Gaussian Processes as an Alternative to Polynomial Gaze Estimation Functions

Laura Sesma-Sanchez*
Public University of Navarra

Yanxia Zhang†
Lancaster University

Hans Gellersen‡
Lancaster University

Andreas Bulling§
Max-Planck-Institute for Informatics

Abstract

Interpolation-based methods are widely used for gaze estimation due to their simplicity. In particular, feature-based methods that map the image eye features to gaze, are very popular. The most spread regression function used in this kind of method is the polynomial regression. In this paper, we present an alternative regression function to estimate gaze: the Gaussian regression. We show how the Gaussian processes can better adapt to the non-linear behavior of the eye movement, providing higher gaze estimation accuracies. The Gaussian regression is compared, in a simulated environment, to the polynomial regression, when using the same mapping features, the normalized pupil center-corneal reflection and pupil center-eye corners vectors. This comparison is done for three different screen sizes. The results show that for larger screens, where wider gaze angles are required, i.e., the non-linear behavior of the eye is more present, the outperformance of the Gaussian regression is more evident. Furthermore, we can conclude that, for both types of regressions, the gaze estimation accuracy increases for smaller screens, where the eye movements are more linear.

Keywords: gaze estimation, Gaussian process, polynomials

Concepts: •**Human-centered computing** → **User interface design;** •**Theory of computation** → **Gaussian processes;**

1 Introduction

Gaze estimation can be defined as the process that infers gaze from eye image data. The information of where someone is gazing can be very useful for many applications in the field of psychology, advertising or web design among others. Gaze can also be used to control a device such as a tablet, a computer or a television. In this case, the accuracy requirements are very high. There are many research goals in the eye tracking community and, although it is not always necessary, high accuracy is a key point for many applications.

Gaze estimation techniques can be classified in model-based and interpolation-based [Hansen and Ji 2010]. The majority of existing gaze tracking methods require an explicit procedure for calibrating the model or regression parameters to fit individual users. Interpolation- or regression-based methods are widely used for gaze estimation due to their simplicity. They model the relation between eye image data and gaze via a regression and they do not require a calibrated hardware setup. In particular, feature-based methods, which rely on local features such as pupil centers, eye corners and reflections, are very popular. The most spread regression function used in this kind of method is the polynomial regression and the most common mapping feature that relies on infrared (IR) illumination is the pupil center-corneal reflection (pc-cr) vector [Cerrolaza et al. 2012]. When no IR light is present, the pupil/iris center-eye corners (pc-ec) can be used [Sesma et al. 2012].

Finding an optimal mapping function for gaze estimation is a research objective. Cerrolaza et al. [2012] analyzes a large set of polynomial functions with different number of terms and degree. They conclude that neither the use of higher orders or complete mathematical expressions significantly improves the accuracy of a gaze tracking system. Blignaut [2013] studies the relationship between gaze targets and pc-cr vectors to propose derived polynomials that optimize the gaze mapping. Furthermore, more recently, Blignaut [2014] compares different polynomial functions that appear in the literature, in terms of gaze estimation accuracy. According to this study, the polynomial function that achieves the highest accuracy is one proposed by Blignaut [2013].

Polynomial functions are not necessarily the most reasonable to map gaze due to the spherical surface of the eye. Lately, other researchers in the field have studied other types of regression for feature-based gaze estimation such as generalized regression neural networks [Zhu and Ji 2004], support vector machines [Zhu et al. 2006] and Gaussian regression [Zhang et al. 2014]. The Gaussian regression in the latter work is merely for horizontal movements.

In this paper, we present an alternative regression function to estimate gaze: the Gaussian regression. The Gaussian regression is compared, in a simulated environment, to the polynomial regression when using the same mapping functions for three different screens: a 10-inch tablet, a standard 22-inch monitor and a 40-inch screen. We show how the Gaussian processes can better adapt to the non-linear behavior of the eye movement, providing higher gaze estimation accuracies and how the accuracy, in general, increases for smaller screens, where the eye movements are more linear.

