Ergonomic-Centric Holography: Optimizing Realism, Immersion, and Comfort for Holographic Display: supplemental document
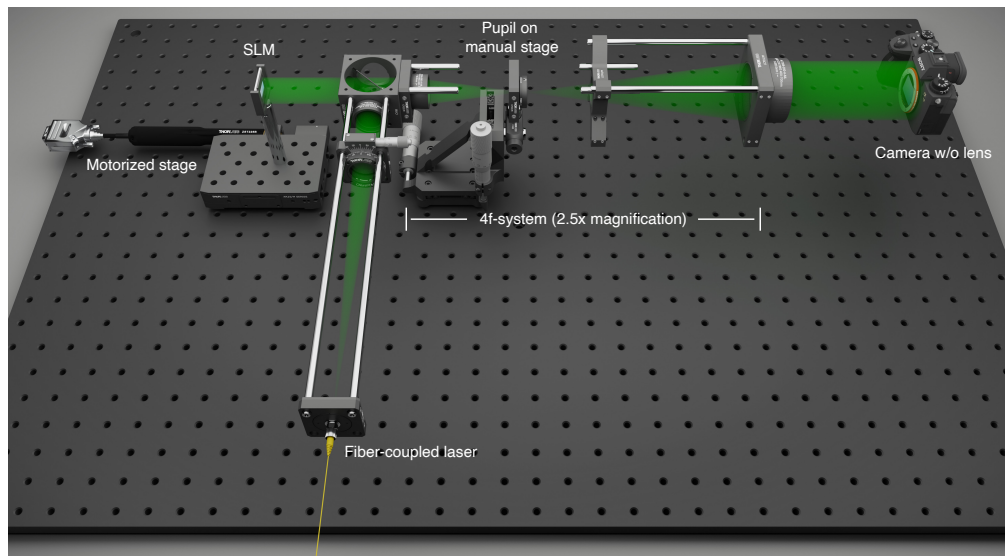
1. ADDITIONAL HARDWARE IMPLEMENTATION DETAILS

Fig. S1. Schematic rendering of the experimental setup. Only green laser is visualized. The control boxes of the SLM and laser are omitted.

Figure S1 shows a schematic rendering of our experimental setup. The labeled components match the ones used in the setup to ease reproduction. The fiber-coupled laser is collimated by a lens, linearly polarized, and directed to the SLM by a beamsplitter. The virtual 3D image, relayed by a 4f system, is directly imaged by the camera sensor without a lens attached to the camera body. The $2.5\times$ transverse magnification is set to match the image roughly to the size of the camera sensor to fully utilize the sensor resolution. Refocusing through the 3D scene is achieved by translating the SLM with the motorized stage. Miniaturization of such a setup (the parts before the pupil on the manual stage) has been demonstrated by Kim et al. [1] in a compact eyeglass form factor.

2. PRINCIPLE AND RENDERING OF LAYERED DEPTH IMAGE

Layered depth image (LDI) is an advanced representation for 3D graphic data [2]. Unlike a conventional RGB-Depth image that stores a single pixel at each spatial coordinate, it stores a sequence of RGB-D images along the line of sight originating from each spatial coordinate. As illustrated in Fig. S2, its first pixel records the first front-facing surface intersecting with the line of sight, and the second pixel records the second intersection assuming the line of sight pierces through the scene, and so forth till a maximum hit (layer) count is met. LDI can efficiently record near-complete foreground and background information using a few layers (e.g., 5) and be used to generate realistic incoherent focal stacks. More importantly, its simple data structure is very deep learning friendly compared to other representations such as general point clouds or meshes,

which are much more difficult to deal with due to their scene-dependent length. This can greatly facilitate the migration of EC-H from iterative optimization to neural network prediction.

Layered depth images can be efficiently rendered using depth peeling [3]. To generate a N-layer LDI, depth peeling runs the rendering pipeline N times and simultaneously maintains two depth (Z) buffers. One works conventionally, while the other remains a constant at each rendering pass and sets the minimum distance at which a fragment can be drawn without being discarded. For each pass, the previous pass's conventional Z-buffer is used as the current minimal Z-buffer to render the content right behind the previous pass. The pseudocode of depth peeling is outlined in Algorithm S1, which can be easily implemented using OpenGL.
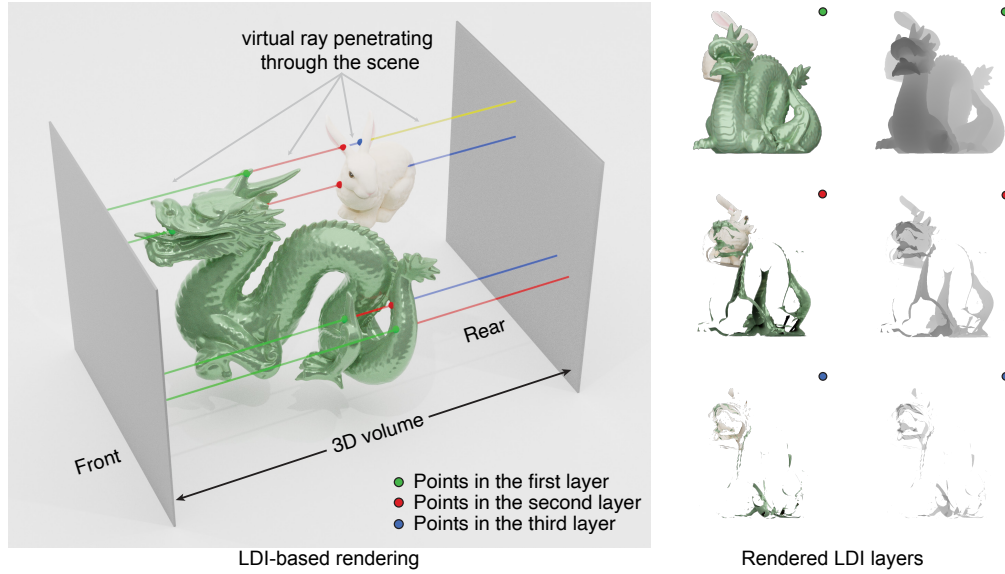

LDI-based rendering   Rendered LDI layers

Fig. S2. The principle of layered depth image (LDI) rendering. In the rendering schematic (left), exemplary rays originate from the front of the volume, traverse through the scene, and terminate at the end of the volume. To aid visualization, the ray's color is modified after each intersection. Every hit with a scene object is recorded with the corresponding depth. All intersections sharing the same index count are consolidated into a single image (see right images).

