Taxonomic Study of Polynomial Regressions Applied to the Calibration of Video-Oculographic Systems

Juan J. Cerrolaza*
Electrical and Electronics Department
Public University of Navarra

Arantxa Villanueva†
Electrical and Electronics Department
Public University of Navarra

Rafael Cabeza‡
Electrical and Electronics Department
Public University of Navarra

Abstract

Of gaze tracking techniques, video-oculography (VOG) is one of the most attractive because of its versatility and simplicity. VOG systems based on general purpose mapping methods use simple polynomial expressions to estimate a user's point of regard. Although the behaviour of such systems is generally acceptable, a detailed study of the calibration process is needed to facilitate progress in improving accuracy and tolerance to user head movement. To date, there has been no thorough comparative study of how mapping equations affect final system response. After developing a taxonomic classification of calibration functions, we examine over 400,000 models and evaluate the validity of several conventional assumptions. The rigorous experimental procedure employed enabled us to optimize the calibration process for a real VOG gaze tracking system and, thereby, halve the calibration time without detrimental effect on accuracy or tolerance to head movement.

CR Categories: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Input devices and strategies; I.4.9 [Image Processing and Computer Vision]: Applications

Keywords: gaze tracking, video-oculographic, taxonomic classification, calibration, polynomial equation, multiple linear regression.

1 Introduction

In recent years the number of applications of gaze tracking systems has increased considerably as the accuracy and performance of the trackers have improved. Of existing methods, those based on video-oculography are specially attractive because of their versatility, simplicity and non-intrusiveness. However, in gaze tracking technology there are still problems to overcome and improvements to be made. Despite improvements in feature detection and tracking [Morimoto et al. 2000], eye image analysis still presents problems in outdoor applications or when room lighting varies. Tolerance to head movements is another issue. To better detect tracking features, most systems need high resolution eye images, which require cameras with a narrower field. The system, thus, becomes more sensitive to head movements. Some researchers have tackled this problem by tracking the head and eye separately. For example, in the works of Ohno [Ohno and Mukawa 2004] and Beymer [Beymer and Flickner 2003] a stereo system is used to detect head position, and a narrow-field camera is steered to the head position detected to focus on the eye at high resolution. However, head movement can result in situations beyond the calibration scenario, and so the accuracy of most systems decreases with changes in the position of the camera relative to the head.

视频眼震

Gaze estimation is the process that infers gaze from the eye image. Once the image has been analyzed, and the features such as: pupil center or glint(s), have been extracted; gaze is deduced as a function of the image features. Assuming a fixed head position, the accuracy of a system is largely dependent on the gaze estimation method. The connection between image and gaze has been modeled in various ways, but the methods can be divided into two main groups depending on whether they are based on geometry or general purpose mapping.

Geometry based models express gaze as a function of the configuration of the system and subject. They explore the underlying geometry: such as, the position of the camera, screen size, and subject's characteristics, and their geometrical connections in order to construct mathematical expressions that describe gaze as a function of system parameters and variables. These models provide essential information about tracker behaviour and performance. They enable evaluation of accuracy, of sensitivity with respect to specific parameters, and of the number of calibration points needed. There are two drawbacks to these models. First, they are more difficult to construct. Second, and more importantly, they require additional calibration of part or the whole of system geometry: i.e., screen, lighting and camera [Guestrin and M.Eizenman 2006] [Hennessey et al. 2006] [Villanueva and Cabeza 2007].

Mapping methods are general purpose polynomial expressions that describe the gazed point as a function of image features and a set of unknown coefficients that are deduced during the subject's calibration process. The coefficients are deduced by means of a numerical fitting process, such as, multiple linear regression. They describe the Point of Regard (PoR), i.e, the point being gazed at on the screen (p_x, p_y), as a polynomial expression using selected image features as input. Compared to geometry based models, mapping methods provide little information about the intrinsic behaviour of the system; they are, however, much simpler to construct and do not require additional hardware calibration, which makes setup much faster for the system user. Ramanauskas has demonstrated that, in terms of the accuracy, the two approaches (i.e., geometry and mapping) can be very similar [Ramanauskas 2006].

There are three critical aspects to the design of a mapping function: the image features to be used as input to the estimation expression, the degree of the polynomial, and the number of terms. These aspects determine essential characteristics of the system, such as, the number of calibration points needed (to which subject calibration time is directly proportional), the accuracy, and the range of head

1.特征2.次数3.项数

*e-mail: juanjose.cerrolaza@unavarra.es
†e-mail: avilla@unavarra.es
‡e-mail: rcabeza@unavarra.es

ETRA 2008, Savannah, Georgia, March 26–28, 2008.
© 2008 ACM 978-1-59593-982-1/08/0003 $5.00

movement which is tolerable. Despite these relationships, there is a lack of consensus about how to determine the above three design aspects.

That higher order polynomials and that systematic use of complete expressions (i.e. including all terms) improve accuracy are some of the traditionally accepted ideas. However, there are few comprehensive studies on these issues. Morimoto et al. proposes the minimization of error by increasing the order of the polynomial, but just under stationary conditions [Morimoto and Mimica 2005]. The results of White et al. suggest that, with head movement, the accuracy with high order expressions was little better or even worse than that with simpler models [White et al. 1993]. Brolly et al. experiment with polynomials up to order three obtaining slight improvements over the equations of order two [Brolly and Mulligan 2004]. However, due to the experimental methodology employed by White et al., or to the three cameras system used by Brolly et al., it is not clear whether the conclusions are applicable to other video-oculography based gaze trackers such as the one under study in this work.

高阶带有头部移动，错误率更高

The work we present here is a thorough empirical study of general purpose polynomial mapping methods for a non-intrusive gaze tracking system, which tries to fill an existing gap in this field. We study the accuracy of the system as a function of the expression used, in terms of polynomial order, number of terms, and image features. Calibration requirements and sensitivity to changes in head position are also studied. In the next section, we present a detailed taxonomical organization of different calibration functions. Section 3 describes the experimental methodology and defines the parameters utilized to evaluate and classify all the equations considered. Finally, the results and conclusions drawn from this study are detailed in sections 4 and 5.

