

Holospeed: High-Speed Holographic Displays for Dynamic Content

Dorian Chan, Oliver Cossairt, Nathan Matsuda, and Grace Kuo

Abstract—Holographic displays are plagued by speckle — noise-like artifacts caused by the coherent interference of laser light. To mitigate this challenge, state-of-the-art systems use time multiplexing on fast spatial light modulators (SLMs) to effectively temporally smooth out these effects. In our work, we observe that such an approach struggles in practice in the context of dynamic content, manifesting motion blur and stroboscopic artifacts thanks to a fundamental mismatch between expected and displayed motion. To tackle this challenge, we propose a paradigm of holographic high-speed display, where we use the underlying fast SLM to reproduce target content that changes at the same framerate. Approaches built using this paradigm mitigate motion blur and strobing, and simultaneously minimize speckle and maximize contrast with the right loss functions. We demonstrate such a methodology in both simulation and a real system.

Index Terms—Computational Display, Holography, Near-Eye Display, VR/AR



1 INTRODUCTION

H OLOGRAPHIC displays have seen significant attention in recent research. By illuminating a spatial light modulator (SLM) with coherent laser light, these devices are theoretically capable of providing accommodation cues and glasses-free vision correction at high resolution. Additionally, their potential compact form factor and light efficiency make them an attractive choice for future augmented/virtual reality (AR/VR) systems.

However, holographic displays are not without their challenges. Chief among them is *speckle* — unavoidable noise-like artifacts that dramatically reduce the perceived quality of displayed 3D content. To this date, a number of approaches have been proposed to tackle speckle, ranging from algorithmic adjustments [1], [2], [3], [4] to more sophisticated hardware systems [5], [6], [7], [8].

One particularly simple and effective solution is *time multiplexing*. In short, if the utilized SLM has a sufficiently high framerate, multiple patterns depicting the same scene can be displayed during the human eye’s persistence-of-vision time. These patterns are effectively averaged together by the brain, minimizing speckle effects. To further improve output quality, the displayed frames can be *jointly* optimized as a whole, resulting in (i) improved contrast and (ii) further despeckling. With the emergence of fast SLMs based on digital-micromirror devices (DMDs), most state-of-the-art holographic displays [9], [10], [11], [12], [13] now use such jointly-optimized time multiplexing for despeckling.

However, this past research has focused on static scenes, and inadequately explored *dynamic content*. Such a case is a must in AR/VR, ranging from simple videos that are streamed to the user, to world-locked virtual content that moves as a user moves. In our work, we demonstrate that users in practice may experience various visual artifacts

with dynamic scenes on time-multiplexed holographic displays, once they are integrated into AR/VR systems.

For example, *motion blur* is a major challenge in time-multiplexed displays. As we’ll discuss, time multiplexing behaves like a *high-persistence display* which creates blur in the presence of eye motion. For example, as the eye rotates to track a moving object, the display continues to multiplex the “moving” object at its original location, resulting in smearing on the retina that visually appears as blur.

To avoid such problems, most modern commercial AR/VR architectures rely on low-persistence displays where displayed content is merely flashed on for a short duration (about 10% of the frame time) every frame [14]. While an effective solution for traditional displays, this modality poses a dilemma for holographic displays. For a given high-framerate SLM, a shorter persistence time equates to fewer frames for temporal multiplexing, resulting in poor image quality but reduced motion blur. Using a longer persistence time would permit more frames for time-multiplexing, resulting in diminished speckle but increased blur. If the number of temporally multiplexed frames is maximized, as described in prior state-of-the-art work on holographic displays, then the user will see about 10× more motion blur compared to commercial VR displays.

In addition, we’ll show that holographic displays, like most display architectures, are susceptible to *stroboscopic* artifacts. Strobing describes when phantom copies of moving objects appear, which can happen if object motion does not match eye motion. Such effects are distracting and nauseating for current AR/VR headset users [14], [15].

In this work, we explore how to effectively despeckle dynamic content using a fast SLM without inducing motion blur or strobing. In particular, we treat a holographic system with a fast SLM as a *high-speed display*, that is capable of displaying unique, albeit speckled frames at extremely high rates — rather than a fast display that merely replicates the same slow image in traditional time multiplexing. Under this paradigm, with priors on eye motion and principled

- D. Chan is with the Computer Science Department at Carnegie Mellon University.
- D. Chan, O. Cossairt, N. Matsuda and G. Kuo are with Reality Labs Research, Meta.

