Journal of Eye Movement Research
7(1):4, 1-11

Mapping the Pupil-Glint Vector to Gaze Coordinates in a Simple Video-Based Eye Tracker

Pieter Blignaut
University of the Free State, South Africa

In a video-based eye tracker, the normalized pupil-glint vector changes as the eyes move. Using an appropriate model, the pupil-glint vector can be mapped to gaze coordinates. Using a simple hardware configuration with one camera and one infrared source, several mapping functions – some from literature and some derived here – were compared with one another with respect to the accuracy that could be achieved. The study served to confirm the results of a previous study with another data set and to expand on the possibilities that are considered from the previous study. The data of various participants was examined for trends which led to derivation of a mapping model that proved to be more accurate than all but one model from literature. It was also shown that the best calibration configuration for this hardware setup is one that contains fourteen targets while taking about 20 seconds for the procedure to be completed.

Keywords: Eye tracking, gaze estimation, polynomial models, calibration targets, accuracy, validation

Introduction

A wide variety of gaze tracking systems are available commercially, each tailored for a specific set of applications. Video-based eye-tracking is based on the principle that when near infrared (NIR) light is shone onto the eyes, it is reflected off the different structures in the eye to create four Purkinje reflections (Crane and Steele, 1985). The vector difference between the pupil centre and the first Purkinje image (PI) (also known as the glint or corneal reflection (CR)), is tracked.

Tracking a person's gaze with a video-based system involves a number of steps. These steps can be loosely grouped into two sets, namely those involved with the detection of the eyes and eye features (e.g. pupil and glint centres) in the video frames, and those which map the detected features to gaze coordinates or Point of Regard (PoR) on the stimulus. For purposes of this paper, it is assumed that the location of features in the eye video is known and the focus is on the challenge to use these as input to determine a person's Point of Regard.

A simple video-based eye tracking system was developed with one camera and one infrared light source. The accuracy obtained with various combinations of calibration set and mapping model is evaluated for this system.

This study aims to replicate and confirm the results of an earlier study (Blignaut, 2013). The data set in the earlier study showed clear trends with regard to the relationships between the gaze target coordinates and the pupil-glint vectors in the eye video. In that study it was found that the number and arrangement of calibration targets as well as the mapping function is critically important to ensure good accuracy of a video-based eye tracker. A mapping model was derived (see below for details) and it was proven to provide an accuracy of less than 0.5° when used with a 24-point grid.

It was also indicated in Blignaut (2013) that accuracy cannot be based on a few validation points only as their might be areas of bad accuracy between the validation points. This study aims to see if these results hold for another data set.

DOI 10.16910/jemr.7.1.4
This article is licensed under a
Creative Commons Attribution 4.0 International license.
ISSN 1995-8692

The results of this study should contribute towards the drive for cheaper and simpler eye trackers as it aims to find the optimal polynomial model to map features in the eye video of an eye tracker with one camera and one infrared source to gaze coordinates. While the principles of deriving the model might be applied to more complicated trackers, it is highly unlikely that the resulting models will apply to other configurations.

Gaze Estimation

In a video-based eye tracker the pupil-glint vector changes as the eyes move (Figure 1). The model-based gaze estimation approach determines the 3D line of sight and calculates a person's Point of Regard (PoR) as the point where the line of sight intersects some object in the field of view, usually a computer monitor. It follows that for this approach to be implemented, the positions and orientations of the cameras, infrared lights and monitor (or other object(s) that the user might view) has to be known to a high accuracy.

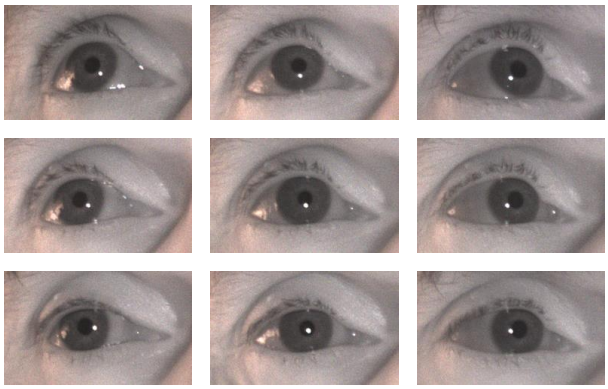


Figure 1. Corneal reflection at various gaze positions

Regression-based systems use polynomial expressions to determine the Point of Regard as a function of the pupil-glint vector in the eye image. Polynomial models should include two independent variables (x and y components of the pupil-glint vectors) which may or may not interact with each other for each one of the dependent variables (X and Y of the Point of Regard) separately. The coefficients for each term in the model need to be determined for every individual through a calibration process.

Using an appropriate polynomial model, the difference in x-coordinates (x') and difference in y-coordinates (y') between the pupil and glint can be mapped to screen coordinates (X,Y). Corrections for head movement can be done by normalising the pupil-glint vector in terms of the distance between the glints (if there are more than one IR source) or between the pupils (inter-pupil distance (IPD)), e.g. $x = x' / IPD$ (if there is only on IR source as in this study).

Polynomial Models in General

A set of n points can be approximated with a polynomial of n or less terms

$$X = \sum_{k=0}^{n-1} a_k x^i y^j \ , i \in [0,k], j \in [0,k]$$

where x and y refer to the normalised x and y components of the pupil-glint vector of a specific eye at a specific point in time and X refers to the X-coordinate of the PoR for the specific eye on the two dimensional plane of the screen. A similar, but not necessarily identical, model can be used for the Y-coordinate of the PoR for the specific eye.

According to Hennessey et al. (2008), the polynomial order may vary but is most often of first order:

$$X = a_0 + a_1x + a_2y + a_3xy$$
$$Y = b_0 + b_1x + b_2y + b_3xy$$

The coefficients a_k and b_k, $k \in (0,n-1)$, are determined through a calibration process which requires the user to focus on a number of dots (also referred to as calibration targets) at known angular positions while storing samples of the measured quantity (Abe et al., 2007).