2 Gaussian processes for gaze estimation

A Gaussian process is a statistical distribution, where any finite linear combination of samples has a joint Gaussian distribution. A key fact of the Gaussian process is that it can be fully defined by a mean function and a covariance function [Rasmussen and Williams 2006]. Therefore, when a zero mean can be assumed, the process is completely specified by estimating the covariance function. A commonly used covariance function is the sum of a squared exponential and independent noise that has the following form:

$$k_y\left(\mathbf{x}_p, \mathbf{x}_q\right) = \sigma_f^2 e^{-\frac{1}{2l^2}\left(\mathbf{x}_p - \mathbf{x}_q\right)^2} + \sigma_n^2 \delta_{pq}, \quad (1)$$

where k_y is the covariance function for the noisy targets y, x_p and x_q are input vector samples, l is the characteristic length-scale, σ_f is the signal variance and σ_n the noise variance. The last three parameters are call hyperparameters and they can be optimized by

*e-mail:laura.sesma@unavarra.es
†e-mail:yazhang@lancaster.ac.uk
‡h.gellersen@lancaster.ac.uk
§andreas.bulling@acm.org

ETRA '16,, March 14-17, 2016, Charleston, SC, USA
ISBN: 978-1-4503-4125-7/16/03
DOI: http://dx.doi.org/10.1145/2857491.2857509

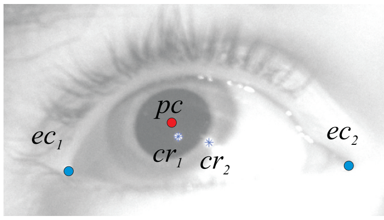


Figure 1: *Tracking Features.*

maximizing the marginal likelihood using the following derivatives:

$$\frac{\partial}{\partial\theta}\log p(\mathbf{y}|X,\theta) = \frac{1}{2}\operatorname{Tr}\left((\boldsymbol{\alpha}\boldsymbol{\alpha}^T - K^{-1})\frac{\partial K}{\partial\theta_j}\right), \qquad (2)$$

where $\theta\,(l,\sigma_f,\sigma_n)$ are the set of hyperparameters to optimize, $\boldsymbol{\alpha} = K^{-1}\mathbf{y}$ and K is the covariance matrix. Note that as we have assumed a zero-mean Gaussian process, the output may need to be centered. Moreover, when multiple inputs, it is important that they have similar scales.

In the case of gaze estimation, the Gaussian process maps the relation between the input mapping features and the output point of regard. A Gaussian process for each axis, i.e., vertical and horizontal, is required. The mapping features are computed using the tracking features that appear in Figure 1: the pupil center pc, eye corner centers ec_1 and ec_2 or corneal reflections cr_1 and cr_2. Two different mapping features are evaluated in this work, both two dimensional. One is the widely used normalized pc-cr vector [Cerrolaza et al. 2012] $v_1(v_{1_x}, v_{1_y})$:

$$v_1 = \frac{1}{2}\cdot\left(\frac{pc - cr_1}{\|cr_1 - cr_2\|} + \frac{pc - cr_2}{\|cr_1 - cr_2\|}\right), \qquad (3)$$

where pc, cr_1 and cr_2 are the image coordinates of the pupil center and the corneal reflection and $\|cr_1 - cr_2\|$ is the Euclidean distance between cr_1 and cr_2. And the other one is the normalized pc-ec [Sesma et al. 2012] $v_2(v_{2_x}, v_{2_y})$, which does not required IR light:

$$v_2 = \frac{1}{2}\cdot\left(\frac{pc - ec_1}{\|ec_1 - ec_2\|} + \frac{pc - ec_2}{\|ec_1 - ec2\|}\right), \qquad (4)$$

where pc, ec_1 and ec_2 are the image coordinates of the pupil center and the eye corners and $\|ec_1 - ec_2\|$ is the Euclidean distance between ec_1 and ec_2.

The Gaussian process needs to be characterized to estimate gaze. For this, the hyperparameters of the covariance function are estimated via a calibration procedure similar to the one required to calibrate the coefficients of the polynomial regressions. The user looks at different points from a calibration grid. The eye data obtained and the known gaze points are used to optimize the hyperparameters. Once the hyperparameters are learned, the covariance function can be used to predict and estimate gaze. The more training samples, the more accurate estimates. The GPML toolbox [Rasmussen and Nickisch 2010] is used for the implementation.

