A General Framework for Extension of a Tracking Range of User-Calibration-Free Remote Eye-Gaze Tracking Systems

Dmitri Model
University of Toronto
dmitry.model@gmail.com

Moshe Eizenman
University of Toronto
eizenm@ecf.utoronto.ca

Abstract

Stereo-camera Remote Eye-Gaze Tracking (REGT) systems can provide calibration-free estimation of gaze. However, such systems have a limited tracking range due to the requirement for the eye to be tracked in both cameras. This paper presents a general framework for extension of a tracking range of stereo-camera user-calibration-free REGT systems. The proposed method consists of two distinct phases. In the brief initial phase, estimates of eye-features [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 eye-features in images from *any* one of the systems' cameras to compute the Point-of-Gaze (PoG). Experiments were conducted with a system that includes two cameras in a horizontal plane. The experimental results demonstrate that the tracking range for horizontal gaze directions can be extended by more than 50%: 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. By adding more cameras to the system, the proposed framework allows further extension of the tracking range in both horizontal and vertical direction, while preserving a user-calibration-free status of a REGT system.

Keywords: *Eye Tracking, Remote Gaze Estimation, Extended Range, Calibration-Free, Distributed Eye-gaze Tracker.*

1 Introduction

There are many applications that require gaze tracking relative to fixed objects in space under a wide range of gaze directions. Extended tracking range can be achieved by using head-mounted eye-tracker combined with a head tracker [Allison et al. 1996]. Head-mounted eye-trackers, however, require extensive user-calibration procedures and are invasive (require head gear). In applications for which user calibration procedures and/or head-mounted gear is not desirable (e.g., covert monitoring or studies with infants), 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 (glints) in pairs of images taken by a stereo pair of video cameras [Guestrin and Eizenman 2006; Model and Eizenman 2010a; Model and Eizenman 2010b; Nagamatsu et al. 2010]. 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°) [Guestrin and Eizenman 2006; Villanueva and Cabeza 2007]. In a stereo-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, and the typical tracking range is limited to ±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 [Model 2011; Model et al. 2011; Model and Eizenman 2011a].

This paper describes a general framework that allows extension of a tracking range of a user-calibration-free REGT system. The proposed method can be applied to any REGT system that consists of two or more cameras, with at least 2 of these cameras calibrated as a stereo pair and having a [partially] overlapping field of view. The tracking range can be increased to the union of tracking ranges of each individual camera using the following two-step procedure. 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 preliminary version of this paper has been presented in [Model and Eizenman 2011b]. This paper provides an improved mathematical formulation that enhances the convergence of optimization procedure (Section 2), as well as experimental results with more subjects (Section 3).

2 Method

Figure 1 shows an eye model for the estimation of the point-of-gaze[1]. 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 Figure 1, the point of gaze, **g**, can be written as:

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

[1] In this paper, all points are represented as 3-D column vectors (bold font) in a right-handed Cartesian World Coordinate System (WCS).

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

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

As was shown in [Guestrin and Eizenman 2006; Shih and Liu 2004], $\mathbf{c}$ and $\mathbf{p}$ 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 $\mathbf{c}_{2CAM}$ and $\mathbf{p}_{2CAM}$, respectively. To estimate $\mathbf{c}$ and $\mathbf{p}$ 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 $\mathbf{c}$ and $\mathbf{p}$, d, is required [Guestrin and Eizenman 2006]. Let's call these estimates $\mathbf{c}_{1CAM}(r)$ and $\mathbf{p}_{1CAM}(r,d)$, respectively.

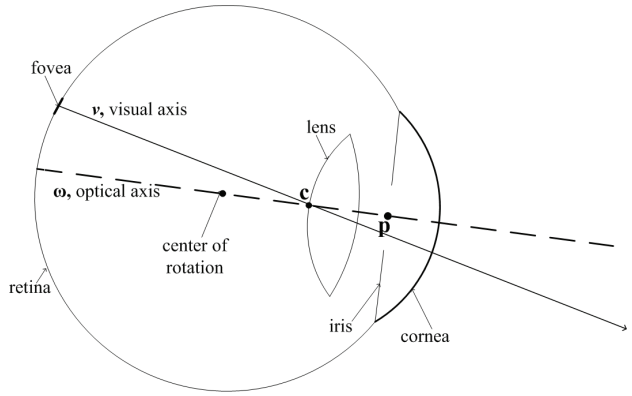

Figure 1. *A simplified schematic diagram of the eye. The optical axis, $\boldsymbol{\omega}$ (the axis of symmetry of the eye) passes through the center of curvature of the cornea, $\mathbf{c}$, and the center of the pupil, $\mathbf{p}$. The visual axis of the eye (the line-of-sight), $\mathbf{v}$,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

