Real-Time, Training-Free Ocular Refraction Sensing with a Thin Diffuser for Near-Eye Displays

Futa Fujimoto*
Institute of Science Tokyo
Tokyo, Japan
fujimoto.f.659b@m.isct.ac.jp

Yuichi Hiroi*
Cluster Metaverse Lab
Tokyo, Japan
Institute of Science Tokyo
Tokyo, Japan
y.hiroi@cluster.mu

Yuta Itoh
Institute of Science Tokyo
Tokyo, Japan
itoh@comp.isct.ac.jp

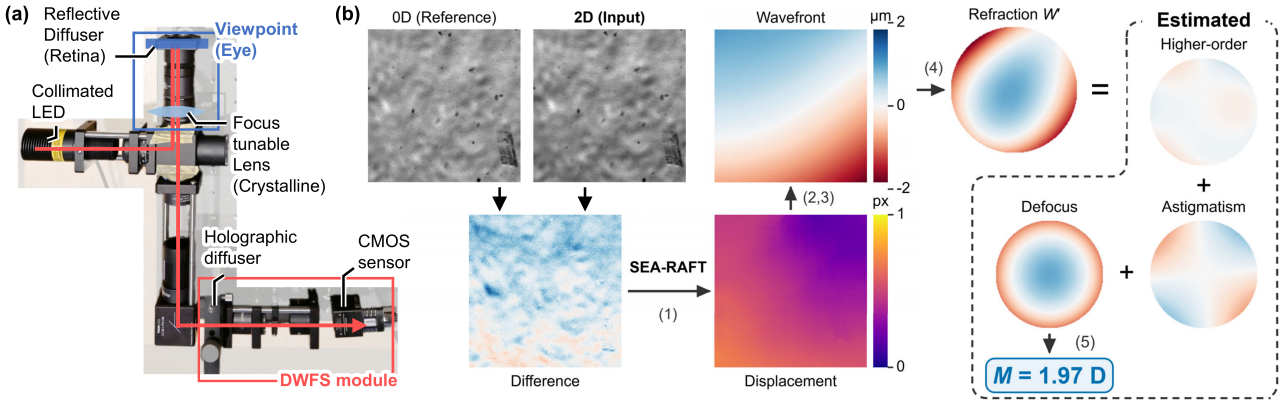


Figure 1: Zero-shot optical flow recovers ocular refractive power from two diffuser caustics without task-specific training. (a) The bench: a holographic diffuser before a bare CMOS sensor images a model eye set by a focus-tunable lens. (b) From the 0/2 D caustics and their difference, (1) SEA-RAFT estimates the displacement, (2,3) a fixed physics stage integrates the wavefront W, (4) a Zernike fit gives the refraction wavefront W' (defocus + astigmatism + higher-order), and (5) its defocus gives $M = 1.97$ D in 13 ms.

Abstract

Focus-adaptive near-eye displays counter the vergence–accommodation conflict by matching focus to the viewer's eye, which requires measuring the eye's refractive state. Clinical aberrometers are too large and slow for a headset, and compact wavefront sensors trade away range and accuracy. Diffuser-based wavefront sensing (DWFS) is a compact, wide-range alternative in which a thin diffuser makes a bare image sensor a refraction sensor, since the caustic speckle it casts shifts with refractive error. Estimating that shift field fast enough has been the bottleneck. We recast the DWFS analysis as a differentiable, GPU-resident back-end and show that a pre-trained optical-flow network, SEA-RAFT, applied zero-shot without fine-tuning on caustics, can drive it in place of the slow registration. With a compact three-parameter per-device calibration it reaches a mean absolute error of 0.093 D against the nominal lens diopter, about 2.7 times inside the 0.25 D clinical tolerance, and a compressed configuration runs at 13 ms and 75 Hz at 0.075 D, the rate a focus-adaptive display needs.

CCS Concepts

• **Human-centered computing** → **Displays and imagers**; • **Computing methodologies** → *Computer vision.*

Keywords

Ocular aberrometry, wavefront sensing, near-eye displays, optical flow, autorefraction

ACM Reference Format:
Futa Fujimoto, Yuichi Hiroi, and Yuta Itoh. 2026. Real-Time, Training-Free Ocular Refraction Sensing with a Thin Diffuser for Near-Eye Displays. In *Proceedings of 32nd ACM Symposium on Virtual Reality Software and Technology (VRST '26).* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3822517.3849910

1 Introduction

Head-mounted displays fix imagery at one focal plane, and the resulting focus–vergence mismatch, the vergence–accommodation conflict, causes discomfort [3]. A focus-adaptive display counters it by adapting focus to the eye, using depth from binocular vergence or a direct measurement of the eye's refractive state. The direct route needs no gaze geometry. EyeAR refocuses AR content from eye measurements [5], and Kaneko et al. paired a retinal projection display with ocular wavefront sensing [2]. Both need a refraction sensor that fits in a headset and runs in real time.

*Both authors contributed equally to this research.

VRST '26, Sendai, Japan
2026. ACM ISBN 979-8-4007-2811-2/2026/11
https://doi.org/10.1145/3822517.3849910

Embedding such a sensor is hard. Clinical aberrometers are large and slow, and compact wavefront sensors such as lenslet Shack–Hartmann limit range and accuracy. Diffuser-based wavefront sensing (DWFS) avoids both. Berto et al. showed that a thin diffuser before a bare sensor casts a caustic whose local shift is proportional to the wavefront gradient [1], and McKay et al. built a low-cost, wide-range ocular autorefractor on it [4]. Integrating the shift field gives the wavefront, whose defocus and astigmatism form the clinical power vector (M, J_0, J_{45}). The bottleneck is shift estimation. The multi-level registration DWFS uses is accurate but slow, about two minutes per measurement [4], which has kept it from real-time use.