A least squares regression is then done to determine the polynomials such that the differences between the reported PoRs and the actual PoRs (the positions of the dots on the monitor) are minimised. The regressions are done separately for the left and right eyes and an interpolated PoR is calculated as the average of the (X_{left},Y_{left}) and (X_{right}, Y_{right}) coordinates.

A set of n points can be fitted with a polynomial of n terms in which case R^2 (indication of goodness of fit) will be 1. Regression with less terms or through more points will result in a polynomial that does not necessarily pass through any of the points and R^2 will be less. This does, however, not necessarily mean that the approximations will be worse. In the simplified example of Figure 2 with a single independent variable, a 6th order polynomial (7 terms $a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 + a_5x^5 + a_6x^6$) fits

DOI 10.16910/jemr.7.1.4
This article is licensed under a
Creative Commons Attribution 4.0 International license.
ISSN 1995-8692

the seven points exactly, but interpolation at *x=17* will obviously be bad. A 5th order approximation of the data has a lower R^2, but interpolations between the points will be much better. Because of the end effects, even this polynomial will provide bad approximations at *x* values larger than 18.

The decision of how the regression model should look is therefore not obvious. Draper and Smith (1981) indicates that, apart from the brute-force method (i.e. testing all the possible equations), there is no systematic procedure that can provide the most suitable mapping equation.

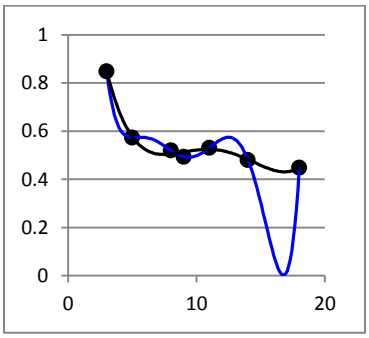


Figure 2. 5th order (black) and 6th order (blue) regressions of seven data points

Specific Models

The simplest model would be to map the gaze coordinates in terms of a linear relationship with the normalised pupil-glint vector without considering interactions between the two dimensions:

$$X = a_0 + a_1x$$
$$Y = b_0 + b_1y$$

A second order polynomial in *x* and *y* with first order interactions are used by Mitsugami, Ukita and Kidode (2003) and Cerrolaza et al. (2012):

$$X = a_0 + a_1x + a_2x^2 + a_3y + a_4y^2 + a_5xy$$
$$Y = b_0 + b_1x + b_2x^2 + b_3y + b_4y^2 + b_5xy$$

The above model can be extended to include second order interactions:

$$X = a_0 + a_1x + a_2x^2 + a_3y + a_4y^2 + a_5xy + a_6x^2y^2$$
$$Y = b_0 + b_1x + b_2x^2 + b_3y + b_4y^2 + b_5xy + b_6x^2y^2$$

Zhu and Ji (2005) used the following model:

$$X = a_0 + a_1x + a_2y + a_3xy$$
$$Y = b_0 + b_1x + b_2y + b_3y^2$$

Cerrolaza and Villanueva (2008) generated a large number of mapping functions, varying the degree and number of terms of the polynomial. They found that, apart from some of the simplest models, increasing the number of terms or the order of the polynomial had almost no effect on accuracy. A preferred model was chosen as one that showed good accuracy across all configurations in addition to having a small number of terms and being of low order:

$$X = a_0 + a_1x + a_2x^2 + a_3y$$
$$Y = b_0 + b_1x^2 + b_2y + b_3xy + b_4x^2y$$

In a previous study (Blignaut and Wium, 2013), we examined the accuracy of 625 polynomials and found the following model to provide the best results for all participants as long as at least 8 calibration points are used:

$$X = a_0 + a_1x + a_2x^3 + a_3y^2 + a_4xy$$
$$Y = b_0 + b_1x + b_2x^2 + b_3y + b_4y^2 + b_5xy + b_6x^2y$$

In another previous study (Blignaut, 2013) we used specific trends with regard to the relationships between the gaze target coordinates and the pupil-glint vectors in the eye video to derive a set of polynomials.

$$X = a_0 + a_1x + a_2x^2 + {}_3x^3 + a_4y + a_5xy + a_6x^2y + a_7x^3y$$
$$Y = b_0 + b_1x + b_2x^2 + b_3y + b_4y^2 + b_5xy + b_6x^2y$$

This model provided very good accuracy (< 0.5°) for the simple one camera, one IR source eye tracker, given that enough calibration points were used to facilitate regression of the multi-term polynomials (Blignaut, 2013).

Calibration Targets

Besides the polynomials to use for interpolation, a mapping model also entails the number and arrangement of calibration targets. To limit the end effects of polynomial interpolation, it is important that there are targets on the edges of the display area. More calibration targets will allow polynomials with more terms which might result in better accuracy but takes more time and might be strenuous on the participant's eyes. The ideal situation would be to find a set of polynomials that is very accurate while using only a small number of calibration targets.

Accuracy

Accuracy is measured as the distance, in degrees, between the position of a known target point and the average position of a set of raw data samples, collected from a participant looking at the point (Holmqvist et al., 2011).

DOI 10.16910/jemr.7.1.4
This article is licensed under a
Creative Commons Attribution 4.0 International license.
ISSN 1995-8692

This error may be averaged over a set of target points that are distributed across the display. The interest in this paper is in minimising the error that occurs due to the calibration process, specifically for a simple one light, one camera configuration.

While an accuracy of 0.3° has been reported for tower-mounted high-end systems operated by skilled operators (Holmqvist et al, 2011), remote systems are usually less accurate. Hansen and Ji (2010) provided an overview of remote eye-trackers and reported the accuracy of most model-based gaze estimation systems to be in the order of 1° - 2°.

Methodology

Experimental Configuration

An eye tracker was developed with a single CMOS camera with USB 2.0 interface and a 10 mm lens together with a single infrared light source. The UI-1550LE-C-DL camera from IDS Imaging (http://en.ids-imaging.com/) has a 1600×1200 sensor with pixel size of 2.8 µm (0.0028 mm). The images were rendered on a screen with a pixel size of 0.364 mm. The calibration was done on a 1360×768 (495 mm×280 mm) section of the screen at a distance of 800 mm from the participants. The camera was positioned 280 mm in front of the screen with an eye-camera distance of 600 mm. See Figure 3 for details.