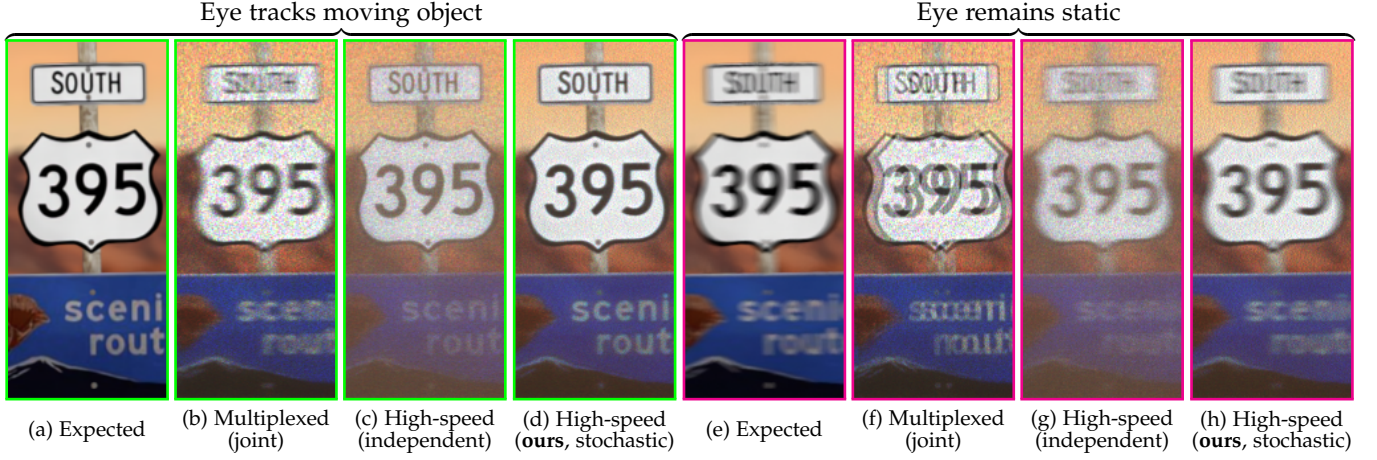


Fig. 1. **Holographic time-multiplexing and motion (simulation).** When observing dynamic content, holographic time-multiplexed displays manifest distracting visual artifacts. For example, the human eye often instinctively tracks moving objects such that they remain visually sharp (a). However, traditional time-multiplexing approaches for despeckling computer-generated holograms result in motion blur, as well as significant speckle (b). Conversely, if the eye remains static, we expect a motion-blurred image (e). However, traditional time-multiplexing produces distracting stroboscopic effects (f). Using a different paradigm of *high-speed* display mitigates both the blur as well as the strobing — however, naive application still results in noticeable speckle and loss of contrast (c), (g). By properly accounting for persistence-of-vision and eye motion, contrast can be preserved and speckle mitigated (d), (h).

loss functions for hologram optimization, we can use such a display to reproduce visually correct output without motion blur or strobing, while preserving the improved contrast and despeckling of traditional joint multiplexing. Given that 1800 Hz displays may be required for lifelike experiences [16], we believe that fundamentally displaying faster content will reap benefits for visual realism.

Before we begin, we note that our work is primarily a theoretical one in which we rely on computational models of human perception to evaluate the prominence of spatiotemporal artifacts, like motion blur and strobing. Full validation of our approach would require a user study, ideally with wide field-of-view, world-locked content in a head-mounted display, similar to AR/VR scenarios. Unfortunately, while these are areas of active research [17], [18], [19], [20], [21], [22], current holographic display prototypes lack the necessary computational speed, form factor, and étendue to support meaningful application under these conditions. However, motion blur and stroboscopic effects are well understood in the context of traditional AR/VR displays [14], validating our computational models of perception and demonstrating the importance of this research area.

Given that the ultimate goal of holographic display research is to show virtual content indistinguishable from the real world, we hope our findings will help inform future holographic architectures to be designed with motion artifacts in mind. In summary, our contributions are as follows:

- We show that standard time multiplexing produces motion blur and strobing artifacts in dynamic content.
- We propose a paradigm of high-speed holographic display to tackle motion blur and strobing effects.
- We propose novel loss functions for high-speed display that properly account for persistence-of-vision and eye motion, preserving contrast and minimizing speckle.
- We explore how such a framework can be applied to real AR/VR systems while minimizing latency.
- We validate our methodologies in both simulation and a real setup on focal stack videos.

2 RELATED WORK

Speckle is a fundamental challenge for holographic displays because coherent laser light is used to replicate incoherent human-facing content. To mitigate this speckle, one line of work proposes using partially coherent light sources in holographic displays, such as LEDs [6], [7], [23], laser grids [5], [24], [25] and comb lasers [8]. While effective, such methodologies introduce extra hardware complexity and cost. On the algorithmic end, one class of techniques attempts to reduce speckle by producing smooth phase holograms [1], [2], [20], [26], [27], such that image points always constructively interfere. However, such techniques produce unrealistic blur that does not drive accommodation [13]. Other approaches attempt to hide speckle noise according to the spatial sensitivity of the eye [3], [4], but such approaches may require careful retinal modeling.

More relevant to this work, many approaches leverage fast SLMs and time multiplexing to minimize speckle. In general, most past work has focused on producing differently speckled versions of the target image [9], [10], [11], [12], [13], [28], [29], [30]. Another line of work proposes splitting an image into disjoint sets of sparse points, each of which is independently displayed by a separate SLM pattern [31], [32]. However, no past research has explored the motion artifacts that result from time multiplexing, and how the underlying fast SLM could address them. Outside of displays, a number of systems in the sciences have used fast SLMs and computer-generated holography to create high-speed content for various applications, such as structured light [33], beam forming [34], [35], [36], optical tweezers [37], [38], microscopy [39], nanomanufacturing [40], and aberration correction [41]. We believe we are the first to apply this paradigm to human-facing holographic displays.

