Towards Convenient Calibration for Cross-Ratio based Gaze Estimation

Nuri Murat Arar Hua Gao Jean-Philippe Thiran
Signal Processing Laboratory (LTS5),
École Polytechnique Fédérale de Lausanne, Switzerland
murat.arar,hua.gao,jean-philippe.thiran@epfl.ch

Abstract

Eye gaze movements are considered as a salient modality for human computer interaction applications. Recently, cross-ratio (CR) based eye tracking methods have attracted increasing interest because they provide remote gaze estimation using a single uncalibrated camera. However, due to the simplification assumptions in CR-based methods, their performance is lower than the model-based approaches [8]. Several efforts have been made to improve the accuracy by compensating for the assumptions with subject-specific calibration. This paper presents a CR-based automatic gaze estimation system that accurately works under natural head movements. A subject-specific calibration method based on regularized least-squares regression (LSR) is introduced for achieving higher accuracy compared to other state-of-the-art calibration methods. Experimental results also show that the proposed calibration method generalizes better when fewer calibration points are used. This enables user friendly applications with minimum calibration effort without sacrificing too much accuracy. In addition, we adaptively fuse the estimation of the point of regard (PoR) from both eyes based on the visibility of eye features. The adaptive fusion scheme reduces accuracy error by around 20% and also increases the estimation coverage under natural head movements.

1. Introduction

Eye gaze movements have a crucial role in people's visual attention, cognitive processes, emotional states, and interpersonal interactions [15]. They are also suitable to interact with a computer vision system either as a unimodal user interface or as a modality for multi-modal interfaces since they are natural and fast. Therefore, robust estimation and tracking of gaze is of great interest for the development of human-computer interaction (HCI) applications. Recently, gaze tracking systems with a wide variety of applications have attracted much attention, and promising advancements have made the idea of gaze-based computer vision applications more and more realistic.

The main goal of a gaze-based interface is to accurately map the user's gaze to the screen coordinates. For interactive applications, remote video-based gaze trackers are preferred since they are non-intrusive and achieve satisfactory accuracy. Remote video-based gaze tracking methods can be classified mainly into two groups [8]: interpolation-based methods [6] and model-based methods [1, 16, 12]. Interpolation-based methods map image features to gaze points. Model-based methods mostly estimate three-dimensional (3D) gaze direction by modeling the eye in 3D. The intersection between scene geometry and gaze direction is computed as the PoR. System requirements of interpolation-based methods tend to be smaller than model-based methods but they are suited to particular applications due to their limitations regarding precision and head movements. Model-based methods offer greater freedom of movement, however, they require more complex system setups such as camera and geometric calibration. Contrary to these methods, CR-based methods [20, 19, 3, 11, 7, 4, 21, 9] share advantages from both interpolation and model-based methods. For instance, they do not require camera or geometry calibration and they allow free head motion. Unfortunately, the performance of CR-based methods might be limited in accuracy and robustness due to the simplifications assumed. There are two major sources of estimation bias in CR-based methods [10]. First, the model assumes that the pupil center and the corneal reflections (glints) lie on the same plane. They are, in fact, not coplanar because the cornea has a spherical surface. Second, the model computes the PoR by considering eye ball's optical axis rather than the visual axis, the real line of sight.

In the original CR method introduced by Yoo et al. [20], no error offset compensation is performed. They later refined their method by several enhancements in feature detection and non-coplanarity compensation using an additional glint [19]. Their suggested calibration improves the accuracy even though the correction for the axes difference is not considered. In a similar approach, Coutinho and Morimoto [3] proposed a method to compensate for the axes

difference for the first time, and showed superior performance. Kang et al. [11] proposed a homography-based correction. This method simplified the calibration procedure by eliminating the fifth glint. In addition, it modelled the error vectors better to compensate for the axes difference. Similarly, Hansen et al. [7] proposed a normalized homography mapping to further improve the robustness against perspective distortions. When users gaze their monitor under normal conditions, most of the time no abrupt change is observed in head pose or head location. For such HCI scenarios homography-based calibration methods [11, 7] work well when there is sufficient number of calibration points. On the other hand, different approaches [4, 9] have recently been proposed to bring robustness against head movements for non-generic HCI scenarios. Recently, Zhang and Cai [21] introduced binocular fixation constraint to jointly estimate the CR homography matrix. They, for the first time, have utilized information from both eyes to improve the estimation accuracy. However, the drawback of their system is that both eyes need to be present to output a PoR, which significantly constrains the working coverage.