Algorithm S1. Depth peeling for LDI rendering

N: the number of layers in the output LDI
X, Y: spatial resolution of the output LDI
C: an RGB image buffer
D_0, D_1: two depth buffers
F_{xy}: the set of all fragments defined at location (x, y)
1: $D_1 \leftarrow 0$
2: **for** $i \leftarrow 0$ to $N - 1$ **do**
3:     CLEAR(C)
4:     $D_{i\%2} \leftarrow \infty$
5:     **parfor** $x \leftarrow 0$ to $X - 1, y \leftarrow 0$ to $Y - 1$ **do**   ▷ shader
6:         **for** f **in** F_{xy} **do**
7:             **if** $f.$`frontfacing` **and** $f.$`depth` $> D_{(i+1)\%2}(x, y)$ **and** $f.$`depth` $< D_{i\%2}(x, y)$ **then**
8:                 $C(x, y) = f.$`color`
9:                 $D_{i\%2}(x, y) = f.$`depth`
10:     **end parfor**
11:     SAVE($C, D_{i\%2}$)   ▷ save i-th layer of LDI

In rendering engines like Blender [1], Algorithm S1 can be implemented by modifying each material node to exhibit depth-dependent behavior. This augmentation involves providing a depth image as input to the material node. At each pixel coordinate, the node samples the depth input and determines if the elements in the associated meshes are positioned in front of the

---

[1]www.blender.org

sampled depth. If so, the material becomes transparent; otherwise, it behaves as intended. To obtain the LDI, the rendering process is repeated N times, with the depth from the previous pass being fed into the augmented materials in each iteration. Figure S3 visualizes the LDI rendered for the "PartyTug 6:00AM" scene using the above procedure in Blender.
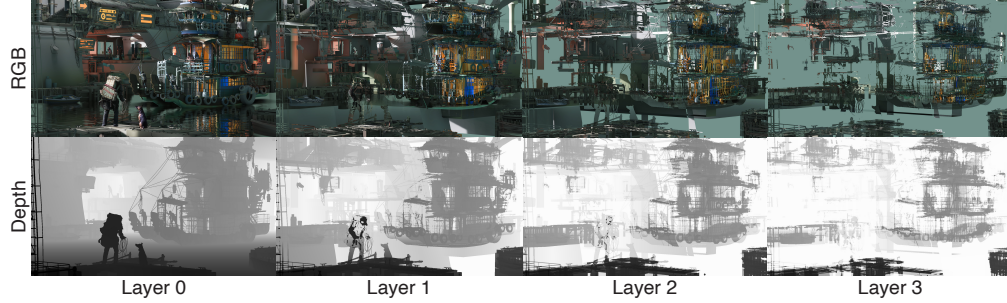

RGB				
Depth				
Layer 0	Layer 1	Layer 2	Layer 3

Fig. S3. The first 4 layers of the LDI rendered for the "PartyTug 6:00AM" scene, which contains very complex spatial structures. A higher value in the depth image represents a further depth.

3. INITIALIZATION OF GLOBAL SCALE s_g

The optimizable global scale parameter s_g bridges the intensity mismatch between the phase-only hologram and the target 3D image. In practice, the initialization of s_g influences convergence speed and image quality. Because free-space propagation conserves the total energy (intensity), we initialize s_g with the following formula to encourage a minimal change of s_g during optimization

$$s_g^{i=0} = \frac{\frac{1}{N}\sum_{n=1}^{N}\overline{a_n^2}}{\frac{1}{N}\sum_{n=1}^{N}\overline{|u(\phi;z_n)|^2}}, \quad \text{(S1)}$$

where $\overline{\cdot}$ is the mean operator. This formula scales the mean intensity of the phase-only hologram to match the mean intensity of the target focal stack. This approach differs from the initialization formula used in the original HOGD algorithm [4]

$$s_g^{i=0} = \frac{\frac{1}{N}\sum_{n=1}^{N}\overline{a_n|u(\phi;z_n)|}}{\frac{1}{N}\sum_{n=1}^{N}\overline{|u(\phi;z_n)|^2}}, \quad \text{(S2)}$$

which roughly scales the mean

Fig. S4. Comparison of our s_g initialization method with the one in the original HOGD algorithm. The convergence of image quality in peak signal-to-noise ratio (PSNR) and s_g are shown for two different scenes. The plots are reported for TM-1 (i.e., no time multiplexing). Adding time-multiplexed frames preserve the trend of convergence.

amplitude of the phase-only hologram to match the mean amplitude of the target focal stack, a property yet doesn't hold as a constant during propagation. As show in Fig. S4, we find Eq. S1 achieves superior performance over Eq. S2 in terms of image quality and provides a significantly closer initialization to the optimized final values.
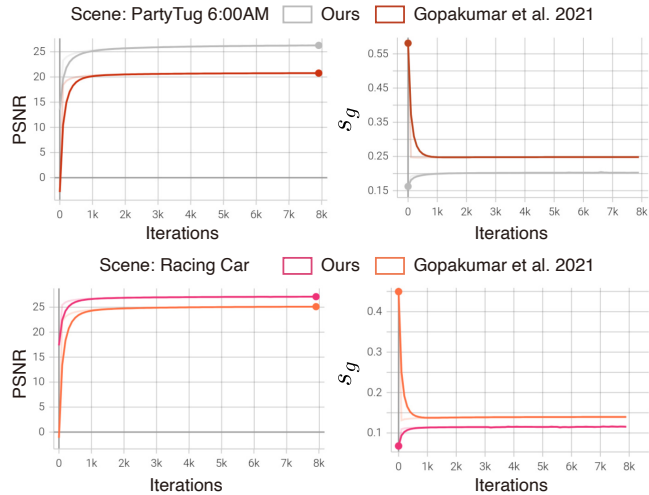
4. IMPLEMENTATION DETAILS OF DOWNSAMPLING OPERATION IN HIGH-ORDER DIFFRACTION PROPAGATION

When numerically implementing Eq.(3), the Fourier domain replication $U\left(f_x, f_y; \phi_t\right)$ changes the output field resolution from $[H, W]$ to $[\alpha H, \alpha W]$, where H and W are the width and height (spatial resolution) of the input field. However, the spatial resolution of amplitude images in the target focal stack a_n in Eq.(4) remains at $[H, W]$, demanding a dimension reduction of Eq.(3) output to $[H, W]$ to evaluate the objective function (loss). According to Fourier optics, increasing the Fourier domain resolution doesn't change the support of the field in the space domain, but increases the spatial resolution within the support; thus, the dimension reduction shall be implemented as downsampling instead of cropping.

In practice, the downsampling can be implemented as a strided convolution operation with a stride α and a uniform kernel of size $[\alpha, \alpha]$ and value $\frac{1}{\alpha^2}$ (a normalization factor that eases s_g initialization). Physically, this mimics the process of virtually placing an imaging sensor with resolution $[H, W]$ at the output plane and integrating the electric field stimulated by the photons arriving at each sensor pixel. This integration shall be implemented over the intensity (incoherently) instead of the complex amplitude (coherently) of the output field since every group of $[\alpha, \alpha]$ pixels arrive at different areas of their corresponding virtual sensor pixel and trigger photoelectric effect independently. Mathematically, denote the complex amplitude of the digitized output field $u(\phi; z) \in \mathbb{C}^{\alpha H \times \alpha W}$, the amplitude of the virtual sensor captured field $\hat{u}(\phi; z) \in \mathbb{R}^{H \times W}$, their relation is given by

$$\hat{u}(\phi; z)[x, y] = \sqrt{\sum_{i=1}^{\alpha} \sum_{j=1}^{\alpha} \frac{1}{\alpha^2} |u(\phi; z)[\alpha x + i, \alpha y + j]|^2} \quad \text{(incoherently, correct)} \quad \text{(S3)}$$

instead of

$$\hat{u}(\phi; z)[x, y] = \sum_{i=1}^{\alpha} \sum_{j=1}^{\alpha} \frac{1}{\alpha^2} u(\phi; z)[\alpha x + i, \alpha y + j] \quad \text{(coherently, wrong)} \quad \text{(S4)}$$

