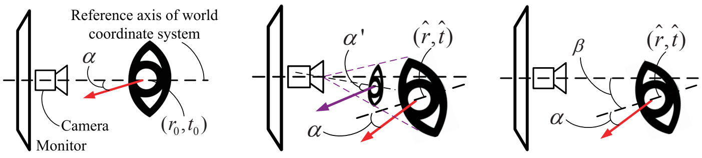
A Head Pose-free Approach for Appearance-based Gaze Estimation

Feng Lu
Takahiro Okabe
Yusuke Sugano
Yoichi Sato
{lufeng,takahiro,sugano,ysato}@iis.u-tokyo.ac.jp

Institute of Industrial Science
the University of Tokyo
Tokyo, Japan



(a) (b) (c)

Figure 1: 2-D illustration of relationship between gaze direction and head pose. (a) Under a fixed head pose $(\boldsymbol{r}_0,\boldsymbol{t}_0)$, gaze direction α can be estimated from appearance. (b) To obtain α under another head pose $(\hat{\boldsymbol{r}},\hat{\boldsymbol{t}})$, the estimated α' should be corrected because of captured eye appearance distortion. (c) Under head pose $(\hat{\boldsymbol{r}},\hat{\boldsymbol{t}})$, gaze direction under WCS should be further compensated for head rotation β.

We aim at solving the appearance-based gaze estimation problem under free head motion without significantly increasing the cost of training. According to recent surveys [1, 2], the limitation of the appearance-based methods lies in that a fixed head pose should be assumed to avoid large number of eye appearance training samples under head motion. We propose to solve this problem with a novel approach:

1. A decomposition scheme is introduced to decouple the original problem into subproblems, namely initial estimation and subsequent compensations.
2. Geometric priors are introduced in appearance-based estimation. Specifically, the combination of 3-D geometric-based and learning-based methods reduces the number of required training samples.
3. The gaze estimation bias caused by eye appearance distortion is learnt effectively using training samples obtained from a 5-seconds video clip.

The generalized appearance-based gaze estimation problem can be formulated as using training data $\mathcal{T}$ to map the eye appearance feature $\hat{\boldsymbol{e}}$ to the gaze direction unit vector $\hat{\boldsymbol{g}}$ under head pose $(\hat{\boldsymbol{r}},\hat{\boldsymbol{t}})$:

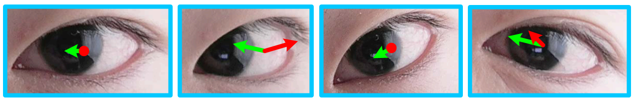
$$\hat{\boldsymbol{g}} = \mathcal{M}(\hat{\boldsymbol{e}},\hat{\boldsymbol{r}},\hat{\boldsymbol{t}}|\mathcal{T}) \qquad (1)$$

We propose to first solve the problem by assuming a fixed head pose $(\boldsymbol{r}_0,\boldsymbol{t}_0)$ as shown in Fig. 1(a) and then compensating for the estimation bias by taking into account the true head pose $(\hat{\boldsymbol{r}},\hat{\boldsymbol{t}})$. The bias under WCS mainly depends on two factors: 1) the estimation error caused by eye appearance distortion (see α' and α in Fig. 1(b)) in accordance with specific capture direction; and 2) the eye orientation variation in accordance with head rotation (see β in Fig. 1(c)). In fact, the problem in Eq. (1) is decomposed into:

$$\hat{\boldsymbol{g}} \simeq \mathcal{M}_{\boldsymbol{r}_0,\boldsymbol{t}_0}(\hat{\boldsymbol{e}}|\mathcal{T}_0^e,\mathcal{T}_0^g)\otimes\mathcal{C}^D_{\boldsymbol{r}_0,\boldsymbol{t}_0}(\hat{\boldsymbol{r}},\hat{\boldsymbol{t}}|\mathcal{T})\otimes\mathcal{C}^R_{\boldsymbol{r}_0}(\hat{\boldsymbol{r}}) \qquad (2)$$