The main contribution of this paper is the introduction of a regularized least-squares regression (LSR) based subject-specific calibration in order to achieve decent accuracy given few number of calibration points. We demonstrate that the proposed method is more generalizable than the state-of-the-art calibration techniques. In addition, our proposed system consists of a rather simple setup while still obtaining high estimation accuracy. Unlike most of the previous efforts in the literature, the system does not require high-resolution eye data, which is captured by directing the camera to the subject's eye, to reach high estimation accuracy and precision. Instead we capture video frames of the whole face with visible but lower resolution of the eye pair. The disadvantages of low-resolution eye data are compensated by a novel adaptive fusion scheme which allows for the calculation of the overall PoR from both eyes, instead of using the dominant eye of the user, as often performed in the previous literature.

In this initial proof-of-concept effort, we have targeted a generic HCI environment where the users were not particularly asked to move or standstill their heads with respect to the monitor. We collected ground truth data separately for subject-specific calibration and testing in a natural manner (no use of chin rest). We have focused on a less tedious subject-specific calibration approach for the users. Also contrary to previous works, we introduce a new evaluation scheme where the test points are not chosen among the calibration points but are generated randomly covering the whole screen. This reduces the possibility of overfitting in addition to creating a more natural and realistic test condition.

The rest of the paper is organized as follows: Section 2 explains a detailed description of the proposed system. Experimental results and discussions are given in Section 3. Finally, Section 4 concludes the paper.

2. Proposed System

The proposed gaze estimation system consists of five main processes, namely, eye detection/tracking, blink and gaze features detection, precise gaze estimation, subject-specific bias correction and adaptive fusion. The details of the system is explained in the following sections.

2.1. Hardware Setup

Our system consists of one PointGrey Flea3 monochrome camera for video capturing, 5 groups of near-infrared (NIR) LEDs for the illumination and a controller unit for the synchronization. The camera has a resolution of 1280×1024, and a 12 mm lens is used. The camera is located below the monitor and slightly closer to the user. In order to create the glints, 4 groups of NIR LEDs with 850 nm wavelength are placed on the corners of the monitor. A band-pass filter around 850 nm is mounted in front of the lens in order to get rid of the ambient light in other wavelengths. The fifth group of LEDs is placed as ring around the lens of the camera to create the bright pupil effect. A micro-controller is programmed to obtain interlaced dark and bright pupil images at 30 frames per second. Besides, we synchronize the LEDs with the camera's shutter to turn on lighting as short as possible for eye safety purpose. In the current setup, the user sits approximately 70 cm away from a 24-inch monitor with a resolution of 1920×1200. The head is not fixed, therefore users are allowed to perform natural head movements.

2.2. Eye Detection/Tracking

Our system starts with eye localization where existence of eyes is determined. In order to localize and track the eyes we utilize a robust non-rigid face tracker based on supervised decent method (SDM) [18] in order to detect, localize and track the eyes. SDM method assumes that an accurate final shape can be estimated with a cascade of regression models given an initial shape. Viola & Jones face detector [17] is used to initialize the shape. The tracker first fits the mean shape in the initial frame and continues the fitting in the succeeding frames. Once the shape is fitted accurately, we extract eye regions by considering the landmarks representing eyes. We do not perform any registration or scaling on the extracted eye regions to ensure any particular eye region resolution. On the extracted eye regions, first we detect whether there is any eye blink or not. If there is no eye blink, we then detect image features for the gaze estimation. The image features include the pupil center and corneal reflections of NIR LEDs, i.e. glints.

2.2.1 Eye Blink Detection

For the eye blink detection, we check the positioning of the landmarks around the eyes. We measure vertical opening (height) of both eyes relative to the eye width. As illustrated in Figure 1, if the average of the ratio of eye height over eye width for both eyes is significantly lower (<0.15) than the open eye form (~ 0.5), we determine that a natural eye blink occurs. Since the average eye blink is completed 100 to 200 milliseconds after the peak closure of eyelids, we do not output any PoR for the corresponding number of frames once an eye blink is detected.

