Constraint-based Gaze EstimationWithout Active Calibration

William Maio, Jixu Chen and Qiang Ji
Dept. Of Electrical, Computer, and Systems Engineering
Rensselaer Polytechnic Institute
Troy, NY
maiow@rpi.edu, chenj4@rpi.edu, qji@ecse.rpi.edu

***Abstract*—Existing eye gaze tracking systems typically require an explicit personal calibration process in order to estimate certain person-specific eye parameters. For natural human computer interaction, such a personal calibration is often cumbersome and unnatural. In this paper, we introduce a new method that estimates a person's gaze without active personal calibration. By exploiting the binocular constraint that the gaze point is produced by the intersection of visual axes of two eyes and some generic person-independent constraints on eye parameters, our method is able to estimate the required person-specific parameters implicitly and naturally without active participation from the user and without use of any special calibration object. Experiments with different subjects show that the proposed method achieves good gaze estimation comparable to the conventional 9 point method.**

I. INTRODUCTION

Gaze tracking is the procedure of determining the point-of-gaze, or the visual axis of the eye. Gaze tracking systems are primarily used in Human Computer Interaction (HCI) and in visual scanning pattern analysis. In HCI, the eye gaze can serve as an advanced computer input [1] to replace the traditional input devices such as a mouse pointer [2]. Also, the graphic display on the screen can be controlled by the eye gaze interactively [3]. Since visual scanning patterns are closely related to the person's attentional focus, cognitive scientists use the gaze tracking system to study human's cognitive processes [4], [5].

In general, video-based eye gaze estimation algorithms can be classified into two groups: 2D mapping based gaze estimation methods [3], [6], [7], [8] and 3D gaze estimation methods [9], [10], [11], [12] which estimate the 3D visual axis of the subjects. Recently, the 3D methods are getting more popular because of its high accuracy under free head movement. However, in current advanced 3D gaze estimation, a calibration procedure is required for each subject to estimate his/her specific eye parameters. In this work, we propose a novel method to estimate eye parameters without explicit calibration procedure. In contrast to the traditional calibration procedure which asks the subject to fixate on several points on the screen, we propose to track both eyes of the subject when he/she naturally looking at the screen, and then the eye parameters can be estimated based on various constraints in binocular vision.

II. RELATED WORK

In traditional gaze estimation method from a single camera, 2D mapping approach learns a polynomial mapping function from the 2D features (e.g. 2D pupil glint vector) [3], [7], [8], [13] or 2D eye images [6] to the gaze point on the screen.

However, the 2D mapping approach has two common drawbacks: First, in order to learn the mapping function, the user has to perform complex experiment to calibrate the parameters of the mapping functions. For example, in the calibration procedure of [14], the subject needs to gaze at 9 evenly distributed points on the screen or gaze at 12 points for higher accuracy. Second, because the extracted 2-D eye image features vary significantly with head position, the gaze mapping function is very sensitive to head motion. In [7], they report detailed data showing how the gaze tracking systems decay as the head moves away from original calibration position. Hence, the user has to keep his head unnaturally still in order to achieve good performance. Methods have also been proposed to handle head pose changes using Neural Network [3] or SVM [8]. These methods, however, either only consider the in-plane head translation [3] or need complex stereo cameras to obtain the 3D eye position [8].

Different from the 2D mapping-based gaze estimation, the 3D gaze estimation is based on the high resolution stereo cameras [9], [10], [11], [12] or single camera with multiple calibrated light sources [15] to estimate 3D eye features (the corneal center, the pupil center, and the optical axis connecting them) directly by the 3D reconstruction technique. Then, the visual axis is estimated from the 3D features, and the gaze point on the screen is obtained by intersecting the visual axis with the screen. However, this kind of method still needs the person-specific calibration to estimate the eye parameters. For example, Guestrin et al. [10] propose to estimate 3D gaze with the use of stereo camera and multiple IR lights. Their method starts with the reconstruction of the optical axis of the eye. The visual axis can be estimated by adding a constant angle to the optical axis. However, the angle between visual axis and optical axis needs to be estimated beforehand through a personal calibration procedure.