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 ($\mathbf{g}_{1CAM}$) and a stereo-pair of images ($\mathbf{g}_{2CAM}$). $\hat{r}$ and $\hat{d}$ can be obtained by solving the following optimization problem:

$$\left[\hat{r},\hat{d}\right] = arg\min \sum_{t=\mathbf{t}} \left\|\mathbf{g}_{2CAM} - \mathbf{g}_{1CAM}(r,d)\right\|_2^2 \qquad (2)$$

where the summation is for time instances, $\mathbf{t}$, during which $\mathbf{g}_{2CAM}$ is available.

After substitution of (1) into (2):

$$\left[\hat{r},\hat{d}\right] = arg\min \sum_{t=\mathbf{t}} \left\| \begin{array}{l} \left(\mathbf{c}_{2CAM} + \mu_{2CAM}\mathbf{R}\boldsymbol{\omega}_{2CAM}\right) - \\ \left(\mathbf{c}_{1CAM}(r) - \mu_{1CAM}\mathbf{R}\boldsymbol{\omega}_{1CAM}(r,d)\right) \end{array} \right\|_2 \qquad (3)$$

where

$$\boldsymbol{\omega}_{2CAM} = \left(\mathbf{p}_{2CAM} - \mathbf{c}_{2CAM}\right) \Big/ \left\|\mathbf{p}_{2CAM} - \mathbf{c}_{2CAM}\right\| \qquad (4)$$

is a normalized direction of the optical axis estimated in a two-camera mode, and

$$\boldsymbol{\omega}_{1CAM}(r,d) = \frac{\mathbf{p}_{1CAM}(r,d) - \mathbf{c}_{1CAM}(r)}{\left\|\mathbf{p}_{1CAM}(r,d) - \mathbf{c}_{1CAM}(r)\right\|} \qquad (5)$$

is a normalized direction of the optical axis estimated in a one-camera mode. Note that $\boldsymbol{\omega}_{1CAM}$ is a function of r and d.

Assuming $\mu_{1CAM} \cong \mu_{2CAM} = \mu$ and reorganizing (3) yields

$$\left[\hat{r},\hat{d}\right] = arg\min \sum_{t=\mathbf{t}} \left\| \begin{array}{l} \left(\mathbf{c}_{2CAM} - \mathbf{c}_{1CAM}(r)\right) + \\ \mu\mathbf{R}\left(\boldsymbol{\omega}_{2CAM} - \boldsymbol{\omega}_{1CAM}(r,d)\right) \end{array} \right\|_2^2 \qquad (6)$$

The first term in (6) represents the mismatch between the estimates of $\mathbf{c}$ by one-camera and two cameras modes (the linear offset error). The second term in (6) represents the mismatch between the directions of the optical axes (the angular error), which is scaled by the distance between $\mathbf{c}$ and the PoG, μ. Since $\mu \gg 1$, the second term is weighted more heavily during the optimization procedure. Separating the linear and angular errors yields

$$\left[\hat{r},\hat{d}\right] = arg\min \sum_{t=\mathbf{t}} \left(\begin{array}{l} \left\|\left(\mathbf{c}_{2CAM} - \mathbf{c}_{1CAM}(r)\right)\right\|_2^2 + \\ \mu^2 \left\|\mathbf{R}\left(\boldsymbol{\omega}_{2CAM} - \boldsymbol{\omega}_{1CAM}(r,d)\right)\right\|_2^2 \end{array} \right) \qquad (7)$$