2 Objectives

The main objective of this work is to provide an exhaustive and detailed review of general purpose mapping methods used in VOG gaze tracking systems. Regression theory [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. The brute-force procedure, although rather cumbersome and very computationally demanding, allows both the study of specific regression models (according to certain criteria discussed later) and the investigation of preconceived ideas about how equation order and number of parameters affect accuracy. To achieve our objective in a rigorous and ordered way, we will first propose a taxonomic classification of the relevant polynomial equations. The taxonomic properties are presented below.

PoR Coordinate. Each coordinate of the PoR (PoR_X, PoR_Y) is mapped independently. Generally, the same polynomial expressions are used for each coordinate, although, if the problem is regarded as an issue of regression alone, there is no mathematical necessity for this. We will study both cases, PoR_X and PoR_Y.

Tracking Features. These are segmented features of eye images. The pupil center and the centers of the glint produced by the infrared light source are one of the most usual combinations of tracking features. Considering the pupil as an ellipse introduces new potential parameters: i.e., the eccentricity and the normalized coefficients of the conic equation. Since the choice of tracking features has implications for both image processing and hardware elements (multiple light sources), in this study we will consider four different system architectures: pupil center with one IR source (I), pupil center with two IR sources (II), pupil ellipse with one IR source (III), and pupil ellipse with two IR sources (IV).

离心率

Mapping Features. Working with certain derived variables can significantly improve the robustness of the system. In this study, several subsets of mapping features will be considered in order to analyze not only the best polynomial equation but also the most appropriate mapping variables. The mapping feature subsets are

a) *Basic features*: Direct use of the tracking parameters.

b) *Pupil center-glint vector*: The use of an additional reference point is an efficient strategy to compensate for head movements. The vector (or vectors if more than one light source is used) between the pupil center and the corneal reflection will be considered.

c) *Pupil-glint vectors with independent calibrations*: If more than one lighting source is used, each of the two pupil-glint vectors can be calibated indepedently. The PoR can be determined as the average of the two corresponding estimations.

d) *Normalized pupil-glint vectors*: When two corneal reflections are available, it is possible to normalize the pupil center-corneal reflection (PC-CR) vector with respect to glint separation distance. This additional nonlinear transformation reduces the effect of head movement.

e) *Normalized pupil-glint vectors with independent calibrations*: Two PC-CR normalized vectors can be calibrated independently and the PoR determined as an average.

f) *Average pupil center-corneal reflection vectors*: Instead of using two different vectors the two pupil-glint vectors can be averaged. This strategy reduces the number of mapping variables, and thereby enables the use and analysis of higher order polynomials.

g) *Average pupil-glint vectors and glints distance*: The separation between glints is used as an independent mapping feature instead of as a normalization variable.

h) *Average normalized pupil-glint vectors*: Use of the mean vector normalized with respect to glint separation distance.

To the features considered in the above mapping subsets, the ellipse parameters must be added when working with tracking sets *III* and *IV*. Obviously, not all the mapping subsets can be studied within a single-source configuration, and so tracking sets *I* and *III* are limited to mapping subsets *a* and *b*.

Polynomial Order. The potential effect that the order of the mapping function has on the accuracy or robustness of the gaze tracker will be analyzed by considering different orders. Although there are reports of many attempts to improve systems by increasing the order of the mapping functions, the validity of this approach has not been demonstrated, especially when head movement is possible.

Number of Terms. Given a set of variables and the order of the polynomial, it is common practice to make use of the most complete mathematical expression available. There is, however, no statistical basis for this practice. Furthermore, the systematic inclusion of terms to the mapping expression can lead to overfitting, a consequence which is too often ignored. An interesting advantage of working with shorter equations is the potential reduction in the number of calibration points, since less coefficients must be calculated. To investigate these matters, in this study we will not only consider the full polynomial expression but also any relevant simpler equations.

In view of the above detailed model classification, a huge number of polynomial equations are considered in this study with the objective of performing a complete revision of the general purpose mapping methods for VOG systems. The computational cost of this task is enormous, as is detailed in the methodology section. Thus, a filter is necessary to reduce the number of cases. Because the number of possible equations depends as much on the number of tracking features as on the polynomial order used in each one of the feature subsets, we have used two approaches to reduce their number. First, we have selected the most suitable elliptical parameters of the pupil. Second, we have limited the order considered for each of the feature subsets. Both processes are described in section 3. It is important to stress that the main objective of this work is to study the gaze estimation systems based on polynomial calibrations, being the study of the different images features extraction techniques out of its scope.

3 Methodology

Some previous works, such as Morimoto et al. [Morimoto and Mimica 2005], study different calibration functions of VOG systems by using synthetic images or computer simulation of subject eye movement. The theoretical nature of such studies makes it difficult to be sure whether all the results are applicable to real systems with real subjects. Other researchers, such as Ramanauskas or Cherif [Ramanauskas 2006] [Cherif et al. 2002], report work with real eye-trackers, but in which the number of experimental subjects was limited: five and three, respectively. The present study is based on a sample of 11 subjects using a real VOG system. The careful experimental procedure employed allows us to extract significant conclusions from the tests performed. A complete description of the methodology employed is given below.