Most recently, gaze estimation methods are proposed without active calibration. Sugano et.al [16], [17] proposed "implicit" calibration methods for 2D based gaze estimation. These methods still need to collect the gaze points as training data. However it collects the gaze point implicitly, i.e. the user is unconscious of this process while working normally and naturally with the eye tracking system. In [16], they collect gaze points implicitly by assuming the subject is looking at the cursor on the screen when he/she click the mouse. However, it still needs active user participation. In [17], gaze points are collected when the subject is watching an image or video in the monitor. Given the saliency map of the image, gaze points

are sampled from it and used as training data to train a mapping function (Gaussian Process Regresser) between eye image and gaze point. However, because of the large uncertainty in saliency map, the accuracy of this system is pretty low ($\sim$ 6 degree) comparing with state-of-the-art ($<$ 1 degree). Furthermore, since this is basically a 2D mapping method which doesn't consider head pose, a chin rest is used to fix the head.

Without collecting any gaze points, Model and Eizenman [18] propose to estimate the eye parameters directly from anatomical knowledge. However, because of the noise in optical axis, it's difficult to achieve accurate result. For standard 40 cm x 30 cm flat monitor, when the noise in optical axis is 0.4 degree, the estimation error is 2.6 degree; when the optical noise is 1 degree, the error is over 5 degree. Although they propose to use larger monitor (160 cm x120 cm) or pyramid observation surface to reduce the error, they are usually not available in real application.

In this paper, we propose to incorporate various constraints to estimate eye parameters automatically and accurately, when the subjects are naturally looking at the standard screen.

III. 3D GAZE ESTIMATION

A. Eye Model

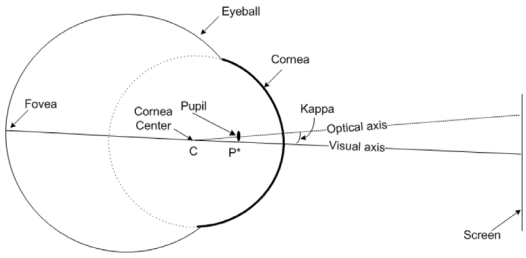


Figure 1. Diagram of the eyeball. The optical axis is defined as the vector formed with the Cornea Center and the Center of the Pupil. The visual axis (gaze) is defined as the vector formed with the Fovea and the Cornea Center.

Fig. 1 presents the model that is used in the eye-gaze estimation system described in this paper. In this model, all the 3D position data are relative to the camera coordinate system. We also assume the eye is looking at a monitor screen. The intersection of the gaze and the screen is represented by 2D position data relative to the left top corner of the screen. The visual axis, a 3 by 1 unit vector $\boldsymbol{v}$ formed by the fovea and cornea center, is defined to be the human gaze. Since the fovea is not observable, optical axis, a 3 by 1 unit vector, $\boldsymbol{p}$, is estimated first via [15], the visual axis is then estimated from the optical axis. The parameter that quantifies the relationship is called Kappa which is a constant angle for each subject. In this paper, the following coordinates are assumed to be known: the camera, the screen, optical axis. The optical axis can be expressed in the following manner with Kappa split into two parameters when referenced to the camera coordinate system:

$$\boldsymbol{p} = \boldsymbol{c} + K\begin{pmatrix} \cos(\varphi)\sin(\theta) \\ \sin(\varphi) \\ -\cos(\varphi)\cos(\theta) \end{pmatrix} \quad (1)$$

$\boldsymbol{p}$ is the optical axis , $\boldsymbol{c}$, is the cornea center relative to the camera coordinate system, K is the line parameter of the optical axis, and φ and θ are the angle representation of the optical axis vector. From the optical axis angles and cornea center, we can be expresses the visual axis as the following:

$$\boldsymbol{v} = \boldsymbol{c} + \lambda\begin{pmatrix} \cos(\varphi+\beta)\sin(\theta+\alpha) \\ \sin(\varphi+\beta) \\ -\cos(\varphi+\beta)\cos(\theta+\alpha) \end{pmatrix} \quad (2)$$

$\boldsymbol{v}$ is the visual axis, $\boldsymbol{c}$ is the cornea center, λ is the line parameter of the visual axis. α and β are the subject specific parameters that account for the deviation angles from the optical axis to the visual axis.