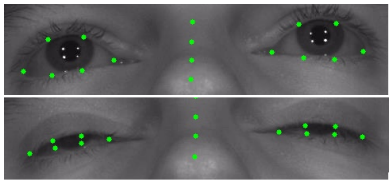

Figure 1. The positioning of facial landmarks when there is no eye blink (top) and an eye blink (bottom).

2.2.2 Glints and Pupil Center Detection

We employ simple image processing algorithms to precisely localize the glints. First of all, histogram equalization is performed followed by thresholding on the input image which results in a binary image. We use adaptive thresholding to avoid tuning the threshold parameter. Then the binary image is processed by morphology operations to get rid of the small blobs caused by noise. In the resulting binary image, we expect to find four blobs which should form a trapezium since they emerge by the reflections of four NIR LEDs located on the corners of the computer monitor (Figure 2.d). Hence, we get the candidate glints by performing connected component analysis. If there are four or more candidate glints remaining, we consider the shapes formed by any four-glints combination. The set of candidates whose convex hull has the highest match with a template shape representing the screen are considered as the final glint features.

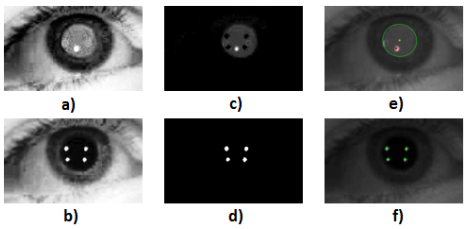

Figure 2. Input and preprocessed images for feature detection: (a) pupil reflection and bright-eye effect, (b) corneal reflection and dark-eye effect, (c) difference image, (d) thresholded dark pupil image, (e,f) output images, detected pupil and glints.

For the pupil center detection, a more sophisticated technique is required since the intensity of the pupil is more similar to its surrounding pixels. For this purpose, we use the robust pupil detection method suggested by Ebisawa [5]. The method is based on bright pupil effect which is obtained when an NIR LED is located in the optical axis of a camera as shown in Figure 2.a. In a similar approach, to robustly detect the pupil, we use two images: one is taken when the corner LEDs on the monitor are turned on and the LEDs on the camera axis are turned off, the other is taken when the monitor LEDs are off and the camera LEDs are on. If these images are obtained from the camera in a very short interval, then the intensity difference of the pupil region in two images is large and that of the region outside of the pupil is very small. Therefore, the difference image has high intensities in the pupil region. The pupil region can be extracted by a segmentation method that is very similar to glint detection, and the center of gravity is considered as the pupil center. Figure 2 illustrates the feature detection processes and outputs of the system.

2.3. Cross-Ratio Gaze Estimation

We employ the original CR method [20] for the estimation of the PoR. It is based on the cross-ratio, the only invariant of projective space. Figure 3 shows a schematic diagram of the CR method. The four LEDs placed on the corners of the screen are projected onto the cornea, producing four corneal reflections. A virtual plane tangent to the cornea surface is assumed to exist, and the four corneal reflections (v_1, v_2, v_3, v_4) lie on this reflection plane. The polygon formed by the four corneal reflections is therefore the projection of the screen. A second projection takes place from the corneal plane to image plane, obtaining the points (g_1, g_2, g_3, g_4) and the projection of the pupil center, p.

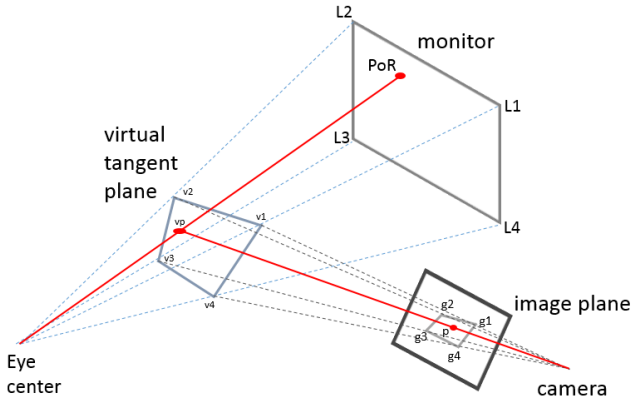

Figure 3. The four light sources are projected onto a reflection plane. The corneal reflections are then projected onto the image plane.

