paper #3254

Computational coherent imaging for accommodation-invariant near-eye displays

Jani Mäkinen, Erdem Sahin, Ugur Akpinar, Atanas Gotchev

Tampere University, Finland

Problem formulation

- Conventional near-eye displays (NED) provide only binocular vergence and disparity as depth cues, missing monocular accommodation and retinal defocus blur, thus resulting in vergence-accommodation conflict (VAC)
- One approach for solving VAC is to make the display accommodation-invariant (AI): depth-independent defocus blur, accommodation mainly driven by disparity
- Let us assume a computational display consisting of a panel, optics and pre-processing algorithm: what is the optimal set of optics and preprocessing algorithm such that the display produces as sharp images as possible?

Pre-processing algorithm

- Processes the input image before it is fed to the display panel
- Compensates for the display optics such that the AI response is achieved (jointly with the optics)
- Analogous to the post-processing deblurring step of similar computational camera designs
- Here the U-net architecture chosen as the pre-processing CNN

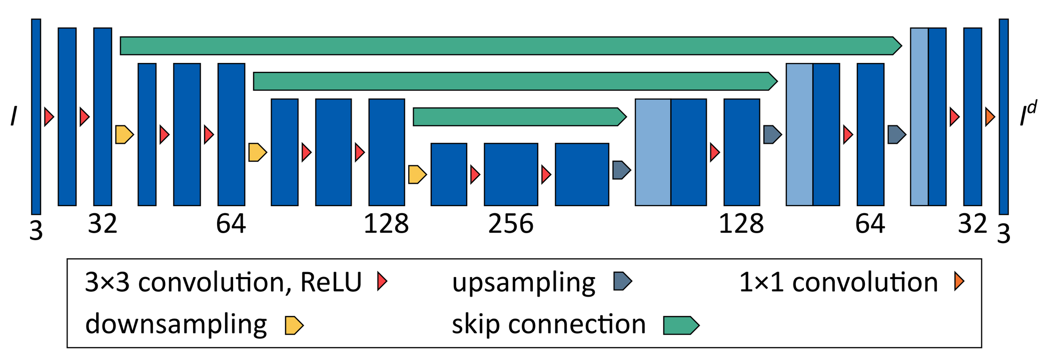

Figure 2: The pre-processing CNN (U-net) structure

Proposed design

- Our proposed design procedure utilizes a learning based approach to jointly optimize the display optics and the pre-processing algorithm (here convolutional neural network, CNN)
- The sharp input image is fed to the CNN, which outputs the display amplitude
- Given the current state of the optics, we simulate the viewer perceived image, which is then compared against the ground truth sharp image using a loss function
- The loss function drives the optimization procedure based on its gradient, optimizing the entire display system

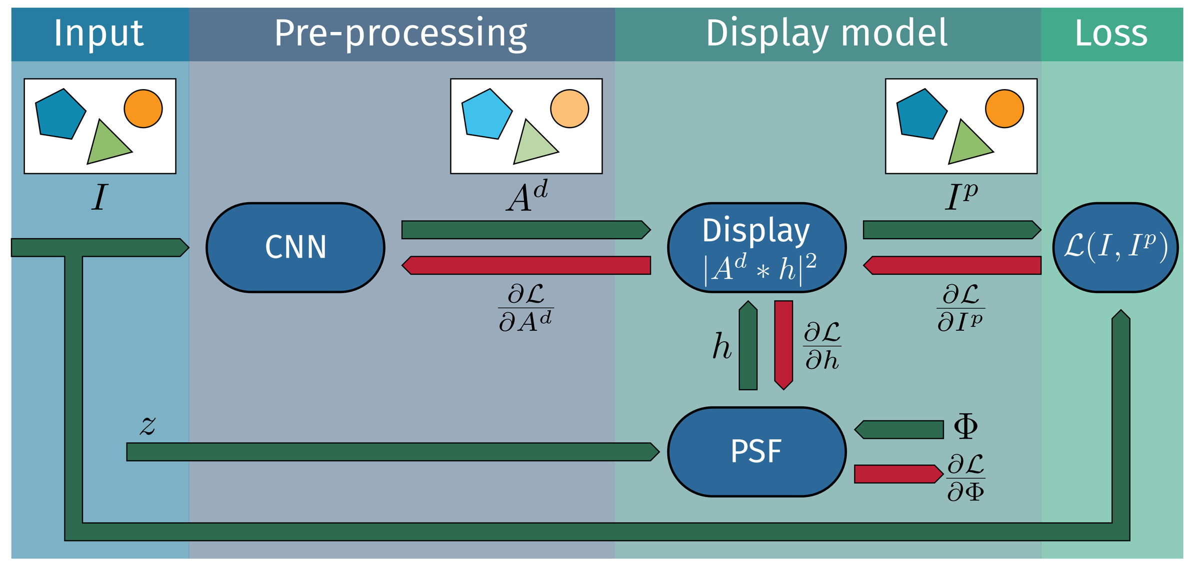

Figure 1: The proposed learning based computational NED system design approach

Display model and image formation

- Viewing process simulated by assuming an aberration-free eye accommodating at distance z
- Image formed on the reference (conjugate) plane as

$$I^p_{\lambda,z}(x,y) = \Big|\underbrace{A^d_\lambda(x,y)}_{\text{Display amplitude}} * \underbrace{h_{\lambda,z}(x,y)}_{\text{Coherent system point spread function}}\Big|^2$$