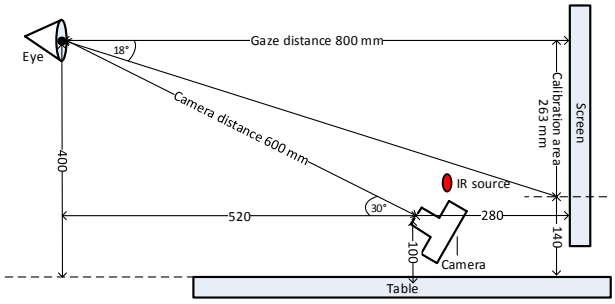


Figure 3. Hardware configuration

Data capturing

The calibration area was divided into a 15×9 grid to have the same width:height ratio as that of the display area (1360×768 pixels) (Figure 4). Twenty six (26) participants were presented with a series of 135 targets that covered the entire grid. Five data sets were captured for every participant.

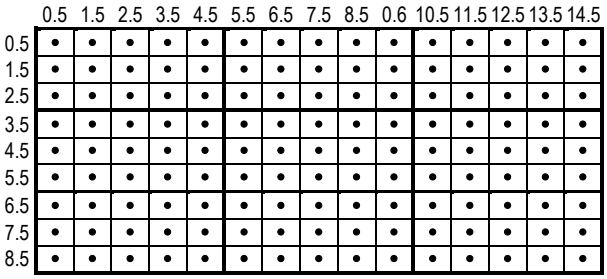


Figure 4. Calibration grid with 15×9 cells

No chin rest was used, but participants were requested to keep their heads as still as possible at a constant distance from the eye camera. Participants could at any time look away from the screen and rest their eyes. A target would only be accepted if the gaze was stable for a minimum period of 1 second. For each target, the pupil-glint vector of the last 500 ms of the period of stable gaze was saved to a database. The saved data were used afterwards to simulate the calibration procedure for various combinations of calibration target arrangements and mapping models.

To ensure that participants focused on a target, an initial linear regression was done after the 135 targets were accepted (Figure 5). A specific target was displayed again if the pupil-glint difference did not fall within 6σ of the average distance from the line. This procedure was repeated until all data points were within this tolerance from the regression line.

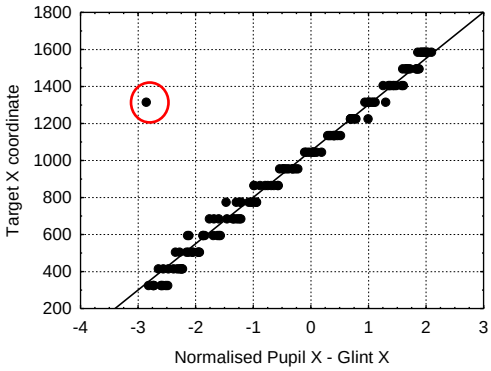


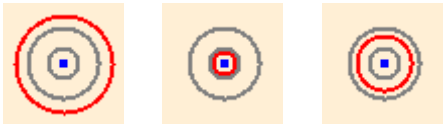
Figure 5. Initial linear regression to determine outliers during data capture

Despite the technique of normalising the pupil-glint vectors in terms of the inter-pupil distance (IPD), accuracy can still be affected by gaze distance. Therefore, in order to compare the accuracy of various polynomials with one another, the gaze distance was controlled.

DOI 10.16910/jemr.7.1.4
This article is licensed under a
Creative Commons Attribution 4.0 International license.
ISSN 1995-8692

Gaze distance was approximated in terms of the inter-pupil distance (IPD). The system was calibrated at 600 mm where the IPD in pixels was equated to the average IPD of 63 mm for adults (Dodgson, 2004). Although the inter-pupil distance differs slightly from one person another, the effect is not more than 30 mm. It is not important that the value is 100% accurate – the distance indicator only serves to ensure that the gaze distance for a specific person remains constant.

Two fixed concentric circles were displayed around each target, representing camera distances of 597 and 603 mm respectively. A third concentric circle was displayed with varying radius as the participant moves his head backwards or forwards (Figure 6). The participant had to move his head until the dynamic circle is between the two fixed circles, i.e. in the range 598 mm - 602 mm. For this study, the focus was solely to find the most appropriate mapping model at a specific gaze distance. Future studies will include gaze distance as a separate independent variable for the regression polynomial.



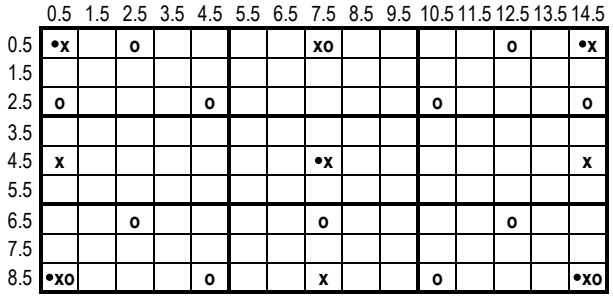
Too near Too far Correct

Figure 6. Real-time feedback on gaze distance with a circle with varying radius (red) around the target that should be between two fixed (grey) circles

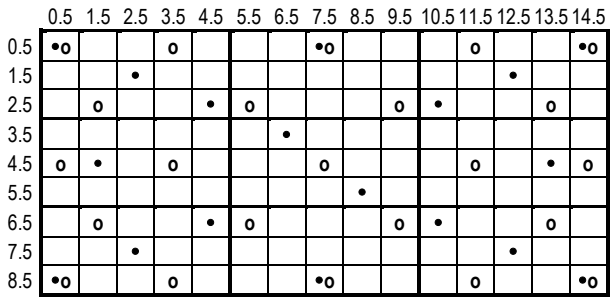
Calibration and Validation

Calibration. Data for the pupil-glint vectors at selected cells in the 15×9 grid (Figure 4) was selected to serve as calibration targets. The accuracy for six different calibration sets, which varies with regard to the number and arrangement of targets (Figure 7), were compared with one another. One of these arrangements is the full 135-point set that were used as benchmark for the best possible accuracy that can be attained.