3.1 Selection of Elliptical Parameters

修正4×4
测试
8×8

Some of the mapping equations involve variables related to treatment of the pupil image as an ellipse. The selection of parameters was carried out after detailed study based on the mathematical implementation of the geometry-based models proposed by Villanueva [Villanueva 2005], which provide a fully controlled environment for the study of alternative aspects of a VOG system. In particular, we studied the behaviour of the elliptical parameters for 27 head locations in a $10cm^3$ space in a virtual VOG system where the user was located at a distance of 60cm from a 17-inch screen. The elliptical parameters considered are the five conic equation coefficient normalized by the x^2 coefficient (1), the two semi-axes, the center, and the eccentricity.

$$x^2 + \frac{A_{12}}{A_{11}}xy + \frac{A_{22}}{A_{11}}y^2 + \frac{B_1}{A_{11}}x + \frac{B_{12}}{A_{11}}y + \frac{A_{12}}{A_{11}} = 0 \quad (1)$$

In order to compare parameter behaviour as a function of user head movements and select the parameters which are most robust to head movements, we used the coefficient of variation:

原始数据
标准差/
原始数据
平均数

$$CV(x) = \frac{\sigma_x}{\mu_x} \quad (2)$$

Defining a 7 × 7 regular grid on the screen, the average *CV* of a given elliptical feature, $CV(EF)$, is calculated as:

$$CV(EF) = \frac{1}{49}\sum_{i=1}^{49} CV_i(EF) \quad (3)$$

$$CV_i(EF) = \frac{\sigma_{EFi}}{\mu_{EFi}} = \frac{\sqrt{\frac{1}{27}\sum_{j=1}^{27}(EF_{i,j} - \overline{EF_i})^2}}{\overline{EF_i}} \quad (4)$$

$$\overline{EF_i} = \frac{1}{27}\sum_{j=1}^{27} EF_{i,j} \quad (5)$$

where $EF_{i,j}$ is the parameter value when the user looks at the i^{th} grid point in the j^{th} head position. It is important to realize that this parameter, $CV(EF)$, gives information about the variation of a particular feature when the user moves the head, but not when he look at a different screen point (sensitivity to eye movement). This is, the lower the value of CV, the better the robustness against head movements. The results of this selection process are presented in section 4.

3.2 Experimental Methodology

In order to obtain real data, we used a VOG system developed in our labs. This system used a Pentium IV computer (3.6GHz, 1GB RAM) with a 17-inch color monitor. A standard camera, with a 35mm objective, and set to a resolution of 1024 × 768 pixels was placed under the monitor to collect BW images at 30 fps. An infrared light source, an LED operating at 880nm, was located on each side of the screen. The selection of the optimal calibration expression depends on the system configuration employed, consequently one of the most standard configurations has been selected for the study.

As our objective was to test the robustness and accuracy of many different calibration functions, two types of grid were used. The first one, the calibration grid, was used to solve the linear regression system (i.e., to calculate the polynomial coefficients of the mapping functions) and consisted of a 4 × 4 regular distribution of points. Once the calibration process had been completed, each model was tested with a second, denser, 8 × 8, grid. This latter grid provided detailed information about the behaviour of each model when tracking over the whole screen.

In practice, head movement is an important issue in VOG systems, and, for this reason, we studied three different locations of the subject. In position 1 the subject was located centrally at approximately 63 cm from the camera. In positions 2 and 3 the subject was displaced 5 cm forwards and backwards in the direction perpendicular to the screen, respectively. We did not investigate head movements along the plane parallel to the monitor because, as concluded by other authors ([Morimoto and Mimica 2005]), VOG systems are sufficiently robust to such movements.

Eleven subjects participated in our data collection process. Each subject was placed at position 1 and was asked to look at the sequence of points on the 4 × 4 calibration grid. The resulting data were used to calibrate the system. Once the mapping function was adjusted, the system accuracy was tested for the 8 × 8 grid points, with the subject in each of the three positions, i.e., *1-3*. Each point was displayed on the screen for a period of time long enough to acquire 30 images. Throughout the data collection session, an exact subject head position was established by means of a chin rest; this enabled us to create the defined test environment needed to extract valid and consistent conclusions.

3.3 Processing of Data

This study involves over 400,000 models, and so it was not feasible to test them all, real-time, in a stand-alone, physical VOG system. To conduct our tests, then, we used the two-step procedure described below. First, we set up the system to record, for all the experimental test sessions, the relevant image characteristics (i.e. all the potential tracking features) while the user was looking at the different grid points. Second, the processing of all this information as well as the evaluation of mapping functions was performed off-line, as detailed below.

3.3.1 Filtering of Data

For each grid point, the system processes a total of 30 images, extracting and logging their tracking features. Despite the use of a chin rest, it is reasonable to assume that not all of these images are equally apt; consequently, some kind of data filtering is needed. However, since there is no subset of features common to all the mapping equations, specific filtering criteria are indicated.

Each processed image can be considered as a n-dimensional vector. Let X_{ik} be the k^{th} vector containing the *n* mapping features of the i^{th} grid point evaluated in a certain head position:

$$X_{ik} = \{x_{ik1}, x_{ik2}, ..., x_{ikn}\}^T \tag{6}$$

where x_{ikz} is the z^{th} mapping feature and $k = 0, ..., 30$ since 30 images are acquired each time. The mean, $\overline{X_i}$, and the $n \times n$ covariance matrix, S_i, are calculated as:

$$\overline{X_i} = \frac{1}{30} \sum_{k=1}^{30} X_{ik} \tag{7}$$

$$S_i = \frac{1}{30} \sum_{k=1}^{30} (X_{ik} - \overline{X_i})(X_{ik} - \overline{X_i})^T \tag{8}$$

Mahalanobis distance, defined as

$$D_{ik} = \sqrt{(X_{ik} - \overline{X_i})^T S_i^T (X_{ik} - \overline{X_i})} \tag{9}$$

is a typical distance measurement between variables in a multidimensional space. This distance allows us to sort the n vectors according to their distance from the mean. Data filtering was, in this way, achieved by selecting the 60% of the images closest to the mean.

3.3.2 Evaluation of Models

From the real data obtained with the physical VOG system, the evaluation of different models was carried out off-line. The task was performed with *Mathematica*^{TM}, which was run simultaneously on 15 desktop computers and required 36 hours. The study of each mapping equation can be considered as a two-phase process: calibration and evaluation.