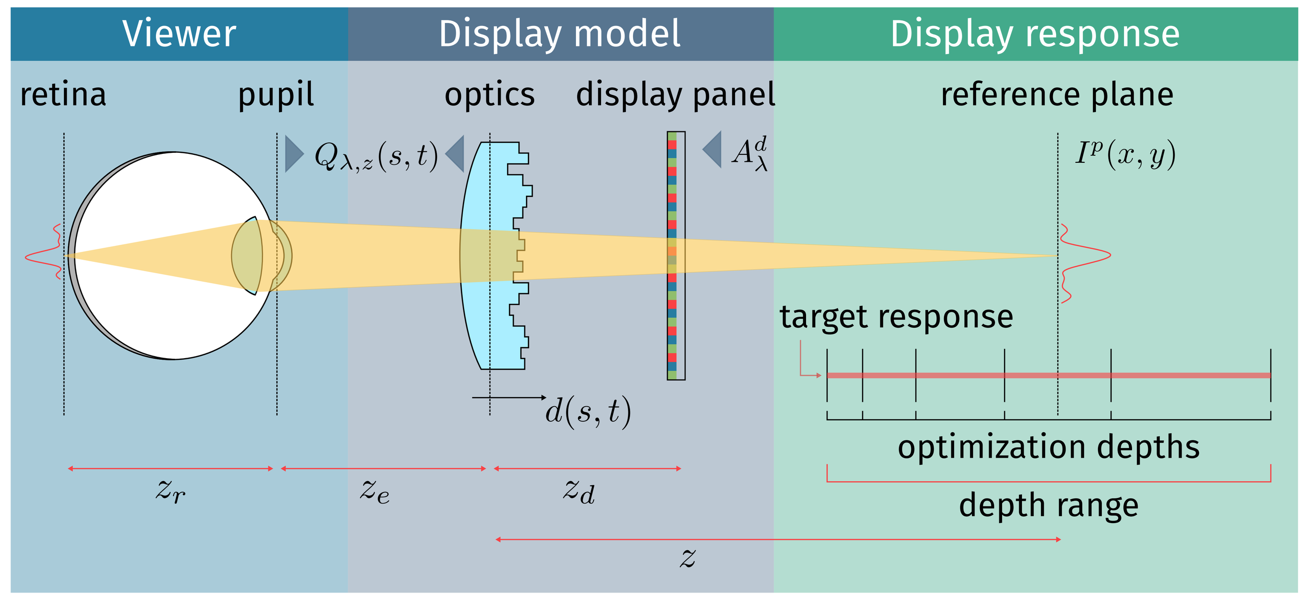

Figure 3: Display design and imaging model

Loss

- Our choice for the loss function combines two metrics: the L_1 distance and the structural similarity index measure (SSIM)
- The aggregated loss function attempts to minimize L_1 and maximize SSIM as a compromise between texture detail quality and the perceived change in structural information
- Overall loss calculated as

$$\mathcal{L}(I,I^p) = \overbrace{\mathcal{L}_{L_1}(I,I^p)}^{L_1\text{ distance}} + \overbrace{\mathcal{L}_{\mathrm{S}}(I,I^p)}^{\text{SSIM-based loss}} + \underbrace{\alpha\mathcal{R}(I,I^p) + \gamma\mathcal{R}^d(I^d)}_{\text{Regularization terms}}$$

Reconstruction results

- Simulation results at five different accommodation depths for each method, proposed design compared against two state-of-the-art methods in computational AI NEDs [1, 2]

