Pupil-Canthi-Ratio: A Calibration-Free Method for Tracking Horizontal Gaze Direction

Yanxia Zhang
Lancaster University
Lancaster, United Kingdom
yazhang@lancaster.ac.uk

Andreas Bulling
Max Planck Institute for Informatics
Saarbrücken, Germany
andreas.bulling@acm.org

Hans Gellersen
Lancaster University
Lancaster, United Kingdom
hwg@comp.lancs.ac.uk

ABSTRACT

Eye tracking is compelling for hands-free interaction with pervasive displays. However, most existing eye tracking systems require specialised hardware and explicit calibrations of equipment and individual users, which inhibit their widespread adoption. In this work, we present a light-weight and calibration-free gaze estimation method that leverages only an off-the-shelf camera to track users' gaze horizontally. We introduce *pupil-canthi-ratio* (PCR), a novel measure for estimating gaze directions. By using the displacement vector between the inner eye corner and the pupil centre of an eye, PCR is calculated as the ratio of the displacement vectors from both eyes. We establish a mapping between PCR to gaze direction by Gaussian process regression, which inherently infers averted horizontal gaze directions of users. We present a study to identify the characteristics of PCR. The results show that PCR achieved an average accuracy of 3.9 degrees across different people. Finally, we show examples of real-time applications of PCR that allow users to interact with a display by moving only their eyes.

Categories and Subject Descriptors

H.5.m [**Information Interfaces and Presentation (e.g. HCI)**]: Miscellaneous; I.5.4 [**Pattern Recognition**]: Applications—*Computer Vision*

General Terms

Human Factors, Measurement

Keywords

Calibration-free, Eye Tracking, Vision-based, Gaussian Regression, Gaze-based Interaction, Pervasive Displays

1. INTRODUCTION

Gaze is an attractive modality for pervasive displays, such as gaze-contingent displays, digital advertisement, webpage analysis, etc. We are efficient at using our eyes, as we naturally look at objects that we are interested in, and we can move our eyes faster and with less effort than other body parts. However, eye gaze is difficult to harness for use in everyday activities. Many existing eye tracking systems offer fine-grained detection of eye movements. However, they are mostly designed for lab-based environments, where the systems employ specialised hardware that requires additional illumination (e.g. infrared) [2]. To achieve accurate detection, these systems also require an explicit procedure for calibrating system parameters to fit individual users. The calibration process hampers natural interaction, as it is often cumbersome and unnatural [1, 3, 6]. Prior research proposed the use of visual saliency for auto-calibration using monocular images, but the setup requires a chin rest [5]. For gaze interaction to be useful in everyday scenarios, we need a system that works out-of-lab and is readily available for wide deployment.

In this work, we present a lightweight vision-based method that detects horizontal gaze directions from camera images. We contribute a calibration-free solution that requires no specialised hardware, works across different users and is suitable for deployment in uncontrolled environments. We propose *Pupil-Canthi-Ratio* (PCR), a novel measure for tracking horizontal eye movements based on the symmetry of our eyes. Our method uses the inner eye corner

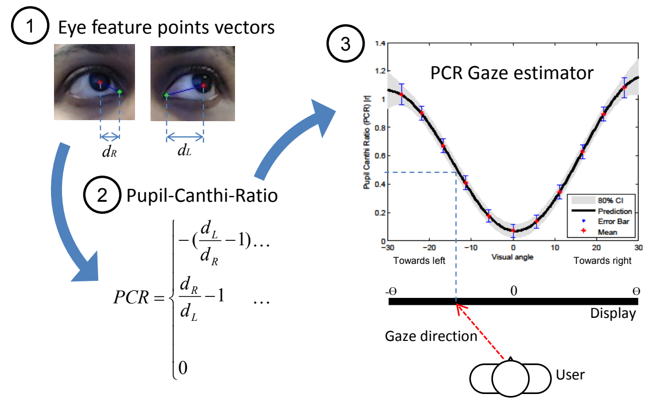


Figure 1: PCR measure. (1) the system extracts inner eye corners and pupil centers from camera images for determining the displacement vectors of left (d_L) and right eyes (d_R). (2) PCR is defined as the ratio of the vectors from both eyes, describing the degree of deviation from looking straight ahead. (3) we establish a mapping between PCR and gaze direction on a display by Gaussian process regression.

