USER-CALIBRATION-FREE REMOTE EYE-GAZE TRACKING SYSTEM WITH EXTENDED TRACKING RANGE

Dmitri Model[1], and Moshe Eizenman[1,2,3]

[1]Department of Electrical and Computer Engineering, [2]Institute of Biomaterials and Biomedical Engineering, [3]Department of Ophthalmology and Vision Sciences, University of Toronto

ABSTRACT

A novel general method to extend the tracking range of user-calibration-free remote eye-gaze tracking (REGT) systems that are based on the analysis of stereo-images from multiple cameras is presented. The method consists of two distinct phases. In the brief initial phase, estimates of the center of the pupil and corneal reflections in pairs of stereo-images are used to estimate automatically a set of subject-specific eye parameters. In the second phase, these subject–specific eye parameters are used with estimates of the center of the pupil and corneal reflections in images from any one of the systems' cameras to compute the Point-of-Gaze (PoG). Experiments with a system that includes two cameras show that the tracking range for horizontal gaze directions can be extended from ±23.2° when the two cameras are used as a stereo pair to ±35.5° when the two cameras are used independently to estimate the POG.

Index Terms— Eye Tracking, Remote Gaze Estimation, Extended Range, Calibration-Free, Distributed Eye-gaze Tracker.

1. INTRODUCTION

Gaze estimation systems (eye-trackers) are used in a myriad of application. Some applications, such as pilot training [1], driving safety research [2-4] and virtual environments [5, 6] require gaze estimation relative to some fixed objects or surfaces under a wide range of head movements and gaze directions. Extended tracking range for these applications is usually achieved by using head-mounted eye-trackers that require user-calibration procedures combined with a head tracker [7, 8]. In applications for which the use of user calibration procedures and/or head-mounted gear is not desirable, such as interactive advertising [9, 10] or interactive museum exhibits [11], user-calibration-free Remote Eye-Gaze Tracking (REGT) systems with extended tracking range can provide a feasible solution.

The current state-of-the-art user-calibration-free REGT systems are based on the estimation of the center of the pupil and corneal reflections (virtual images of light sources that illuminate the subject's face that are created by the cornea) in pairs of images taken by a stereo pair of video cameras [12-14]. These systems can estimate the PoG accurately over a limited range of gaze directions, since the model used for gaze estimation is valid only for a limited range of angles between the subject's direction of gaze and the optical axis of each camera (typically within ±30°) [12, 15]. Therefore, in a two-camera system, the range of gaze directions is limited by the *larger* of the two angles between the subject's direction of gaze and the optical axes of the two cameras (typical tracking range of ±20°). As such, calibration free REGT systems are not suitable for applications that require the estimation of the PoG on multiple monitors, on big murals in a museum, or in studies of eye-misalignment in babies [16, 17] in which the angle between two eyes can often exceed 40°.

This paper describes a new method to estimate the PoG that combines the capacity to estimate the PoG without explicit user-calibration procedures with the extended tracking range of two independent cameras (for two independent cameras the tracking range is limited by the *smaller* of the angles between the subject's direction of gaze and the optical axes of the two cameras). The method consists of two steps. In the first step, the coordinates of eye features (center of the pupil and corneal reflections) in pairs of stereo-images are used to estimate *automatically* a set of subject-specific eye parameters. In the second step, the subject–specific eye parameters are used with estimates of eye-features from *any one* of the images from the systems' cameras to compute the point-of-gaze. This allows for an extension of the tracking range while the system remains calibration-free for the user.

The paper is organized as follows. The proposed approach is presented in Section 2. Experimental results are presented in Section 3. Finally, discussion and conclusions are presented in Section 4.

2. METHOD

In the following analysis, all points are represented as 3-D column vectors (bold font) in a right-handed Cartesian

World-fixed Coordinate System (WCS).

Fig. 1 shows an eye model for the estimation of the point-of-gaze. The front surface of the cornea is modeled as a spherical section. The line connecting the center of curvature of the cornea, **c**, and the pupil center, **p**, defines the optical axis of the eye (**ω**). The line connecting the fovea with **c** defines the visual axis of the eye, **v**, and the angle between the visual and optical axes is κ (angle kappa).

From Fig. 1, the point of gaze, **g**, can be written as:

$$\mathbf{g} = \mathbf{c} + \mu\mathbf{R}(\mathbf{p} - \mathbf{c}) \quad (1)$$

where μ is a line parameter, proportional to the distance from the eye to the monitor and **R** is a rotation matrix which depends on the angle κ.

