Accurate Eye-Tracking from Deflectometric Information using Deep Learning

Jiwon Choi (1), Jiazhang Wang (2), Tianfu Wang (3), Florian Willomitzer (1)

1 : Wyant College of Optical Sciences, University of Arizona, USA
2 : Department of Electrical and Computer Engineering, Northwestern University, USA
3 : Department of Computer Science, ETH Zürich, Switzerland

Abstract: We introduce an accurate eye-tracking method that exploits deflectometric information and uses deep learning to reconstruct the gaze direction. We demonstrate real world experiments with evaluated gaze errors below 1°.

Keywords: Eye-Tracking, Deflectometry, Artificial intelligence, Deep Learning, Virtual Reality, Computational Imaging

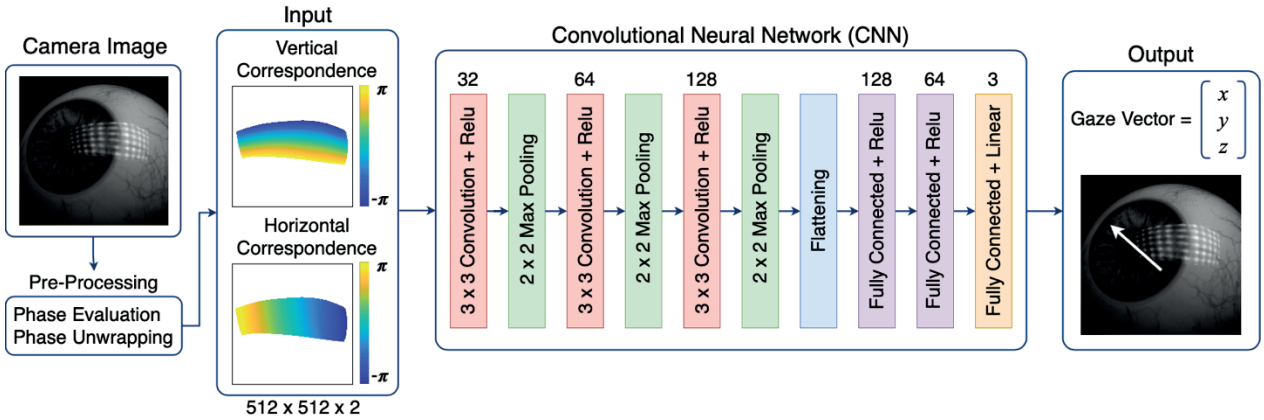

Fig. 1. Deep learning-based eye-tracking using deflectometric information. The eye is illuminated by a screen displaying a cross-sinusoidal pattern. From the single captured eye image under screen illumination, we evaluate the vertical and horizontal screen correspondence maps in single-shot [1], then feed the correspondences into the Convolutional Neural Nets (CNN) to reconstruct the gaze direction. Training is performed with a set of 2,000 simulated eye images under screen illumination, using different shapes, translations and rotation angles of the simulated eye. As our algorithm works on phase maps instead of intensity images, an elaborate photorealistic simulation of eye images is not necessary, and real images are not required for training.

1. Introduction

A robust, accurate, and fast solution to eye-tracking has widespread applications, spanning from psychology and neuroscience research to Virtual Reality (VR). "Image-based" eye-tracking methods calculate gaze direction by analyzing features extracted from 2D eye images [1-3]. Yet, the limited density of detected eye features constrains the accuracy of gaze direction estimation to a few degrees. "Reflection-based" methods (e.g., 'glint tracking') exploit cornea reflections from a few point light sources [1-3]. These methods typically can reach higher accuracy, but their performance is limited by the limited number of point source reflections (~12 in the current state-of-the-art). Recently, our group has introduced a series of approaches that reconstruct the gaze direction from dense deflectometric information [1,2]. By observing the reflection of a full screen (>1 million dense pixels), the number of point reflections from the eye surface can be significantly increased (compared to ~12 sparse glints), which translates to an accurate evaluation of the gaze vector. We direct the reader to [1] for more information.

However, our previous methods are not without limitations. The 'Classic Stereo Deflectometry method' [1] delivers the highest accuracy (gaze error <0.25°) but requires two cameras to achieve the best performance. Our 'Optimization-based inverse rendering approach' [2] uses only one camera but exploits prior knowledge about the measured eye to provide a good starting value for the optimization process. In this contribution, we introduce a novel flavor of our deflectometric eye tracking research: We train a Convolutional Neural Network (CNN) model to predict the gaze direction from the captured deflectometry correspondence maps. Our evaluated gaze error is <1°. Compared to our previous deflectometric approaches, this method requires only one camera (and a screen), and the shape of the measured eye does not need to be known in advance (as long as the training dataset is diverse enough). As our algorithm works on phase maps instead of intensity images, our training dataset does not need to contain photorealistic simulated eye images nor real captured data, which is in stark contrast to other learning-based approaches to eye-tracking.

2. Method

Deflectometry is a well-known optical metrology principle for measuring specular surface [4], which recently has been applied for eye tracking as follows: The reflection of a screen with a known (fringe-) pattern

Paper presented at the 14th International Conference on Optics-Photonics Design & Fabrication, Tucson, July 10th-12th, 2024
Submitted: Feb. 14th 2024, Accepted: Apr. 4th, 2024, Published in Abstract Digest: Aug. 15th, 2024 (https://doi.org/10.1117/12.3048202)

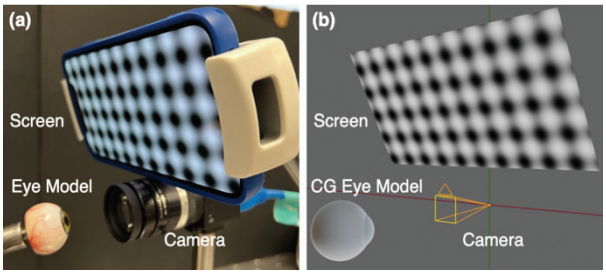

Fig. 2. (a) Experimental setup with one camera and a screen. (b) Digital twin setup used to generate training data in simulation.

is observed over the (partially) specular surface of cornea and sclera. From the deformation of the pattern in the camera image, the correspondence between screen and camera can be evaluated. In standard deflectometry, this information is used to calculate the normal map and eventually shape of the specular surface. As shown in the literature [4], the process of finding accurate deflectometric correspondences can be performed analytically with high robustness to noise and ambient lighting.

