SIMPLIFYING THE CROSS-RATIOS METHOD OF POINT-OF-GAZE ESTIMATION

J. J. Kang[1,3], E. D. Guestrin[1,3], W.J. MacLean[1], M. Eizenman[1,2,3]
University of Toronto/Department of Electrical and Computer Engineering[1], Department of Ophthalmology[2], Institute of Biomaterials and Biomedical Engineering[3], Toronto, Canada

INTRODUCTION

Point-of-gaze estimation systems estimate the point within the visual field that is imaged on the highest acuity region of the retina known as the fovea. A new method has been introduced that estimates point-of-gaze by exploiting the invariance property of cross-ratios in projective transformations [1]. The projective transformation is formed by two perspective projections. The first projection has a projective center at the center of corneal curvature, and projects four light sources placed around a plane to their corneal reflections in the eye. These corneal reflections are virtual images formed when light from the sources reflects off of the outer corneal surface, which acts as a convex mirror. The second projection is a camera projection, centered at the nodal point of a camera that images the eye, and projects the corneal reflections to their images on the camera's imaging plane. These two perspective projections form a single projective transformation relating the plane defined by the four light sources (the scene plane) to the camera's imaging plane. The same two perspective projections also project the point-of-gaze on the scene plane to the image of the pupil center on the imaging plane.

When the positions of the light sources relative to the scene plane are known (Figure 1), and the image coordinates of the corneal reflections are measured from the image of the eye (Figure 2), their cross-ratios can be calculated. These cross-ratios can then be used to map the image coordinates of the pupil center to the point-of-gaze on the scene plane.

The main advantage of this new cross-ratios method is that it does not require an accurate model of the subject's eye or the camera's optics. However, the estimation accuracy of this new cross-ratios method varies significantly from subject to subject, necessitating the use of error correction techniques to compensate for subject-specific sources of error.

Yoo and Chung [2] proposed an error correction technique that was later refined by Coutinho and Morimoto [3]. Both techniques require the placement of a fifth light source on the optic axis of the camera and a calibration procedure to establish subject-specific correction parameters. To perform calibration under the technique proposed by Yoo and Chung [2], subjects are required to fixate on each of the four original light sources. Coutinho and Morimoto [3] describe a similar calibration procedure, whereby the subjects fixate on a sequence of points in the scene

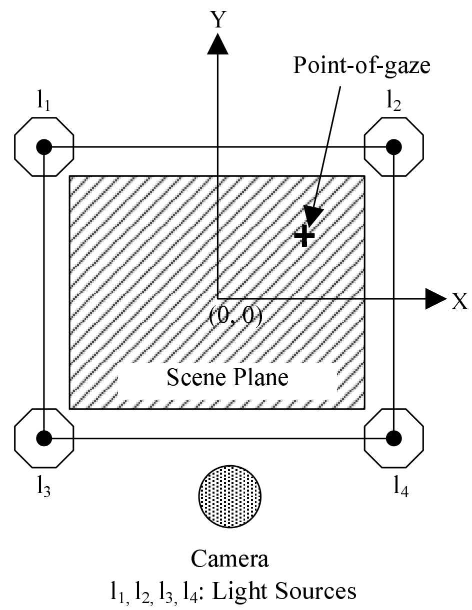

Figure 1: System configuration of the point-of-gaze estimation system

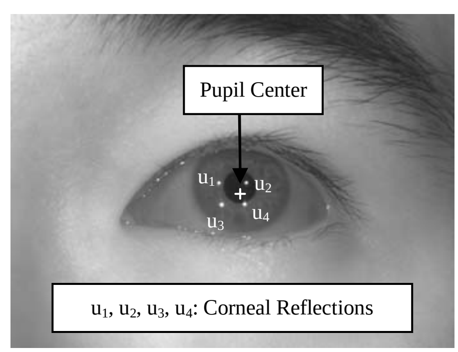

Figure 2: Image of the eye showing the corneal reflections and pupil center

plane (not necessarily corresponding to the positions of the four original light sources).