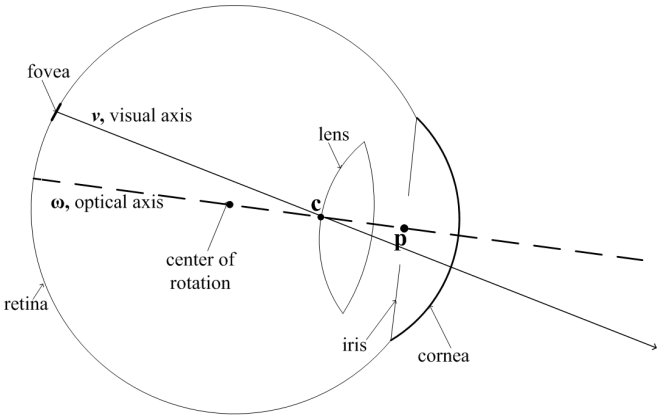

Fig. 1. A simplified schematic diagram of the eye. The *optical* axis, **ω** (the axis of symmetry of the eye) passes through the center of curvature of the cornea, **c**, and the center of the pupil, **p**. The *visual* axis of the eye (the line-of-sight), **v**, connects the fovea (the region of the highest visual acuity on the retina) with **c**. The point-of-gaze, **g**, is given by the intersection of the visual axis with the scene.

As was shown in [12, 18], **c** and **p** can be estimated without any subject calibration procedure using a pair of images from a stereo pair of video cameras. Let's denote these estimates $\mathbf{c}_2$ and $\mathbf{p}_2$, respectively. To estimate **c** and **p** using features detected in an image of a single camera, the knowledge of the radius of curvature of the cornea, r, and the distance between **c** and **p**, d, is required [12]. Let's call these estimates $\mathbf{c}_1(r)$ and $\mathbf{p}_1(r,d)$, respectively.

The optimal values $\hat{r}$ and $\hat{d}$ for parameters r and d, respectively, are those values that minimize the difference in the point-of-gaze estimates from a single image ($\mathbf{g}_1$) and a stereo-pair of images ($\mathbf{g}_2$). $\hat{r}$ and $\hat{d}$ can be obtained by solving the following optimization problem:

$$\left[\hat{r},\hat{d}\right] = arg\min \sum_{t=\mathbf{t}} \left\| \mathbf{g}_2 - \mathbf{g}_1(r,d) \right\|_2^2 \quad (2)$$

where the summation is for time instances, **t**, during which $\mathbf{g}_2$ is available.

After substitution of (1) into (2):

$$\left[\hat{r},\hat{d}\right] = arg\min \sum_{t=\mathbf{t}} \left\| \begin{array}{l} \left(\mathbf{c}_2 - \mathbf{c}_1(r)\right) + \\ \mu_t \mathbf{R}\left(\left(\mathbf{p}_2 - \mathbf{c}_2\right) - \left(\mathbf{p}_1(r,d) - \mathbf{c}_1(r)\right)\right) \end{array} \right\|_2^2 \quad (3)$$

As $\mu \gg 1$, and $\mathbf{R}^{\mathrm{T}}\mathbf{R}=\mathbf{I}$ (where **I** is the identity matrix and $^{\mathrm{T}}$ denotes transpose), the term $\left(\mathbf{c}_2 - \mathbf{c}_1(r)\right)$ can be neglected and (3) can be approximated by:

$$\left[\hat{r},\hat{d}\right] \cong arg\min \sum_{t=\mathbf{t}} \left\| \left(\mathbf{p}_2 - \mathbf{c}_2\right) - \left(\mathbf{p}_1(r,d) - \mathbf{c}_1(r)\right) \right\|_2^2 \quad (4)$$

Since the stability of the solution to the minimization problem in (4) can be affected by the presence of outliers (e.g., due to blinks), outlier estimates of **p** and **c** have to be removed prior to the minimization procedure. Instead of using the computationally demanding iteratively re-weighted least squares method to remove outliers [19] (which requires a repeat of the optimization procedure for the entire set of available estimates), a more computationally efficient approach was developed. The approach is based on the fact that it is possible to estimate r_t and d_t from a single stereo-pair of images collected at a time instance t right after the images become available (there is no need to wait until the images for all the time instances are collected). After the collectiodmitryn of all T samples is complete, a histogram-based outlier removal procedure is applied. Algorithm I summarizes the estimation procedure for r and d.