A conventional method to estimate α and β is the 9-point calibration process. This process requires the users to gaze at 9 different points on the screen. Then, optimization techniques are applied to find α and β that minimize the following objective function:

$$F(\alpha,\beta) = \sum_i \|G_i(\alpha,\beta) - CP_i\|_2^2 \quad (3)$$

G_i is the estimated gaze and CP_i is the calibration points which are vectors that consist of x and y components relative to the left top corner of the screen. i is the reference of the 9 calibration points .

IV. Implict Constrained Eye Calibration

A. Binocular Constraint Criteria

Model and Eizeman [18] proposed using optimization algorithms to estimate the subject specific parameters using anatomical knowledge The binocular constraint assumes that the 2 visual axes intersects at 1 point on the observed object. This assumption can be expressed in the following objective function:

$$F(\alpha^L,\beta^L,\alpha^R,\beta^R) = \sum_i \|G_i^L(\alpha^L,\beta^L) - G_i^R(\alpha^R,\beta^R)\|_2^2 \quad (4)$$

G_i is the gaze points that consist of x and y component relative the top left corner of the screen, the subscript i denotes the different samples of the gaze, α and β are the subject specific parameters that is described in the previous section. The superscripts L and R denote L for left eye and R for right eye.

B. Eye Constraints

In addition to the binocular constraint, we outlined the novel constraints that we impose in this section.

1) Screen Constraint

Assuming the user is looking at a surface, we can impose the constraint that all of the visual axes intersect points, G_{yi}^R , $G_{x_i}^R$, which lay on the surface of interest. In our experiment the screen size is 1280 by 1024 pixels and 360 by 287 mm. The following inequality constraint can be imposed on the two gazes if our assumption is correct.

$$0 < G_{x_i}^L, G_{x_i}^R < 1280 \quad (5)$$

$$0 < G_{y_i}^{L} G_{y_i}^{R} < 1024 \quad (6)$$

2) Subject Independent Parameter Range Constraint

Similar to many parameters in the human body, the eye parameter should follow similar distribution. Using this knowledge, we can construct a range using previously obtained eye parameters from the 9-Point calibration process [14]. This simple but powerful constraint bound improves the optimization by filtering unreasonable outcomes. Using parameter data set from previous studies we have found that subject parameter falls in the following range:

$$-3 \le \alpha^{L} \le 1.5 \quad (7)$$
$$-2 \le \alpha^{R} \le 1.5 \quad (8)$$
$$-3.5 \le \beta^{L} \le 1 \quad (9)$$
$$-3 \le \beta^{R} \le 3 \quad (10)$$

C. Constrainted Eye Prameter Estimation

Given the binocular constraint and eye constraints, an inequality-constrained optimization problem is formed. This problem can be solved using quadratic programming. [19] The binocular constraint and our novel constraints can be expressed in the following equations:

$$S(\alpha^{L}, \beta^{L}, \alpha^{R}, \beta^{R}) = \begin{bmatrix} G_{x_i}^{L,R} - \chi \\ -G_{x_i}^{L,R} \\ G_{y_i}^{L,R} - \psi \\ -G_{y_i}^{L,R} \end{bmatrix} \quad (11)$$

$$P(\alpha^{L}, \beta^{L}, \alpha^{R}, \beta^{R}) = \begin{bmatrix} \nu_1 - \alpha^{L} \\ \alpha^{L} - \xi_1 \\ \nu_2 - \beta^{L} \\ \beta^{L} - \xi_2 \\ \nu_3 - \alpha^{R} \\ \alpha^{R} - \xi_3 \\ \nu_4 - \beta^{R} \\ \beta^{R} - \xi_4 \end{bmatrix} \quad (12)$$