2.1 Comparison to polynomial regressions

The motivation to propose the Gaussian regression for gaze tracking is that it can adapt better to the non-linearity behavior of the eye movement than the polynomial regressions. Although polynomial regressions are the most spread regression functions in feature-based gaze estimation, we think that Gaussian regressions are an alternative. In this work, two polynomial regressions are implemented and compared to the Gaussian regression. The complete order two equation, which is used in many commercial systems and research prototypes [Morimoto and Mimica 2005], $poly$ 1:

$$\begin{pmatrix} PoR_x \\ PoR_y \end{pmatrix} = K \begin{pmatrix} 1 \\ v_x \\ v_y \\ v_x^2 \\ v_y^2 \\ v_x v_y \end{pmatrix},$$

where PoR_x and PoR_y are the point of regard in screen coordinates, v_x and v_y are the normalized vectors and K is the unknown coefficient matrix. And one proposed by Blignaut [2013], which is reported to give the highest accuracy for polynomials, $poly$ 2:

$$\begin{aligned} PoR_x = \; & K_{x0} + K_{x1}v_x + K_{x2}v_y + K_{x3}v_x v_y + K_{x4}v_x^2 \\ & + K_{x5}v_x^2 v_y + K_{x6}v_x^3 + K_{x7}v_x^3 v_y, \end{aligned}$$

$$\begin{aligned} PoR_y = \; & K_{y0} + K_{y1}v_x + K_{y2}v_y + K_{y3}v_x v_y + K_{y4}v_x^2 \\ & + K_{y5}v_y^2 + K_{y6}v_x^2 v_y, \end{aligned}$$

where PoR_x and PoR_y are the point of regard in screen coordinates, v_x and v_y are the normalized vectors and $K_{x0}\cdots K_{x7}$ and $K_{y0}\cdots K_{y6}$ are the unknown coefficients.

Note that there is one polynomial equation for each axis and that, for comparison purposes, the mapping features used are the same as the ones used in the Gaussian regression (see Equations 3 and 4).

3 Evaluation Framework

The aim of this work it to evaluate the Gaussian regression as alternative to polynomial regression for feature-based gaze estimation. For this purpose, a study was carried out in a simulated environment. For the analysis, the eye simulator framework from Böhme et al. [2008] was used. It consists of a MATLAB library that permits to simulate a complete eye tracking environment. It provides the pupil and glint positions in the image plane with sub pixel precision and gives the possibility to add Gaussian noise with zero mean and certain standard deviation. The work from Böhme et al. [2008] was enhanced to include the eye corners to be able to obtain the normalized pc-ec vector. It was done in a similar way as in the work of Sesma et al. [2012]; assuming that the eye corners are fixed for a certain head position. The distance between corners is 3 cm and an offset of 1 mm is applied to the other axes so that the two corners are not aligned. For the eye model, the standard parameters of the simulator are used, such as a 7.98 mm cornea radius and 2 and 6 $degrees$ fovea angles.

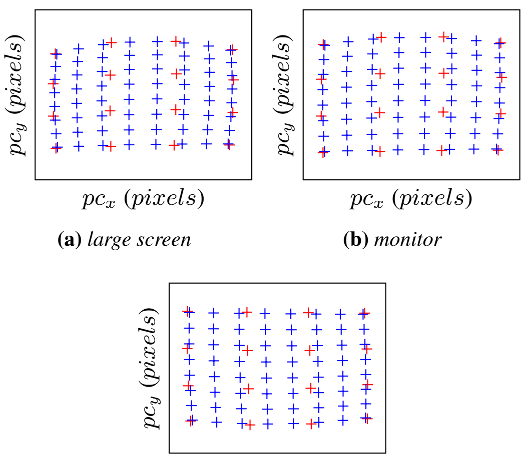
Three standard working scenarios are simulated for three different screens: a 10-inch tablet, a standard 22-inch monitor and a 40-inch screen. The camera which captures single eye images is the same for all the three setups, with an image resolution of 1600x1200 $pixels$ and a 1/3" image sensor. Two different camera focal lengths have been tested: 6 and 35 mm. Furthermore, two evaluations have been carried out, one considering perfect feature detectors and another adding Gaussian noise with zero mean and std 0.1 $pixels$ to the feature positions. In general, the user is sitting in front of the eye tracker. The camera is placed below the monitor and if present, the infrared light sources are attached to both sides of the screen. The characteristics of the screen and working distances used for each of the three different setups can be seen in Table 1. To simulate a realistic scenario, the user is centered on the screen at different working distances depending on the screen size. The calibration procedure consists in gazing at a 4x4 grid on the screen and the