ALGORITHM I
Estimation of Subject-Specific Parameters: r and d

1. Estimate r_t and d_t, by minimizing (4) for a single time sample t. This can be done immediately after the data for each time sample t becomes available.
2. Repeat Step 1 until T samples are collected.
3. Find time instances for inliers $\mathbf{t}_{\mathrm{in}}$, and the average value for inliers $\bar{r}_{\mathrm{in}}$ and $\bar{d}_{\mathrm{in}}$:
   a. Find time instances for inliers in r, $\mathbf{t}_{\mathrm{r}}$, and the average value, $\bar{r}_{\mathrm{in}}$, of all the inliers:
      i. Partition all r_{t}, into bins of width W, e.g., $W = 0.1$ mm (the actual value of W depends on the noise level in the REGT system).
      ii. Select the bin with the largest count of data points.
      iii. Calculate the average value, $\bar{r}$, of all the data points in the selected bin and two adjacent bins.
      iv. The inliers are r_{t} that satisfy $|\bar{r} - r_t| < W$.
      v. Calculate the average value of all the inliers, $\bar{r}_{\mathrm{in}}$.
   b. Find all time instances for inliers in d, $\mathbf{t}_{\mathrm{d}}$, and the average value, $\bar{d}_{\mathrm{in}}$, of all the inliers using a procedure similar to the one described in Step a.
   c. The time instances for inliers are given by: $\mathbf{t}_{\mathrm{in}} = \mathbf{t}_{\mathrm{r}} \cap \mathbf{t}_{\mathrm{d}}$
4. Re-optimize (4) using all inliers at once ($\mathbf{t}_{\mathrm{in}}$) and $\bar{r}_{\mathrm{in}}$ and $\bar{d}_{\mathrm{in}}$ as an initial guess to obtain the optimal values for $\hat{r}$ and $\hat{d}$.

In parallel to the execution of Algorithm I, the angle between the optical and visual axes (κ) can be estimated from stereo-images without explicit user-calibration procedure by following one of the techniques described in [13, 14, 16].

After the estimation of full set of subject-specific eye-parameters (r,d and κ), the PoG can be calculated by using eye features from a single camera [12].

3. EXPERIMENTS

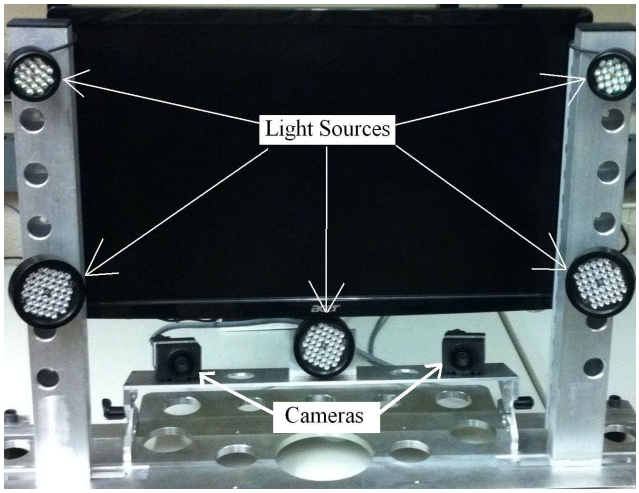

Fig. 2. Prototype REGT system.

The experiments were carried out with the two-camera system shown in Fig. 2. The two cameras were calibrated as a stereo pair, and were used either as a two-camera stereo system (two cameras used as a stereo-pair) or 2 independent one-camera systems (referred to as 'split-mode two-camera system' hereafter). The horizontal angle between the optical axes of the two cameras was 22°.

Performance evaluation was carried out with 3 adult subjects. During the experiments, subjects sat at approximately 70 cm from the system. Subject-specific eye parameters (r and d) were estimated using Algorithm I with T = 50. Angle kappa was estimated using an implicit calibration procedure [16].

To estimate the horizontal gaze tracking range and the RMS error in PoG estimation, subjects were asked to look at a grid of 33 fixation points, arranged in 3 rows (100 mm apart) and 11 columns (100 mm apart), as shown in Fig. 3. This grid of fixation points spanned ±35.5° of horizontal gaze directions. Fifty PoG estimates were collected for each fixation point. The experiment was repeated twice, once in the standard two-camera stereo mode [14] and once in the split-mode. The same system components (camera, light sources) were used for both modes. The results of the experiments are summarized in Table I.