Noting that $\mathbf{R}^T\mathbf{R}=\mathbf{I}$ (where $\mathbf{I}$ is the identity matrix and ^{T} denotes transpose), yields

$$\left[\hat{r},\hat{d}\right] = arg\min \sum_{t=\mathbf{t}} \left(\begin{array}{l} \left\|\left(\mathbf{c}_{2CAM} - \mathbf{c}_{1CAM}(r)\right)\right\|_2^2 + \\ \mu^2 \left\|\left(\boldsymbol{\omega}_{2CAM} - \boldsymbol{\omega}_{1CAM}(r,d)\right)\right\|_2^2 \end{array} \right) \qquad (8)$$

In practice, the knowledge of the exact distance between $\mathbf{c}$ and the PoG is not required, and μ can be approximated by the average distance between $\mathbf{c}$ and the PoG.

Since the stability of the solution to the minimization problem in (8) can be affected by the presence of outliers (e.g., due to blinks), outlier estimates of $\mathbf{p}$ and $\mathbf{c}$ have to be removed prior to the mini-

Algorithm 1
Estimation of Subject-Specific Parameters: r and d

1. Estimate r_t and d_t, by minimizing (8) 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 $\mathbf{t}_{in}$, and the average value for inliers $\overline{r}_{in}$ and $\overline{d}_{in}$:
   a. Find time instances for inliers in r, $\mathbf{t}_r$, and the average value, $\overline{r}_{in}$, 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, $\overline{r}$, of all the data points in the selected bin and two adjacent bins.
      iv. The inliers are r_t that satisfy $\left|\overline{r} - r_t\right| < W$.
      v. Calculate the average value of all the inliers, $\overline{r}_{in}$.
   b. Find all time instances for inliers in d, $\mathbf{t}_d$, and the average value, $\overline{d}_{in}$, of all the inliers using a procedure similar to the one described in Step 3.a.
   c. The time instances for inliers are given by: $\mathbf{t}_{in} = \mathbf{t}_r \cap \mathbf{t}_d$
4. Re-optimize (8) using all inliers at once ($\mathbf{t}_{in}$) and $\overline{r}_{in}$ and $\overline{d}_{in}$ as an initial guess to obtain the optimal values for $\hat{r}$ and $\hat{d}$.

mization procedure. Instead of using the computationally demanding iteratively re-weighted least squares method to remove outliers [Holland and Welsch 1977] (which requires repetition 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 images for all the time instances are collected). After the collection of all T samples is completed, a histogram-based outlier removal procedure is applied. Algorithm 1 summarizes the estimation procedure for r and d.

In parallel to the execution of Algorithm 1, 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 [Model 2011; Model and Eizenman 2010a; Model and Eizenman 2011a; Nagamatsu et al. 2010].

After the estimation of the full set of subject-specific eye-parameters (r,d and κ), the PoG can be calculated by using eye features from a single camera [Guestrin and Eizenman 2006].

3 Experiments

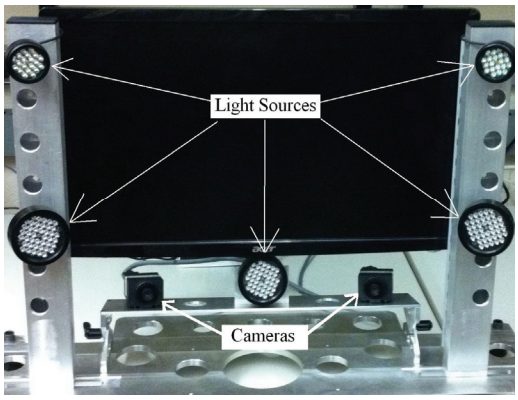

Figure 2. *Prototype REGT system.*

The experiments were carried out with the two-camera system shown in Figure 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 5 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 1. Angle kappa was estimated using an implicit calibration procedure [Model and Eizenman 2011a].

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 Figure 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 and once in the split-mode. The same system components (camera, light sources) were used for both modes. The RMS error in PoG estimation is provided in Table 1.