We make DWFS real-time in two steps. First, we recast its analysis back-end as a single differentiable, GPU-resident module, trainable end-to-end. Second, we drive it with a pre-trained optical-flow network used zero-shot in place of the slow registration. Although it was unclear whether a network trained on natural video would transfer to caustic speckle, which looks nothing like a natural scene, we confirm that it does.

2 Method

A refractive error tilts the light leaving the eye, and the diffuser turns each local tilt into a lateral shift of its caustic speckle. The shift field thus samples the wavefront slope, which integrates to the wavefront and hence the refraction. The pipeline does this in five stages (Fig. 1): (1) estimate the displacement field D between the reference and aberrated caustics; (2) scale D into the slope $\nabla W = D\,p_x/d$; (3) integrate the slope into the wavefront W with a Fourier-domain solver; (4) fit 16 Zernike modes; and (5) read the defocus and astigmatism modes as the power vector (M, J_0, J_{45}), the eyeglass prescription clinicians write. The pixel pitch $p_x = 6.0\,\mu$m and diffuser-to-sensor gap $d = 5.15$ mm are fixed by the optics, and stages 2–5 form a differentiable GPU module.

The first stage departs from prior DWFS. SEA-RAFT [6] replaces its iterative registration, applied zero-shot. The output is proportional to the true power up to a device-dependent response, which a three-coefficient calibration $\hat{M} = aM^2 + bM + c$ removes (the mild quadratic term absorbs the sensor's response curvature); the coefficients are fit once per device from known-diopter captures.

3 Results and Outlook

We built the caustic-capture bench following Kaneko et al. [2] (Fig. 1a), with a model eye whose refraction is set by a focus-tunable lens. We evaluate on 35 caustic pairs from five sessions and seven lens settings, 0.5 to 3.5 D, scoring each against the nominal diopter D with leave-one-set-out calibration. Figure 2 and Table 1 report accuracy and latency.

Zero-shot SEA-RAFT reaches a mean absolute error of 0.093 D (max 0.24 D; Fig. 2), with all 35 samples inside the 0.25 D clinical tolerance; the residual is mostly per-session setup variation. Counter-intuitively, compressing to 160 px and two iterations *improves* accuracy to 0.075 D and cuts latency to 13 ms and 75 Hz on an RTX 3070. Because we read out only low-order defocus and astigmatism, downsampling low-passes the flow and removes high-frequency estimation noise without touching that signal. On the same data, classical Demons registration is both slower (~6.4 s/pair) and less accurate (0.100 D; Table 1).

Table 1: Speed and accuracy on the recapture (RTX 3070). Error is the leave-one-set-out calibrated mean $|M - D|$; zero-shot SEA-RAFT beats classical Demons registration on both.

Configuration	Grid	$\lvert M - D\rvert$ [D]	Latency
Full (12 iter.)	480	0.093	~144 ms
Compressed (2 iter.)	160	0.075	13 ms (75 Hz)
Demons registration	480	0.100	~6.4 s

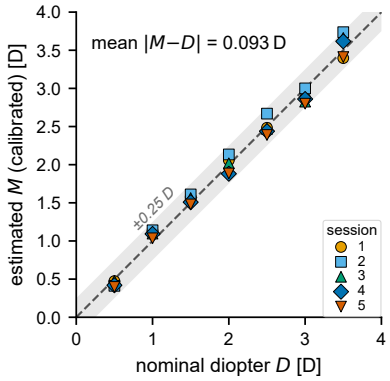


Figure 2: Zero-shot SEA-RAFT stays within the ±0.25 D clinical band (shaded) at every diopter; color and shape denote the five sessions, and the dashed line marks parity.

Distilling the flow front-end can reach the 90 Hz a focus-adaptive display needs [2], and the differentiable back-end enables end-to-end fine-tuning and diffuser co-design.

Acknowledgments

This study was supported by JST ASPIRE Grant Number JPMJAP2327 and JSPS KAKENHI Grant Number 25K03173, Japan.

References

[1] Pascal Berto, Hervé Rigneault, and Marc Guillon. 2017. Wavefront sensing with a thin diffuser. *Optics Letters* 42, 24 (2017), 5117–5120.
[2] Mayu Kaneko, Yuichi Hiroi, and Yuta Itoh. 2021. Focus-Aware Retinal Projection-based Near-Eye Display. In *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE.
[3] Gregory Kramida. 2016. Resolving the vergence-accommodation conflict in head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics* 22, 7 (2016), 1912–1931.
[4] Grace N. McKay, Faisal Mahmood, and Nicholas J. Durr. 2019. Diffuser-based computational imaging for ocular wavefront aberrometry. *Biomedical Optics Express* 10, 4 (2019), 1718–1735.
[5] Damien Constantine Rompapas, Aitor Rovira, Alexander Plopski, Christian Sandor, Takafumi Taketomi, Goshiro Yamamoto, Hirokazu Kato, and Sei Ikeda. 2016. EyeAR: Refocusable augmented reality content through eye measurements. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*.
[6] Yihan Wang, Lahav Lipson, and Jia Deng. 2024. SEA-RAFT: Simple, Efficient, Accurate RAFT for Optical Flow. In *European Conference on Computer Vision (ECCV)*.