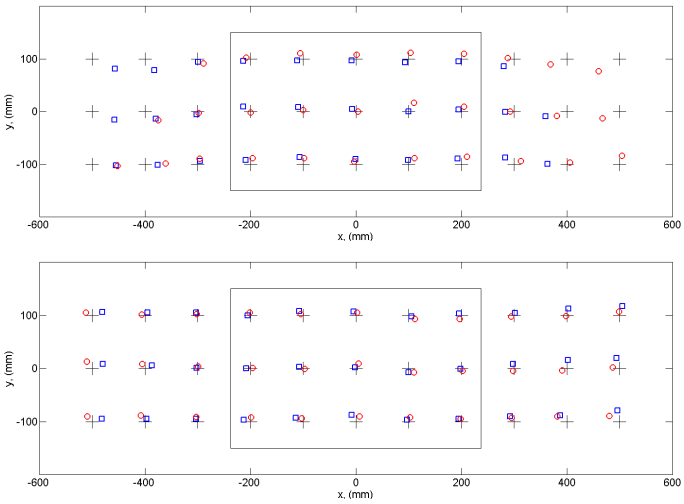

Fig. 3. Point-of-gaze estimation with subject 1 using a two-camera stereo system (top) and a split-mode system (bottom). Crosses indicate the actual positions of the fixation targets; small squares indicate average PoG estimates of the right eye; small circles indicate average PoG estimates of the left eye. The rectangle in the center indicates the viewing area of a wide-screen 22" monitor.

TABLE I
RMS ERROR IN POINT-OF-GAZE ESTIMATION (MM)

	Two-Camera Stereo System (Screen Area)	Split-mode Two-Camera System (Screen Area)	Split-mode Two-Camera System (Off-Screen Area)
Subject 1	13.4	10.4	16.6
Subject 2	12.5	14.6	20.2
Subject 3	15.5	13.1	18.3
Average	**13.8**	**12.7**	**18.4**

Fig. 3 shows the results of the experiments with subject 1. As one can see from Figure 3 and Table I, in the central area (a 22" computer monitor) both systems can estimate the PoG with comparable RMS errors.

As expected, the accuracy of gaze estimation in the stereo mode deteriorates when one of the angles between the optical axis of the subject's eye and the optical axes of the systems' cameras exceeds ≈35° (this angle is reached when the subject looks ≈24° to the left or to the right of the primary position). The deterioration in accuracy is due to the fact that as the angle between the optical axis of the subject's eye and the optical axis of the system's cameras increases, the corneal reflections tend to be formed on the peripheral (non-spherical) part of the cornea. Since the eye model for gaze estimation assumes that corneal reflections are created by a spherical section of the cornea [12, 18], corneal reflections that are formed by non-spherical sections of the cornea can introduce large biases in the estimation of the PoG. In the 'split mode', accurate PoG estimates for the full range of fixation points could be obtained, since the angle between the direction of the optical axes of the

subject's eyes and one of the system's cameras never exceeds 25°.

As can be seen from Fig. 3, the tracking range of a stereo system is limited to about ±300 mm horizontally (or, equivalently, ±23.2°), whereas a split-mode system can handle gaze direction of up to ±35.5° (±500 mm) horizontally. The estimates at ±500 mm of a split mode system suffer from increased noise, but the bias is still less than 2°.

4. DISCUSSION AND CONCLUSIONS

A novel approach for user-calibration-free REGT system with extended tracking range has been presented. During a brief initial start-up phase, a stereo pair of cameras with overlapping fields of view is used to estimate subject-specific eye-parameters without any explicit user-calibration procedure. These eye parameters are then used with images from any of the system's cameras ('split' mode) to estimate the point-of-gaze. Therefore, no user calibration is required.

As was demonstrated by the experiments (see Table I), there is no deterioration in the accuracy of PoG estimation when the system is used in a split mode compared to the original two-camera stereo system. As expected, the split-mode system enables tracking over a larger range of gaze directions. In essence, by adopting the suggested approach, one can extend the tracking range of REGT system from the area of a single monitor to the area of two monitors without adding any new hardware.

Finally, the suggested approach can be readily extended from 2 cameras to *N* cameras, as long as at least 2 of the cameras have overlapping fields of view. Adding more cameras to the system will extend the tracking range without a need for any subject calibration procedures. Thus, the suggested approach enables a scalable, user-calibration-free, distributed eye-gaze tracking system.