Figure S5 displays a significant difference in results derived from Eq.(S3) and Eq.(S4) for the "PartyTug 6:00AM" scene. The term "coherent integration" here refers to scenarios where the hologram is optimized using coherent integration, while the simulated reconstruction employs incoherent integration to accurately and physically mirror experimental behavior. The results of coherent integration exhibit pronounced high-order artifacts and diminished contrast due to the mismatch with reality. For both results, when the simulation utilizes incoherent integration, a precise correspondence between the simulation and experiment is observed, which validates the accuracy of Eq. (S3).
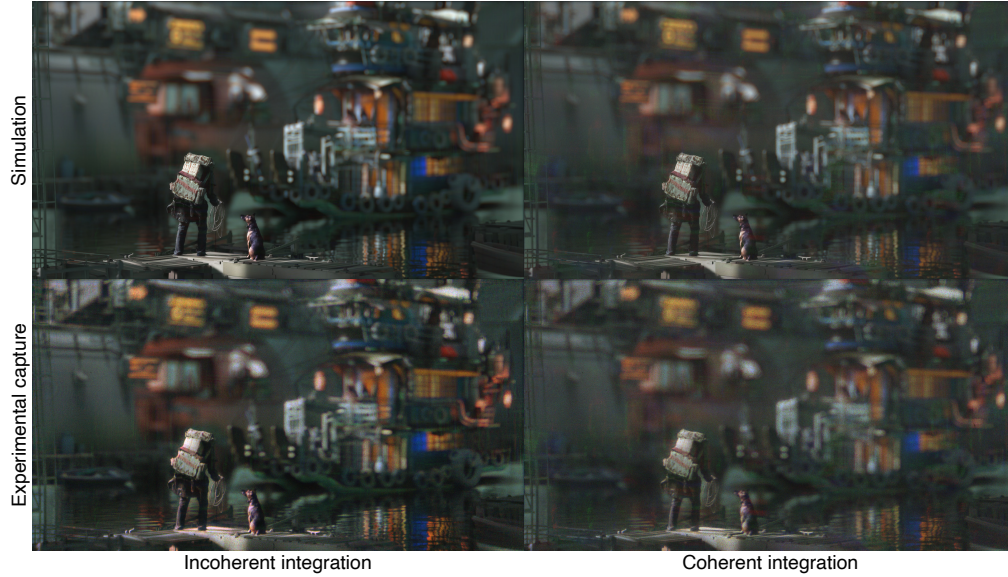
Fig. S5. Comparison of simulation and experimental results obtained using coherent integration and incoherent integration.

5. EVALUATION OF NUMBER OF TIME-MULTIPLEXED FRAMES FOR SPECKLE REDUCTION AND INCOHERENT DEFOCUS REPRODUCTION
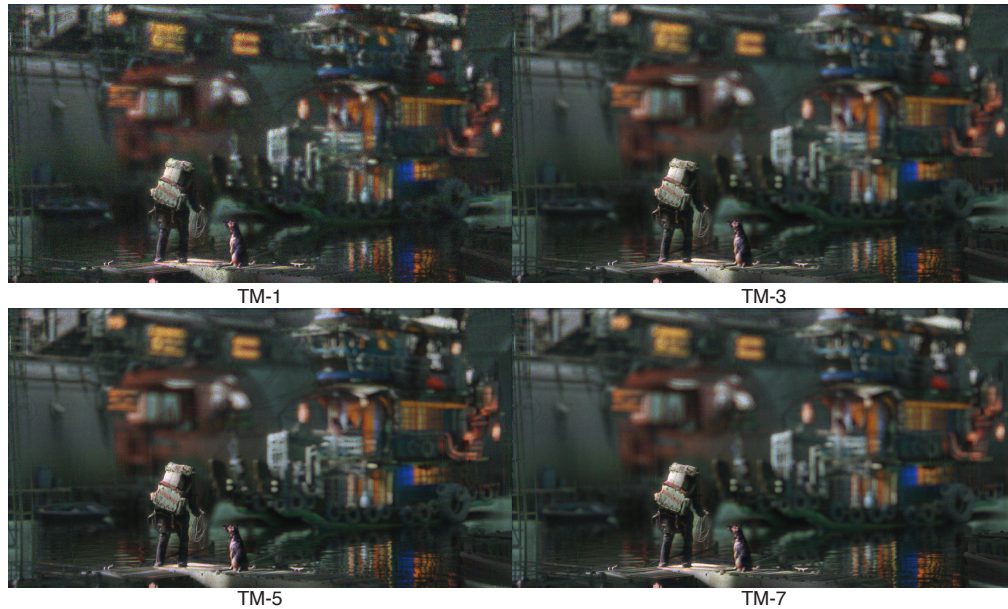

Fig. S6. Comparison of experimentally captured EC-H holograms using different numbers of time-multiplexed frames.

Figure S6 demonstrates the influence of the number of time-multiplexed frames on speckle reduction and incoherent defocus reproduction. We refer readers to Lee et al. [5] for a detailed speckle analysis. In brief, the speckle contrast is inversely proportional to the square root of the number of time-multiplexed frames. In this example, the gain visually diminishes after adding more than 5 subframes. The number also dictates the quality of incoherent defocus reproduction since the image formation for every subframe is yet governed by the coherent imaging model, and the model mismatch has to be bridged by the incoherent accumulation across subframes. Note

that we use 8-bit phase SLM compared to the 4-bit phase SLM used in Choi et al. [6], potentially reducing the number of subframes needed to achieve the same noise level and increasing the contrast of the experimental results due to better reproduction of dark pixels.

6. PUPIL SAMPLING AND SOFTWARE IMPLEMENTATION DETAILS

We implement ray-tracing for incoherent focal stack rendering in Taichi programming language (version 1.4), which offers Python-style syntax for parallel computing. For a full HD resolution LDI with 5 layers, it takes 2.2 seconds to render one depth of field image on an NVIDIA RTX A6000 GPU. EC-H optimization is implemented in PyTorch 1.13. For CGH optimization, we use a propagation distance of 60mm for both unfiltered and pupil-mimicking results with a 6mm wide 3D volume. To handle the large memory cost of optimizing at 3× spatial resolution (5760×3240 px) for multiple holograms at multiple focal planes and pupils (e.g., naive implementation takes 100+GB memory for 3 holograms at 6 depths planes and 9 pupils), we implement gradient accumulation [2]

Fig. S7. Visualization of fixed pupils and random pupils used in pupil optimization. The fixed pupils are underlaid in faint orange in the visualization of random pupils.

to split optimization into mini-batches such that the forward/backward evaluation of each batch fits in the GPU memory. For optimizing RGB hologram with 6 depth planes, one iteration takes ∼1.5 seconds on an NVIDIA A6000 GPU. In pupil-mimicking mode, the fixed pupils form a uniform grid over the entire eye box, while the random pupils sample other pupil positions and sizes within the lattice of the grid. We graphically visualize the mixture of fixed pupils and random pupils in Fig. S7 to facilitate the understanding of this scheme. For all presented results, we run 8000 iterations for unfiltered results and 2000 iterations for pupil-optimization results, after which the losses typically plateau.