Since the regression is done separately for the X and Y dimensions, it was thought that the arrangement of calibration targets should be such that they cover as many distinct X and Y values as possible. Having 9 points on a 3×3 grid effectively limits the regression to 3 distinct X and 3 distinct Y values. In Figure 7 below, 18 calibration targets are, for example, arranged to cover 11 distinct X and 9 distinct Y coordinates.



• 5 *targets with 3 distinct X and 3 distinct Y coordinates*
x *9 targets with 3 distinct X and 3 distinct Y coordinates*
o *14 targets with 7 distinct X and 4 distinct Y coordinates*



• *18 targets with 11 distinct X and 9 distinct Y coordinates*
o *23 targets with 9 distinct X and 5 distinct Y coordinates*

Figure 7. Five different sets of calibration targets

Validation. Normally, the accuracy of a system can be determined by requesting a participant to look at a second series of evenly spread targets (the first series being the calibration targets) for which the X and Y coordinates are known. The accuracy is then calculated as the average of the differences between the known and reported positions.

To ensure representative measurement of accuracy, validation targets (to be distinguished from calibration targets) should be spread regularly across the entire range at small intervals. For this study, the saved pupil glint vectors for all 135 cells in the 15×9 grid were used towards this purpose (cells marked with • in Figure 4).

For every participant/calibration set combination, the pupil-glint vectors in the calibration cells were used to determine the regression coefficients for each one of 12 polynomial models (Table 1). Models 9-12 in Table 1 will be derived later. A high level algorithm of the process is given in Figure 8.

DOI 10.16910/jemr.7.1.4
This article is licensed under a
Creative Commons Attribution 4.0 International license.
ISSN 1995-8692

Table 1. Summary of models for which the accuracies are determined.

No.	Reference	Polynomial for X	Polynomial for Y
1	Simple model	x,1	y,1
2	Hennessey, 2008	x, y, xy, 1	x, y, xy, 1
3	Mitsugami, 2003	x^2, x, y^2, y, xy, 1	x^2, x, y^2, y, xy, 1
4	Mitsugami, 2003	x^2, x, y^2, y, x^2y^2, xy, 1	x^2, x, y^2, y, x^2y^2, xy, 1
5	Zhu & Ji (2005)	x, y, xy, 1	x, y, y^2, 1
6	Cerrolaza & Villanueva (2008)	x^2, x, y, 1	x^2, y, x^2y, xy, 1
7	Blignaut & Wium (2013)	x^3, x, y^2, xy, 1	x^2y, xy, y, x^2, x, y^2, 1
8	Blignaut (2013)	x^3y, x^3, x^2y, x^2, x, xy, y, 1	x^2y, xy, y, x^2, x, y^2, 1
9	Derived below	x^3y, x^3, x^2y, x^2, x, xy, y, 1	x^2y, xy, y, x^2, x, 1
10	Derived below	x^3y, x^3, x^2y, x^2, x, xy, y, 1	x^5y, x^4y, x^3y, x^2y, xy, y, x^5, x^4, x^3, x^2, x, 1
11	Derived below	x^3y, x^3, x^2y, x^2, x, xy, y^3, y^2, y, 1	x^2y, xy, y, x^2, x, 1
12	Derived below	x^3y, x^3, x^2y, x^2, x, xy, y^3, y^2, y, 1	x^5y, x^4y, x^3y, x^2y, xy, y, x^5, x^4, x^3, x^2, x, 1

Notes:

(i) For simplicity and readability, polynomials are written without coefficients. The polynomial $a_0x+a_1y+a_2xy+a_3$ is, for example, written as x,y,xy,1.

(ii) Models 9-12 will be derived later in this paper.

```
For each participant
{
   For each calibration set
   {
      For each model
      {
         Determine mapping polynomials
         Get the average accuracy over the valida-
         tion points
      }
   }
}
```

Figure 8. Algorithm for data analysis

Derivation of alternative models

Relationship between (Pupil X – Glint X) and PoR_x

The models mentioned above express the *X* and *Y* coordinates of the PoR as two bivariate polynomials in terms of the *x* and *y* coordinates of the pupil-glint vector. In an attempt to simplify the relationships, the data that was captured was analysed while controlling for one of the two independent variables (the *x* and *y* coordinates of the pupil-glint vector). In other words, *X* was expressed in terms of *x* only at a specific value for *Y*. This means that for the 15×9 grid, nine relationships could be written for *X* in terms of *x* and 15 relationships could be written for *Y* in terms of *y*.

Figure 9 shows the relationships between the target X coordinate and the average normalized pupil-glint X (referred to as PG_X) of 130 data sets (26 participants with 5 repetitions each) at 9 distinct Y coordinates for the left and right eyes.