In this paper we propose a simplified error correction technique that uses a similar calibration procedure but eliminates the need for the fifth light source. We compare the performance of the new technique with the two previously described error correction techniques.

CORRECTION BY HOMOGRAPHIC MAPPING

We propose a correction technique based on a homographic mapping. The parameters of the homographic mapping are established via a calibration procedure, whereby the subject fixates on a sequence of n calibration points, $\mathbf{c}_j=[c_{jx}\ c_{jy}]^T$ for $j = 1..n$ ($n \geq 4$), located on the scene plane. For each calibration point, the point-of-gaze, $\mathbf{g}_j=[g_{jx}\ g_{jy}]^T$, is estimated using the cross-ratios method. The homographic mapping is calculated to minimize the least squares distance between $\mathbf{c}_j$ and $\mathbf{g}_j$, for the n calibration points. After calibration, point-of-gaze estimates are mapped to new positions using this calibrated homographic mapping, thereby reducing the estimation error.

The homographic mapping is described by **H**, a 3x3 homography matrix:

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}. \quad (1)$$

We solve for **H** using the Direct Linear Transformation (DLT) algorithm [4] by constructing the following equation:

$$\mathbf{Ah} = \mathbf{0}, \quad (2)$$

where **A** is given by:

$$\mathbf{A} = \begin{bmatrix} c_{1x} & c_{1y} & 1 & 0 & 0 & 0 & -c_{1x}g_{1x} & -c_{1y}g_{1x} & -g_{1x} \\ 0 & 0 & 0 & c_{1x} & c_{1y} & 1 & -c_{1x}g_{1y} & -c_{1y}g_{1y} & -g_{1y} \\ c_{2x} & c_{2y} & 1 & 0 & 0 & 0 & -c_{2x}g_{2x} & -c_{2y}g_{2x} & -g_{2x} \\ 0 & 0 & 0 & c_{2x} & c_{2y} & 1 & -c_{2x}g_{2y} & -c_{2y}g_{2y} & -g_{2y} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ c_{nx} & c_{ny} & 1 & 0 & 0 & 0 & -c_{nx}g_{nx} & -c_{ny}g_{nx} & -g_{nx} \\ 0 & 0 & 0 & c_{nx} & c_{ny} & 1 & -c_{nx}g_{ny} & -c_{ny}g_{ny} & -g_{ny} \end{bmatrix} \quad (3)$$

and **h** is a column vector containing the elements of **H**:

$$\mathbf{h} = \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \\ h_{33} \end{bmatrix}. \quad (4)$$

The solution of **h** is the null-space of **A**. In general, for noisy measurements of $\mathbf{c}_j$ and $\mathbf{g}_j$, no exact non-zero solution for **h** exists. Thus, in accordance to the DLT algorithm, we minimize the vector norm $||\mathbf{Ah}||$ with the constraint $||\mathbf{h}||$ = 1 (to avoid the trivial solution $\mathbf{h}=\mathbf{0}$). The solution to **h** is thus the unit singular vector corresponding to the smallest singular value of **A**, which is found by performing the singular value decomposition of **A**. The elements of **h** are then rearranged to reconstruct **H**.

After calibration, given a point-of-gaze estimate **G** converted to 3D homogeneous coordinates, i.e. $\mathbf{G}=[g_x\ g_y\ 1]^T$, the corrected point-of-gaze estimate $\mathbf{G_c}$ (also in homogenous coordinates) is given by:

$$\mathbf{G_c} = \mathbf{HG}. \quad (5)$$

The normalized 2D coordinates for the point-of-gaze estimate are obtained by dividing the 3D vector $\mathbf{G_c}$ by its third element, and then discarding the third element.

EXPERIMENTAL RESULTS

A point-of-gaze estimation system was implemented to evaluate the performance of the proposed correction technique. Four near-infrared (850 nm) light sources were placed around a computer screen in a rectangular formation (w=500 mm, h=320 mm). A camera was placed below the monitor to capture images of the subject's right-eye. To apply the previously proposed correction techniques [2], [3], a ring of light-emitting diodes was placed around the camera to provide a light source on the optic axis of the camera.