Calibration. The calibration coefficients are calculated according to the corresponding subset of tracking features and the function under evaluation. We applied a least squares estimation procedure which used the screen coordinates of the 4×4 grid used in *position 1* as the dependent variable (response variable). For each grid point, the vector of independent variables (explanatory variables) was obtained by averaging the 18 filtered vectors of mapping features. Note that the length of the mapping equations considered is limited by the size of the calibration grid. Consequently, none of the models tested has more than 16 terms.

Evaluation. Once the mapping equation had been obtained, its predictive capability was tested over the 8×8 grid for each of the three user locations. The data obtained during this evaluation process allow us to characterize and evaluate each model according to various parameters which are detailed below.

3.3.3 Validation of Results

Although processing was off-line, all data were obtained from real sessions with a real VOG system (implemented in C++) under real working conditions. In order to validate off-line results (obtained with *Mathematica*^{TM}), we compared the calibration and evaluation processes with their equivalent implementations in the VOG tracker. Both platforms generated almost identical coefficients through the least squares estimation process, and any differences in predicted gaze coordinates (i.e. the screen coordinates) never exceeded two pixels. These results validate the off-line processing approach we adopted.

3.4 Evaluation Parameters

We used several parameters to evaluate and quantify the suitability of models.

Average Error. This is a unique parameter that reflects the global behaviour of a given mapping function. It is obtained by averaging all the prediction errors over the 8×8 grid. $\overline{ErrX_{i,j}}$ and $\overline{ErrY_{i,j}}$ represent the X and Y coordinate average error of the i^{th} user in the j^{th} *position*, where $j = 1, 2$ or 3. Thus, the total average error in a certain position can be obtained as

$$\begin{aligned} \overline{ErrX_j} &= \frac{1}{11} \sum_{m=1}^{11} \overline{ErrX_{m,j}} \\ \overline{ErrY_j} &= \frac{1}{11} \sum_{m=1}^{11} \overline{ErrY_{m,j}} \end{aligned} \tag{10}$$

where 11 is the number of subjects.

Maximum Error. This error parameter gives information about the worst expected behaviour of a model. If $ErrMxX_{i,j}$ and $ErrMxY_{i,j}$ are the biggest error of the i^{th} user in the j^{th} position over the 8×8 grid, the average maximum errors are defined as

$$\begin{aligned} \overline{ErrMxX_j} &= \frac{1}{11} \sum_{m=1}^{11} ErrMxX_{m,j} \\ \overline{ErrMxY_j} &= \frac{1}{11} \sum_{m=1}^{11} ErrMxY_{m,j} \end{aligned} \tag{11}$$

Standard Deviation. Since there were 11 subjects, an additional parameter to evaluate each model is the standard deviation of the average and maximum errors. The standard deviation of the average error in X and Y is given by

$$\begin{aligned} \sigma_{ErrXj} &= \sqrt{\frac{1}{11} \sum_{m=1}^{11} (ErrX_{m,j} - \overline{ErrX_j})^2} \\ \sigma_{ErrYj} &= \sqrt{\frac{1}{11} \sum_{m=1}^{11} (ErrY_{m,j} - \overline{ErrY_j})^2} \end{aligned} \tag{12}$$

The standard deviation of the maximum error is defined as

$$\begin{aligned} \sigma_{ErrMxXj} &= \sqrt{\frac{1}{11} \sum_{i=1}^{11} (ErrMxX_{m,j} - \overline{ErrMxX_j})^2} \\ \sigma_{ErrMxYj} &= \sqrt{\frac{1}{11} \sum_{i=1}^{11} (ErrMxY_{m,j} - \overline{ErrMxY_j})^2} \end{aligned} \tag{13}$$

These parameters give information about a model's stability.

The independent evaluation of coordinates allowed us to select the optimal equation for each axis. This approach is more flexible and more coherent than assigning the same model to both axes.

Number of Terms. As mentioned previously, the number of terms of a model can be considered as an interesting evaluation parameter because the number of calibration points required is directly related to the equation length, and consequently, calibration is easier and quicker for shorter models.

Head Movement Tolerance. The robustness of each model to user head movement is an interesting additional characteristic. If a system is directed to users with disabilities whose movement capacity is highly limited, the model selection process should be more focused on the evaluation of *position 1* parameters. However, if a more general system is desired, the *position 2* and *position 3* data must be taken into account.

3.5 Filtering of Models

The study includes an enormous number of models, which makes it difficult to make a detailed interpretation of the results for each. By filtering equation evaluation parameters, we achieve two things: first, models whose behaviour is clearly inaccurate are removed; and second, equations with similar responses can be grouped under the same model category. Each category is defined by the set of tracking features (*I* to *IV*), the subset of mapping characteristics (*a* to *h*) and the order of the equations.

Based on the evaluation parameters described above, we propose a parallel filtering process. In this way, the representative equation group is obtained as the intersection of the N_1 equations with the lowest mean error, the N_2 with the lowest maximum error, and the N_3 and N_4 equations with the lowest standard deviation of these errors. By controlling the N_{ith} parameters, a weighted filtering is obtained. In particular, only the *position 1* data are considered, and the mean error is defined as the most relevant parameter by setting N_1 as the lowest one.

Once the N models from the intersection have been determined, their representative features are averaged in order to obtain global group descriptors. Where there are enough suitable models in the category, we used N = 10. In the filtering process, we only considered *position I*; tolerance to head movement and number of terms being left to future refinements in the process.

4 Results and Discussion

4.1 Elliptical Parameters

With the objective of improving system tolerance to head movements, elliptical parameters were included in the study. There are a large number of elliptical parameters available for inclusion in the set of predictive features. Since the number of possibles mapping equations depends directly on this set, it is necessary to reduce the number of parameters as far as possible to make the study computationally affordable. We used the elliptical parameter's coefficient of variation (*CV*) as filtering criterion. Figure 1 shows the coefficient of variation, under conditions with head movement, of all the elliptical parameters considered. The parameters with lower *CV* are less dependent on head position.