Because pupil size variation causes a difference in the amount of energy inlet in the Fourier domain, we add an optional amplitude normalization factor to account for the potential large variations in the resulting image. Specifically, we let the base pupil radius as the reference pupil radius r_{ref}, for which we assume the reconstruction amplitude should absolutely match the target amplitude. We then normalize the amplitude obtained from other pupil sizes by $s_p = r_p/r_{ref}$, assuming the image intensity is ideally proportional to the area of the pupil (amplitude proportional to the radius). In practice, this ideal scale is impossible to achieve since it's impossible to distribute the energy evenly in the Fourier domain; therefore, its usage is rather optional, but we find it helps stabilize the convergence.

7. ADDITIONAL EXPERIMENTAL UNFILTERED RESULTS

We present additional experimental results in the unfiltered setup, showcasing a diverse set of target images (see Fig.S8 and Fig.S9). In comparison to TM-NH, EC-H consistently generates higher contrast images by effectively concealing the energy and interference caused by high-order diffractions, the artifact from which is particularly evident in slowly-varying color regions, as observed in the background of the "Tree Creatures" image (Fig.S8, top) and the "Wander" image (Fig.S9, bottom). Compared to HOGD, EC-H produces incoherent natural defocus without ringing or fringe artifacts. The extent of blur is also significantly larger than the unsupervised responses produced by HOGD. The hologram phases and exemplary ground truth focal stacks can be found in Fig. S16.
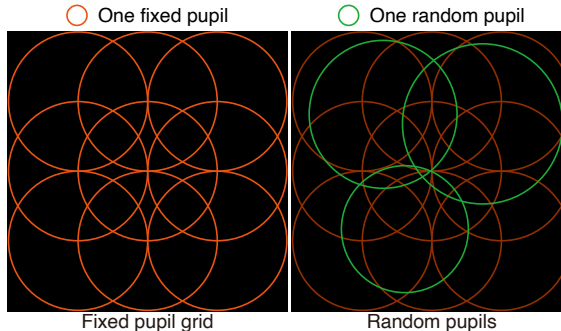
---

[2]https://pytorch-lightning.readthedocs.io/en/stable/advanced/training_tricks.html
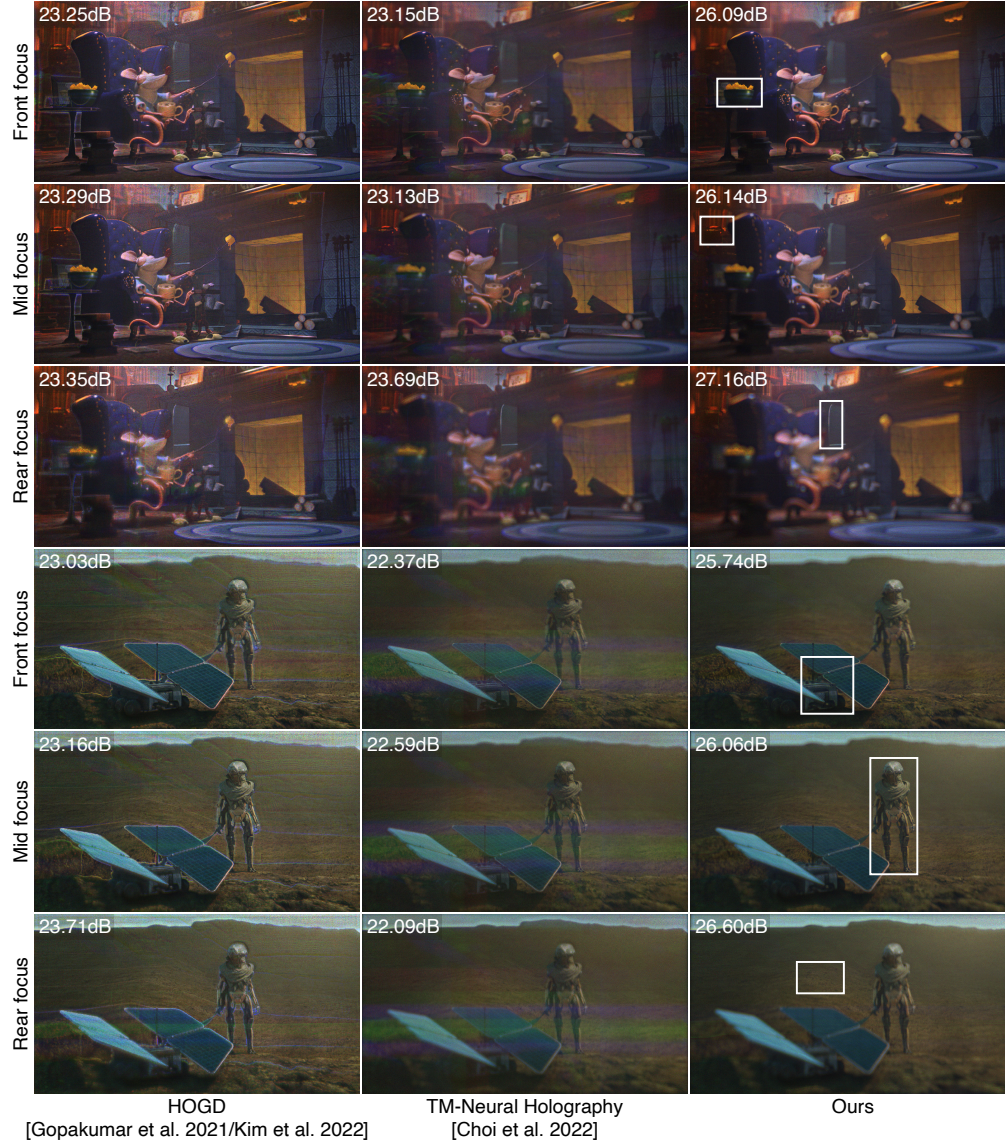
Fig. S8. Additional experimental results captured in the unfiltered mode. The captured results are presented in 900 DPI resolution for close-up examination. Source image: "'Tree Creature" by Daniel Bystedt, and "Racing Car" by Pokedstudio. The "Tree Creature" shows recording planes 9, 17, 28, the "Racing Car" shows recording planes 5, 15, 27.

Fig. S9. Additional experimental results captured in the unfiltered mode. The captured results are presented in 900 DPI resolution for close-up examination. Source image: "The Junk Shop" by Alex Treviño/ original concept by Anaïs Maamar, and "Wanderer" by Daniel Bystedt. The "The Junk Shop" shows recording planes 4, 15, 29, the "Wanderer Car" shows recording planes 4, 10, 15.