AVI '14, May 27 - 30, 2014, Como, Italy
Copyright 2014 ACM 978-1-4503-2775-6/14/05...$15.00.
http://dx.doi.org/10.1145/2598153.2598186

as the stable reference point and calculates a displacement vector from the moving pupil center, and PCR is calculated as a ratio of the displacement vectors of both eyes (see Fig. 1). The changes of PCR describe the degree of users' eye movements towards left or right, and we leverage this characteristic for detecting horizontal gaze directions.

We collect gaze direction data from 12 participants. We establish a mapping between PCR and gaze direction by applying Gaussian process regression on the data. Our results show that PCR achieves an average accuracy of 3.9 degrees across different people. As an added advantage, the minimal hardware requirement (i.e. a single webcam) and the low computational power of our method make PCR an ideal approach for wide deployment. We have developed example applications to showcase the utility of the PCR measure. Using this measure, gaze direction can be sensed in real-time and applied for selecting or scrolling content on a large display.

2. CHARACTERISATION OF THE PUPIL-CANTHI-RATIO

Our proposed system requires a single webcam for capturing images of users standing in front of a display. We employ Zhang et al.'s image processing techniques for extracting eye feature points [7]. When the system detects a user's face, it crops the left and the right eye regions, and then the system extracts the inner eye corner and the pupil center as feature points (see Fig. 1(1)).

From the extracted points, the system calculates a horizontal distance for each eye. Let C_i denotes the x coordinate of the inner eye corner and P_i denotes the x coordinate of the pupil center, the horizontal distance d_i is defined as $d_i = |C_i - P_i|$, where $i \in \{R, L\}$ represents the right and the left eye respectively. Using the horizontal distances of the right and the left eyes, we define *Pupil-Canthi-Ratio* (PCR) as a distance ratio as follow:

$$PCR = \begin{cases} -(\frac{d_L}{d_R} - 1), & \text{if } d_L - d_R > 0; \\ \frac{d_R}{d_L} - 1, & \text{if } d_L - d_R < 0; \\ 0 \end{cases} \quad (1)$$

A negative PCR indicates a gaze direction of looking towards left, while a positive PCR indicates looking towards right.

PCR indicates how far a user's gaze is averted from the center of his visual field. In order to characterise PCR as an estimator for horizontal gaze, we collected data from 12 participants in a controlled study. The purpose of this study is to establish a mapping relationship between PCR and the maximum range in visual angle of horizontal gaze that we can cover with our eyes.

2.1 Data Collection

In our study, we used a 55 inch (121cm×68.5cm) display, with a resolution of 1920x1080 pixels, mounted on a wall at the height of 120cm (lower bezel) above ground. To record eye images, we used a Logitech HD Pro Webcam C920 with a resolution of 1280x720 pixels and a frame rate of 30Hz. The camera was mounted on a tripod and positioned at a distance of 50cm in front of the display.

The study was conducted in an office environment under normal lighting conditions. We recruited 12 participants (seven female), aged between 19 and 33 years. None of them wore glasses during the study. The participants stood at a distance of 1.2m in front of the display (hence, the visual angles of the display are 53.5° horizontal and 31.9° vertical). The captured eye image resolution was 80×70 pixels, which varied slightly across the participants.