Figure 3 shows the results of the experiments with subjects 1 to 5. As one can see from Figure 3 and Table 1, for the central gaze

Table 1. RMS Error in Point-of-Gaze Estimation (degrees)

Subject #	Two-Camera Stereo System		Split-Mode Two-Camera System	
	Central Gaze (±20°)	Off-Central Gaze (<-20° ; >20°)	Central Gaze (±20°)	Off-Central Gaze (<-20° ; >20°)
1	0.67	6.44	0.70	1.13
2	1.00	6.26	0.94	1.46
3	0.98	4.41	0.84	1.10
4	0.84	3.50	0.74	1.40
5	0.80	3.77	0.70	1.09
Mean	**0.86**	**4.88**	**0.78**	**1.24**

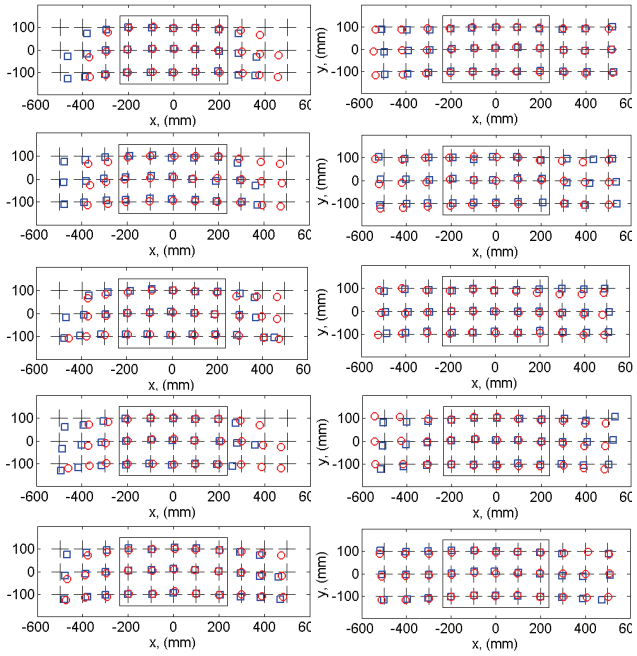

Figure 3. *PoG estimation with subjects 1-5 using a two-camera stereo system (left column) and a split-mode system (right column). Crosses indicate the actual positions of the fixation targets; small squares indicate average PoG estimates of the right eye; small circles indicate average PoG estimates of the left eye. The rectangle in the center indicates the central area (±20° horizontally, ±12°vertically).*

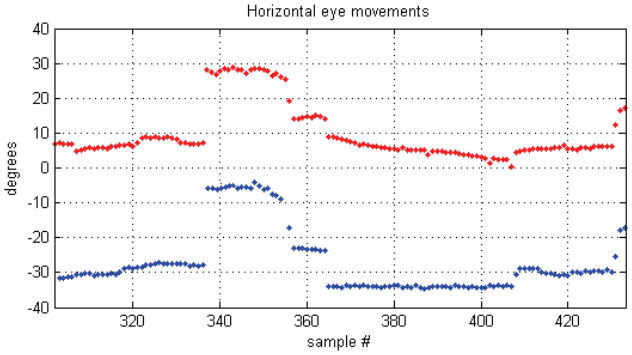

Figure 4. *Illustration of horizontal eye movements of a 9 months old baby with strabismus (red: left eye, blue: right eye). Due to large angle of deviation (≈35°), the simultaneous tracking of both eyes was possible only in the split-mode.*

directions (±20°) both systems can estimate the PoG with comparable RMS errors. The stereo system, however, has much higher gaze estimation errors for off-central gaze directions (average error of 4.88° in stereo mode vs. 1.24° in split mode).

As expected, the accuracy of gaze estimation in the stereo mode deteriorates when the angle between the subject's **ω** and the optical axis of one 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 subject's optical axis 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 [Guestrin and Eizenman 2006; Shih and Liu 2004], 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', the angle between the direction of gaze and the optical axes of one of the system's cameras never exceeds 25° degrees and therefore reasonable accuracy could be obtained for the full range of fixation points.

As can be seen from Figure 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 provide relatively accurate estimates of gaze directions of up to ±35.5° (±500 mm) horizontally. The estimates at ±35.5° of a split mode system suffer from increased bias, but the bias is still less than 2°.

4 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.

Experimental results suggest (see Table 1 and Figure 3) that for central gaze there is no deterioration in the accuracy of PoG estimation when the system is used in a split mode compared to the stereo-tracking mode. As expected, the split-mode system enables tracking over a larger range of gaze directions. In essence, by adopting the suggested approach, the horizontal gaze tracking range of REGT system is extended from ±23° to ±36° without adding any new hardware.