Outside of holography, a number of works have leveraged other traditionally time-multiplexed systems as dynamic high-speed displays. In one line of research, fast DMDs are used to help replicate three-dimensional content in the context of light-field [42], [43], [44] or multifocal

displays [45]. Other works rely on such systems for fast active vision [46], [47], or to help embed hidden information into projected content [46], [48], [49]. We are not aware of any past work that leverages these fast SLMs to mitigate the motion artifacts of time multiplexing.

3 TIME-MULTIPLEXED HOLOGRAPHIC DISPLAYS AND VISUAL QUALITY

Time-multiplexed displays assume that sequentially-displayed SLM frames will be averaged together by the brain into a single frame, thanks to persistence-of-vision. Let N_{TM} be the number of SLM frames available to be displayed for some target frame I_{target} , and τ_{SLM} the SLM framerate; without loss of generality, we assume that the duration of persistence-of-vision matches N_{TM} . Then, after N_{TM} frames are displayed, the perceived image can be expressed as:

$$\mathcal{P}_{\text{TM}}\{\phi\} = \sum_{i=0}^{N_{\text{TM}}} \mathcal{M}\{\phi[i]\} \quad (1)$$

where $\phi[i]$ denotes the i th SLM pattern and $\mathcal{M}\{\cdot\}$ the holographic forward model, *e.g.*, the model from Choi *et al.* [9] (see supplement for our model). For simplicity, let \mathcal{M} output a focal stack.

Then, to display some desired focal stack I_{target} , a standard approach involves separately optimizing each individual SLM pattern to best reproduce the target focal stack using random instantiations (we omit the scaling factor required to normalize $\mathcal{M}\{\cdot\}$):

$$\forall i, \min_{\phi[i]} \|\mathcal{M}\{\phi[i]\} - I_{\text{target}}\|^2. \quad (2)$$

While simple to implement and easily parallelizable [28], such “independent” time multiplexing has two key limitations. First, solving Equation (2) often results in biased outputs, *i.e.*, multiple random instantiations do not, in expectation, result in output matching I_{target} . An illustrative scenario is dark regions with zero brightness in the target scene — every $\mathcal{M}\{\phi[i]\}$ must produce a non-negative brightness for these regions, and thus the cumulative brightness will typically be greater than zero. As a result, Equation (2) typically results in reduced contrast, decreasing visual quality. Furthermore, solving this expression produces roughly uncorrelated speckle between different frames. While effective for despeckling over many patterns, convergence can be poor under fewer frames.

To mitigate these effects, current state-of-the-art systems leverage “joint” time multiplexing. Instead of independently optimizing each frame, these approaches attempt to match the perceived image over all displayed frames with the ground truth:

$$\min_{\phi} \|\mathcal{P}_{\text{TM}}\{\phi\} - I_{\text{target}}\|^2. \quad (3)$$

By optimizing all patterns simultaneously, speckle convergence is improved and contrast errors can be reduced, *e.g.*, a dark speckle in one frame can be compensated by a bright speckle in another frame. In practice, the recovered $\phi^*[i]$ in both Equation (2) and Equation (3) produce approximate replicas of I_{target} . Thus, we can write:

$$\mathcal{M}\{\phi^*[i]\} \approx I_{\text{target}}. \quad (4)$$

3.1 Time-multiplexing and moving content

While effective for static scenes, these techniques face inherent challenges when applied to dynamic content. For instance, under phenomena such as smooth pursuit or the vestibulo-ocular reflex (VOR), the human eye continuously rotates to track moving objects, in order to maintain a stationary image on the retina — these physiological mechanisms allow us to see fine details on moving objects (smooth pursuit), or on stationary objects as we move (VOR).

To see why this might be a problem for holographic time multiplexing, consider the case where the display aims to reproduce some real object that does not remain static, but continues to translate relative to the viewer over the next N_{TM} SLM frames. The user then rotates their eye to track this motion, as visualized in Figure 2(a). Formally, let $\mathcal{T}^{\mathbf{x}}\{\cdot\}$ denote a 2D translation by $\mathbf{x} = (x, y)$. Then, the real scene over time will be given by $\mathcal{T}^{\mathbf{x}_s(t)}\{I_{\text{target}}\}$, where $\mathbf{x}_s(t)$ represents the shift from object motion at time t . Modeling eye rotation as a 2D translation of the perceived image, if the user was actually viewing the real target the eye would perceive:

$$I_{\text{real}} = \int_t \mathcal{T}^{\mathbf{x}_e(t)} \left\{ \mathcal{T}^{\mathbf{x}_s(t)} \{I_{\text{target}}\} \right\} dt, \quad (5)$$

where $\mathbf{x}_e(t)$ is the shift imposed by the eye. Assuming that the eye perfectly tracks the scene, $\mathbf{x}_e = -\mathbf{x}_s$ and therefore $\mathcal{T}^{\mathbf{x}_e(t)}$ and $\mathcal{T}^{\mathbf{x}_s(t)}$ cancel out, leaving:

$$I_{\text{real}} \approx I_{\text{target}}. \quad (6)$$

In short, a user will ideally see a sharp image of the target despite scene motion, as the eye moves to compensate for it.