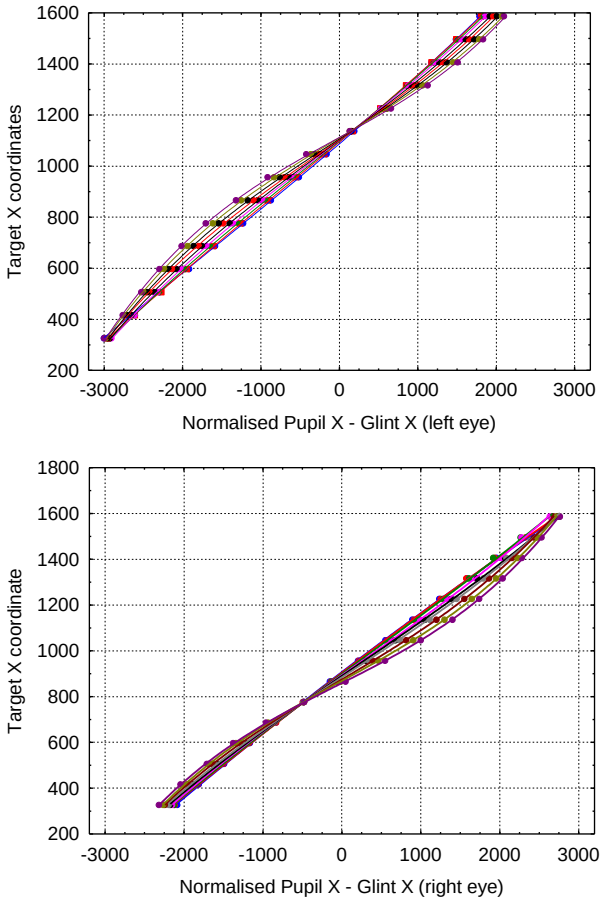


Figure 9. Target X coordinate against average normalized pupil-glint X of 26 participants at 9 distinct Y coordinates for the left eye (top) and right eye (bottom).

From Figure 9 it can be inferred that a third degree polynomial should fit each one of the 9 relationships very well. The following relationship was found between the normalised pupil-glint differences and the X-coordinate of the target for the left eye at Y=200:

$$X = 1.24x^3 + 6.77x^2 + 260.4x + 1087.8,\ R^2 = 0.9999.$$

The R^2 values at the other Y-positions of the targets are given in Table 2.

Table 2. R^2 values for the regression formulas for X at specific values of Y.

Eye	Target Y coordinate								
	200	285	370	455	540	625	710	795	880
Left	0.9999	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998
Right	0.9999	0.9998	0.9997	0.9998	0.9998	0.9996	1.0000	1.0000	0.9999

In general terms, the relationship can be written as

$$X = ax^3 + bx^2 + cx + d. \quad (1)$$

DOI 10.16910/jemr.7.1.4

This article is licensed under a
Creative Commons Attribution 4.0 International license.

ISSN 1995-8692

Figure 10 shows plots of the coefficients *a*, *b*, *c* and *d* against the average normalized pupil-glint Y for the left eye. Similar trends were found for the right eye. Table 3 shows the respective R^2 values for the left and right eyes.

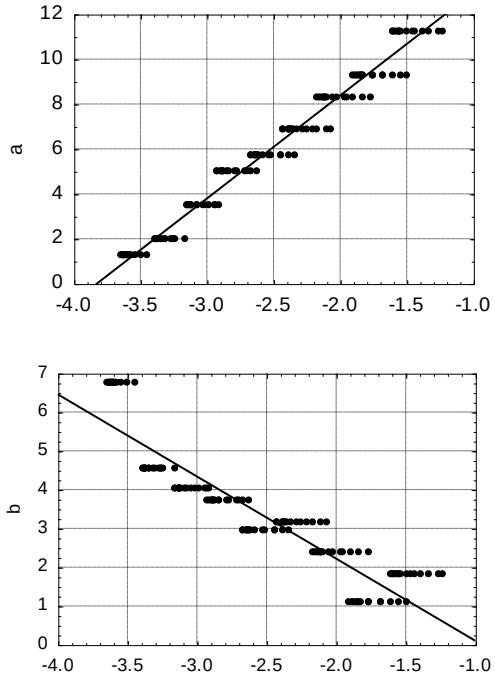
The relationships could thus be written as follows:

$a = a_1y + a_0$
$b = b_1y + b_0$
$c = c_1y + c_0,$
$d = d_3y^3 + d_2y^2 + d_1y + d_0$

Substituting these into the general equation (1) above, we get

$$X = a_1x^3y + a_0x^3 + b_1x^2y + b_0x^2 + c_1xy + c_0x + d_3y^3 + d_2y^2 + d_1y + d_0. \quad (2)$$

This is a small difference from the model used in Blignaut (2013) as in the latter study a linear relationship was used between d and the target Y coordinate.



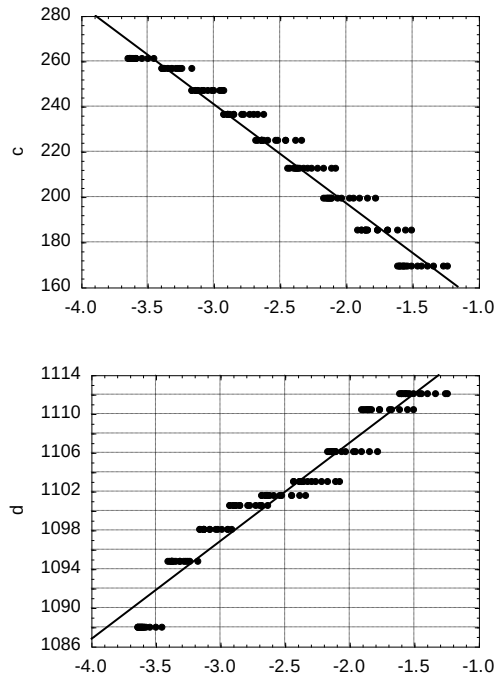


Figure 10. Coefficients a, b, c and d in the regression formula for the target X coordinate against the normalized Pupil-Glint Y for the left eye.

Relationship between (Pupil Y – Glint Y) and PoR$_y$

Figure 11 shows the relationships between the target Y coordinate and the average normalized pupil-glint Y (PG_Y) of 130 data sets (26 participants with 5 sets each) at 15 distinct X coordinates for the left and right eyes.

From Figure 11 it can be inferred that a straight line should fit each one of the 15 relationships very well. The following relationship was found between the normalised pupil-glint differences and the Y-coordinate of the target for the left eye at X=326:

$Y = 287.85y + 1220.7, \; R^2 = 0.9979.$

The R^2 values at the other X-positions of the targets are given in Table 4. In general terms, the relationship can be written as

$$Y = ay + b \quad (3)$$

In Blignaut (2013) this relationship between the target Y coordinate and *PGy* was expressed as a second order polynomial.