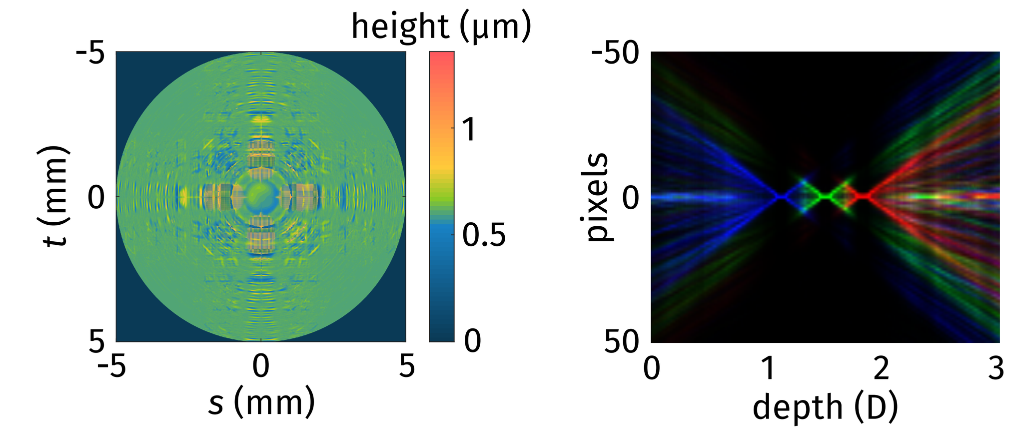

Figure 4: Optimized DOE height map and 1D cross-sections of PSF intensities

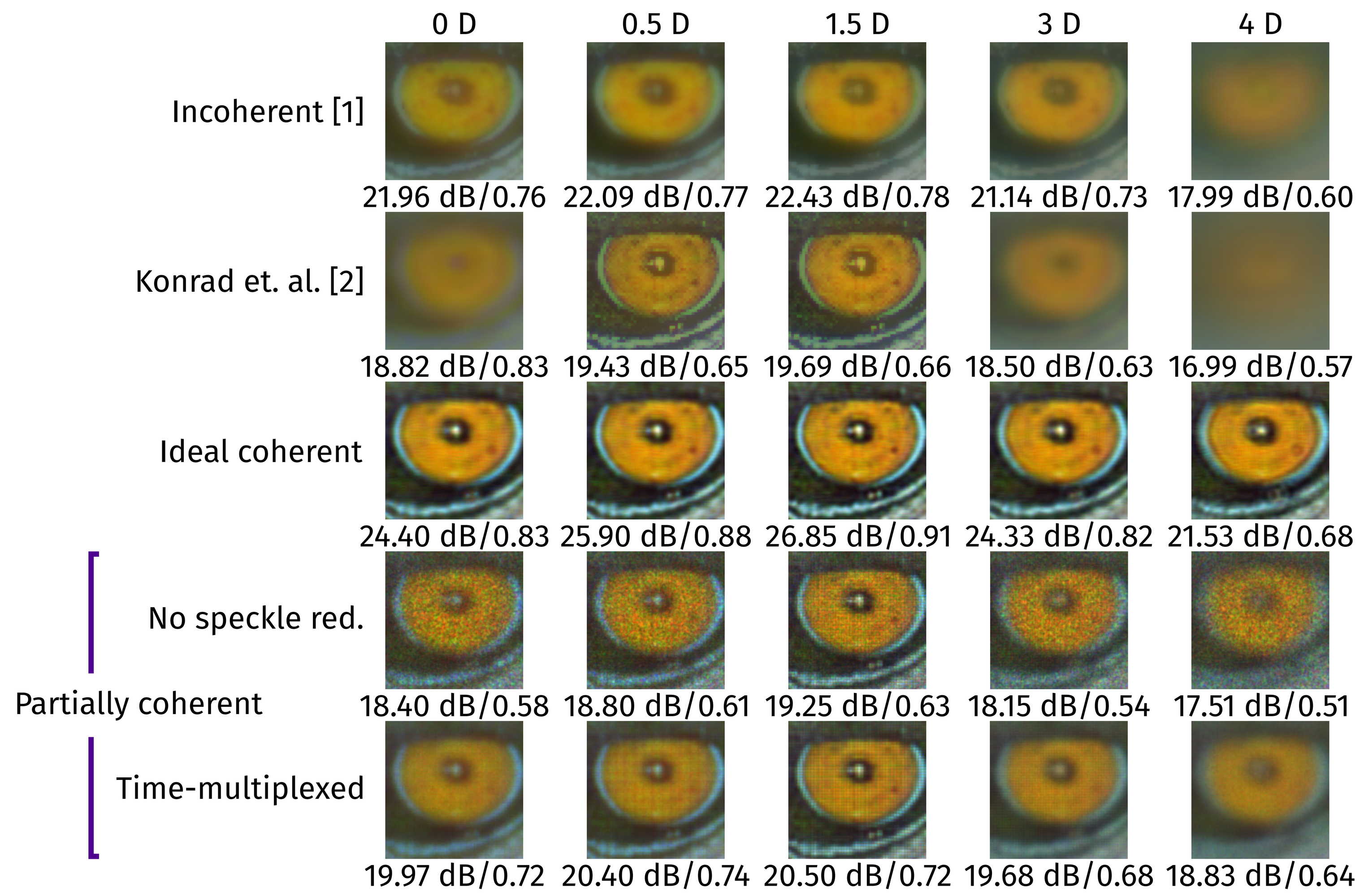

Figure 5: Zoomed-in simulation results for the Baboon test image, PSNR/SSIM values of the full size reconstruction below each image

[1] U. Akpinar, E. Sahin, and A. Gotchev, “Phase-coded computational imaging for accommodation-invariant near-eye displays,” in 2020 IEEE International Conference on Image Processing (ICIP), 2020, pp. 3159–3163.

[2] R. Konrad, N. Padmanaban, K. Molner, E. A. Cooper, and G. Wetzstein, “Accommodation-invariant computational near-eye displays,” ACM Trans. Graph., vol. 36, no. 4, July 2017.