$$\min_{\alpha^{L},\beta^{L},\alpha^{R},\beta^{R}} f(\alpha^{L}, \beta^{L}, \alpha^{R}, \beta^{R})$$
$$subject\ to\ S_j(\alpha^{L}, \beta^{L}, \alpha^{R}, \beta^{R}) \le 0 \quad j = 1, \dots 4$$
$$P_j(\alpha^{L}, \beta^{L}, \alpha^{R}, \beta^{R}) \le 0 \quad j = 1, \dots 8 \quad (13)$$

(11) is the screen constraint using the parameters in (5) and (6) for χ and ψ, and (12) is the subject independent parameter range constraints using parameters in (7),(8),(9),(10) for ν and ξ. The initialization used for this non-linear optimization problem is the mean subject parameters from the 9-point calibration algorithm. It can also be obtained from anthropological statistics:

α^{L}=-0.1528^{0}, β^{L}=-0.9468^{0}, α^{R}=-0.3995^{0}, β^{R}=-0.5832^{0}

V. Experimental Results

We used the 9-point calibration data set to test our implicit subject specific parameter estimation. To compare our results with previously proposed method, we also run the data set using the method in [18]. Table I. shows the results of the estimated eye parameters using the proposed method with only the angle range constraints, and the 9-point method. Due to the noise in the gaze estimation system, for some subjects, the parameters do not quite converge to the results for 9-point calibration method. However, when compared to other methods in Table II, the improvement is significant.

TABLE I. Subject Specific Parameter Estimation Result

Subjects:	Subject-specific angles (°)							
	9-Point Calibration				*Parameter Range Constraint*			
	α^{L}	α^{R}	β^{L}	β^{R}	α^{L}	α^{R}	β^{L}	β^{R}
1	1.3	-0.2	-1.2	-1.5	1.5	0.0	-3.0	-2.7
2	-3.0	1.3	-3.5	-3.0	-3.0	1.5	-2.8	-3.0
3	1.0	-1.9	0.9	-0.6	-1.5	-0.6	-3.0	-0.4
4	0.2	-0.8	0.0	2.9	-1.5	-2.0	-3.0	-0.4

Table II summarizes the gaze point estimation results with the constrained subject specific parameter estimation. 9P column is the result for the 9-point calibration. E2 column is the result for the method in [18] applied on a conventional screen[1]. E2A column is the result with the subject independent parameter range constraint. E2S column is the result with the screen constraint. The result is in terms of the accuracy of the estimated gaze angles. In all the experiments, the subject sat 500 mm away from the screen.

TABLE II. Gaze Angle Estimation Result using different Subject Speicifc Parameters

Subject	Error from the ground truth gaze (°)							
	9P		*E2*		*E2A*		*E2S*	
	L	*R*	*L*	*R*	*L*	*R*	*L*	*R*
1	2.6	2.4	12	12	2.9	2.4	11	8.9
2	2.8	2.4	8.2	3.4	2.9	2.7	8.2	3.4
3	2.1	3.9	5.6	22	4.0	4.4	6.5	8.6
4	3.8	3.6	7.9	19	5.6	5.5	6.3	5.4

We tried different combinations of constraints and found that the subject independent parameter range constraint produced results that were closest to the 9-point calibration process. We also found that the subject independent parameter range constraint is a stricter constraint than the screen constraint. When the optical axis already intersects with the screen, then the screen constraints does not affect the optimization.

Table III. shows the detailed results for subject 1. Figure 2 presents the scan pattern results of the 9-point calibration process and our method. The solid line is gaze estimated using the parameter obtained from the subject independent parameter range constraint implicit calibration method, and the dash-dot line is gaze estimated using the parameter obtained from the 9-point calibration method. From Figure 2 we can conclude that the implicit parameter estimation method produces a scan pattern that closely follows the 9-point calibration result.