DOI 10.16910/jemr.7.1.4
This article is licensed under a
Creative Commons Attribution 4.0 International license.
ISSN 1995-8692

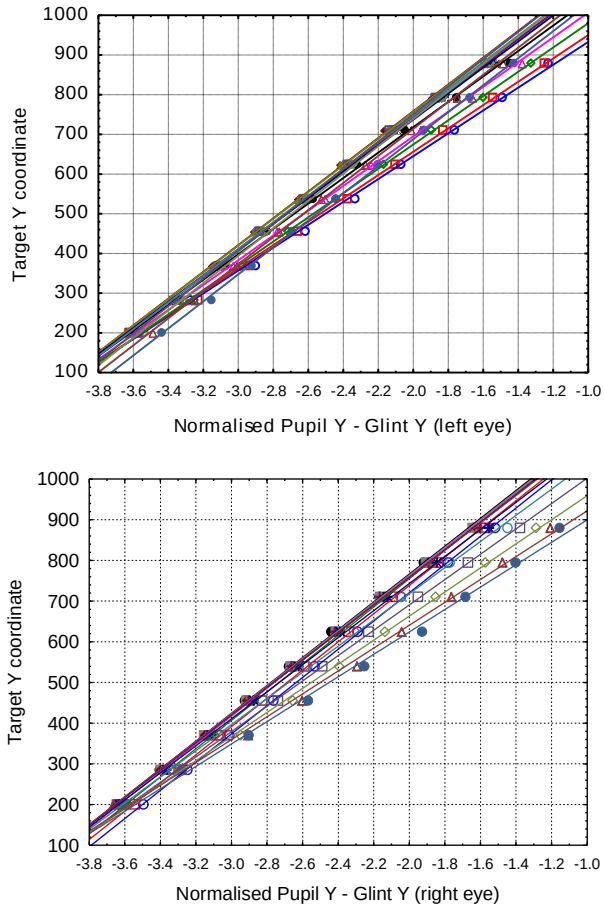


Figure 11. Target Y coordinate against average normalised pupil-glint Y of 26 participants at 15 distinct X coordinates for the left eye (top) and right eye (bottom).

Table 3. R^2 values of the polynomial fits for the coefficients in the regression formula for the target X coordinate against the normalised pupil-glint Y for the left and right eyes.

	a	b	c	d
Order of fit	1	1	1	3
R^2 left eye	0.97	0.84	0.97	0.96
R^2 right eye	0.91	0.88	0.92	0.94

Table 4. R^2 values for the regression formulas for Y at specific values of X.

Eye	Target Y coordinate							
	326	416	506	596	686	776	866	956
Left	0.9979	0.9983	0.9997	0.9995	0.9989	0.9991	0.9983	0.9993
Right	0.9997	0.9995	0.9993	0.9991	0.9994	0.9998	0.9997	0.9993

	1046	1136	1226	1316	1406	1496	1586
Left	0.9994	0.9994	0.9992	0.9992	0.9991	0.9996	0.9996
Right	0.9991	0.9991	0.9989	0.9996	0.9982	0.9956	0.9961

Figure 12 shows plots of the coefficients *a* and *b* against the average normalized pupil-glint X for the left and right eyes. It is evident that the trends of the plots change at when the normalised pupil-glint vector changes from negative to positive or when the glint moves to the other side of the pupil (cf. Figure 1). Therefore, it might be an option to have separate polynomial fits for $PG_X > 0$ and $PG_X < 0$, but that could cause sudden transition jumps when the glint moves to the other side of the pupil.

Table 5 shows the R^2 values of second and fifth order polynomial fits for the left and right eyes. An important consequence of a fifth order polynomial is that we will need at least six distinct X values for a calibration procedure and that there are calibration points at the edges of the area of interest to limit the impact of end-effects.

Table 5. R^2 values of polynomial fits for the coefficients in the regression formula for the target Y coordinate against the target X coordinate for the left and right eyes.

	a	b	a	b
Order of fit	2	2	5	5
R^2 left eye	0.906	0.910	0.962	0.952
R^2 right eye	0.949	0.963	0.981	0.979

The fifth order relationships could be written as follows:

$$a = a_5x^5 + a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0$$
$$b = b_5x^5 + b_4x^4 + b_3x^3 + b_2x^2 + b_1x + b_0$$

Substituting these into the general equation (3) above, we get

$$\begin{aligned} Y &= ay + b \\ &= a_5x^5y + a_4x^4y + a_3x^3y + a_2x^2y + a_1xy + a_0y \\ &\quad + b_5x^5 + b_4x^4 + b_3x^3 + b_2x^2 + b_1x + b_0 \end{aligned} \qquad (4a)$$

Since fifth order polynomials would need a large number of calibration points, the relationships were also approximated with second order polynomials:

$$a = a_2x^2 + a_1x + a_0$$
$$b = b_2x^2 + b_1x + b_0$$

yielding

$$\begin{aligned} Y &= ay + b \\ &= a_2x^2y + a_1xy + a_0y + b_2x^2 + b_1x + b_0 \end{aligned} \qquad (4b)$$

DOI 10.16910/jemr.7.1.4
This article is licensed under a
Creative Commons Attribution 4.0 International license.
ISSN 1995-8692