As the virtual tangent plane on the cornea has the same

planar projective transformation of the screen and image planes, the pupil center on image plane corresponds to the PoR on screen, that can be computed by equality of the cross-ratios.

2.4. Subject-Specific Calibration

Cross-ratio based gaze estimation algorithms have some assumptions that limit the performance. There are two major sources of error: i) non-coplanarity of the pupil and glints planes, and ii) the angular offset between visual and optical axes of the eye. Since the cornea curvature and the angular offset are subject-specific, a calibration needs to be performed to compensate for the estimation bias. This procedure is performed once, prior to the use of the system. The users are asked to look at N calibration points on the monitor for K frames long. Subject-specific bias correction can be learnt by minimizing the distances between the estimated gaze positions and the corresponding calibration points on the monitor. As described in Section 1, many techniques have been proposed. The performance is improved given enough training data for calibration. However, augmenting the amount of data by increasing the number of calibration points could be tedious and thus harms the user experience. Moreover, the performance of CR methods is highly sensitive to feature detection errors by their nature. As opposed to many of the previous work, our system deals with low-resolution eye data and hence the robustness of feature detection is reduced. These motivated us to further investigate different methods to model the error vectors more robustly against outliers and noise using few number of calibration points.

We first consider a linear transform for the subject dependent error compensation. In order to estimate the linear transform, we employ a regularized least-squares regression (also known as ridge regression) because they have better generalization capabilities than homography methods due to reduced model parameters and relaxed constraints.

The transform $\boldsymbol{\beta}$ is defined with a 3×2 matrix, where the first row corresponds to the offset parameters. Assuming $\mathbf{X}$ as the input data, which stacks the PoR coordinates:

$$\mathbf{X} = \begin{bmatrix} 1 & \cdots & 1 \\ \mathbf{x}_1 & \cdots & \mathbf{x}_n \end{bmatrix}.$$

The corresponding output data $\mathbf{Y}$ stores the target coordinates for calibration. The cost function $E(\boldsymbol{\beta})$ for the regularized least square problem is defined as:

$$E(\boldsymbol{\beta}) = \|\boldsymbol{\beta}^T \mathbf{X} - \mathbf{Y}\|^2 + \lambda \|\boldsymbol{\beta}\|_F^2. \quad (1)$$

λ is the regularization shrinkage and $\|\|_F$ stands for the Frobenius norm. A closed form solution can be found by setting the first order derivative of the cost function $E(\boldsymbol{\beta})$ to zero, and we obtain:

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}\mathbf{X}^T + \lambda \mathbf{I})^{-1} \mathbf{X}\mathbf{Y}^T. \quad (2)$$

Using the learned model $\widehat{\boldsymbol{\beta}}$, we can predict a calibrated coordinate giving an input PoR $\mathbf{x}$:

$$\widehat{\mathbf{y}} = \widehat{\boldsymbol{\beta}}^T \mathbf{x}. \quad (3)$$

Note that in the LSR method, the number of the model parameters is less than in the homography estimation. In homography estimation, in total 9 parameters need to be recovered. The problem might be under determined when less points are used for calibration. In the experiments, we demonstrate that a simpler LSR model does not suffer this problem and hence generalizes better on unseen test points.

2.5. Adaptive Fusion Scheme

Although our hardware setup causes a big disadvantage in terms of resolution, it provides a more realistic experience to users allowing free head movement. Besides, the availability of the data for both eyes enables us to get two PoRs for the same frame. In order to output an overall PoR per frame, we propose an adaptive fusion scheme which improves the overall estimation accuracy and precision compared to the performance achieved using single eye. Adaptive fusion scheme ideally performs a weighted averaging of individual PoRs obtained from both eyes as follows:

$$PoR_{overall} = \sum_i PoR_i * W_i, \;\; \sum_i W_i = 1, \;\; i \in \{L, R\},$$

where W_R and W_L are the weights for the right and left eye's PoRs respectively. In case one of the PoRs could not be calculated for a given frame, then the weight of the missing PoR is set to zero. We don't report an overall PoR in case both PoRs are unavailable for a given frame. In this initial work, we assign equal weights to both eyes, and achieved improved overall estimation accuracy and precision. However, the scheme also allows for different weightings of the PoRs. For instance, the weights can be assigned by the feature detection module considering the reliability of the detected features and the eye dominance of the user's. We leave the feature detection reliability and eye dominance based weighting as our future work.

