Adaptive Eye-Camera Calibration for Head-Worn Devices

David Perra[1], Rohit Kumar Gupta[2], Jan-Micheal Frahm[2]
[1]Google Inc. [2]The University of North Carolina at Chapel Hill.

Most of the current calibration schemes solve for a globally optimal model of the eye-device transformation by performing calibration on a per-user or once-per-use basis. These schemes are impractical for real-world applications because they do not account for changes in calibration during the time of use. In this paper, we present a novel, continuous, locally optimal calibration method for use with head-word devices. Our calibration scheme allows a head-worn device to calculate a locally optimal eye-device transformation by computing an optimal model from a local temporal window of previous frames. In addition, we also proposed an automatic calibration technique based on naturally occurring interest regions within user's environment which avoids user's active participation.

Shape based eye tracking methods track portions of the eye anatomy such as corneal reflection, pupil contour, and iris contour. Corneal reflection and pupil contour methods require infrared ray (IR) active illumination. Our system contains a user-facing camera that is used to capture user's eye movement, 2 IR LEDs that are positioned beside the eye camera to produce glints on the eye, and a scene facing camera that captures the user's environment. These cameras are calibrated and have non-overlapping views. Our method (shown in Figure 1) estimates the user's gaze from each pair of frames received. With two glints location and the projected pupil center, along with a coarse approximation of the physical parameters we estimate the user's gaze in coordinate system of user-facing camera.

User's gaze is projected upon scene-facing camera's image plane and might contain some errors due to two reasons: 1) the user's eye parameters are initially unknown, and 2) the transformation between the eye and the device and between the two cameras are only initial approximations. To remedy this we performed an automatic calibration using naturally occurring regions of interest found within the scene image. In recent years, lot of work has been done to obtain better-quality saliency maps. Among these GBVS [4], AWS [6], and ImgSig [5] usually exhibit the best performance. Chen *et al.*[2] estimated the probability distribution of the eye parameters and eye gaze by combining saliency map with the 3D eye model. Our method also uses a similar approach with differs from their incremental learning framework. Our method is also motivated by Alnajar *et al.*[1] who proposed auto-calibration of gaze estimators in an uncalibrated setup.

The Gullstrand model [3] is used in the geometric model-based gaze estimation in our method. In this model, exterior corneal surface is approximated by a spherical convex mirror and optical axis is defined by 2 predefined rotations around the visual axis. Other details of the model is explained in the paper. It provides six main parameters to optimize over: the corneal radius, R_c, the eye's index of refraction, η_1, the distance from the pupil center to the corneal center of the cornea, $d_{c,p}$, the distance between the user-facing camera and the eye, $d_{o,r}$, and the two angular offsets used to find the visual axis from the optic axis: α and β. In addition, given the internal camera calibration the head-worn device has six degrees of freedom itself: three dimensions of translation (x, y, and z) and three dimensions of rotation (ϕ, θ, and ψ). Hence, the parameters of an eye-camera model at a time t can be described as

$$P_t = < x, y, z, \phi, \theta, \psi, d_{o,r}, R_c, d_{c,p}, \eta_1, \alpha, \beta > \qquad (1)$$

In order to optimize the eye-camera transformation, and to ensure that our calibration remains causal, our system minimizes the sum of absolute differences in the x and y directions between the projected visual axis, v_j, and the nearby salient interest point, s_j, for all frames within a short window of frames preceding time t. For a window of size k, the cost function, $C(p_t)$ can be written as:

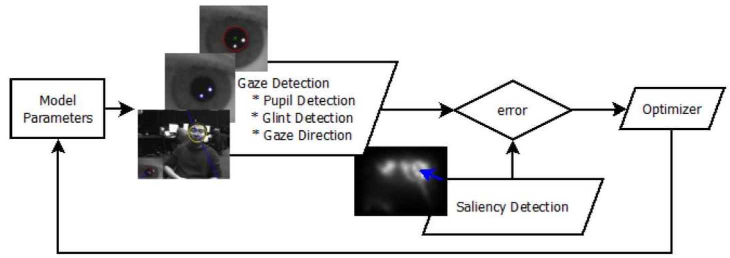

Figure 1: The proposed framework for eye-camera parameters estimation pipeline

$$C(p_t) = \begin{bmatrix} d_x(v_{t-1}, s_{t-1}) \\ d_y(v_{t-1}, s_{t-1}) \\ d_x(v_{t-2}, s_{t-2}) \\ d_y(v_{t-2}, s_{t-2}) \\ \cdots \\ d_x(v_{t-k}, s_{t-k}) \\ d_y(v_{t-k}, s_{t-k}) \end{bmatrix} \qquad (2)$$

where $d_x(v_j, s_j)$ is the absolute difference in the x-direction and $d_y(v_j, s_j)$ is the absolute difference in the y-direction for the projection of point s_j on the gaze direction v_j.

We obtain the locally optimal solution, P_t, by solving the following least-squares optimization problem:

$$P_t = \underset{p_t}{\text{argmin}} \sum_{j=1}^{k} C_j(p_t)^2 \qquad (3)$$

Optimizing over the eye-camera model parameters in the window yields an estimate of the locally optimal parametric model, P_t, at time t. The resulting parametric model effectively maximizes calibration accuracy while minimizing gaze estimation error. P_t is further used as an initial estimate of the model parameters for the next window. As k increases, the optimization approaches a single globally optimal eye-camera model. Gaze estimation error for our system was as low as 0.15-0.20 degrees running at 4-5 fps. We concluded that our continuous, automatic calibration provides a more realistic and better gaze estimation technique.

[1] Fares Alnajar, Theo Gevers, Roberto Valenti, and Sennay Ghebreab. Calibration-free gaze estimation using human gaze patterns. In *15th IEEE International Conference on Computer Vision*, 2013.

[2] Jixu Chen and Qiang Ji. Probabilistic gaze estimation without active personal calibration. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 609–616, June 2011. doi: 10.1109/CVPR.2011.5995675.

[3] E.D. Guestrin and E. Eizenman. General theory of remote gaze estimation using the pupil center and corneal reflections. *Biomedical Engineering, IEEE Transactions on*, 53(6):1124–1133, June 2006. ISSN 0018-9294. doi: 10.1109/TBME.2005.863952.

[4] Jonathan Harel, Christof Koch, and Pietro Perona. Graph-based visual saliency. In *ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 19*, pages 545–552. MIT Press, 2007.

[5] Xiaodi Hou, Jonathan Harel, and Christof Koch. Image signature: Highlighting sparse salient regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(1):194–201, 2012. ISSN 0162-8828. doi: http://doi.ieeecomputersociety.org/10.1109/TPAMI.2011.146.

[6] V Leborn Alvarez, A Garca-Daz, X. Fdez-Vidal, and X. Pardo. Dynamic saliency from adaptative whitening. In *Natural and Artificial Computation in Engineering and Medical Applications*, volume 7931 of *Lecture Notes in Computer Science*, pages 345–354. Springer Berlin Heidelberg, 2013. ISBN 978-3-642-38621-3.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.