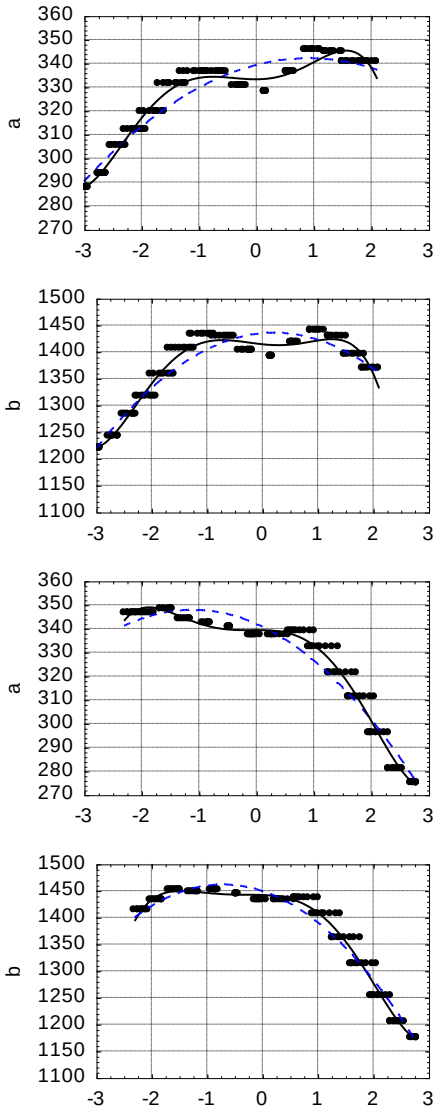


Figure 12. Coefficients a and b in the regression formulas for the target Y coordinate against the normalized Pupil-Glint X for the left eye (top two) and right eye (bottom two). Second and fifth order polynomial fits are shown.

Results

Comparison of polynomial models and calibration configurations

Table 6 shows the average accuracy over 130 data sets (26 participants with 5 data sets each) for selected calibration configurations (See Table 1 for details about each model). The second Mitsugami model (Mitsugami et al., 2003) gives the best results for the 9-point calibration configuration (0.87°) while no model gives acceptable results with 5 calibration points. The original Blignaut model provides the best accuracy with all calibration configurations with 14 or more points. The small improvement of 23 points over 14 points might not justify the extra effort.

Table 6. Average accuracy over 130 data sets (26 participants with 5 repetitions each) per model and calibration configuration. The headers refer to the model number and configuration of points, e.g. 14 (3434) means 14 points configured in 4 rows with 3, 4, 3 and 4 points respectively (see Figure 7). The best accuracy for every calibration configuration is boldfaced.

No.	Calibration configuration					
	5 (212)	9 (3x3)	14 (3434)	18(11X,9Y)	23 (54545)	135 (15x9)
1	1.40	1.29	1.15	1.16	1.14	1.11
2	**1.37**	1.09	0.96	0.94	0.94	0.90
3	2.94	0.91	0.76	0.75	0.73	0.68
4	4.93	**0.87**	0.72	0.70	0.68	0.64
5	1.50	1.09	0.95	0.96	0.94	0.90
6	1.72	0.93	0.84	0.83	0.81	0.78
7	3.43	0.80	0.62	0.62	0.60	0.56
8	6.17	2.06	**0.58**	**0.56**	**0.53**	**0.48**
9	6.24	2.09	0.62	0.59	0.57	0.53
10	7.76	4.74	0.70	0.67	0.62	0.55
11	7.32	3.07	0.62	0.60	0.57	0.52
12	8.74	5.53	0.70	0.67	0.62	0.54

Distribution of Accuracy

Figure 13 shows a bubble chart of the average error over all 130 data sets (26 participants with 5 data sets each), represented by the size of a bubble, against the X and Y coordinates of the validation targets for model 4 in combination with the 9-point calibration configuration. Although the average accuracy over the entire display area is reasonable (0.87°), there is a large area at the bottom of the screen where the accuracy is unacceptably bad. Figure 14 shows similar bubble charts for model 8 in combination with the 14- and 23-point calibration configurations. Not only is the average accuracy considerably better (0.58° and 0.53° respectively), the errors are also distributed more evenly across the entire screen.

Discussion

Using a data set with 26 participants, watching five sets of 135 gaze targets in a 15×9 grid, the effectiveness of a mapping function that was derived in an earlier study (Blignaut, 2013) was confirmed.

DOI 10.16910/jemr.7.1.4
This article is licensed under a
Creative Commons Attribution 4.0 International license.
ISSN 1995-8692

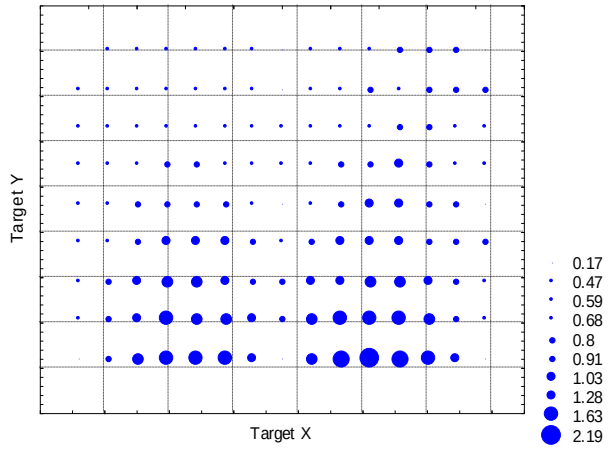


Figure 13. Bubble chart showing accuracy for model 4 in combination with the 9-point calibration configuration.

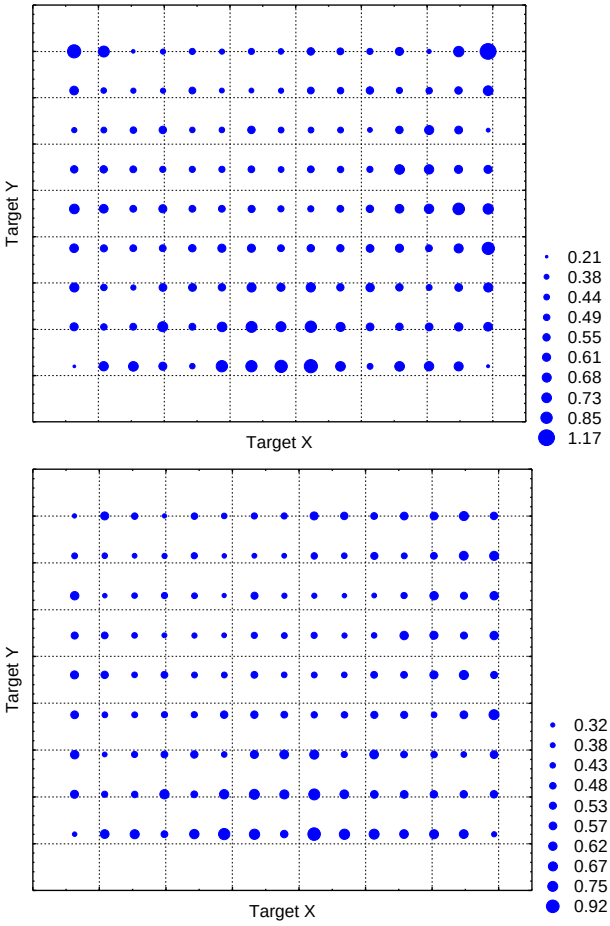


Figure 14. Bubble charts showing accuracy for model 8 in combination with the 14- (top) and 23-point (bottom) calibration configurations.