3. Experiments and Results

3.1. Evaluation Data and Protocol

To evaluate the performance of the proposed system, we conducted user experiments. Ten users, nine of whom had no previous experience with any gaze tracking system, participated in our experiments. Since we targeted a generic and natural HCI environment, the ground truth data is collected in a natural manner where the users were asked to look at the target stimulus points naturally the way they feel comfortable. Therefore, we did not require the use of chin rest to keep the user's head still and to keep user's one of

the eyes within the field of view of the camera in order to capture high resolution eye data. The statistics of the user data regarding the head pose variation during the experiments are illustrated in Table 3.1. Note that these statistics are obtained by the head pose estimation provided by the SDM face tracker [19].

	Yaw Angle		Pitch Angle	
	Cal	**Test**	**Cal**	**Test**
Min	-19.11	-11.18	-18.51	-19.5
Max	23.06	16.52	7.95	3.88
Mean	2.37	2.09	-6.92	-7.23
Stdv	4.28	3.22	2.78	1.79

Table 1. Head pose variation statistics (in degree) obtained by the face tracker on the collected experimental data.

The calibration data and test data are acquired in two separate sessions. In calibration data acquisition, users were asked to look at 25 uniformly distributed target points on the screen. The target stimulus points were displayed in a left to right and top to bottom sequence in a 5×5 grid. In test data acquisition, as opposed to the previous studies, we introduce a new evaluation scheme where the test points are independent from calibration points. To achieve this, users were asked to look at 18 target stimulus points in a 3×3 grid covering the whole screen. The positions of the target stimulus points in a region were randomly determined. We ensure 2 stimulus points have to be shown in each region in order to cover the whole screen. The display order of the regions and the points is also randomly determined. This way, we reduce the possibility of overfitting as well as creating a more natural and realistic test condition.

Each target point is displayed for 100 frames (3.33 seconds), and the data of both eyes during this period is captured. To keep the attention of the user on the target stimulus points, the size of the circular target varies continuously from an initial radius of 30 pixels to a final radius of 20 pixels. For testing, we discard the first 20 frames of each target point and keep the latter 80 frames for the evaluation in order to avoid saccadic gaze movement at the beginning of each point display. We report our eye tracker's performance as gaze estimation accuracy, which is defined as the average displacement in degrees between the real stimuli point and the estimated PoR.

3.2. Results

For our evaluation, we first run our face tracker on the captured data to extract eye regions. Due to the limited resolution of the eye region, the size of the extracted eye region is around 90×60 pixels and size of the polygon formed by the glints is around 12×7 pixels. On the detected features we apply CR-based gaze estimation to calculate the initial PoR. In the calibration process, we model the subject-specific error vectors by minimizing the distances between the initial PoRs and the real target points using the calibration data. In the test process, we apply the learnt model to correct the initial PoRs estimated on the test data.

The results achieved by the proposed system on the test data are shown in Table 2. We report the mean of the average estimation accuracy error in degree over all subjects by altering the number of calibration points. We list the results obtained using individual eyes as well as both eyes combined with adaptive fusion. The rightmost column, **Coverage**, shows the percentage of frames in which we are able to output a PoR for the given eye data.

Eye Data	**Calibration**		**Coverage (%)**
	5 Points	**25 Points**	
Left Eye	1.44	1.29	94.47
Right Eye	1.38	1.29	91.55
Overall	**1.15**	**1.03**	**96.83**

Table 2. Average gaze estimation accuracy errors (in degree).

The results demonstrate that the estimation error reduces with increasing number of calibration points used. In addition, it validates the effectiveness of the proposed adaptive fusion scheme. Firstly, it improves the estimation accuracy by about 20% over using only left eye or using only right eye. Secondly, it increases the coverage, the working volume, of the system compared to using single eye data. As shown in Table 2, the system outputs a PoR for 96.83% of all frames while eye blink is detected for 2.41% of all the frames. Therefore, the system could not output a PoR only 0.76% of all the frames due to missing features. Obviously the reason is that the data obtained from a single eye may not be sufficient to calculate a PoR for some of the test points, especially those positioned close to the right or left border of the monitor.