Although the two center coordinates have the highest value of *CV* of all the geometrical features, i.e., center, semiaxes and eccentricity, they cannot be excluded from the study because they are necessary to calculate the PC-CR vector. Eccentricity has a low *CV* value and contains information about the two semiaxes, and is, therefore, kept as a tracking feature whilst the semiaxes parameter is dropped.

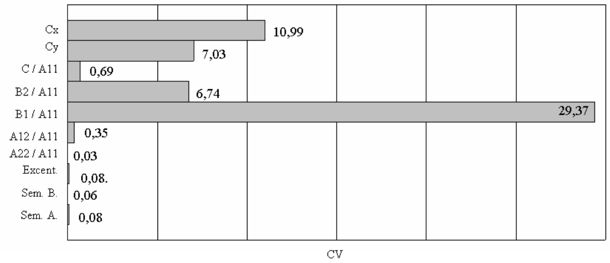

Figure 1: *Coefficient of variation (CV) of elliptical features.*

With regard to the behaviour of the conic normalized equation, the second order coefficients present lower values of *CV* than the first order ones. However, an additional simplification can be obtained by using a standard technique for model optimization: the relationships between the coefficients of each predictive parameter employed indicate that the two second order equation coefficients can be combined by addition (because their individual regression coefficients are practically identical).

Therefore, the new tracking features incorporated into the mapping functions when the pupil is considered as an ellipse, i.e., in *configuration III* and *configuration IV*, are the ellipse center, ellipse eccentricity and the sum of the two normalized second order coefficients.

4.2 Results

Description of the detailed analysis of each mapping equation included in this study would run to many pages and is, therefore, beyond the scope of this article. However, by adequate filtering of the available data as explained above in section *3.5*, it is possible to obtain results for each category of mapping equation. Tables 1 to 4 contain the error data for each subcategory, i.e., mapping features and order for the four system configurations: *I, II, III* and *IV*. Each row in these tables contains the average and maximum error of a mapping function category, obtained by considering the best N equations according to the filtering criteria (average error, maximum error and standard deviations). In practice, these N models gave almost identical results.

Configuration I. The models obtained with the basic set of features, i.e., without difference vectors or normalization of any kind, have considerable error levels, even in absence of head movement (*position I*). The use of the derived PC-CR vector (mapping subset *b*) significantly improves behaviour under static conditions: average errors of 10 and 11 pixels for PoRX and PoRY, respectively. However, errors in *position 2* (101 and 69 pixels) and *3* (62 and 44 pixels) remain high.

Configuration II. Despite the various transformations tested (mapping subsets *b, c, f* and *g*), the errors with simple PC-CR vectors in configuration II were comparable to those obtained in *configuration I*. However, the use of glint separation as a normalization parameter (*d, e,* and *h*) gave models whose maximum errors and robustness to user head movement were significantly better. The average errors in *positions 2* were now 16 and 32 pixels (for PoRX and PoRY, respectively). In *positions 3* these errors were 19 and 33 pixels. An interesting result is that, apart from normalization, none of the proposed transformations applied to the basic mapping features (two independent calibrations or averaging the two PC-CR vectors) has any important effect on the errors.

Configuration III. In this configuration we only studied models up to order 2 because of the increased number of mapping features. However, the incorporation of elliptical features to the mapping function produces no improvements with respect to *configuration I*.

Configuration I

	position 1				*position 2*				*position 3*			
	X		Y		X		Y		X		Y	
	Avg	Max	Avg	Max	Avg	Max	Avg	Max	Avg	Max	Avg	Max
a1	44	104	46	101	419	563	763	888	266	434	497	590
a2	30	89	28	94	325	664	516	876	215	519	453	668
b1	13	51	15	66	100	248	70	182	68	220	45	126
b2	11	41	12	56	101	225	69	176	64	222	45	121
b3	10	39	11	52	101	219	69	167	62	227	44	121
b4	10	39	11	52	101	225	69	169	63	229	44	121

Table 1: *Configuration I: Pupil center and one IR source. Average and maximum error (pixels)*

Configuration II

	position 1				*position 2*				*position 3*			
	X		Y		X		Y		X		Y	
	Avg	Max	Avg	Max	Avg	Max	Avg	Max	Avg	Max	Avg	Max
a1	23	70	45	101	203	316	771	889	149	310	505	593
b1	11	47	13	68	76	197	89	202	53	215	54	135
b2	9	42	11	60	53	155	75	166	36	200	47	123
c1	9	42	16	67	45	138	70	178	30	147	44	116
c2	8	36	11	55	44	129	68	163	29	147	43	107
c3	8	36	10	52	43	126	67	159	28	146	43	105
c4	8	35	10	51	43	128	67	160	29	148	43	106
d1	10	44	15	73	17	81	34	90	20	176	36	127
d2	10	45	11	59	17	80	33	87	21	194	33	104
e1	10	41	18	71	16	77	36	94	19	163	38	109
e2	8	39	13	60	16	76	33	81	19	163	34	102
e3	8	38	11	56	16	77	32	80	19	175	33	90
e4	8	36	11	55	16	77	32	81	19	176	33	90
f1	10	43	16	68	46	144	70	180	30	184	45	125
f2	9	38	12	57	45	137	68	166	30	187	44	116
f3	9	37	11	53	45	133	68	164	30	188	43	114
f4	8	35	11	53	44	131	68	171	30	187	44	116
g1	10	42	17	71	55	162	109	226	37	197	65	148
g2	9	38	11	58	49	146	67	166	31	192	44	119
h1	10	42	18	73	17	78	36	90	20	172	38	114
h2	10	41	13	60	16	78	34	83	19	173	35	102
h3	8	38	11	58	16	78	33	82	19	189	33	101
h4	8	38	11	56	17	78	32	82	18	188	33	101

Table 2: *Configuration II: Pupil center and two IR sources. Average and maximum error (pixels)*