8. ADDITIONAL EXPERIMENTAL PUPIL-MIMICKING RESULTS

We present additional experimental results in the pupil-mimicking setup, where images are captured with varying pupil positions and sizes for a diverse set of target images (refer to Fig. S10 and Fig. S11). We diversify the pupil positions by offsetting them into different quadrants or along different axes, using both positive and negative distances. We observe that the artifacts associated with each baseline method, as analyzed in the main paper text, generally persist across examples. For all scenes, the image intensity produced by EC-H uniformly and gracefully decays as the pupil size reduces and the location moves away from the eye box center. A study of the results in the Fourier domain can be found in the following section for the Fig. S10 top example.
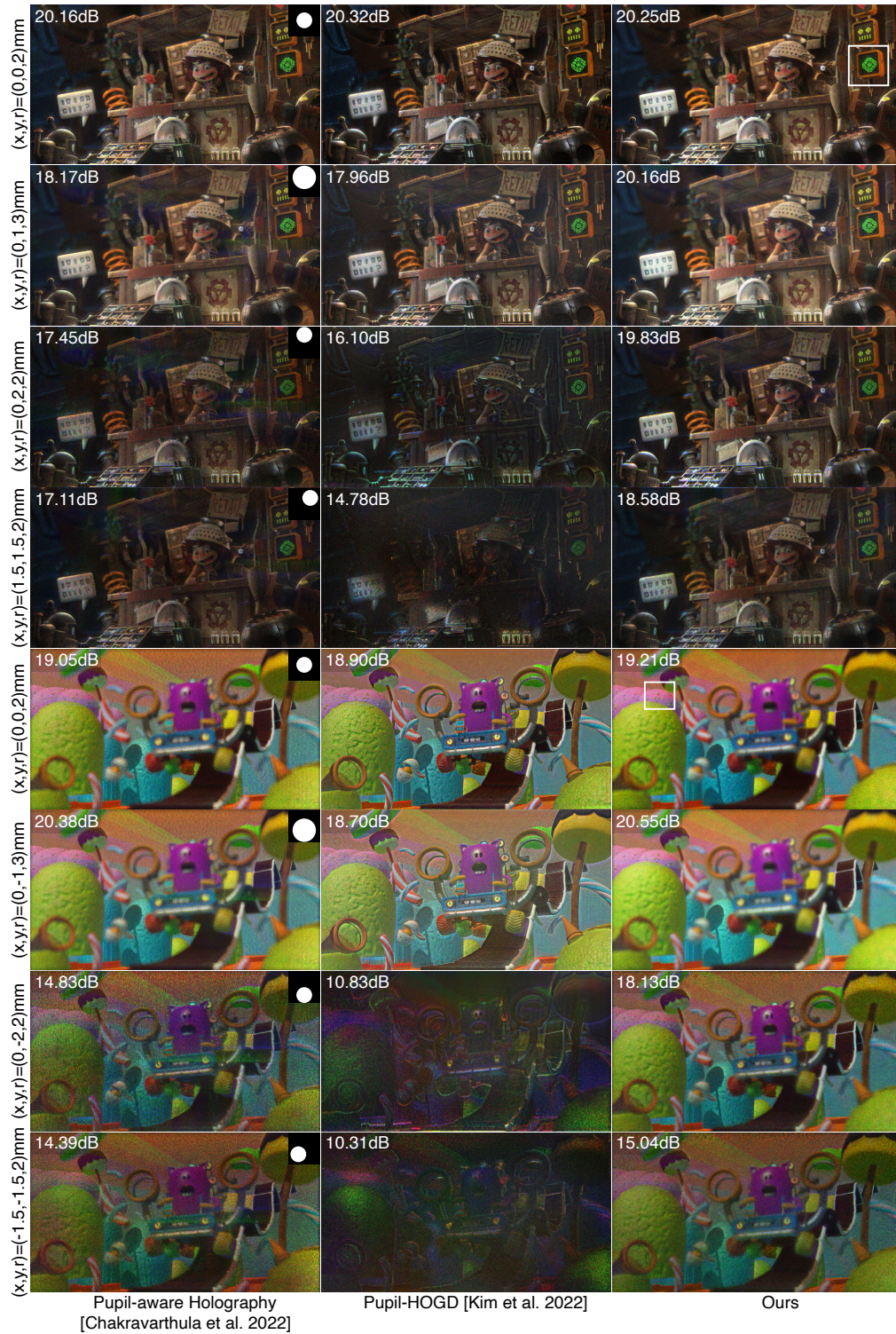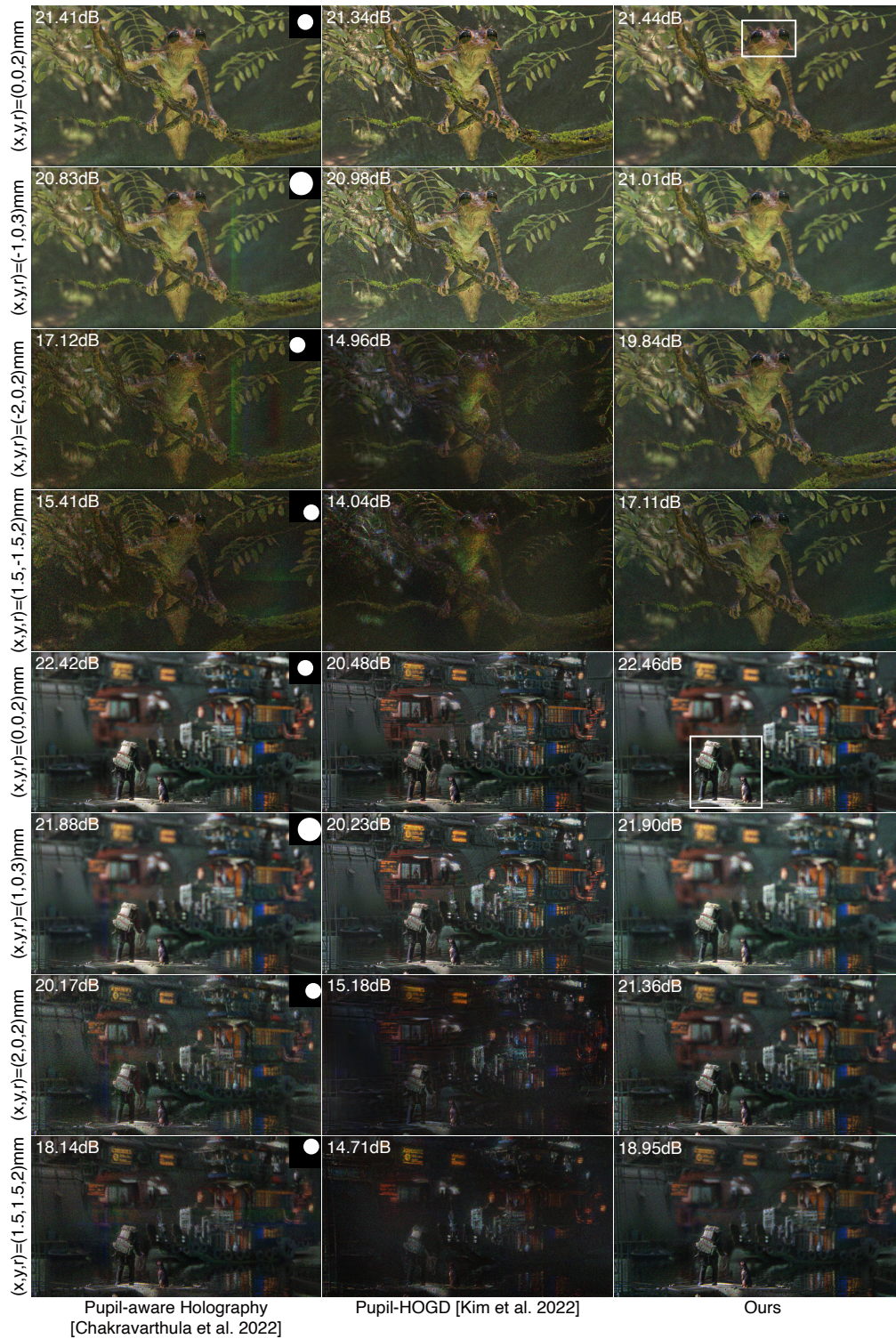
Fig. S10. Additional experimental results captured in the pupil-mimicking mode. The captured results are presented in 900 DPI resolution for close-up examination. Source image: "Mr. Elephant" by Glenn Melenhorst, and "Racing Car" by Pokedstudio.