[1] In [18], the optimization criterion works with reasonably with a pyramid surface. Eizenman's method applied on a conventional display is expected to have large error.

TABLE III. GAZE POINT ESTIMATION RESULT ON DIFFERENT POINTS FOR SUBJECT 1

	Error from the ground truth gaze (mm)					
	9-Point Calibration			***Parameter Range Constraint***		
Left	16.6	24.7	11.8	16.0	29.4	10.4
	19.1	28.4	31.1	21.7	34.1	27.8
	25.5	24.8	26.0	28.2	24.4	25.4
Right	18.0	31.6	13.5	18.2	30.0	13.3
	28.9	27.2	16.8	27.6	26.1	16.7
	14.8	19.4	17.0	14.4	21.0	18.1

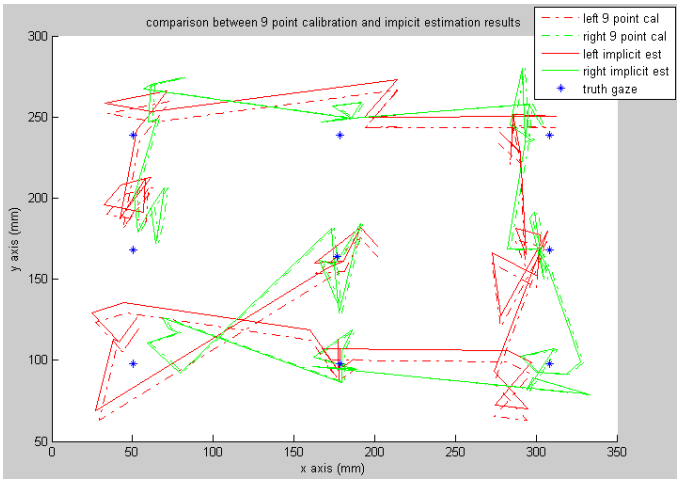
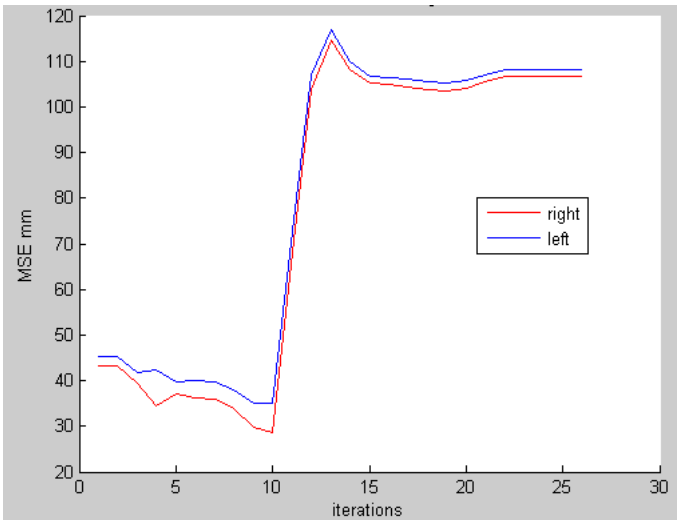


Figure 2. The gaze result of subject 1. The green and red are the left gaze and right gaze. The dash-dot line is the gaze using the 9-point calibration parameter and the solid line is the gaze using the implicit parameter estimation. The blue star is the ground truth.

To further compare between the different combination of the constraints and methods, the optimization progress of using only the binocular constraint is presented in figure 3 and the optimization progress with constraints is presented in figure 4. The first plot is the result of the right and left gaze mean square error deviation at each of the optimization iteration. The second plot is the results of the Optimization criteria of eq. (4) at each of the iteration. Note that the binocular constraint only method initially minimizes the objective function and reduces the mean square error, but after 10 iterations, the algorithm produces in a larger mean square error.