中心点
正确率
更高

Configuration III

	position 1				*position 2*				*position 3*			
	X		Y		X		Y		X		Y	
	Avg	Max	Avg	Max	Avg	Max	Avg	Max	Avg	Max	Avg	Max
a1	50	114	43	103	271	438	756	958	196	374	489	698
b1	37	143	35	150	130	476	99	394	100	67	74	364
b2	10	43	11	58	98	236	70	267	60	223	48	254

Table 3: *Configuration III: Pupil ellipse and one IR source. Average and maximum error (pixels)*

Configuration IV

	position 1				*position 2*				*position 3*			
	X		Y		X		Y		X		Y	
	Avg	Max	Avg	Max	Avg	Max	Avg	Max	Avg	Max	Avg	Max
a1	25	76	37	110	126	286	697	953	108	278	435	648
b1	11	53	12	92	64	202	68	281	44	217	47	284
c1	11	50	13	63	50	176	68	226	31	182	45	231
c2	8	39	10	56	46	154	65	232	27	173	47	253
d1	11	51	14	99	20	103	32	223	23	195	40	291
e1	12	49	15	70	19	98	34	179	23	184	41	230
e2	9	44	10	58	18	140	30	206	23	235	37	249
f1	12	51	13	65	51	180	68	243	31	195	46	237
f2	8	41	10	62	48	162	67	254	29	191	48	255
g1	11	52	12	91	58	191	84	258	38	212	54	282
h1	12	51	15	75	19	101	34	194	23	188	41	238
h2	9	47	11	62	19	137	31	217	24	232	38	251

Table 4: *Configuration IV: Pupil ellipse and two IR sources. Average and maximum error (pixels)*

Configuration IV. As with *configuration III*, use of elliptical features does not improve the behaviour of the system. The results obtained with these models are very similar to those obtained with *configuration II*, although the maximum errors are higher.

One of the justifications for the equation filtering proposed in section *3.5* was the absence in each configuration of particular optimal functions for a subset of mapping features and a given order. In other words, equations can be grouped, and within each group the behaviour is very similar. The results presented in the tables, show that the differences within certain subcategories are minimal (1-2 pixels). In particular, it is interesting to observe how increasing the order of the model has scarcely any effect. This similarity of behaviour within subcategories, makes it possible to construct a graphical comparison of the four system configurations by averaging the best subcategories of each one (Fig. 2).

In Fig. 2, whose source information is detailed in Tables 1 to 4, it can be seen that all configurations can provide similar average and maximum errors under static conditions, i.e., position 1. These errors are approximately 9 and 12 pixels for PoRX and PoRY, respectively. Consequently, the simplest configuration for a VOG gaze tracker whose tolerance to head movements is not essential is a single IR-LED; the PC-CR vector should be used as mapping feature. Although the differences between *I* and *III* are minimal, the latter requires a more complex software implementation to accurately detect the pupil ellipse, and so *configuration I* is the simplest option. On the other hand, if a more versatile system is desired, allowing the user certain freedom of head movement, a configuration with two light sources (*II* or *IV*) is required in order to attain acceptable average errors (16 and 32 pixels in X and Y axes, respectively).

Any configuration that employs the normalized PR-CR vector for calibration variables provides satisfactory results, even in *positions 2* and *3*. However, maximum error values do vary between models; although there are no significant differences between the average errors of configuration II and IV (about 2 pixels of difference), the maximum errors are clearly higher if the elliptical parameters are incorporated into the mapping. The increased maximum error with elliptical parameters may be related to the processing complexity required to detect the elliptical pupil contour, processing which is more sensitive to deformation due to glints or partial eyelid occlusions than that corresponding to center-detection. Further research is required to determine whether more robust pupil contour detection procedures can reduce the maximum errors or even achieve better average errors