The participants' task was to look at eleven visual stimuli shown on the display one at a time, in a randomised order. Each stimulus consisted of a red circle with a diameter of 40 pixels (i.e. 1° of visual angle) shown on a light grey background. The center of the red circle was marked with a small black dot with a diameter of 5 pixels. The eleven locations were horizontally distributed across the display (12cm apart), which corresponds to the horizontal visual angles of 26.6°, 21.8°, 16.7°, 11.3°, 5.7°, 0°, -5.7°, -11.3°, -16.7°, -21.8°, and -26.6°. The stimulus location changed every five seconds. To minimise errors, images collected during the first and the last second were discarded for each stimulus location.

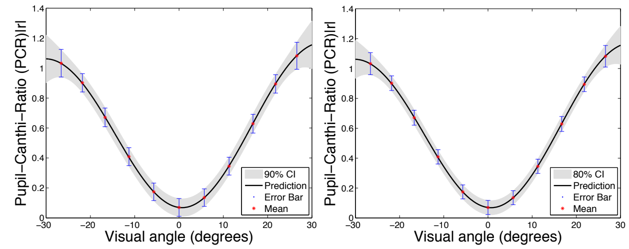


Figure 2: Mapping relationship between horizontal gaze direction and PCR. The solid line indicates an estimation of PCR for 400 values of visual angle θ evenly distributed in the interval $(-30°, 30°)$. The values are represented in the absolute form. Pointwise 90% (left) and 80% (right) confidence intervals are shaded. The error bars indicate the confidence interval around the mean of all observations.

2.2 Data Analysis Using Gaussian Process Regression

We denote θ as the visual angle variable, which corresponds to the horizontal gaze point on the display that a user is looking at, and r as the observation value of PCR. We are interested in the underlying mapping relationship of θ and PCR along the continuous horizontal space. Since we obtained observations of discrete points, we applied an interpolation method to investigate the mapping relationship.

For interpolation, we made an assumption that there is an underlying Gaussian process that represents $r = f(\theta)$. We used the GPML toolbox[1] and computed the predictions using a Gaussian process with 400 test input points evenly distributed on the interval $(-30°,30°)$. We used a covariance function that was a sum of a squared exponential covariance term and independent noise. We initialized the log of the hyper-parameters to be all minus one. All the hyper-parameters were learned by optimizing the marginal likelihood. Thereafter, we made predictions using the learned hyper-parameters. Figure 2 illustrates the results by showing the regression model and the mean and error bars of all observations. The solid lines represent the estimation of PCR $|r|$ (in absolute form) for 400 values of visual angle $\theta \in (-30°, 30°)$. Figure 2 shows two different confidence intervals (CI), 90% and 80% are shaded. The red symbol in Figure 2 indicates the estimated mean across all participants at different stimulus gaze points. The error bars indicate the confidence interval around the mean of all observations.

2.2.1 Interpretation

As illustrated in Figure 2, the gradient level decreases around the centre ($|\theta| < 6°$) and also as the visual angle becomes too large ($|\theta| > 21°$); hence, PCR becomes less discriminative. PCR is most discriminative at region corresponding to $|\theta| \in (6°, 21°)$. We observed that with different confidence interval, it is possible to dis-

[1] **http://www.gaussianprocess.org/gpml/code/**

tinguish the regions where the user is looking at. In the figure that represents 90% CI, when $|\theta| \in (6°, 21°)$, the error bars for PCR estimate do not overlap. When there is no overlap, the screen space can be divided into regions. Hence, we can divide the total screen space into 7 regions (3 regions from the left-half of the screen, another 3 from the right-half, and 1 from the central region). Within 80% CI, when $|\theta| \in (6°, 27°)$, the error bars for PCR estimate do not overlap. Hence, the screen space can be divided into 9 regions.