However, a time-multiplexed holographic display is agnostic to this motion, and will continue to display the moving object at its original location $\mathcal{M}\{\phi^*[i]\} \approx I_{\text{target}}$ (Equation (4)), even as the eye rotates. Thus, the user would actually perceive on such a system:

$$I_{\text{perceived}} \approx \sum_{i=0}^{N_{\text{TM}}} \mathcal{T}^{\mathbf{x}_e(t_i)} \{ \mathcal{M}\{\phi^*[i]\} \} \approx B * I_{\text{target}}, \quad (7)$$

where $t_i = \tau_{\text{SLM}}i$, and B denotes some blur kernel corresponding to the motion of the eye. In other words, a time-multiplexed holographic display will result in a *motion-blurred* version of the same image.

In incoherent displays, such a phenomenon is called *sample-and-hold* blur, and is a known artifact of the *persistence* of a display. In short, for a given framerate, persistence refers to the proportion of the total framerate which a display holds the desired pattern. Because time-multiplexed holographic displays effectively show the same pattern over N_{TM} subframes, they behave as high-persistence displays, and therefore suffer from the same motion artifacts.

To address these challenges, incoherent AR/VR systems rely on low-persistence displays where desired content is simply flashed on for a short duration [14]. However, in holographic displays, such flashing would result in little time multiplexing, as very few SLM frames could be shown in such a short duration. More specifically, most headsets today aim for less than 2 ms of persistence [14]. On a 1 kHz SLM, this equates to two frames of holographic time-multiplexing. Clearly, new approaches are needed to display speckle-free dynamic content in holographic systems.

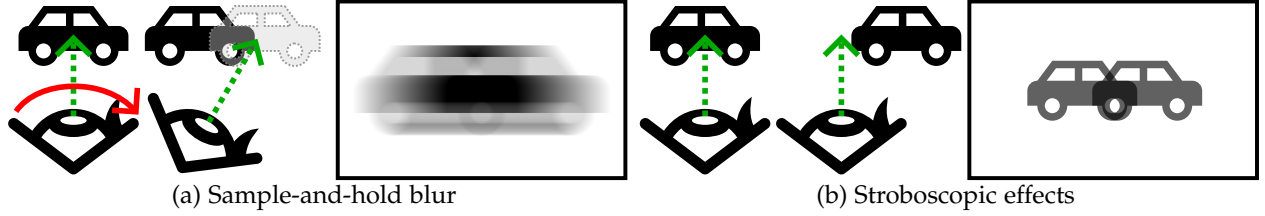


Fig. 2. **Visualizing sample-and-hold blur and stroboscopic effects.** Consider a car moving to the right, displayed on a time-multiplexed holographic display. If our eye fixates and tracks this car (a), it will continuously rotate following the expected location of the car. However, since the display continues to show time-multiplexed versions of the car at its old location, our eye will perceive a motion-blurred version with persistence-of-vision. Now, consider when the eye is fixed (b). Then, the car will be displayed at multiple discrete locations over time to our eye, creating ghosting effects. Both of these artifacts reduce visual realism, and may induce visual discomfort [14], [15].

Additionally, like any other low-speed display, such holographic displays also suffer from *stroboscopic* artifacts when eye motion does not match an object's motion. To see when this might occur, consider when the eye is static while the object translates as before (Figure 2(b)). In this case, the user should ideally perceive a motion-blurred image:

$$I_{\text{real}} = \int_t \mathcal{T}^{\mathbf{x}_s(t)} \{I_{\text{target}}\} dt = B * I_{\text{target}}. \quad (8)$$

However, on a time-multiplexed holographic display, the user would instead perceive a sharp image:

$$I_{\text{perceived}} \approx \sum_{i=0}^{N_{\text{TM}}} \mathcal{M}\{\phi^*[i]\} \approx I_{\text{target}}, \quad (9)$$

introducing another mismatch between real-world visuals and actual time-multiplexed output. In the context of a full video, where every consecutive non-overlapping set of N_{TM} frames produces a different target frame, Equation (9) effectively produces a phantom array of identical sharp objects, rather than smooth motion blur. For instance, in the aforementioned case where the scene translates, the first set of N_{TM} frames would reproduce I_{target} , while the next set would reproduce the scene at its next shifted location $\mathcal{T}^{\mathbf{x}_s(t_{N_{\text{TM}}})} \{I_{\text{target}}\}$. If the displayed content is viewed between frames $0.5N_{\text{TM}}$ and $1.5N_{\text{TM}}$, a user would simultaneously observe both the first and second target frame:

$$I_{\text{perceived}} \approx 0.5 \left(I_{\text{target}} + \mathcal{T}^{\mathbf{x}_s(t_{N_{\text{TM}}})} \{I_{\text{target}}\} \right). \quad (10)$$

In AR/VR settings, such stroboscopic effects become noticeable with head motion over 5-10 degrees/second, far below the speed of a leisurely head turn [14]. This results in nausea and visual discomfort for the user [14], [15]. Along with sample-and-hold blur, such artifacts may limit visual realism in real-world holographic displays.

4 HIGH-SPEED HOLOGRAPHIC DISPLAYS

To tackle these unique challenges of dynamic scenes, we observe that such blurring and strobing artifacts are fundamentally due to the low framerate of the underlying target content. Intuitively, if we had a perfect display with infinite framerate, we would be able to directly reproduce any desired real-world scene, and therefore avoid any such undesirable effects. Thus, in the context of time-multiplexed holographic displays, we propose a paradigm of *high-speed* display. In short, the SLMs used for holographic temporal multiplexing actually display patterns at thousands of hertz.