The subject was positioned at a distance of one meter from the monitor and the subject's head was supported by a chin-rest to keep the right eye within the camera's field of view. The subject was asked to fixate on a sequence of nine points presented on the computer screen. For each fixation point, the image coordinates of the corneal reflections and pupil center were measured for 100 consecutive image frames. The estimation rate was 30 frames per second. The measured coordinates were used to estimate the point-of-gaze using the cross-ratios method without correction. The above procedure was repeated for three subjects, and the results are shown in Figure 3.

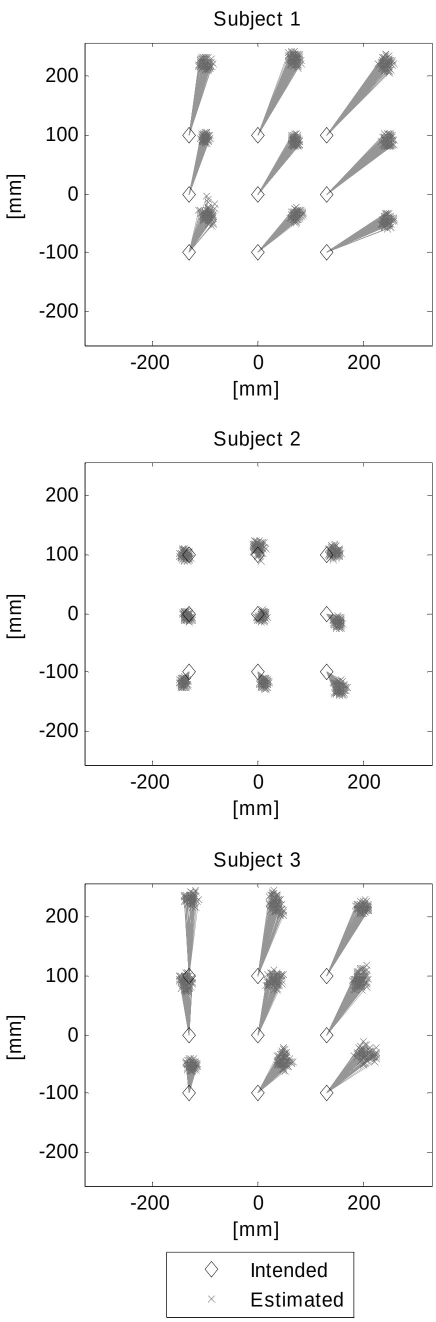

Figure 3: Point-of-gaze estimates without correction

As expected, the accuracy of the point-of-gaze estimates, without correction, varied significantly between the subjects. The RMS estimation errors for Subjects 1, 2, and 3 were 123.3 mm, 22.0 mm, and 104.5 mm, respectively. This error combines contributions from estimation bias and dispersion of the estimates around the biased mean for each fixation point. The RMS value of the dispersion varies from 6.4 mm to 9.3 mm, and is due to the noise in the measurements of the image coordinates of the pupil center and corneal reflections.

Next, the three correction techniques (Yoo and Chung [2], Coutinho and Morimoto [3], and the proposed homography technique) were each applied to the uncorrected experimental results. Each method required a calibration procedure whereby the subjects fixated on a set of calibration points. For the Yoo and Chung [2] correction method, the calibration points were the four light sources placed at the corners of the scene plane. For the Coutinho and Morimoto [3] technique and the homography technique, four calibration points located on the scene plane were used: $\mathbf{c}_j=[\pm 130\ \pm 100]^T$. Figure 4 shows the point-of-gaze estimates after applying the new homography technique. Table 1 presents the RMS error of the corrected point-of-gaze estimates obtained using each technique. The results show that the performance of the new homography technique is better than the previously proposed techniques. Furthermore, the RMS estimation error for each subject after correction using the homography technique is within 15% of the dispersion error ascribed to measurement noise, suggesting that the estimation bias has been mostly eliminated.