In the particular case of this contribution, we obtain the horizontal and vertical screen-camera correspondence information in single-shot, using a crossed sinusoidal pattern paired with a wavelet-transform-based evaluation approach [1]. Eventually, we use the obtained correspondence information as input for our learning-based algorithm (see Fig 1). As discussed, using correspondence instead of intensity image input has several advantages for creating a training dataset. While it is very hard and cumbersome to simulate and render a diverse set of realistic eye images under different ambient lighting and screen illumination, correspondence maps for different eye shapes can be always simulated very close to real captured correspondence maps. Moreover, restricting ourselves to correspondence maps prevents the network from learning secondary features of the eye or periocular region (veins, wrinkles, etc.) which might not be present in the real images.

To train our network, we generate 2,000 correspondence pairs (vertical and horizontal correspondence) by rendering images with different eye shape, translation, and rotation. Our used renderer is a virtual copy of our setup (see Fig. 2 and [2]) and only assumes a purely specular eye surface (texture-less) under screen illumination. 1,800 of the simulated correspondence pairs (together with the respective gaze vectors) are used for training while the rest is used for validation. Our model uses the angular difference between ground truth gaze vector and predicted gaze vector as loss. The network structure is depicted in Fig. 1. We train for 130 epochs with a batch size of 32, which takes ~35mins on a single NVIDIA GeForce RTX 3060 GPU (see Fig. 3 for loss plot). Predicting the gaze direction takes ~0.7ms from a (scaled down) 512 x 512 pix correspondence pair.

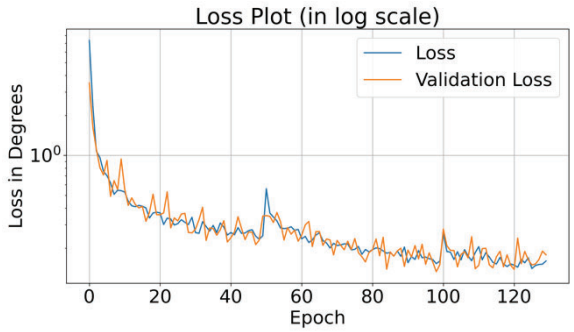

Fig. 3. Loss plot in log scaled degrees. Convergence trend shown for 130 epochs.

3. Experiment and Result

To quantitatively validate our deep learning-based gaze estimation method, we conduct real-world experiments on a realistic eye model mounted on a high-precision rotation stage. We define an arbitrary rotation angle to 0° and rotate the eye model to different rotation positions (−4°, −2°, 0°, 2°, 4°). We use the procedure outline above and in [1] to display crossed fringe patterns and retrieve the correspondence pairs on the eye surface in single-shot, which are eventually fed into the trained model to predict the gaze. For each rotation position, we capture 20 correspondence pairs, while we always rotate the eye model between two consecutive measurements. Since the absolute gaze direction for this experiment is unknown, we calculate the *relative gaze angle* between two rotation positions and compare the result with the angle we rotated the eye model ("ground truth"). All angles are calculated w.r.t. the gaze at position 0° which serves as reference. We then evaluate the mean relative error $\boldsymbol{\varepsilon_0}$ at each rotation position $\boldsymbol{a}$ w.r.t. the $\boldsymbol{0°}$ position (see [2] for more details). The evaluation result in Tab. 1. shows that our evaluated gaze error $\boldsymbol{\varepsilon_0}$ is always <1°.

Table 1. Gaze angle estimation

Rotation Position *a*	**-4°**	**-2°**	**0°**	**2°**	**4°**
Mean Relative Error ε_0	0.47°	0.98°	0°	0.80°	0.89°

In summary, we have developed a deep learning-based eye tracking method using deflectometric information and demonstrated promising first results. In future work, we will evaluate our method on real human eyes in vivo, will increase the size of the training dataset and explore potential factors such as pattern coverage and type of correspondence input.

4. References

[1] J. Wang, T. Wang, B. Xu, O. Cossairt, and F. Willomitzer. "Accurate Eye Tracking from Dense 3D Surface Reconstructions using Single-Shot Deflectometry." Preprint arXiv:2308.07298 (2023).

[2] T. Wang, J. Wang, O. Cossairt, and F. Willomitzer. "Optimization-Based Eye Tracking using Deflectometric Information." Preprint arXiv:2303.04997 (2023).

[3] S. Ghosh, A. Dhall, M. Hayat, J. Knibbe, and Q. Ji. "Automatic gaze analysis: A survey of deep learning based approaches." *IEEE PAMI* , *46*(1), 61-84. (2023).

[4] L. Huang, M. Idir, C. Zuo, and A. Asundi, "Review of phase measuring deflectometry," Opt. Lasers Eng. 107, 247–257 (2018).

Paper presented at the 14th International Conference on Optics-Photonics Design & Fabrication, Tucson, July 10th-12th, 2024
Submitted: Feb. 14th 2024, Accepted: Apr. 4th, 2024, Published in Abstract Digest: Aug. 15th, 2024 (https://doi.org/10.1117/12.3048202)