Thus, a holographic display need not solely reproduce differently speckled versions of the same low-speed frame with each underlying high-speed SLM pattern — we can reproduce shifted versions, blurred versions, or altogether unique content with every SLM pattern, as we fundamentally have a modulator that produces high-speed frames.

A simple realization of this idea is to directly optimize each $\phi[i]$ to reproduce the desired high-speed scene $V_{\text{HS}}(\cdot)$ at corresponding time t_i :

$$\forall i, \min_{\phi[i]} \|\mathcal{M}\{\phi[i]\} - V_{\text{HS}}(t_i)\|^2. \quad (11)$$

We term such an approach *independent high-speed display*. While effective in reducing sample-and-hold blur and stroboscopic effects, we observe in practice that analogous to Equation (2), independent high-speed display results in reduced contrast and poor speckle convergence as each frame is optimized independently of the others. How then, can we preserve contrast and improve speckle convergence, without introducing blur or strobing like Equation (3)?

To mitigate these challenges, we formally model the images perceived by a viewer under eye motion and persistence-of-vision from high-speed content:

$$\mathcal{P}\{V, \mathbf{x}_e\} = k *^{(t)} \left(T^{\mathbf{x}_e} *^{(\mathbf{x})} V \right), \quad (12)$$

where $T^{\mathbf{x}_e}(\mathbf{x}, t) = \delta(\mathbf{x} - \mathbf{x}_e(t))$, $k(t)$ represents the temporal kernel of interest thanks to persistence-of-vision, and $*^{(\mathbf{x})}$ and $*^{(t)}$ denote convolutions over the spatial and temporal dimensions, respectively. In words, the model shifts every frame according to eye motion, and then blurs the input video over time to model the temporal processing of the eye [50], [51], [52], [53].

Then, an ideal holographic display should solve the following optimization problem, matching the images perceived from the SLM with the ideal perceived images from the desired target sequence:

$$\min_{\phi} \sum_{t \in T} \|\mathcal{P}\{\mathcal{M}\{\phi\}, \mathbf{x}_e\}(t) - \mathcal{P}\{V_{\text{HS}}, \mathbf{x}_e\}(t)\|^2, \quad (13)$$

where T denotes the set of timestamps for which input content is available — we assume every τ_{step} for simplicity. We call solving this expression *motion-aware high-speed display*.

Conventional joint time-multiplexing (Equation (3)) can be viewed as a special case of the above framework. First, traditional joint multiplexing assumes that the eye remains static as it views dynamic content, e.g., $\mathbf{x}_e(t)$ is assumed to remain constant. In contrast, Equation (13) *explicitly models*

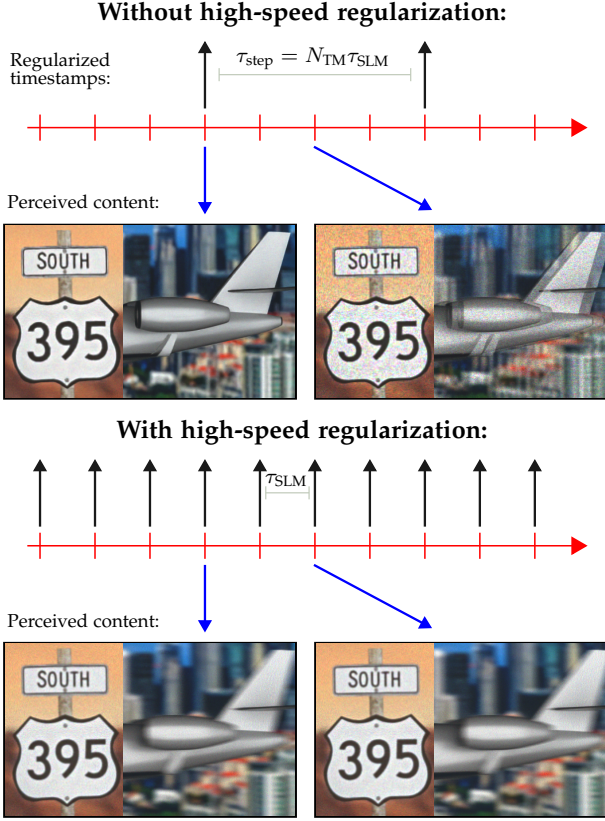


Fig. 3. **The impact of high-speed regularization.** Traditional joint time multiplexing regularizes the perceived image at only intermittent times, e.g., every $N_{TM}\tau_{SLM}$. Thus, at other “misaligned” points in time, the perceived image can behave in undesirable ways. More speckle appears (the sign), and when eye motion does not match an object’s motion, strobing artifacts appear instead of natural motion blur (the plane). When using high-speed regularization, because the perceived content is optimized more frequently every τ_{SLM} , such artifacts disappear.

eye motion on top of persistence-of-vision, avoiding any potential sample-and-hold blur while preserving the improved contrast and speckle convergence of joint time multiplexing.