It was proven again that a simple system with one USB camera and a single infrared source is capable of achieving accuracy values that compare well with that of industrial systems (see discussion above). It is also not necessary to apply a complex mathematical model for gaze estimation. Mapping the pupil-glint vector to gaze coordinates can work as long as the mapping function is carefully selected along with the optimum configuration of calibration targets.

A systematic approach towards the selection of an appropriate mapping function was illustrated. Visual inspection of calibration data in one dimension while controlling for the other, can be used to discover relationships between the dependent (PoR) and independent (pupil-glint vector) variables. This approach can be used to derive a mapping function that will result in good (≅0.5°) accuracy values, provided that 14 or more calibration points are used. It was shown that complex, high-order, relationships do not necessarily provide better results than a model that was derived in an earlier study (Blignaut, 2013).

It is important to note that accuracy should not be sacrificed for the sake of calibration speed. It is believed that a few more seconds for the calibration routine is a worthwhile investment towards better results during studies. A proper implementation of the calibration procedure should allow participants to blink or even look away during the process in order to reduce eye fatigue. While capturing data for this study, participants completed a set of 135 gaze targets within three minutes, meaning that a calibration procedure with 23 targets should take less than 30 seconds to complete.

The best combination of mapping function and calibration configuration should ensure an even distribution of inaccuracy. The bubble chart in Figure 13 shows that there might be sudden occurrences of bad accuracy as well as regions where the accuracy is much worse than the average.

It was also motivated above that a good average accuracy might not be acceptable. Application of the often used 9-point calibration grid in combination with the Mitsugami model on the data in this study resulted in an average error 0.87° which is within the typical range of video-based systems (Hansen and Ji, 2010). The uneven distribution of error (Figure 13) might make the system unusable for gaze interaction as certain targets might be inaccessible.

DOI 10.16910/jemr.7.1.4
This article is licensed under a
Creative Commons Attribution 4.0 International license.
ISSN 1995-8692

As a consequence of the results in this study it follows that the accuracy of an eye tracking system cannot be expressed in absolute terms and depends on several factors that include hardware configuration as well as software algorithms, procedures and parameters. Manufacturers may, therefore, provide regular updates to the so-called firmware – the software that resides inside the machine itself and which is actually responsible for the eye tracking data that is provided to the researcher.

In other words, stating the accuracy of an eye tracker without reference to the calibration procedure that was used can be misleading. Researchers should also specify the version of firmware along with the make and model of the eye tracker that was used for a study.

References

Abe, K., Ohi, S., & Ohyama, M. (2007). An eye-gaze input system using information on eye movement history. In C. Stephanides (Ed.), *Universal access in HCI*, Part II. HCII2007, LNCS 4555, (pp 721-729). Berlin Heidelberg: Springer-Verlag.

Blignaut, P.J. (2013). A new mapping function to improve the accuracy of a video-based eye tracker. *Proceedings of the South African Institute of Computer Scientists and Information Technologists* (SAICSIT), 8-9 October 2013, East London, South Africa.

Blignaut, P.J., & Wium, D.J. (2013). The effect of mapping function on the accuracy of a video-based eye tracker. *Proceedings of the Eye Tracking South Africa Conference* (ETSA), 29-31 August 2013, Cape Town, South Africa.

Cerrolaza, J., & Villanueva, A. (2008). Taxonomic study of polynomial regression applied to the calibration of video-oculographic systems. *Proceedings of the 2008 Symposium on Eye Tracking Research and Applications (ETRA)*, 26-28 March 2008, Savannah, Georgia.

Cerrolaza, J.J., Villanueva, A., Villanueva, M., & Cabeza, R. (2012). Error characterization and compensation in eye tracking systems. *Proceedings of the 2012 Symposium on Eye Tracking Research and Applications (ETRA)*, 28-30 March 2012, Santa Barbara, California.

Crane, H.D., & Steele, C.M. (1985). Generation-V dual-Purkinje-image eye tracker. *Applied Optics*, 24, (pp 527-537).

Dodgson, N.A. (2004). Variation and extrema of human interpupillary distance. In J.O. Merritt, S.A. Benton and M.T. Blas (eds.) *Stereoscopic Displays and Virtual reality Systems XI*, SPIE Vol 5291, 19-22 January 2004, San Jose, California, ISSN 0277-786X.

Draper. N.R., & Smith, H. (1981). *Applied Regression Analysis.* John Wiley and Sons.

Hansen, D.W., & 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(3), (pp 478-500).

Hennessey, C., Noureddin, B., & Lawrence, P. (2008). Fixation precision in high-speed noncontact eye-gaze tracking. *IEEE Transactions on Systems, Man and Cybernetics – Part B: Cybernetics*, 38(2), April 2008.

Holmqvist, M., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H. and Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford: Oxford University Press.

Mitsugami, I., Ukita, N., & Kidode, M. (2003). Estimation of 3D gaze position using view lines. *Proceedings of the 12th International Conference on Image Analysis and Processing (ICIAP'03)*, IEEE.

Zhu, Z., & Ji, Q. (2005). Eye gaze tracking under natural head movements. *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*

DOI 10.16910/jemr.7.1.4
This article is licensed under a
Creative Commons Attribution 4.0 International license.
ISSN 1995-8692