Fig. S11. Additional experimental results captured in the pupil-mimicking mode. The captured results are presented in 900 DPI resolution for close-up examination. Source image: "Tree Creature" by Daniel Bystedt, and "PartyTug 6:00AM" by Ian Hubert.

9. EVALUATION OF ENERGY DISTRIBUTION IN THE FOURIER DOMAIN
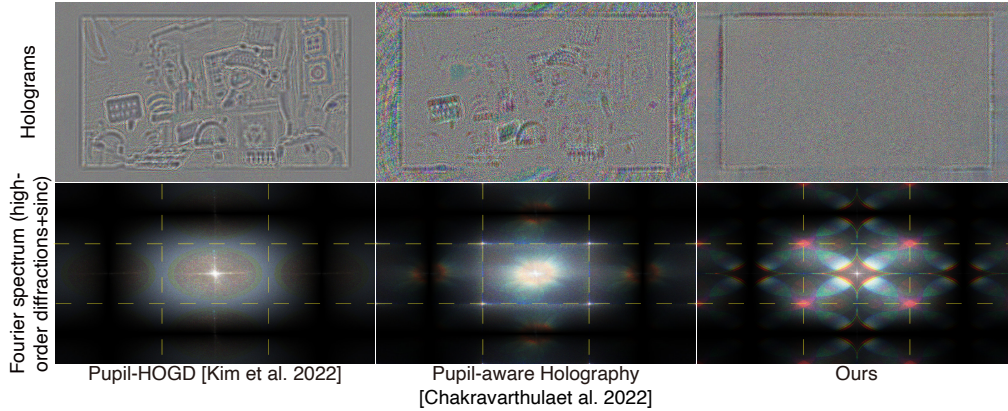

Fig. S12. Comparison of phase-only holograms and their energy distributions in the frequency domain. The borders between the orders are indicated by the semi-transparent yellow dashed lines. Note that the frequency spectrum is normalized for visualization based on its own value distribution. Consequently, identical pixel values in different images do not equal the same absolute value.

In the pupil-mimicking mode, EC-H uses a fixed pupil sampling grid to structurally spread the energy spectrum away from low-frequency regions to effectively use the entire eye box. EC-H also includes per-iteration dynamic pupils to ensure high-quality reconstruction at other pupil locations within the lattice of the fixed pupil grid. In Fig. S12, the effect of this mixed pupil optimization can be readily seen when compared with the prior state-of-the-art methods in the Fourier domain. For pupil-HOGD (column 1), a significantly greater portion of the energy is concentrated in the low-frequency band near the DC, forming an extended bright "+" shape. By modeling a dynamic pupil (column 2), the concentrated energy is noticeably disseminated toward the mid-frequency, forming a bright oval around the center. Pupil-shaped patterns can also be vaguely seen, which match the spectrum visualization in the original pupil-aware holography paper. However, because of the absence of high-order diffractions modeling, bright spots and contour lines are formed at the intersection and borders between different orders. They contribute to the rainbow-like artifacts observed in the reconstructed images and prevents the extended use of the eye box. When changing to the proposed mixed pupil optimization with high-order diffractions modeling, the bright spots and contour lines disappear, and the energy in the low-frequency "+" further reduces. The extra energy, instead of uniformly disseminated to the mid-frequency, is more structurally distributed to both the mid and high-frequency regions, preventing significant intensity drop at the periphery of the eye box.

In addition to the comparison with prior works, we conduct an ablation study by comparing using 25 random pupils and 25 fixed pupils in Fig. S13 to more systematically analyze the benefit of using mixed pupils. It is evident that the utilization of fully randomized pupils, coupled with high-diffraction modeling, significantly mitigates the rainbow-like artifacts commonly seen in pupil-aware holography. Nonetheless, image intensity at peripheral pupils still suffers from a faster decay of intensity as well as a non-uniform decay, particularly in this example. On the other hand, when 25 fixed pupils are employed, the image intensity is more consistently maintained across the pupils, as more energies are pushed to the mid and high frequencies (i.e., center low-frequency "+" reduces from fully random to mixed to fully fixed.). However, this approach generates more noticeable rainbow-like artifacts, serving as a trade-off. Ultimately, determining the optimal configuration for pupils remains an open question and is user-dependent based on their eye sensitivity to different artifacts. Thus, we advocate for more extensive and rigorous exploration in constructing optimal pupil batches for future research.
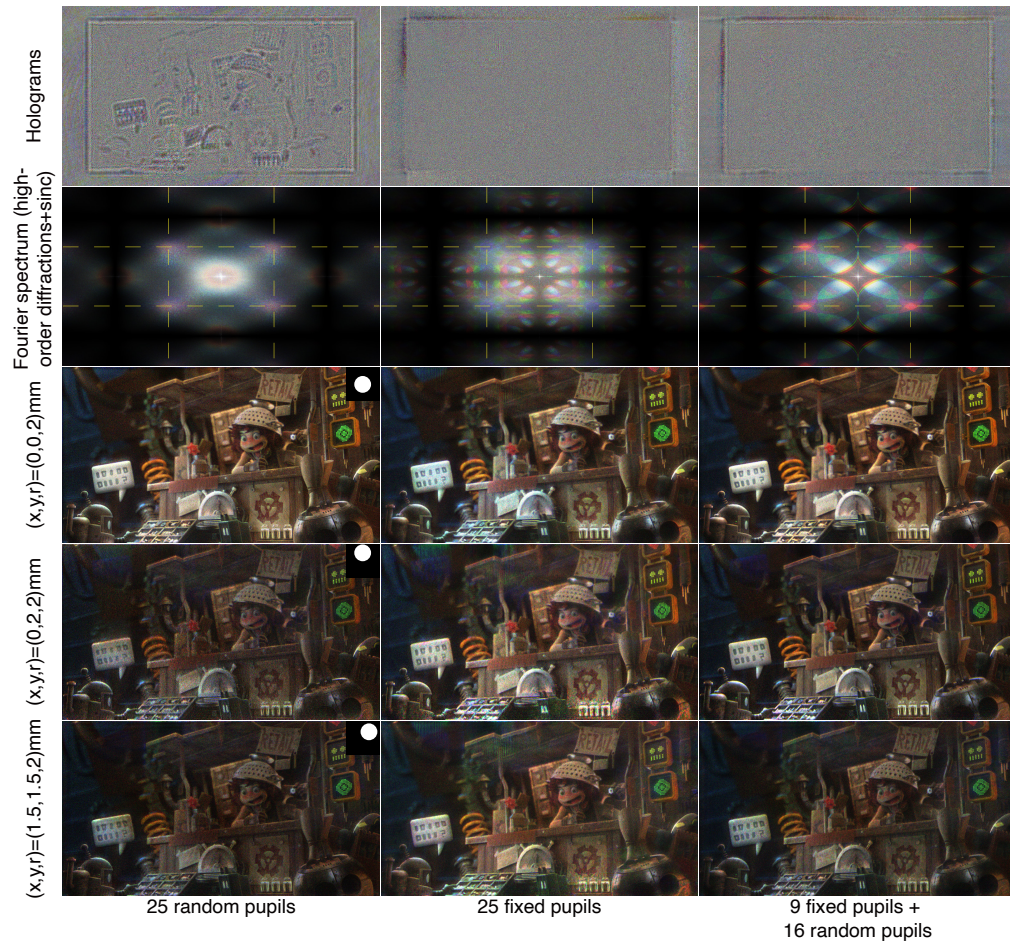
Fig. S13. Ablation study of pupil-mimicking optimization. Here, we compare mixed pupil optimization with using only random pupils or fixed pupils at the same number of total pupils.