Second, in traditional multiplexing, every consecutive, non-overlapping set of N_{TM} frames is used to reproduce a single desired image. In other words, it only optimizes the perceived image at intermittent intervals $\tau_{step} = N_{TM}\tau_{SLM}$. While the regularized images can be very high quality, the perceived image at other timestamps, e.g., $0.5N_{TM}\tau_{SLM}$ can again suffer from speckle as their outputs are no longer constrained. Furthermore, such an approach can be prone to stroboscopic effects as it still reproduces low-speed content. In contrast, our framework introduces a paradigm of *high-speed regularization*, where the perceived image is regularized much more frequently with high-speed content, e.g., $\tau_{step} = \tau_{SLM}$. This ensures that the perceived image has low speckle at any point in time, and that any scene motion that the eye does not follow gracefully degrades into physically correct motion blur (see Figure 3). By producing perceptually correct high-speed content, we can hopefully increase visual quality.

Without considering content generation, our framework does not, in theory, add significant computational expense over traditional multiplexing — one iteration of gradient descent on Equation (13) requires the same number of for-

ward and backward propagations through the holographic forward model as Equation (3). In practice, there is some additional computational expense from computing the perceived image at more timestamps after holographic propagation. However, given the convolutions required by coherent propagation [54] as well as the increasing complexity of modern neural modeling [2], [9], [21], such linear factors may become less significant with larger images and SLMs.

Finally, even if high-speed regularization is not used, solving Equation (13) still produces a form of high-speed display that mitigates sample-and-hold blur, as each SLM pattern essentially produces a shifted version of the target scene. Put another way, such an approach temporally up-samples input content to the framerate of the SLM by shifting it according to the expected eye motion. This inherent high-speed display is what mitigates sample-and-hold blur.

To solve Equation (13), a direct approach would simultaneously optimize all to-be-displayed SLM frames, such that, perceptually speaking, they jointly reproduce all desired content at all timestamps. However, such an approach becomes less attractive for real-time, interactive settings where output target frames are not known apriori. We discuss this aspect in further detail in Section 4.2.

4.1 Modeling distributions of eye motion

Solving Equation (13) requires a specific eye motion $\mathbf{x}_e(t)$. Although a precise eye trajectory can, in theory, be obtained using an eye tracker, in practice, it is more common to have access to a distribution of possible eye movements rather than a single deterministic path. For instance, eye tracker output is often inaccurate or delayed, and thus probabilistic modeling may be required [55]. If no eye tracking is available, a distribution of eye motions could be estimated directly from the target content if the eye is assumed to be tracking objects in the scene [56]. For example, by computing optical flow over the input video, a distribution of eye motion can be estimated by computing a histogram over the scene flow. In the case where SLM frames need to be precomputed for a desired video, such an approach may be the only option available since exact eye motion is not known apriori. Thus, handling distributions of eye motion may need required in practice to leverage an approach like Equation (13).

To address this need, we propose two approaches that account for the distribution of possible eye motions. For the short temporal window that contributes to the image perceived at time t , we assume that the trajectory of the eye $\mathbf{x}_e(t_i)$ can be linearly approximated by $\mathbf{x}_m(t_i) = \mathbf{c} + \mathbf{m} \cdot (t_i - t)$, where \mathbf{m} is drawn from a distribution with a probability density function $p_m(\mathbf{m}, t)$. Without loss of generality, we can also safely set \mathbf{c} to 0 without impacting any downstream optimization. Then, with this probabilistic motion model in hand, we first propose simply taking a weighted average of Equation (13) over all possible eye motions and minimizing:

$$\sum_{t \in T} \int_{\mathbf{m}} \|\mathcal{P}\{\mathcal{M}\{\phi\}, \mathbf{x}_m\}(t) - \mathcal{P}\{V_{HS}, \mathbf{x}_m\}(t)\|^2 p_m(\mathbf{m}, t) d\mathbf{m}. \quad (14)$$

To avoid excessive computational expense per iteration if there are many possible motions, inspired by stochastic

gradient descent, we optimize a single \mathbf{m} per iteration sampled from $p_m(\mathbf{m}, t)$. We label this approach *stochastic motion-aware holographic display*.

While an effective solution, in compute-constrained scenarios, not enough optimization iterations may be possible to sufficiently cover the space of potential motions. Thus, we propose what we call *kernel-based motion-aware holographic display* based on moving the integral over eye motions in Equation (14) inside the norm:

$$\min_{\phi} \sum_{t \in T} \|\mathcal{P}_{\text{kernel}}\{\mathcal{M}\{\phi\}\}(t) - \mathcal{P}_{\text{kernel}}\{V_{\text{HS}}\}(t)\|^2, \quad (15)$$

$$\mathcal{P}_{\text{kernel}}\{V\} = \int_{\mathbf{m}} \mathcal{P}\{V, \mathbf{x}_{\mathbf{m}}\}(t) \cdot p_m(\mathbf{m}, t) d\mathbf{m}.$$

Note that if the distribution of eye motions remains constant over time, $\mathcal{P}_{\text{kernel}}\{V\}$ can be expressed as a convolution $K * (\mathbf{x}, t) V$ for efficient computation. Intuitively, this method can be viewed as computing and minimizing the “average” speckle over all potential eye motions. In practice, this approach can produce slight haloing especially when used without high-speed regularization, as will be seen in Section 5.1.