Figure 4 shows an example of eye movements of a nine months old infant with strabismus (eye misalignment). Due to large angle of deviation between the optical axes of the two eyes, when left eye (red line) is oriented centrally, the right eye (blue line) is rotated ≈35° away from the center. The overall range of gaze directions in this example is from -35° to +30° relative to the center. Therefore, simultaneous binocular tracking of both eyes (which is required for the estimation of angle of eye misalignment [Model 2011; Model et al. 2011; Model and Eizenman 2011a]) in this patients is possible only using the split-mode method that is described in this paper.

Finally, the suggested split-mode method can be readily extended from 2 cameras to *N* cameras, as long as at least 2 of the *N* cameras have overlapping fields of view. The cameras with the overlapping fields of view will allow automatic estimation of subject-specific eye parameters, which in turn will enable calibration-free tracking using each of the system's *N* cameras. Thus, the suggested approach enables a scalable, user-calibration-free, distributed eye-gaze tracking system.

Acknowledgements

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

ALLISON, R.S., EIZENMAN, M. and CHEUNG, B.S.K. 1996. Combined head and eye tracking system for dynamic testing of the vestibular system.*IEEE Trans on Bio. Eng., 43(11)*, 1073-82.

GUESTRIN, E.D. and EIZENMAN, M. 2006. General theory of remote gaze estimation using the pupil center and corneal reflections. *IEEE Trans. Biomed. Eng, 53, 6*, 1124-33.

HAMBURG, J.V.M. 2009. Amnesty International Domestic Violence Ad http://blogs.amnesty.org.uk/blogs_entry.asp?eid=3460.

HOLLAND, P.W. and WELSCH, R.E. 1977. Robust Regression Using Iteratively Reweighted Least-Squares. *Communications in Statistics: Theory and Methods A6*, 813-827.

KNEP, B. 2003. Big Smile http://www.blep.com/bigSmile/index.htm.

MILEKIC, S. 2010. Gaze-tracking and museums: current research and implications. In *Procs of the Museums and the Web*, Denver, CO, USA, 31 March, 2010.

MODEL, D. 2011. A Calibration Free Estimation of the Point of Gaze and Objective Measurement of Ocular Alignment in Adults and Infants. *PhD Thesis,* Dept. of Electrical & Computer Engineering, University of Toronto, Toronto, ON, Canada.

MODEL, D., BUNTING, H., SACCO, J., KRAFT, S. and EIZENMAN, M. 2011. An Objective and Automated Method to Measure Eye Alignment. In *proc. of CMBEC'34*, Toronto, ON, Canada.

MODEL, D. and EIZENMAN, M. 2010a. An Automatic Personal Calibration Procedure for Advanced Gaze Estimation Systems. *IEEE Trans Biomedical Engineering, 57, 5*, 1031-1039.

MODEL, D. and EIZENMAN, M. 2010b. User-calibration-free remote gaze estimation system. In *Proc. of ETRA 2010*, 29-36.

MODEL, D. and EIZENMAN, M. 2011a. An automated Hirschberg test for infants. *IEEE Trans Biomed Eng, 58, 1*, 103-109.

MODEL, D. and EIZENMAN, M. 2011b. User-Calibration-Free Remote Eye-Gaze Tracking System With Extended Tracking Range. In *proc. of IEEE CCECE*, Niagara Falls, ON, 1268-71.

NAGAMATSU, T., SUGANO, R., IWAMOTO, Y., KAMAHARA, J. and TANAKA, N. 2010. User-calibration-free gaze tracking with estimation of the horizontal angles between the visual and the optical axes of both eyes. In *Proc of ETRA 2010,* 251-4.

SHIH, S.W. and LIU, J. 2004. A novel approach to 3-D gaze tracking using stereo cameras. *IEEE Tran Systems, Man, and Cybernetics, Part B: Cybernetics 34, 1*, 234-245.

VILLANUEVA, A. and CABEZA, R. 2007. Models for gaze tracking systems. *J. Image Video Process. 2007, 4*, 1-16.