10. REPRODUCTION OF INTRA-PUPIL MOTION PARALLAX

A unique advantage of the extended eye box enabled by the EC-H is the ability to achieve noticeable intra-pupil depth-dependent motion parallax. In Figure S15 and Figure S14, we present horizontal and vertical motion parallax visualizations based on the experimentally captured results. We want to emphasize that readers should refer to Supplement Video 3 for a much better perception of the motion parallax, as it demonstrates the alternation of different captures over time on the same canvas.

We would like to emphasize that for the most accurate representation of motion parallax, the depth of field images used to supervise the optimization process should ideally undergo a depth-dependent shift. This shift is necessary to reflect the observed motion parallax at a non-center pupil location, which occurs due to changes in the observation position. To achieve this, one can shift the camera position and perspective during the LDI rendering process, generating a grid of LDIs across the full eye box to create position-dependent focal stacks.

However, due to the current SLM specification, the motion parallax induced by the extended eye box is still tiny. Therefore, we decided to omit this additional step, while acknowledging that it deviates from the most physically accurate implementation. Nonetheless, we discovered that even when using the same focal stack to supervise the optimization at different pupil positions, the motion parallax effect can still be induced. This effect arises from the fundamental modeling of holographic image formation with pupil masking, which has been the standard technique to simulate the large motion parallax effect induced by fabricated ultra-high resolution hologram (see Zhang et al. [7] Fig.7/8). In our case, the optimizer attempts to extrapolate the content accordingly. As a result, artifacts in the upward, downward, leftward, and rightward edges can be observed due to the extrapolation process. It will be really interesting to see if image quality and realism can be further improved after closing this gap. We defer this fine-grained optimization to future work.
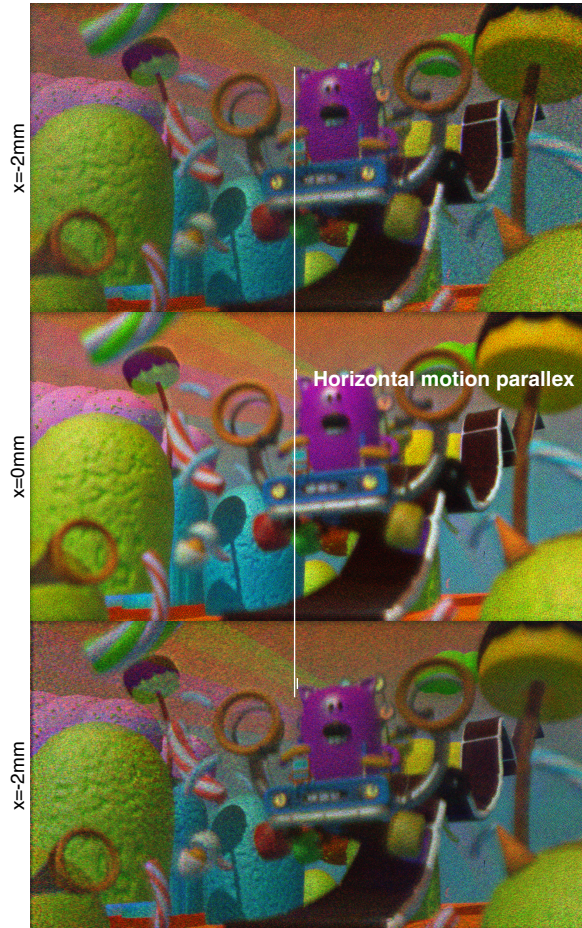

Fig. S14. Experimentally captured horizontal motion parallax for the "Racing Car". The vertical white line and small white bars help perceive the motion parallax.
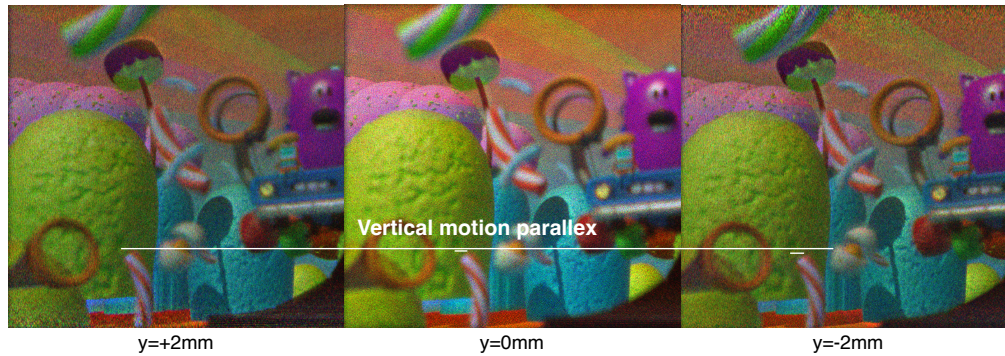
Fig. S15. Experimentally captured vertical motion parallax for the "Racing Car". The horizontal white line and small white bars help perceive the motion parallax.

11. ADDITIONAL PHASE PATTERNS AND FOCAL STACK EXAMPLES

Figure S16 shows additional phase-only holograms optimized under unfiltered and pupil-mimicking modes. Due to statistical similarity, only one subframe from all time-multiplexed frames is visualized for each scene. The visual appearance of the produced holograms undergoes significant changes when incorporating pupil optimization, as analyzed in the previous section. When optimized under the unfiltered mode, the holograms present low-frequency patterns that vaguely outline the shape of a dilated target image. This becomes invisible in holograms optimized under the pupil-mimicking mode as the design of the pupil batch structurally distributes the energy distribution towards the mid and high-frequency regions. Supplement Video 1 shows additional focal sweeps of the rendered incoherent focal stacks along with captured holograms for our and baseline methods.

Fig. S16. Visualization of exemplary focal stack images and holograms optimized for unfiltered and pupil-mimicking modes for scenes used in this paper.