4.2 Motion-aware high-speed displays on-the-fly

In reality, dynamic content is generated on-the-fly in many significant applications like AR/VR and teleconferencing. In an ideal implementation of such systems, whenever target frames arrive at the display, they should be shown as soon as possible to minimize latency. How can we replicate such a modality given the previously-described framework?

Without high-speed regularization, an input target arrives every $\tau_{\text{step}} = N_{\text{TM}}\tau_{\text{SLM}}$. Let τ_l be the maximum allowed latency — for simplicity, let $\tau_l = N_{\text{TM}}\tau_{\text{SLM}}$. Thus, for every input target frame I_{target} , we directly solve the next set of N_{TM} frames to produce the target at τ_l . For stochastic display, this is mathematically:

$$\min_{\phi} \int_{\mathbf{m}} \|\mathcal{P}\{\mathcal{M}\{\phi\}, \mathbf{x}_{\mathbf{m}}\}(\tau_l) - I_{\text{target}}\|^2 \cdot p_m(\mathbf{m}) d\mathbf{m}. \quad (16)$$

This can be viewed as a motion-aware analogue of conventional joint time-multiplexing, where all SLM frames reproducing a single macro-frame are optimized together.

High-speed regularization can be applied with a slightly different approach that minimizes latency. Consider the case where target frames arrive at the same framerate as the SLM, i.e., $\tau_{\text{step}} = \tau_{\text{SLM}}$. In this setting, as the display system steps through each SLM pattern one at a time, a new desired perceptual target is also made available on-the-fly for each individual frame. Therefore, to minimize latency, the display system must optimize the *current* SLM pattern such that the *current* image perceived from the SLM matches the *current* desired perceived image. For pattern i , the stochastic display case therefore minimizes:

$$\mathcal{L}_{\text{seq}}(\phi[i], \phi_h, V_{\text{HS}}) = \int_{\mathbf{m}} \|\mathcal{P}\{\mathcal{M}\{\phi_h \oplus \phi[i]\}, \mathbf{x}_{\mathbf{m}}\}(t_i) - \mathcal{P}\{V_{\text{HS}}, \mathbf{x}_{\mathbf{m}}\}(t_i)\|^2 \cdot p_m(\mathbf{m}, t_i) d\mathbf{m}, \quad (17)$$

where ϕ_h denotes a history of past SLM frames, and \oplus denotes concatenation. In other words, we sequentially optimize and display fast SLM patterns one at a time to match the concurrent desired content, while keeping track of previous patterns. See supplement for pseudocode.

5 EXPERIMENTS

We use a Texas Instruments DLP6750Q1EVM for our fast SLM, with 1358×800 pixels with $10.8 \mu\text{m}$ pitch and 4-bit phase modulation at a maximum framerate of 1440 Hz. Our laser is a 520 nm FISBA READYBeam. To highlight differences in speckle, color results are synthesized with this wavelength. We also use a 4F system with an aperture to block the DC component and higher-order diffraction. Output focal stacks are produced from 10 mm to 20 mm from the SLM, matching a 45° FOV where virtual content is placed from 0 to 4 diopters away from the viewer. To calibrate our real system, we use Adam [57] to learn a forward model that captures SLM and optical non-idealities. To solve the optimization problems from Section 3, we use Adam [57] with soft rounding to model SLM quantization, with 500 iterations unless otherwise noted. Please refer to the supplement for more details.

We assume a 16.66 ms persistence-of-vision time, corresponding to $N_{\text{TM}} = 24$ SLM patterns at 1440 Hz. For simplicity, we assume a rectangular kernel $k(t) = 1, 0 \leq t \leq 16.66 \text{ ms}$ in Equation (12). To test our stochastic and kernel-based approaches, we calculate Farneback optical flow [58] on the target videos, which we then histogram to estimate a distribution of scene and therefore eye motion. Please see the supplement for more discussion.

To simulate what a person would perceive in the real system, we capture focal stacks via a FLIR BFS-U3-123S6M-C sensor mounted on a translation stage. To model eye motion, we eschew mechanically rotating the camera, and instead digitally shift and sum captured images to avoid any problems with étendue.

5.1 Results

5.1.1 Methods

We compare the following methods:

- *Conventional joint time-multiplexing*. The baseline method. Every sequential non-overlapping set of N_{TM} frames is used to reproduce a different target frame (Equation (3)).
- *Independent high-speed display (ours)*. From Equation (11), every high-speed SLM frame is used to reproduce the concurrent moving target scene.
- *Motion-aware high-speed display (ours)*. We solve Equation (13) with estimates of eye motion. We test without high-speed regularization ($\tau_{\text{step}} = N_{\text{TM}}\tau_{\text{SLM}}$) and with high-speed regularization ($\tau_{\text{step}} = \tau_{\text{SLM}}$).
- *Stochastic motion-aware high-speed display (ours)*. We solve Equation (14) with an eye-motion distribution from optical flow, with and without high-speed regularization.
- *Kernel-based motion-aware high-speed display (ours)*. We solve Equation (15) with the same motion distribution from above, with and without high-speed regularization.

All motion-aware approaches are solved using the sequential methodology from Section 4.2.