3. HORIZONTAL GAZE ESTIMATION USING PCR

We assume a noisy observation model $\theta = f(r) + N(0, \sigma_n^2)$ with independent noise function. $f(r)$ is assumed to be a zero-mean Gaussian process with a squared exponential covariance function $k(r, r') = \sigma_f^2 \exp(\frac{-|r-r'|^2}{2l^2})$ [4]. Given a set of labeled training samples $\{(r_i, \theta_i)|i = 1, ..., M\}$, we wish to infer gaze visual angle θ^* for unseen PCR r^* calculated from an input test eye image. The parameters $\{l, \sigma_f, \sigma_n\}$ were learned during the training process by maximising the marginal likelihood. With this assumption, the best estimate for θ^* is:

$$\theta^* = k(r^*)^\intercal (K + \sigma_n^2 I)^{-1} \theta \tag{2}$$

where $K_{ij} = k(r_i, r_j)$, $k_i(r^*) = k(r_i, r^*)$ and θ represents the vector of observations $\{\theta_i | i = 1, M\}$ from the training samples.

As a final step, we convert gaze direction in visual angles to display coordinates. If the user is at a distance d from the display (with a width of W mm and a horizontal resolution of HR pixels), the screen coordinates can be approximated by intersecting the gaze direction characterised by the visual angle θ and the screen plane as

$$p_\theta = \frac{HR \times d \times \tan\theta}{W} \tag{3}$$

For consistency with the visual angle representation, we denote the screen coordinate from left to right as $(-\frac{W}{2}, \frac{W}{2})$ in millimeters and $(-\frac{HR}{2}, \frac{HR}{2})$ in pixels.

3.1 Leave-one-out Cross Validation

To evaluate the accuracy of our proposed method, we adopt the leave-one-out cross validation methods over the data of the 12 subjects. We first use training data of 11 subjects to learn the parameters of the covariance function, and then test all the frames from the remaining subject. We vary different training data sizes (3 iterations of random sampling) and achieve mean accuracy of 3.9°. Table 1 summarises the gaze estimation error. We further illustrate the observations over 11 visual angles fit in the learned model in Figure 3.

Table 1 illustrates that the proposed method converges very fast and requires few training samples for learning the model. The training is only performed once for obtaining a set of parameters for the model. Furthermore, PCR is a single measure scale, which makes our method computational inexpensive and suitable for real-time applications. Figure 3 shows the regression model provides consistent prediction accuracy for the entire screen space. This figure also shows that, in our setup, PCR is a robust person-independent measure for gaze direction in the range of no more than 23° towards the left or the right directions. This is partially due to the fact that we only got sample points in a limited range of visual angles $(-26.6°, 26.6°)$. Collecting training points distributed over a wider range could lead to more accurate estimates, but it becomes unrealistic for fixation of extreme gaze angles.

Table 1: Summary of gaze estimation errors of 12 subjects, standing at a distance of 1.2m away from the display, with different fraction of training data for learning. Gaussian models are trained through batch training over data from 11 subjects, and we used leave-one-subject-out cross validation

	Training data frame size (max. 4000 frames)									
	100 (2.5%)		200 (5%)		300 (7.5%)		1000 (25%)		2000 (50%)	
	(deg.)	(mm)	(deg.)	(mm)	(deg.)	(mm)	(deg.)	(mm)	(deg.)	(mm)
S1	4.73	105.77	4.62	103.60	4.69	104.96	4.63	103.74	4.61	103.33
S2	3.95	89.12	4.17	93.98	4.06	91.73	4.01	90.73	3.96	89.73
S3	4.36	99.77	4.33	99.63	4.28	99.22	4.33	100.08	4.34	100.12
S4	3.86	85.44	3.93	87.05	3.84	85.34	3.93	87.07	3.93	87.14
S5	3.64	83.07	3.51	80.34	3.49	79.90	3.47	79.43	3.45	79.14
S6	3.77	84.00	3.78	84.53	3.64	81.42	3.80	85.01	3.86	86.20
S7	3.75	85.21	3.94	89.99	3.82	87.59	3.88	88.71	3.88	88.60
S8	4.10	92.70	3.80	86.12	3.68	83.52	3.71	84.22	3.76	85.34
S9	3.41	78.00	3.48	79.42	3.65	83.23	3.69	84.17	3.65	83.18
S10	3.97	90.22	3.77	86.37	4.45	102.52	3.77	86.62	3.74	85.85
S11	4.44	101.01	4.24	97.14	3.99	89.65	3.80	86.86	3.76	86.30
S12	3.97	90.59	3.88	89.06	3.65	83.43	3.72	85.50	3.68	84.65
M	**4.00**	**90.41**	**3.96**	**89.77**	**3.94**	**89.38**	**3.9**	**88.51**	**3.89**	**88.29**
SD	0.37	8.22	0.33	7.17	0.37	8.16	0.31	6.58	0.31	6.55