12. LIMITATIONS AND DISCUSSIONS

With the proposed improvements, the EC-H still shares two limitations with HOGD [4]. First, we find that time multiplexing and better global scale initialization cannot effectively address the challenge of reproducing sparse targets. Figure S17 illustrates an example of the USAF resolution target with three depth planes. In such cases, concealing the high-order diffractions within the rest of the scene is difficult, as a significant portion of the scene remains black. Time multiplexing, which integrates sub-frames incoherently rather than coherently, does not provide additional freedom to mitigate high-order diffractions as the frame increases (see row 2). One potential solution is to incorporate a second spatial light modulator, enabling the coherent addition of the complex amplitudes of the two SLMs to cancel out the remaining high-order diffractions [8]. We also note that when optimizing under pupil-mimicking mode (see row 3), the artifacts can be largely hidden at the center pupil location since it effectively performs a low-pass filter; however, when shifted toward the periphery of the eye box, localized high-order diffractions can still be observed.
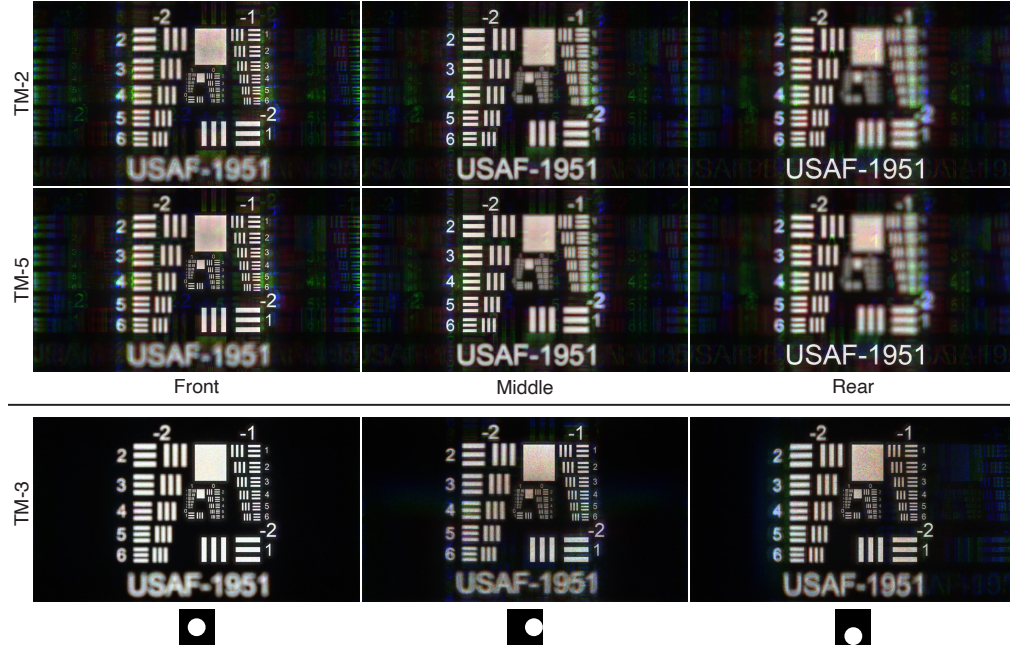
Fig. S17. Simulated reconstruction of the USAF-1951 target under unfiltered and pupil-mimicking modes.

Another limitation is the sensitivity of image quality to the propagation distance. In the absence of filtering and with the SLM plane having a unit amplitude, a certain propagation distance is necessary to diffract the field into the desired target amplitude distribution (e.g., 60mm for the results in this paper). As the propagation distance increases, high-order copies separate further away from the central order, reducing the overhead and difficulty of concealing them within the central diffraction. Figure S18 compares the "PartyTug 6:00AM" image (focused on the man) reconstructed with propagation distances of 20mm, 40mm, and 60mm. The high-order copies of the man are distinctly visible with a 20mm propagation distance, gradually diminishing as the distance extends. This limitation will naturally be resolved as the pixel pitch of the SLM becomes smaller, since a wider diffraction angle will lead to a greater separation of the high-order diffractions from the central order under the same propagation distance.
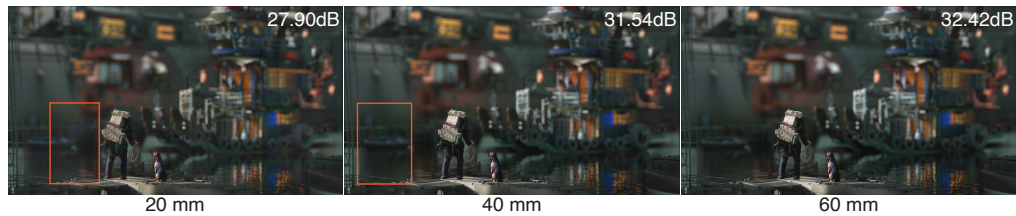

Fig. S18. Simulated reconstruction at different propagation distances with numbers indicating PSNR with respect to the target image.

For the pupil-mimicking mode, we find that achieving results with consistent minor artifacts becomes extremely challenging when further pushing the pupil away from the eye box center (e.g., $(x, y, r = \pm 2, \pm 2, 2)$mm, even when including them in the fixed pupil grid as visualized in Fig. S19, where we show the reconstructions at extreme corners of the four quadrants. Note that while one of the results only has minor artifacts, artifacts in the other three results have become prominently noticeable. This presents a limit on how much we can extend the eye box formed by the central order diffraction using high-order diffractions. As discussed in the main paper, including higher-resolution phase masks during joint optimization may ideally improve the display etendue.

Fig. S19. Simulated reconstructions at pupils shifted far away from the eye box center.

Finally, we want to highlight that the utilization of the camera (hardware)-in-the-loop (CILT) technique has gained popularity in enhancing the quality of experimental results [9]. The proposed mixed-batch pupil optimization method remains compatible with such optimization approaches, as users can leverage a motorized translation stage and a motorized iris to simultaneously control the position and size of the pupil. By accumulating and averaging the gradients computed for each pupil setting, CILT can encompass the contribution of image discrepancy across all pupil positions and sizes.

REFERENCES

1. J. Kim, M. Gopakumar, S. Choi, Y. Peng, W. Lopes, and G. Wetzstein, "Holographic glasses for virtual reality," in *ACM SIGGRAPH 2022 Conference Proceedings,* (Association for Computing Machinery, New York, NY, USA, 2022), no. Article 33 in SIGGRAPH '22, pp. 1–9.
2. J. Shade, S. Gortler, L.-W. He, and R. Szeliski, "Layered depth images," in *Proceedings of the 25th annual conference on Computer graphics and interactive techniques,* (Association for Computing Machinery, New York, NY, USA, 1998), SIGGRAPH '98, pp. 231–242.
3. C. Everitt, "Interactive order-independent transparency," gamedevs.org (2001).
4. M. Gopakumar, J. Kim, S. Choi, Y. Peng, and G. Wetzstein, "Unfiltered holography: optimizing high diffraction orders without optical filtering for compact holographic displays," Opt. Lett. **46**, 5822–5825 (2021).
5. B. Lee, D. Kim, S. Lee, C. Chen, and B. Lee, "High-contrast, speckle-free, true 3D holography via binary CGH optimization," Sci. Rep. **12**, 2811 (2022).
6. S. Choi, M. Gopakumar, Y. Peng, J. Kim, M. O'Toole, and G. Wetzstein, "Time-multiplexed neural holography: A flexible framework for holographic near-eye displays with fast heavily-quantized spatial light modulators," in *ACM SIGGRAPH 2022 Conference Proceedings,* (Association for Computing Machinery, New York, NY, USA, 2022), no. Article 32 in SIGGRAPH '22, pp. 1–9.
7. H. Zhang, L. Cao, and G. Jin, "Computer-generated hologram with occlusion effect using layer-based processing," Appl. Opt. **56**, F138–F143 (2017).
8. S. Choi, J. Kim, Y. Peng, and G. Wetzstein, "Optimizing image quality for holographic near-eye displays with michelson holography," Optica **8**, 143–146 (2021).
9. Y. Peng, S. Choi, N. Padmanaban, and G. Wetzstein, "Neural holography with camera-in-the-loop training," ACM Trans. Graph. **39**, 1–14 (2020).