3.2.1 Comparison with Previous Work

We compare our results (LSR) with the state-of-the-art methods such as normalized homography (N-HOM) [7], Gaussian process regression (GPR) [7], and binocular homography fusion (BHF) [21]. Figure 4 shows the mean and standard deviation of the estimation accuracy error with respect to different number of calibration points. The detailed comparison results are listed in Table 3.2.1. The results demonstrate that the proposed method is superior to the other methods in any configuration.

The estimation error reduces with increasing number of points used for calibration. However, a less tedious and user-friendly system should involve as little effort as possible for the subject-specific calibration. The simplest way to achieve this is to minimize the number of calibration points, without sacrificing too much the estimation accuracy. As il-

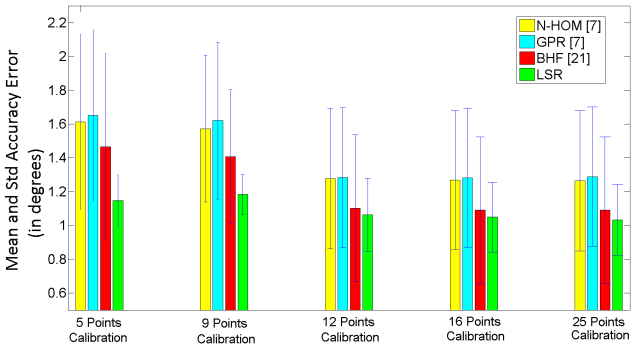

Figure 4. Comparison with the state-of-the-art methods.

lustrated in Figure 4 and Figure 5, the proposed method is less sensitive to the number of calibration points than any of the other methods. The system can still reach reasonable estimation accuracy of 1.15 ± 0.2 with only 5 calibration points.

Moreover, the proposed methodology brings two other advantages. Firstly, a higher coverage is reached through adaptive fusion scheme. As shown in Table 3.2.1, the system outputs a PoR for 97% of all frames while eye blink occurs for 2.41%. The coverage drops when single eye data is used because the features can not be detected for some test points where the viewing angle of the captured eye is extreme. For such cases, the features may still be detected from the other eye and therefore, a PoR may be calculated. On the other hand, [21] proposes to use the data from both eyes simultaneously for improving the accuracy. However, the coverage gets even lower compared to using single eye data since their system restricts the availability of both eyes to output a PoR. The results validate the effectiveness of the proposed methodology in terms of both accuracy and coverage. Secondly, the computational complexity of the proposed calibration method is suitable for real-time gaze tracking while non-linear regression methods such as GPR [7] and kernel methods require much higher computational effort. For instance, it takes on average $\sim$1.5 seconds to apply GPR on a test sample if it is trained with 25 calibration points ($\sim$1250 samples) on a PC with Intel i7 3.2GHz processor, while the training of LSR under the same conditions takes $\sim$0.6 seconds, and once the model parameters are learnt, it only requires a matrix multiplication (3×2) to apply estimation error correction on a test sample.

3.2.2 Comparison with Different Regression Methods

For the calibration, in addition to LSR and homography based methods, we investigate other widely used regression techniques such as partial least-squares regression (PLSR) [14], support vector regression (SVR) [2] and Gaussian process regression (GPR) [13]. The results of investigated methods are shown in Figure 5.

We apply PLSR with linear, polynomial and Gaussian kernels, but we only plot the polynomial kernel as it achieves better performance. Figure 5 indicates that LSR achieves the lowest estimation accuracy error, especially when fewer number of points are used for the calibration. The results confirm that LSR, as a simpler model, generalizes better over the whole screen than the other models. In addition, the slight error increase for 9 points calibration is due to a specific subjects whose calibration data for some points is erroneous. This leads inaccurate calibration for this subject, and eventually brings more negative impact on the overall mean accuracy.