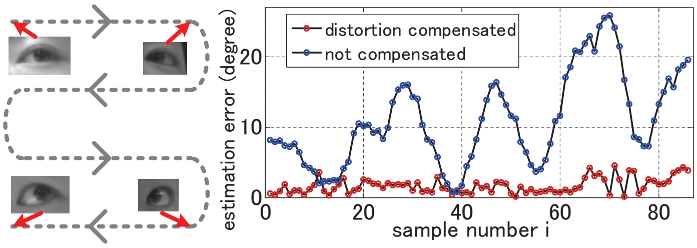
The gaze estimation under fixed head pose $(\boldsymbol{r}_0,\boldsymbol{t}_0)$ by $\mathcal{M}_{\boldsymbol{r}_0,\boldsymbol{t}_0}(\hat{\boldsymbol{e}}|\mathcal{T}_0^e,\mathcal{T}_0^g)$ can be done using conventional fixed head pose methods. The gaze direction compensation for head rotation by $\mathcal{C}^R_{\boldsymbol{r}_0}(\hat{\boldsymbol{r}})$ can be achieved via geometric manipulations such as translations and rotations. Therefore, the key point of this method is to compensate the gaze estimation bias caused by eye appearance distortion under different head poses by $\mathcal{C}^D_{\boldsymbol{r}_0,\boldsymbol{t}_0}(\hat{\boldsymbol{r}},\hat{\boldsymbol{t}}|\mathcal{T})$.

While the eye orientation varies relatively to the camera, distortion exists in the captured eye image. In the eye coordinate system (ECS), this orientation is depicted by the capture direction that is calculated by a vector pointing to the camera centre. Thus we propose to learn a regression $\mathcal{C}^D_{\boldsymbol{r}_0,\boldsymbol{t}_0}(\hat{\boldsymbol{r}},\hat{\boldsymbol{t}}|\mathcal{T})$ by investigating the relationship between the capture direction variation $\Delta\boldsymbol{v}^c$ and the gaze estimation biases $\Delta\boldsymbol{\phi}=[\Delta\phi^x,\Delta\phi^y]$ caused by eye appearance distortions under ECS.



(a) (b) (c) (d)

Figure 2: Eye images from video clip. Green/red arrows indicate capture direction vectors/eye (face) normals. Note that in (a) and (c), capture directions are similar under ECS, thus their appearance distortions and gaze direction biases are also similar.



(a) (b)

Figure 3: Regression for appearance distortion compensation. (a) Distorted eye images captured under head motion and fixed gaze position. (b) Results of leave-one-out experiments with/without appearance distortion compensation.

Subject	Full comp.	No comp.	Training samples
Average	**2.38°**	5.99°	33 + video clip
Sugano *et al.* [3]	4° ∼ 5°		≈ 10^3

Table 1: Estimation accuracy under free head motion.

Training the regression needs adequate training samples with different $\Delta\boldsymbol{v}^c$. Note that there is no requirement of specified gaze positions or head poses for the training samples. Thus we propose an unconventional calibration process that captures a short video clip while the user is gazing at a *fixed but arbitrarily assigned position* on the screen and moving his/her head (just rotating is effective). As there is no change of gaze positions and the user's head motion is free, the procedure can be done within several seconds while obtaining sufficient training samples. Therefore, a tedious calibration is avoided. Fig. 2 shows examples of eye images from the video clip and visualizes the camera directions under ECS.

After $\Delta\boldsymbol{\phi}_i$ and $\Delta\boldsymbol{v}_i^c$ being calculated from obtained training samples, a Gaussian Process (GP) model is used to learn a regression for estimating the gaze direction biases caused by different capture directions.

We first validate the appearance distortion compensation by leave-one-out experiments. Fig. 3(b) plots the experimental results. The average estimation bias reduced from 10.85° to 1.65° after being compensated. Furthermore, efficacy of the complete gaze estimation approach under free head motion is evaluated with multiple subjects. An average estimation accuracy of 2.38° is achieved as shown in Table 1.

[1] D.W. Hansen and Qiang Ji. In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(3):478–500, 2010.

[2] C.H. Morimoto and M.R.M. Mimica. Eye gaze tracking techniques for interactive applications. *Computer Vision and Image Understanding*, 98(1):4–24, 2005.

[3] Yusuke Sugano, Yasuyuki Matsushita, Yoichi Sato, and Hideki Koike. An incremental learning method for unconstrained gaze estimation. In *Proceedings of the 10th European Conference on Computer Vision (ECCV 2008)*, pages 656–667, 2008.