Table 1: *Three different setups.*

	Screen size	Screen resolution	Working distance
Large screen	40″	1920x1080 mm	80 cm
Monitor	22″	1680x1050 mm	55 cm
Tablet	10″	1280x800 mm	30 cm

Table 2: *Mean and Std Euclidean gaze estimation errors in degrees for a simulated user for the test grid when using perfect feature detectors.*

		$poly$ 1		$poly$ 2		Gaussian	
Screen	Method	Mean	Std	Mean	Std	Mean	Std
large screen	pc-cr	0.76	0.30	0.26	0.13	0.07	0.05
	pc-ec	0.82	0.33	0.29	0.14	0.09	0.06
monitor	pc-cr	0.38	0.15	0.15	0.07	0.02	0.01
	pc-ec	0.41	0.16	0.16	0.08	0.03	0.02
tablet	pc-cr	0.23	0.09	0.09	0.04	0.01	0.00
	pc-ec	0.25	0.10	0.09	0.04	0.01	0.01



(a) *large screen* **(b)** *monitor* **(c)** *tablet*

Figure 2: *Pupil center coordinates when looking at the calibration points (red) and test points (blue).*

same procedure is repeated for testing using an 8x8 grid. The calibration and testing grid are created leaving a $0.05 * screen\ size$ and $0.06 * screen\ size$ offset at the edges of the screen. The camera captures 30 frames per grid point and the working distances are the same for both procedures. For the evaluation, the Euclidean gaze estimation errors ge_{ED} for every test point are calculated. Furthermore, the mean and standard deviation of the ge_{ED} are used to have information about the global behavior of the regression methods.

4 Results

The objective of this work is to proof that Gaussian regressions are an alternative for feature-based gaze estimation. The gaze estimation errors obtained with the Gaussian regression and two different polynomial regressions are compared for three setups with three different screen sizes for the camera focal lengths 6 and 35 mm when considering perfect and noisy feature positions.

The outperformance of the Gaussian regression can be seen in Table 2. The table shows the mean and std ge_{ED} in $degrees$ for a simulated user for the test grid for the different setups and regressions considered when perfect feature detectors are used. In this case, the focal length used does not affect the gaze estimation errors. The error differences are higher for the large screen setup because the non-linear behavior of the eye movement appears when looking at some parts of the screen. This can be observed in Figure 2, where the pupil center image coordinates, when looking at the calibration and testing points of the three different setups, are represented. The non-linearity is more evident at the top rows, where we can see that the pupil centers follow a curve. Note how the non-linearity increases at the sides of the screen. For the monitor setup, however, the non-linearity is not as visible and for the smallest screen, it is practically nonexistent. Furthermore, in general, we can see in Table 2 that the gaze estimation accuracy increases as the screen size decreases. This is due to the more linear performance of the eye movements when gazing at smaller screens from a standard working distance. This is an advantage when estimating gaze with mobile devices such as a tablet or a mobile phone.

Table 3: *Mean and Std Euclidean gaze estimation errors in degrees for a simulated user for the test grid when using camera focal lengths 6 and 35 mm and applying a Gaussian noise with zero mean and std 0.1 $pixels$ to the feature centers.*