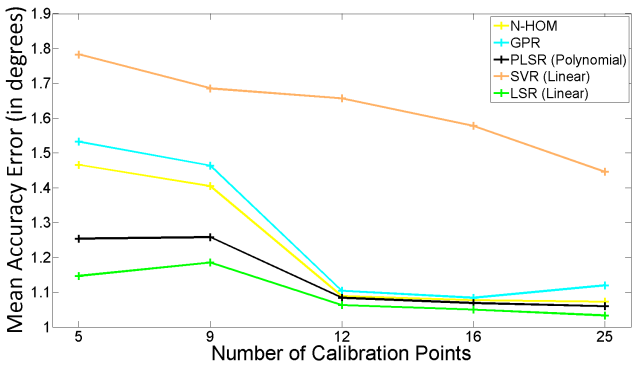

Figure 5. Comparison with the different regression techniques for learning calibration models.

4. Conclusions

In this paper, we present an automatic gaze estimation system which accurately works under natural head movements. The system does not require high-resolution eye data as opposed to most of previous work. Operating with low-resolution data enables the system to output PoRs from each eye simultaneously. A novel adaptive fusion scheme is introduced for achieving improved overall estimation accuracy. In addition, an extensive investigation of different machine learning techniques for subject-specific calibration is carried out to compensate for the major source of errors in CR-based gaze estimation. A novel regularized least-squares regression based calibration method is proposed for the purpose of a more user-friendly calibration process. Besides a new evaluation scheme, where the test points are not chosen among the calibration points, is suggested. The results validate that the proposed method outperforms previous approaches especially when few points are used for calibration. As the future work, a more sophisticated weighting is planned for the adaptive fusion scheme in order to reach higher estimation accuracy.

Method	Eye Data	# of Calibration Points					Coverage
		5	9	12	16	25	(%)
No Calibration	Right Eye	9.03 ± 0.54	9.03 ± 0.54	9.03 ± 0.54	9.03 ± 0.54	9.03 ± 0.54	91
	Left Eye	7.09 ± 0.43	7.09 ± 0.43	7.09 ± 0.43	7.09 ± 0.43	7.09 ± 0.43	94
	Both Eyes	5.69 ± 0.32	5.69 ± 0.32	5.69 ± 0.32	5.69 ± 0.32	5.69 ± 0.32	89
	Both (Adaptive Fusion)	5.88 ± 0.3	5.88 ± 0.3	5.88 ± 0.3	5.88 ± 0.3	5.88 ± 0.3	**97**
N-HOM	Right Eye	1.74 ± 0.57	1.68 ± 0.3	1.29 ± 0.21	1.28 ± 0.21	1.28 ± 0.2	94
	Left Eye (N-HOM [7])	1.61 ± 0.52	1.57 ± 0.43	1.28 ± 0.41	1.27 ± 0.41	1.26 ± 0.42	91
	Both Eyes (BHF [21])	1.46 ± 0.55	1.41 ± 0.4	1.1 ± 0.43	1.09 ± 0.43	1.09 ± 0.43	89
	Both (Adaptive Fusion)	1.47 ± 0.52	1.41 ± 0.28	1.09 ± 0.19	1.08 ± 0.19	1.07 ± 0.18	**97**
GPR	Right Eye	1.85 ± 0.55	1.76 ± 0.3	1.32 ± 0.21	1.29 ± 0.23	1.38 ± 0.21	94
	Left Eye (GPR [7])	1.65 ± 0.51	1.62 ± 0.46	1.28 ± 0.41	1.28 ± 0.41	1.29 ± 0.41	91
	Both Eyes	1.53 ± 0.54	1.47 ± 0.41	1.12 ± 0.43	1.1 ± 0.44	1.14 ± 0.42	89
	Both (Adaptive Fusion)	1.53 ± 0.52	1.46 ± 0.28	1.1 ± 0.19	1.08 ± 0.2	1.12 ± 0.17	**97**
LSR	Right Eye	1.44 ± 0.18	1.48 ± 0.19	1.31 ± 0.22	1.31 ± 0.22	1.29 ± 0.21	94
	Left Eye	1.38 ± 0.43	1.41 ± 0.43	1.31 ± 0.46	1.29 ± 0.47	1.29 ± 0.47	91
	Both Eyes	1.16 ± 0.4	1.2 ± 0.4	1.08 ± 0.46	1.06 ± 0.45	1.05 ± 0.46	89
	Both (Adaptive Fusion)	$\mathbf{1.15 \pm 0.15}$	$\mathbf{1.19 \pm 0.12}$	$\mathbf{1.06 \pm 0.22}$	$\mathbf{1.05 \pm 0.21}$	$\mathbf{1.03 \pm 0.21}$	**97**

Table 3. Comparison with the state-of-the-art calibration techniques with changing number of calibration points and the eye data used for the evaluation. Average gaze estimation accuracy errors (in degree) are reported.

