Supplementary Materials for "Offset Calibration for Appearance-Based Gaze Estimation via Gaze Decomposition"

Zhaokang Chen Bertram E. Shi
The Hong Kong University of Science and Technology
zchenbc@connect.ust.hk eebert@ust.hk

A. Appendix

A.1. Cross-Dataset Evaluation

For cross-dataset evaluation, we trained on the MPIIGaze dataset [4] and tested on the ColumbiaGaze dataset [3]. For each subject, we calibrated at the five images that correspond to the same gaze target ($T = 1, S = 5$), and tested on the remaining 65 images. We compared our proposed method with the fine-tuning the latent parameters method (LP) in [1] and the differential method (DF) in [2], where they were re-implemented using our architecture. We omitted the comparison with fine-tuning the last fully-connected layer and the adaptation methods since their performance was not good for single gaze target calibration in our previous experiments and in theory.

In Table A1, we present the mean angular errors when calibrating at different gaze targets. Our proposed gaze decomposition (GD) performed the best. On average, it outperformed LP by $0.8°$ (14.5%) and DF by $0.6°$ (11.3%).

A.2. Evaluation of the Learned Bias

We evaluated whether the learned biases, i.e., $\hat{b}_i$ in Eq. (2) in the main manuscript, were consistent for the same subject using the data from leave-one-subject-out (15 fold) cross-validation on MPIIGaze.

The mean and SD of each subject are shown in Table A2. Across subjects, the yaw means ranged from $-5.4°$ to $5.4°$ (SDs from $0.1°$ to $0.3°$). The pitch means ranged from $-2.9°$ to $3.9°$ (SDs from $0.1°$ to $0.3°$). We compared the intra-subject variance computed from the 14 folds where the subject was in the training set with the inter-subject variance computed from the means of the estimated biases. For yaw, the average intra-subject variance was $0.03\ \mathrm{deg}^2$ in comparison to the inter-subject variance of $5.40\ \mathrm{deg}^2$. For pitch, the variances were $0.05\ \mathrm{deg}^2$ and $3.66\ \mathrm{deg}^2$. The intra-subject variance was a small percentage (0.56%-1.4%) of the inter-subject variance, indicating that the bias is learned consistently and reliably during training.

Method	Vertical	Horizontal							Average
		$-15°$	$-10°$	$-5°$	$0°$	$5°$	$10°$	$15°$	
LP [1]	$0°$	5.2	5.8	5.6	5.2	5.1	5.5	6.3	5.5
	$-10°$	5.4	5.7	5.5	5.2	5.1	5.2	5.7	
DF [2]	$0°$	5.9	5.4	5.1	5.1	5.1	5.3	5.9	5.3
	$-10°$	5.5	5.1	5.0	5.0	5.0	5.2	5.7	
Ours (w/o GD)	$0°$	5.5	5.1	4.9	5.0	5.0	5.2	5.6	5.2
	$-10°$	5.5	5.1	4.9	4.8	4.9	5.0	5.5	
Ours	$0°$	**5.3**	**4.7**	**4.5**	**4.5**	**4.5**	**4.7**	**5.3**	**4.7**
	$-10°$	**4.9**	**4.5**	**4.4**	**4.4**	**4.5**	**4.6**	**5.1**	

* The calibration-free gaze estimation error of our proposed method with gaze decomposition is $5.5°$.

Table A1. Mean Angular Error ($°$) of Calibrating at Different Gaze Targets on the ColumbiaGaze Dataset.

References

[1] E. Lindén, J. Sjöstrand, and A. Proutiere. Learning to personalize in appearance-based gaze tracking. *IEEE International Conference on Computer Vision Workshops*, 2019.

[2] G. Liu, Y. Yu, K. A. Funes-Mora, J.-M. Odobez, and E. T. SA. A differential approach for gaze estimation with calibration. In *British Machine Vision Conference*, 2018.

[3] B. A. Smith, Q. Yin, S. K. Feiner, and S. K. Nayar. Gaze locking: Passive eye contact detection for human-object interaction. In *Proceedings of the ACM Symposium on User Interface Software and Technology*, pages 271–280. ACM, 2013.

[4] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling. Appearance-based gaze estimation in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4511–4520, 2015.

		P1		P2		P3		P4		P5		P6		P7		P8	
Yaw	Mean(°)	2.8	-2.8	1.2	-1.2	0.3	-0.4	0.2	-0.2	-1.8	1.9	-4.1	4.0	-0.7	0.7	1.4	-1.5
	SD(°)	0.2	0.3	0.3	0.1	0.1	0.2	0.1	0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.1	0.2
Pitch	Mean(°)	0.0	0.0	-0.8	-0.8	3.9	3.9	-1.7	-1.8	0.1	0.1	-0.5	-0.4	-1.5	-1.6	0.3	0.3
	SD(°)	0.2	0.2	0.3	0.3	0.3	0.3	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
		P9		P10		P11		P12		P13		P14		P15			
Yaw	Mean(°)	-3.5	3.4	-0.7	0.9	-0.3	0.2	0.9	-0.9	-0.4	0.3	1.9	-1.8	5.4	-5.4		
	SD(°)	0.1	0.1	0.3	0.2	0.1	0.2	0.2	0.2	0.1	0.2	0.1	0.1	0.1	0.1		
Pitch	Mean(°)	-0.6	-0.7	1.7	1.8	-2.9	-2.9	1.8	1.7	0.2	0.2	-2.8	-2.8	2.9	2.9		
	SD(°)	0.2	0.2	0.3	0.3	0.2	0.3	0.3	0.3	0.1	0.2	0.2	0.2	0.2	0.2		

* For each subject, the left column corresponds to the non-flipped images and the right column corresponds to the horizontally-flipped images.

Table A2. Mean and Standard Deviation (SD) of the Learned Bias $\hat{b}$ for Each Subject in Training on the MPIIGaze dataset.