(a) *f=6 mm*

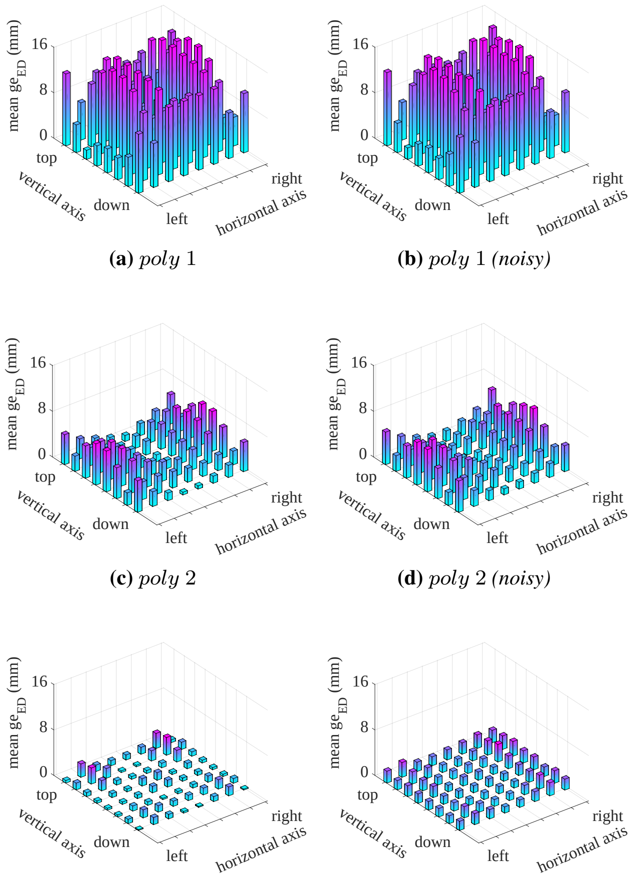
		$poly$ 1		$poly$ 2		Gaussian	
Screen	Method	Mean	Std	Mean	Std	Mean	Std
large screen	pc-cr	0.94	0.21	0.66	0.14	0.62	0.12
	pc-ec	0.86	0.30	0.40	0.11	0.31	0.05
monitor	pc-cr	0.53	0.10	0.41	0.08	0.39	0.07
	pc-ec	0.45	0.14	0.24	0.06	0.19	0.02
tablet	pc-cr	0.30	0.06	0.22	0.04	0.21	0.03
	pc-ec	0.27	0.09	0.13	0.03	0.10	0.01

(b) *f=35 mm*

		$poly$ 1		$poly$ 2		Gaussian	
Screen	Method	Mean	Std	Mean	Std	Mean	Std
large screen	pc-cr	0.77	0.30	0.28	0.13	0.12	0.03
	pc-ec	0.82	0.33	0.29	0.14	0.10	0.05
monitor	pc-cr	0.39	0.14	0.16	0.06	0.07	0.01
	pc-ec	0.41	0.16	0.16	0.07	0.05	0.02
tablet	pc-cr	0.23	0.09	0.09	0.04	0.04	0.01
	pc-ec	0.25	0.10	0.09	0.04	0.03	0.01

The results have shown that the Gaussian regression outperforms the two evaluated polynomial regressions when perfect feature detectors are considered. It is expected, however, that for noisy feature positions, the differences between the regressions decrease; these differences decrease as the error increases. This is demonstrated with the gaze estimation errors achieved with the two evaluated focal lengths. The Gaussian error of zero mean and std 0.1 $pixels$ implies higher errors for the smaller focal length, as the image resolution of the eye area is much smaller. Table 3 shows the mean and std of the ge_{ED} for a focal length of 6 and 35 mm when noisy data is considered. The results show how the setup with focal length 6 mm is more affected by the noise, in such a way that the differences between regressions decreases. In this case, the performance of the $poly$ 2 and the Gaussian is very similar. However, when the feature detector's are very accurate at higher focal lengths, such as 35 mm, the Gaussian regression outperforms the other two polynomial regressions as it happened with perfect feature detectors.