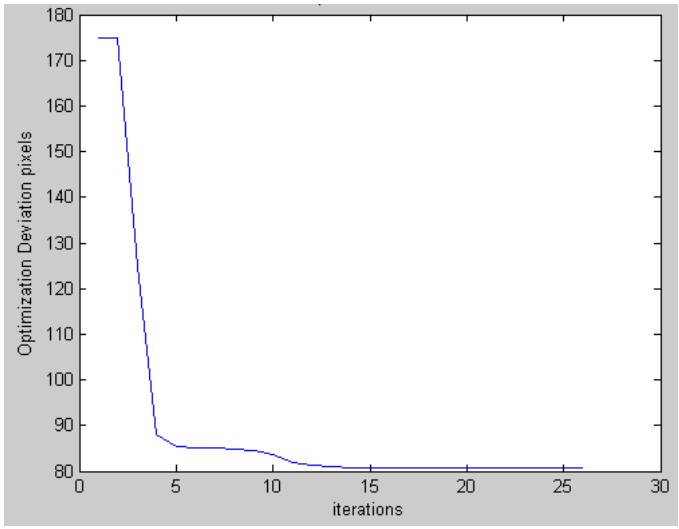
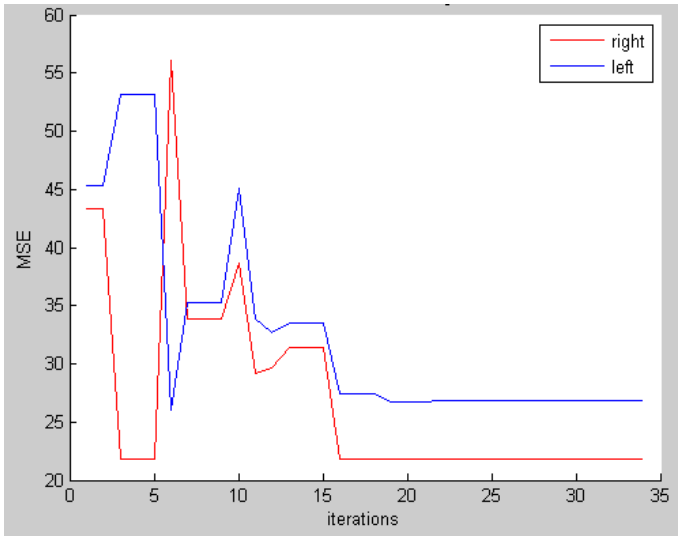


Figure 3. Implicit Calibration with only the binocular constraint [18]
The top plot presents the optimization results of the mean square error on the y-axis and iterations on the x-axis. The bottom plot presents the evaluation of the optimization criteria of eq. (4) with each of the iterations.



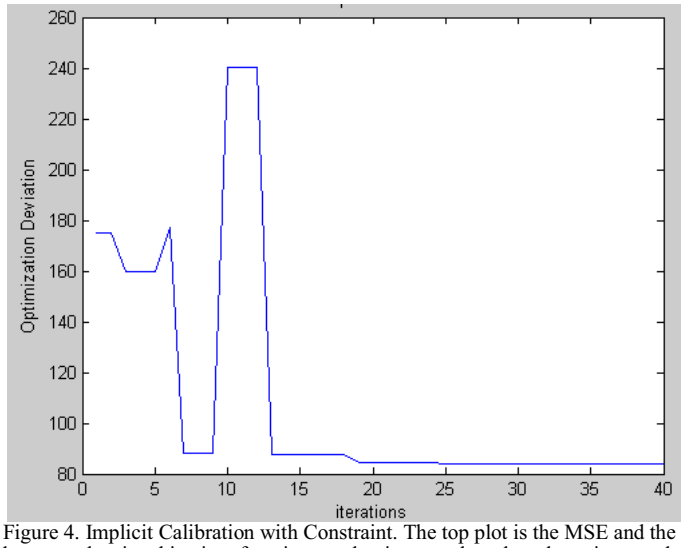


Figure 4. Implicit Calibration with Constraint. The top plot is the MSE and the bottom plot is objective function evaluation results plotted against each iteration.

Adding on the constraints to the objective function, we observe a decrease in both the MSE in mm and the objective function as it settles down around 30 iterations. When the optimization is stable with constraint the result is comparable to the 9-Point calibration algorithm.