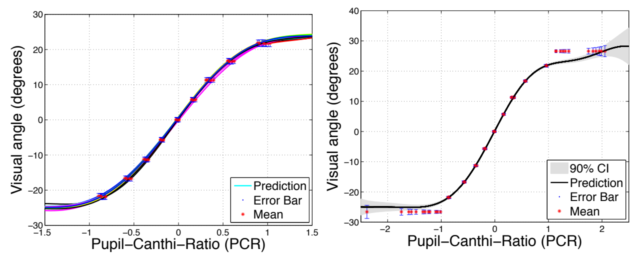


Figure 3: Different colour in the left figure represents the learned regression model per validation over one subject. The overlap of the 12 curves shows consistent model across different subjects, which indicates PCR is a robust person-independent measure. However, the generalised model does not fit when visual angles exceeds around 23° towards left or right side as shown in the right figure.

4. APPLICATIONS

From the study, we learned that PCR can be mapped to a domain of discrete regions or to a domain of continuous space. In this section, we explain two ways of employing PCR for eye-based interaction and illustrate example applications. We implemented the applications using OpenCV and Microsoft Visual Studio and executed the application in real-time.

4.1 Mapping PCR to Screen

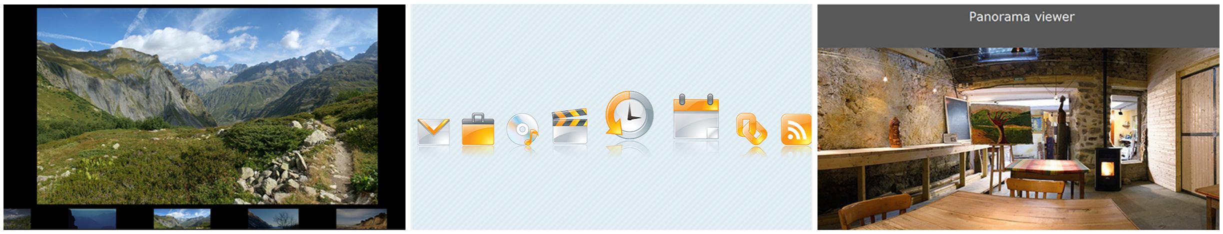
PCR can be mapped directly onto the screen space domain. This is similar to having a cursor on the display, where the cursor follows where a user looks at.

4.1.1 Photo Slideshow Browser

As described in the previous section, the screen space can be divided into a horizontal sequence of distinguishable regions. Content can be illustrated discretely, and users can select a discrete region by looking at it. We implemented a photo slideshow browser. The display shows a row of thumbnail images. A user stares at a thumbnail to make a selection, and an enlarged version of the selected image is displayed above the row of thumbnails (see Fig. 4(a)).

4.1.2 Fisheye View Menu

We implemented an application that accepts continuous gaze input for scrolling through a fisheye view menu (see Fig. 4(b)). Sim-



(a) Slideshow (b) Fisheye Menu (c) Panorama

Figure 4: Screenshots of the example applications.

ilar to the Mac OS X dock, we use PCR to move a cursor along a menu of icons. An icon gets magnified as a user looks at it.