The accuracy of a gaze estimation method cannot only be deter-



(a) $poly$ 1 **(b)** $poly$ 1 *(noisy)*

(c) $poly$ 2 **(d)** $poly$ 2 *(noisy)*

(e) *Gaussian* **(f)** *Gaussian (noisy)*

Figure 3: *Euclidean gaze estimation error in mm for a simulated user for the test grid for a large screen and camera focal length 35 mm when using three different regression functions with the normalized pc-cr vector with perfect feature detector's and with Gaussian noise with zero mean and std 0.1 $pixels$.*

mined by the mean accuracy, as Blignaut [2014] claims. Therefore, Figure 3 shows the $mean\ ge_{ED}$ in mm for the test grid for a large screen and focal length 35 mm when using the three different regression methods with the normalized pc-cr vector with perfect feature detector's and with Gaussian noise with zero mean and std 0.1 $pixels$. We can observe how the error distribution is not uniform and how the Gaussian regression greatly improves the accuracy of the polynomial $poly$ 1 and $poly$ 2. However, the performance of the $poly$ 2 is also very good. The outperformance of $poly$ 2 over $poly$ 1 was already demonstrated with real data in the work of Blignaut [2014]. In this work, the result is validated with simulated data.

5 Conclusions

In this paper, we present an alternative regression function to estimate gaze: the Gaussian regression. The Gaussian regression is compared, in a simulated environment, to the polynomial regression when using the same mapping features for three different screens: a 10-inch tablet, a standard 22-inch monitor and a 40-inch screen. The results show that for larger screens, where wider gaze angles are required, i.e., the non-linear behavior of the eye is more present, the outperformance of the Gaussian regression is more evident. However, note that Gaussian processes can be computational expensive with their computation cost dependent on the input feature vector's dimension. Furthermore, we can conclude that, for both regressions, the gaze estimation accuracy increases for smaller screens, where the eye movements are more linear. This is an advantage when using gaze estimation for mobile devices such as a tablet or a mobile phone. Therefore, the gaze angle range that takes place, is an important fact to take into account when evaluating and characterizing gaze estimation algorithms. The next steps in this research will include the evaluation of Gaussian regressions with real data to validate the promising results that have been obtained within a simulated environment. Moreover, other variables such as head movements and number of calibration points are to be explored.

Acknowledgements

The author would like to thank the Ministry of Economy and Competitiveness of Spain for the FPI Grant BES-2010-031304.

References

BLIGNAUT, P. 2013. A new mapping function to improve the accuracy of a video-based eye tracker. In *Proc. SAICSIT '13*, ACM, NY, USA, 56–59.

BLIGNAUT, P. 2014. Mapping the pupil-glint vector to gaze coordinates in a simple video-based eye tracker. *Journal of Eye Mov. Research 7(1)*, 4, 1–11.

BÖHME, M., DORR, M., GRAW, M., MARTINETZ, T., AND BARTH, E. 2008. A software framework for simulating eye trackers. In *Proc. ETRA '08*, ACM, 251–258.

CERROLAZA, J. J., VILLANUEVA, A., AND CABEZA, R. 2012. Study of polynomial mapping functions in video-oculography eye trackers. *ACM Trans. Comput.-Hum. Interact. 19*, 2 (July), 10:1–10:25.

HANSEN, D., AND JI, Q. 2010. In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Trans. Pattern Anal. Mach. Intell. 32*, 3, 478–500.

MORIMOTO, C., AND MIMICA, M. 2005. Eye gaze tracking techniques for interactive applications. *Computer Vision and Image Understanding 98*, 1, 4–24.

RASMUSSEN, C. E., AND NICKISCH, H. 2010. Gaussian processes for machine learning (gpml) toolbox. *The Journal of Machine Learning Research 11*, 3011–3015.

RASMUSSEN, C. E., AND WILLIAMS, C. K. I. 2006. *Gaussian Processes for Machine Learning*. the MIT Press.

SESMA, L., VILLANUEVA, A., AND CABEZA, R. 2012. Evaluation of pupil center-eye corner vector for gaze estimation using a web cam. In *Proc. ETRA '12*, ACM, New York, NY, USA, 217–220.

ZHANG, Y., BULLING, A., AND GELLERSEN, H. 2014. Pupil-canthi-ratio: A calibration-free method for tracking horizontal gaze direction. In *Proc. AVI '14*, ACM, New York, NY, USA, 129–132.

ZHU, Z., AND JI, Q. 2004. Eye and gaze tracking for interactive graphic display. *Mach. Vision Appl. 15*, 3 (July), 139–148.

ZHU, Z., JI, Q., AND BENNETT, K. 2006. Nonlinear eye gaze mapping function estimation via support vector regression. In *Proc. ICPR 2006*, vol. 1, 1132–1135.