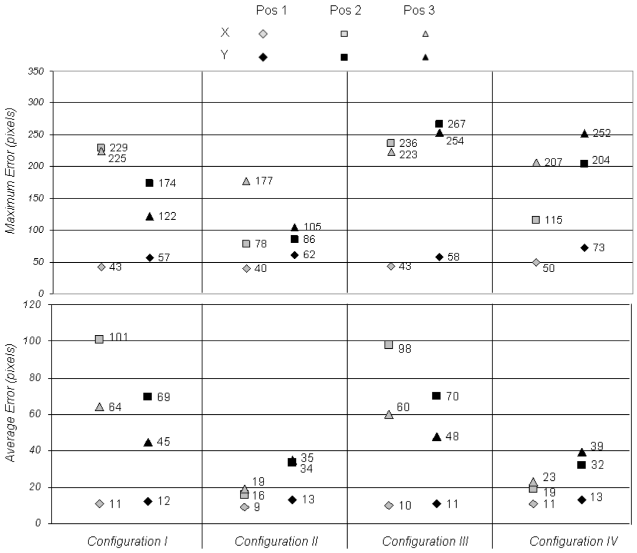

Figure 2: *General comparative of the four system configurations.*

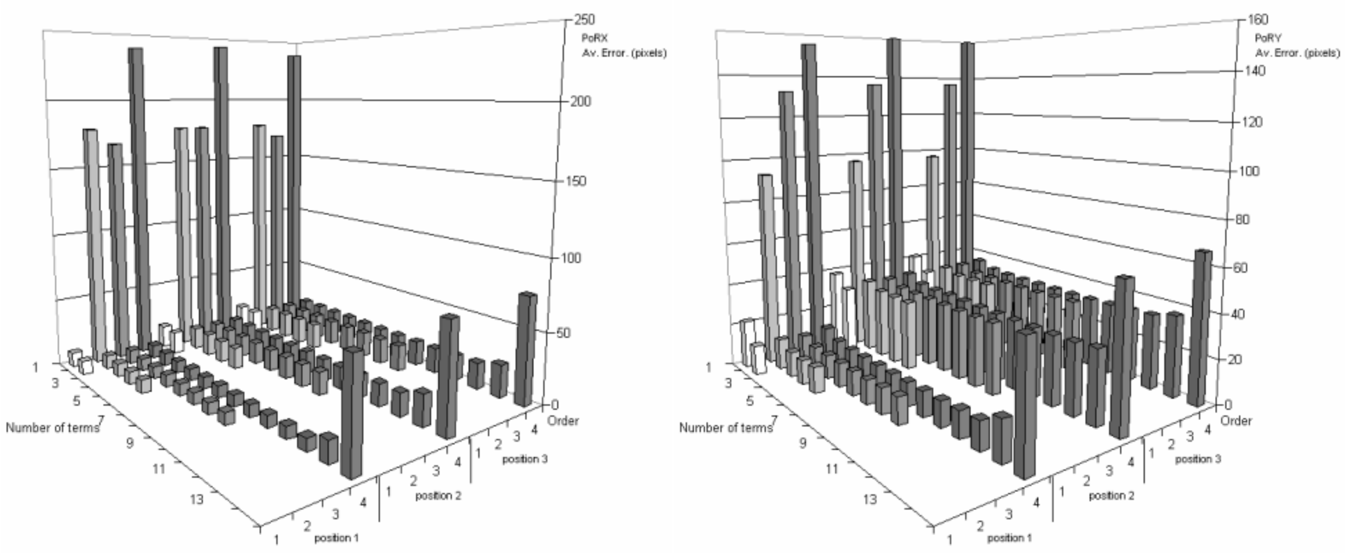

Figure 3: *Average error vs. Order and Number of terms. Models II.h.*

参数太多，
未选练好

4.3 Order and Number of Terms

Tables 1 to 4 indicate the effect of mapping function order. These results show that higher order only reduces the error significantly for certain very simple models (*I.a.1 or III.b.1*). Other models behave quite satisfactorily with first order expressions. Figure 3, which is derived from data for configuration II with mapping feature set h., plots average error response versus the mapping equation order and the number of terms. Data correspond to the best equation (minimum average error in *position 1*) for a given order and length.

As can be seen in the figure, increasing the order does not reduce the error significantly; the average errors remain almost constant (9 and 11 pixels for PoRX and PoRY, respectively, in *position 1*). In fact, the tolerance to head movement is very similar too, and there are no appreciable differences in *positions 1*, *2* (16 and 33 pixels) or *3* (19 and 33 pixels) with changes in order. Leaving aside single-term models, the number of terms has almost no effect on errors. This is an important result because it contradicts systematic use of complete expressions, and confirms that accurate calibration can be achieved through functions with a reduced number of terms.

4.4 Optimizing of Calibration

In view of the above results, optimization of the calibration process, by reducing the user time required whilst maintaining accuracy and robustness to head movement, becomes a valid objective. Since the number of calibration points is directly related to the length of the mapping functions, shorter expressions allow us to test the system behaviour when a reduced calibration grid is used. Although system accuracy is not significantly affected by the number of terms, as was discussed in the previous section, a different distribution of calibration points could affect system behaviour, and thus each proposed model must be studied in detail.

From Tables 1 to 4, the equations that minimize the average and maximum error in both coordinates, PoRX and PoRY, are two functions from II.e.2 and II.e.3, respectively. However, it is hard to select a particular group of better equations since the differences between many of them are statistically insignificant (1-2 pixels). Among the multiple possible combinations, one that also minimizes the number of terms is:

$$\begin{aligned} PoRX &= C_{x0} + C_{x1}\widehat{x} + C_{x2}\widehat{y} + C_{x3}\widehat{x}^2 \\ PoRY &= C_{y0} + C_{y1}\widehat{y} + C_{y2}\widehat{x}^2 + C_{y3}\widehat{x}\widehat{y} + C_{y4}\widehat{x}^2\widehat{y} \end{aligned} \tag{14}$$

where ($\widehat{x}$,$\widehat{y}$) are the coordinates of the normalized PC-CR vector and $C_{x,y}$ are the coefficients to be determined during calibration. Note that models from class II.e use the two available PC-CR vectors to perform two independent calibrations, from which the final PoR is obtained as the average. Thus, equations from (14) are duplicated, which complicates the software implementation. However, since the difference between class II.e and any other that utilizes the normalized PC-CR vectors, like d and h, is negligible, it is possible to obtain an alternative pair of equations which are easier to implement. We have used this approach to develop the calibration model:

对每个vector都有
一对方程式对应

$$\begin{aligned} PoRX &= D_{x0} + D_{x1}\overline{x} + D_{x2}\overline{x}^3 + D_{x3}\overline{y}^2 \\ PoRY &= D_{y0} + D_{y1}\overline{x} + D_{y2}\overline{y} + D_{y3}\overline{x}^2\overline{y} \end{aligned} \tag{15}$$

where ($\overline{x}$,$\overline{y}$) are the coordinates of the average normalized PC-CR vector and $D_{x,y}$ are the coefficients to be determined. As mentioned previously, the differences between some of the equations are minimal so (14) and (15) must be considered merely as examples.

The small number of terms in these expressions allows us to define a new calibration grid of 8 points (Fig. 4). Use of the new grid halves the calibration time associated with the old 4 × 4 one. We studied the response of both pairs of functions, (14) and (15), to the new distribution of calibration points according to the methodology described in section 3. The absolute errors, i.e., considering both coordinates, are given in Table 5. Although with the new, 8-point calibration grid there is a small increase in the error, system behaviour is still fully acceptable.

The proposed configurations have been also implemented and tested in the real VOG gaze tracker employed in this study; the deviation from the off-line processed data was less than two pixels.