4.2 Mapping PCR to Speed

An alternative way of mapping PCR is to apply it as a relative measure for controlling the speed of moving content. The center region is a reference for zero speed, and as the user looks further away from the center, the speed increases in the same direction of the user's gaze.

4.2.1 Panorama Image Viewer

We implemented a panorama image viewer (see Fig. 4(c)). The application shows part of a circular panorama image. The user looks left or right to pan the image. We implemented that the further the user looks away from the center of the display, the faster the image scrolls, and the user can simply stop the panning by looking at the central region.

5. DISCUSSION AND LIMITATIONS

The PCR method is a lightweight software solution that detects gaze directions from camera images. The method is based on a relative measurement of users' eye feature points, which requires no prior calibration of the users for system training. This makes it particularly suitable for public environments, such as train stations and airports, where we have no control or prior information of the users. Besides public displays, mobile devices with integrated camera, such as tablets, can also support PCR inputs. Nonetheless, the method is less suitable for mapping PCR to displays of small-screen devices (e.g., mobile phones), as the available screen space limits the visual angle of the users. PCR is more suitable for tasks that map PCR to speed. For example, users could look at different directions for controlling audio volume, scrolling music albums or panning pictures. However, it would be interesting to investigate issues that users encounter with PCR applications on mobile devices, such as how far away from the screen that users need to look for triggering inputs; and how movements of mobile devices influence system performance.

An extension of PCR for tracking vertical eye movements would further increase expressiveness of the interface. This would allow scrolling of 2D content, such as maps. However, vertical tracking based on eye image features is significantly more challenging. PCR could be defined by the difference of y coordinates of the eye corners and the pupil centers. However, PCR becomes less distinguishable in vertical direction as the eyes are easily occluded by the eye lids.

As PCR is based on the horizontal symmetry of eye movements, which requires both eyes on the same horizontal level, the performance of PCR can be influenced by head tilting. Similar to existing gaze estimation methods [2], PCR works best when a user is facing forward towards the camera. Future work could focus on combining head pose and gaze direction to extend PCR such that head movement and eye movement are seamlessly accommodated.

6. CONCLUSION

In this paper, we presented a novel measure Pupil-Canthi-Ratio and defined a lightweight and calibration-free method for estimating horizontal gaze direction. PCR is based on the symmetry of our eyes and can be extracted from eye images. It is a relative measure which describes the degree of users' gaze towards left or right from looking ahead. We mapped PCR to gaze direction using Gaussian regression. Our study and evaluation showed that PCR is a robust measure for person-independent horizontal gaze estimation and achieved an average accuracy of 3.9 degrees. We also illustrated examples of interaction applications and discussed limitations and potential extension of PCR. As the proposed method requires only images from a normal camera, it provides a readily deployable solution for gaze interaction for out-of-lab applications.

7. ACKNOWLEDGEMENTS

We thank Ming Ki Chong, Ken Pfeuffer and Eduardo Velloso for helping us to improve this paper. This work was supported by the EU Marie Curie Network iCareNet under grant number 264738.

8. REFERENCES

[1] J. Chen and Q. Ji. Probabilistic gaze estimation without active personal calibration. In *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '11, pages 609–616. IEEE Computer Society, 2011.

[2] D. Hansen and Q. Ji. In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(3):478–500, 2010.

[3] C. H. Morimoto and M. R. M. Mimica. Eye gaze tracking techniques for interactive applications. *Comput. Vis. Image Underst.*, 98(1):4–24, 2005.

[4] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2005.

[5] Y. Sugano, Y. Matsushita, and Y. Sato. Calibration-free gaze sensing using saliency maps. In *Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2667–2674. IEEE Computer Society, 2010.

[6] Y. Zhang, A. Bulling, and H. Gellersen. Towards pervasive eye tracking using low-level image features. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA '12, pages 261–264. ACM, 2012.

[7] Y. Zhang, A. Bulling, and H. Gellersen. Sideways: A gaze interface for spontaneous interaction with situated displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 851–860. ACM, 2013.