5. Acknowledgments

This project is supported by the Swiss Commission for Technology and Innovation (CTI) under grant number 13594.1 PFFLR-ES. The authors would like to thank Yves Moser from Logitech for his valuable contributions in the user experiments.

References

[1] D. Beymer and M. Flickner. Eye Gaze Tracking Using an Active Stereo Head. In *CVPR*, pages 451–458, 2003.

[2] C. Burges. A tutorial on support vector machines for pattern recognition. *Knowledge Dis. and Data Mining*, 2(2), 1998.

[3] F. Coutinho and C. Morimoto. Free head motion eye gaze tracking using a single camera and multiple light sources. *In Brazilian Symp. Computer Graphics and Image Processing*, pages 171–178, 2006.

[4] F. L. Coutinho and C. H. Morimoto. Improving Head Movement Tolerance of Cross-Ratio Based Eye Trackers. *In IJCV*, 101(3):459–481, 2013.

[5] Y. Ebisawa. Improved video-based eye-gaze detection method. *Trans. on Instrum. Meas.*, 47(4):948–955, 1998.

[6] D. Hansen and A. Pece. Eye Tracking in the Wild. *In CVIU*, 98(1):182–210, 2005.

[7] D. W. Hansen, J. S. Agustin, and A. Villanueva. Homography normalization for robust gaze estimation in uncalibrated setups. In *ETRA*, 2010.

[8] D. W. Hansen and Q. Ji. In the eye of the beholder: a survey of models for eyes and gaze. *In Trans. on PAMI*, 32(3):478–500, 2010.

[9] J.-B. Huang, Q. Cai, Z. Liu, N. Ahuja, and Z. Zhang. Towards accurate and robust cross-ratio based gaze trackers through learning from simulation. In *ETRA*, 2014.

[10] J. J. Kang, M. Eizenman, E. D. Guestrin, and E. Eizenman. Investigation of the cross-ratios method for point-of-gaze estimation. *In Trans. on Biomedical Engineering*, 55(9):2293–302, 2008.

[11] J. J. Kang, E. D. Guestrin, W. J. Maclean, and M. Eizenman. Simplifying the cross-ratios method of point-of-gaze estimation. In *Canadian Medical and Biological Engineering Conference*, 2007.

[12] B. Noureddin, P. Lawrence, and C. Man. A Non-Contact Device for Tracking Gaze in a Human Computer Interface. *In CVIU*, 98(1):52–82, 2005.

[13] C. E. Rasmussen and C. K. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[14] R. Rosipal and N. Krämer. Overview and recent advances in partial least squares. In *Subspace, Latent Structure and Feature Selection: Statistical and Optimization Perspectives Workshop*, 2005.

[15] G. Underwood. *Cognitive Processes in Eye Guidance*. Oxford University Press, 2005.

[16] A. Villanueva and R. Cabeza. Models for Gaze Tracking Systems. *In J. of Image and Video Processing*, (3), 2007.

[17] P. Viola and M. Jones. Robust real-time face detection. *In IJCV*, 57:137–154, 2004.

[18] X. Xiong and F. De la Torre. Supervised Descent Method and Its Applications to Face Alignment. *CVPR*, 2013.

[19] D. H. Yoo and M. J. Chung. A novel non-intrusive eye gaze estimation using cross-ratio under large head motion. *In CVIU*, 98(1):25–51, 2005.

[20] D. H. Yoo, J. H. Kim, B. R. Lee, and M. J. Chung. Non-contact eye gaze tracking system by mapping of corneal reflections. In *AFGR*, 2002.

[21] Z. Zhang and Q. Cai. Improving cross-ratio based eye tracking techniques by leveraging the binocular fixation constraint. In *ETRA*, 2014.