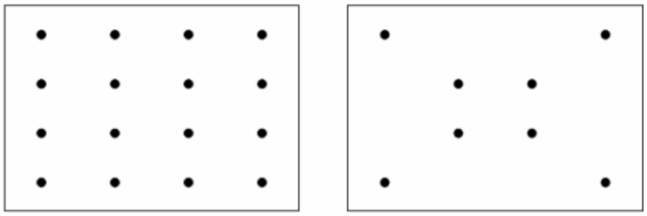

Figure 4: *The original 4×4 calibration grid and the new 8 points one.*

	position 1		position 2		position 3	
	Avg. Err.	Max. Err.	Avg. Err.	Max. Err.	Avg. Err.	Avg. Err.
(e.2 , e.3)	19 / 0.46^o	69 / 1.65^o	40 / 1.04^o	106 / 2.76^o	43 / 0.96^o	185 / 4.10^o
(h.2 , h.2)	19 / 0.46^o	71 / 1.70^o	40 / 1.04^o	107 / 2.79^o	44 / 0.98^o	213 / 4.73^o
8 points calibration grid						
(e.2 , e.3)	16 / 0.35^o	64 / 1.50^o	36 / 0.93^o	104 / 2.71^o	40 / 0.89^o	183 / 4.06^o
(h.2 , h.2)	17 / 0.40^o	67 / 1.60^o	38 / 0.99^o	103 / 2.68^o	41 / 0.91^o	205 / 4.55^o
4x4 calibration grid						

Table 5: *Optimized Calibration Process: Average and maximum total errors using both calibration grids (pixels/(^{o})).*

5 Conclusions

This article presents a detailed and rigorous experimental study of the general mapping methods used in VOG gaze tracking systems. We propose a taxonomic classification, which allows us to study the polynomial calibration expressions used in these systems and to evaluate the validity of several conventional procedures. Four different configurations; eight subsets of mapping features; expressions up to fourth order; and the independent study of both coordinates, PoRX and PoRY, constitute the set of over 400,000 calibration functions which we analyze.

Our results are obtained from a rigorous experimental procedure that combines data obtained from studies involving eleven subjects using a real VOG gaze tracker, and off-line simulations, which had previously been validated. The results demonstrate that there is no single equation whose response is significantly better than the rest for any of the configurations studied. By adequate selection of the best models, according to criteria such as the average and maximum error or the standard deviation, it is possible to extract general conclusions for each type of model.

Under stationary conditions it is possible to obtain valid mapping functions based on the PC-CR vector with any of the hardware configurations. Thus, if a system is being designed for users whose head movement is highly limited, a very simple configuration with a single IR light source will provide acceptable results. However, for a more robust gaze tracker, one that can deal with a certain amount of user head movement; it is necessary to utilize two IR light sources. All configurations that use the PC-CR vectors normalized with respect to glint separation have low average errors for different user head positions. Those models that do not include elliptical parameters in the mapping features subsets have lower maximum errors.

We also studied the effect of increasing the order and number of terms of the calibration function. The results show that higher order polynomials do not noticeably improve system behaviour. Similarly, use of the most complete mathematical expressions does not enhance accuracy; the average error is almost constant and independent of the number of terms. These results are of practical importance in optimization of the calibration process because they demonstrate the existence of simple and robust models that require fewer calibration points.

Acknowledgements

COGAIN (511598): Communication by Gaze Interaction. Network of Excellence FP6-2003-IST-2

References

BEYMER, D., AND FLICKNER, M. 2003. Eye gaze tracking using an active stereo head. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, 451–458.

BROLLY, X. L. C., AND MULLIGAN, J. B. 2004. Implicit calibration of a remote gaze tracker. *Conference on Computer Vision and Pattern Recognition Workshop 8*, 134.

CHERIF, Z. R., NAT-ALI, A., MOTSCH, J. F., AND KREBS, M. O. 2002. An adaptive calibration of an infrared ligth device used for gaze tracking. *IEEE Instrumentation and Measurement Techonology Conference*, 1029–1033.

DRAPER, N. R., AND SMITH, H. 1981. *Applied Regression Analysis*. John Wiley and Sons.

GUESTRIN, E., AND M.EIZENMAN. 2006. General theory of remote gaze estimation using pupil center and corneal reflections. *IEEE Transactions on Biomedical Engineering 53*, 6, 1124–1133.

HENNESSEY, C., NOUREDDIN, B., AND LAWRENCE, P. 2006. A single camera eye-gaze tracking system with free head motion. In *ETRA '06: Proceedings of the 2006 symposium on Eye tracking research & applications*, ACM Press, New York, NY, USA, 87–94.

LI, D. 2006. *Low-cost eye-trackig for human computer interaction*. Master's thesis, Iowa State University Human Computer Interaction Program. http://thirtyixthspan.com/openEyes/MS-Dongheng-Li-2006.pdf.

MERCHANT, J., MORRISSETTE, R., AND PORTERFIELD, J. 1974. Remote measurement of eye direction allowing subject motion over one cubic foot of space. *IEEE Transactions on Biomedical Engineering 21*, 4 (July), 309–317.

MORIMOTO, C. H., AND MIMICA, M. R. M. 2005. Eye gaze tracking techniques for interactive applications. *Comput. Vis. Image Underst. 98*, 1, 4–24.

MORIMOTO, C., KOONS, D., AMIR, A., AND FLICKNER, M. 2000. Pupil detection and tracking using multiple light sources. *Image and Vision Computing 18*, 4, 331–335.

OHNO, T., AND MUKAWA, N. 2004. A free head, simple calibration, gaze tracking system that enables gaze-based interaction. In *Proceedings of the Eye Tracking Research & Applications Symposium*, 115–122.

RAMANAUSKAS, N. 2006. Calibration of video-oculographical eye-tracking system. *ISSN 1392-1215 Electronics and Electrical Engineering*, 8, 65–68.

VILLANUEVA, A., AND CABEZA, R. 2007. Models for gaze tracking systems. *EURASIP Journal on Image and Video Processing*.

VILLANUEVA, A. 2005. *Mathematical Models for Video-Oculography*. PhD thesis, Public University of Navarra.

WHITE, K. P., HUTCHINSON, T. E., AND CARLEY, J. M. 1993. Spatially dynamic calibration of an eye-tracking system. *IEEE Transactions on Systems, Man, and Cybernetics 23*, 4, 1162–1168.