Table 1: Comparison of RMS estimation error for the different correction techniques

Correction Methodology	**RMS Estimation Error (mm)**		
	Subject 1	***Subject 2***	***Subject 3***
Uncorrected	123.3	22.0	104.5
Yoo and Chung	23.5	28.3	23.6
Coutinho and Morimoto	9.0	9.8	12.7
Homography	7.2	8.7	10.3

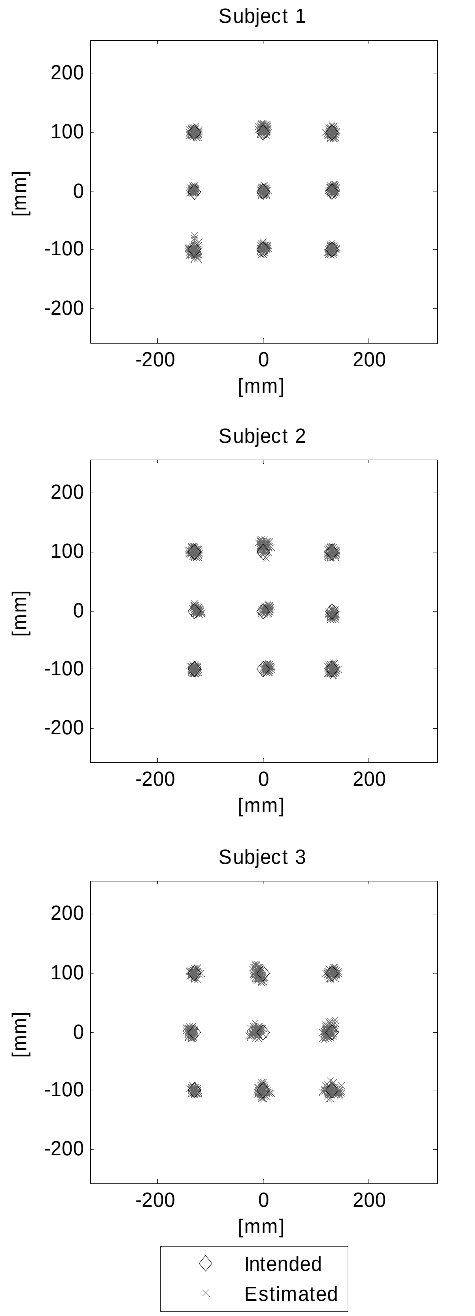

Figure 4: Point-of-gaze estimates after correction using the homography technique

CONCLUSIONS

In this paper, we presented a new correction technique for the cross-ratios method of point-of-gaze estimation based on a homographic mapping. The calibration procedure is very similar to those employed by current correction techniques reported in literature, and requires the subject to fixate on four calibration points. The main advantages of the new homography technique are: (1) reduced point-of-gaze estimation error and (2) a simplified system configuration that eliminates the need for a light source to be placed on the optic axis of the camera. For point-of-gaze estimates on a computer screen positioned one meter away from the subjects, the maximum RMS estimation error after applying the homography technique was 10.3 mm for the three tested subjects.

REFERENCES

[1] D. H. Yoo, J. H. Kim, B. R. Lee, and M. J. Chung, "Non-contact eye gaze tracking system by mapping of corneal reflections," in Proc of *Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 94-99, 2002.

[2] D. H. Yoo and M. J. Chung, "A novel non-intrusive gaze estimation using cross-ratio under large head motion," *Computer Vision and Image Understanding*, vol. 98, no. 1, pp. 25-51, 2005.

[3] F.L. Coutinho and C.H. Morimoto, "Free head motion eye gaze tracking using a single camera and multiple light sources," *Brazilian Symposium on Computer Graphics and Image Processing*, pp. 171-178, 2006.

[4] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd Edition, Cambridge University Press, Cambridge, 2003.