ACKNOLEDGEMENTS

This work was supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC), and in part by scholarships from NSERC and the Vision Science Research Program Award (Toronto Western Research Institute, University Health Network, Toronto, ON, Canada).

REFERENCES

[1] P. A. Wetzel, G. Krueger-Anderson, C. Poprik, and P. Bascom, "An eye tracking system for analysis of pilots' scan paths," United States Air Force Armstrong Laboratory Tech. Rep. AL/HR-TR-1996-0145, Apr. 1997.

[2] L. N. Boyle and J. D. Lee, "Using driving simulators to assess driving safety," *Accident Analysis & Prevention,* vol. 42, pp. 785-787, 2010.

[3] O. Palinko, A. L. Kun, A. Shyrokov, and P. Heeman, "Estimating cognitive load using remote eye tracking in a driving simulator," presented at the Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications, Austin, Texas, 2010.

[4] J. L. Harbluk, Y. I. Noy, P. L. Trbovich, and M. Eizenman, "An on-road assessment of cognitive distraction: Impacts on drivers' visual behavior and braking performance," *Accident Analysis & Prevention,* vol. 39, pp. 372-379, Mar. 2007.

[5] C. Cruz-Neira, D. J. Sandin, and T. A. DeFanti, "Surround-screen projection-based virtual reality: The design and implementation of the cave," presented at the Proceedings of the 20th annual conference on Computer graphics and interactive techniques, Anaheim, CA, 1993.

[6] D. A. Southard, "Viewing model for virtual environment displays," *Journal of Electronic Imaging,* vol. 4, pp. 413-420, 1995.

[7] R. S. Allison, M. Eizenman, and B. S. K. Cheung, "Combined head and eye tracking system for dynamic testing of the vestibular system," *IEEE Transactions on Biomedical Engineering,* vol. 43, pp. 1073-1082, Nov 1996.

[8] J. Barabas, R. B. Goldstein, H. Apfelbaum, R. L. Woods, R. G. Giorgi, and E. Peli, "Tracking the line of primary gaze in a walking simulator: Modeling and calibration," *Behavior Research Methods, Instruments, & Computers,* vol. 36, pp. 757-770, November 2004 2004.

[9] J. v. M. Hamburg, "Amnesty international domestic violence ad," http://blogs.amnesty.org.uk/blogs_entry.asp?eid=3460, 2009.

[10] B. Knep, "Big smile," http://www.blep.com/bigSmile/index.htm, 2003.

[11] S. Milekic, "Gaze-tracking and museums: Current research and implications," presented at the Museums and the Web, Denver, Colorado, USA, 2010.

[12] E. D. Guestrin and M. Eizenman, "General theory of remote gaze estimation using the pupil center and corneal reflections," *IEEE Transactions on Biomedical Engineering,* vol. 53, pp. 1124-1133, Jun. 2006.

[13] T. Nagamatsu, J. Kamahara, and N. Tanaka, "Calibration-free gaze tracking using a binocular 3d eye model," in *Proceedings of the 27th International Conference on Human factors in Computing Systems*, Boston, MA, USA, 2009, pp. 3613-3618.

[14] D. Model and M. Eizenman, "An automatic personal calibration procedure for advanced gaze estimation systems," *Biomedical Engineering, IEEE Transactions on,* vol. 57, pp. 1031-1039, May 2010.

[15] A. Villanueva and R. Cabeza, "Models for gaze tracking systems," *J. Image Video Process.,* vol. 2007, pp. 1-16, 2007.

[16] D. Model and M. Eizenman, "An automated Hirschberg test for infants," *Biomedical Engineering, IEEE Transactions on,* vol. 58, pp. 103-109, 2011.

[17] D. Model, M. Eizenman, and V. Sturm, "Fixation-free assessment of the Hirschberg ratio," *Investigative Ophthalmology & Visual Science,* vol. 51, pp. 4035-4039, August 1, 2010 2010.

[18] S. W. Shih and J. Liu, "A novel approach to 3-d gaze tracking using stereo cameras," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics,* vol. 34, pp. 234-245, Feb. 2004.

[19] P. W. Holland and R. E. Welsch, "Robust regression using iteratively reweighted least-squares," *Communications in Statistics: Theory and Methods,* vol. A6, pp. 813-827, 1977.