VI. CONCLUSIONS

We have presented a constraint-based gaze-estimation without active calibration process. With this technique, the user may look at the screen freely without going through any active calibration process. Compared with similar existing methods, our method is simple and effective. It can lead to smoother interaction with eye-gaze technology.

In the future, we intend to further validate the method with additional subjects and under different operating conditions. In addition, we will identify additional constraints to further improve the accuracy and robustness of this approach.

REFERENCES

[1] R. J. Jacob., "The use of eye movements in human computer techniques: What you look at is what you get.," *ACM Transaction on Information Systems*, vol. 9, pp. 152-161, 1991.

[2] C. Morimoto, and S. Ihde. S. Zhai, "Manual and gaze input cascaded (magic) pointing.," *Proceedings of the SIGCHI conference on Human factors in computing systes,* pp. 246-253, 1999.

[3] Z. Zhu and Q. Ji, "Eye and gaze tracking for interactive graphic display.," in *Machine Vision and Applications*., 2004, ch. 15, pp. 139-148.

[4] S. Liversege and J. Findlay., "Saccadic eye movements and cognition.," *Trends in Cognitive Science*, no. 4, pp. 6-14, 2000.

[5] B.Hood, and C. Macrae M. Mason, "Look into my eyes: Gaze direction and person memory," *Memory*, no. 12, pp. 637-643, 2004.

[6] D. Kriegman, and H. Ahuja. K. Tan, "Apperance based eye gaze estimation," in *Proceedings of IEEE Workshop Applications of Computer Vision*, 2002, pp. 131-136.

[7] C.H. Morimoto and M. R. Mimica, "Eye gaze tracking techniques for interactive applications," *Computer Vision and Image Understanding, Special Issue on Eye Detection and Tracking*, no. 98, pp. 4-24, 2005.

[8] Q. Ji, and K. P. Bennett. Z. Zhu, "Nonlinear eye gaze mapping function estimation via support vector regression," in *International Conference on Pattern Recognition*, 2006.

[9] D. Beymer and M. Flickner., "Eye gaze tracking using an active stereo head," in *IEEE Conference in CVPR03*, 2003.

[10] E.D. Guestrin and M. Eizenman, "Remote point-of-gaze estimation requiring a single-point calibration for applications with infants.," in *Proceedings of the 2008 symposium on Eye tracking research & applications*, 2008.

[11] S.-W. Shih and J. Liu, "A novel approach to 3-d gaze tracking using stereo cameras," *IEEE Transactions on Systems, Man and Cybernetics,* , vol. PartB, no. 34, pp. 234-245, 2004.

[12] Y. Tong, W. Gray, and Q.Ji J. Chen, "A robust 3d eye gaze tracking system using noise reduction," in *Proceedings of the 2008 symposium on Eye tracking research & applications*, 2008, pp. 189-196.

[13] A. Amir, and M. Flickner C. H. Morimoto, "Detecting eye position and gaze from a single camera and 2 light sources," in *Proceedings of the International Conference on Patter Recognition*, 2002.

[14] L. T. Inc. [Online]. http://www.eyegaze.com

[15] E. D. Guestrin and M. Eizenman, "General theory of remote gaze estimation using the pupil center and corneal reflecitons," *IEEE Transactions on Biomedical Engineering*.

[16] Y. Matsushita, Y. Sato, and H. Koike Y. Sugano, "An incremental learning method for unconstrained gaze estimation," in *European Conference on Computer Vision*, 2007.

[17] Y. Matsuhita, and Y. Sato. Y. Sugano, "Calibration-free gaze sensing using saliency maps," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.

[18] D. Model and M. Eizenman, "An Automatic Personal Calibration Procedure for Advanced Gaze Estimation Systems," *IEEE Transactions on Biomedical Engineering*, vol. 57, p. 5, 2010.

[19] Jorge Nocedal, *Numerical optimization*. New York, United States of America: Springer, 1999.