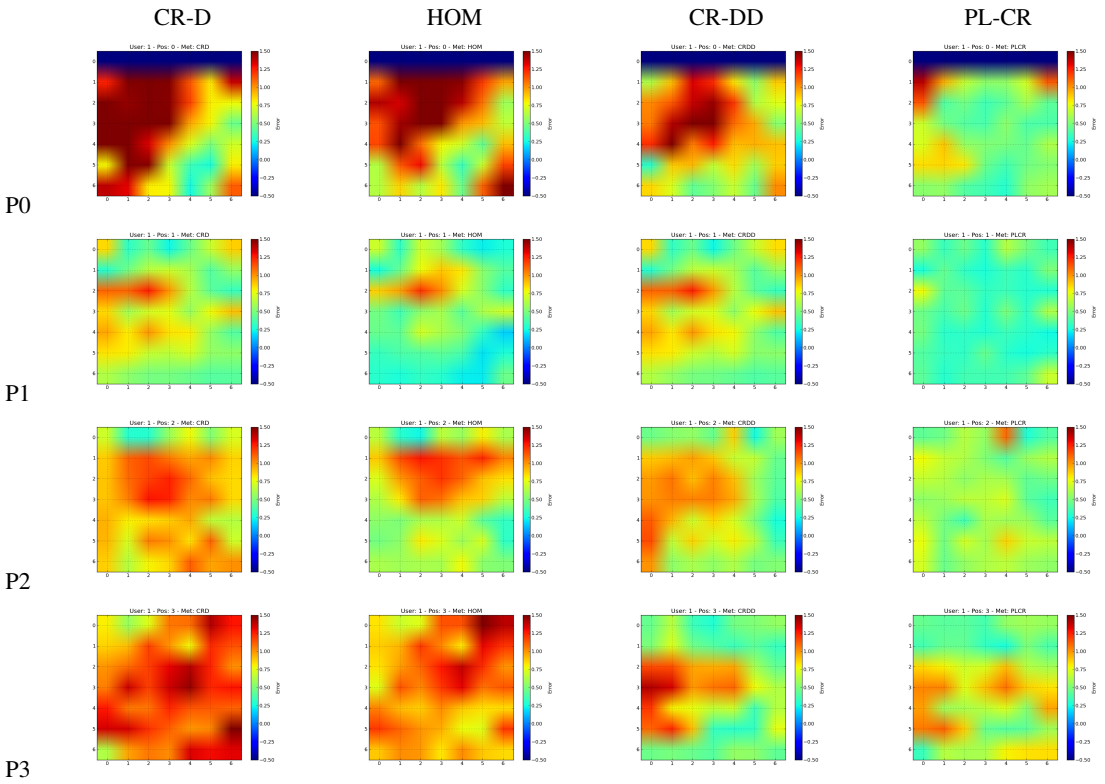

Figure 9: *Distribution of the gaze estimation error for depth translations.*

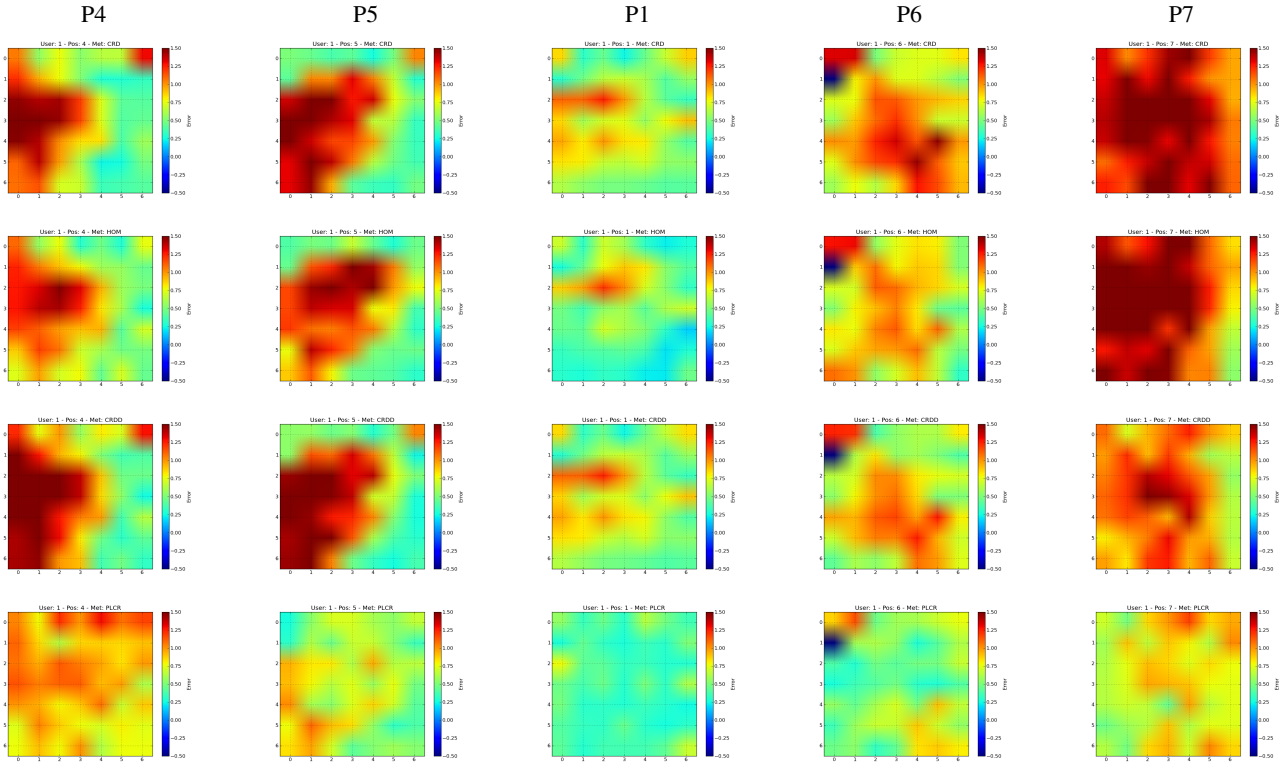

Figure 10: *Distribution of the gaze estimation error for lateral translations.*