To visualize the results, we take the output SLM frames and render estimated perceived images via Equation (12). To highlight stroboscopic effects, we render at timestamps halfway between regularized ones (Figure 3). The figures show both insets for qualitative evaluation as well as PSNR.

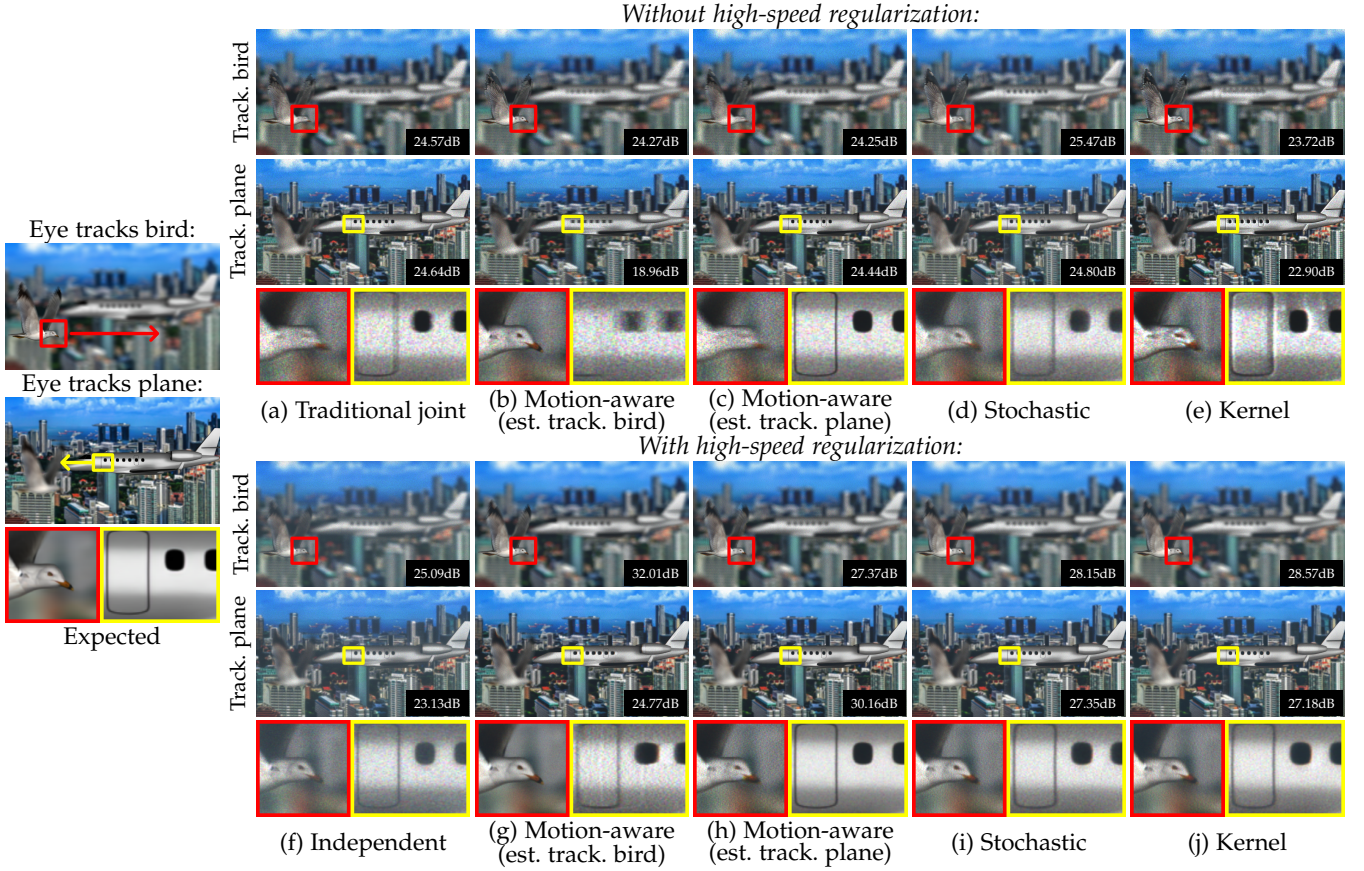


Fig. 4. **Tackling sample-and-hold blur (simulation).** A traditional time-multiplexing approach results in a blurry moving bird and plane, as well as increased speckle (a). Independent high-speed display (f) cleans up these artifacts, but results in loss of contrast while retaining speckle noise. A motion-aware approach optimized for a particular eye motion can mitigate these effects, but suffers a performance hit when the eye moves differently than expected (b), (c), (g), (h). Using models that account for uncertainty can compensate for these effects (d), (e), (i), (j). Adding high-speed regularization reduces visible speckle (g), (h), (i), (j).

5.1.2 Scenes

We built four representative scenes to test these methods:

- *Skyline.* A generic dynamic scene, where a bird flies to the right as a plane flies to the left. A high-contrast city background makes joint optimization important. Shown in Figures 4, 5 and 7.
- *Basketball.* A video meant to represent an AR/VR basketball game, where a ball flies up and to the right towards a static hoop as a blocking arm falls. Shown in Figure 6.
- *Highway.* A scenario with world-locked content, where a user turns their head relative to a road sign, concrete barrier and hilly backdrop. Shown in Figures 1 and 8.
- *Jungle.* A scene with more complex motion, where a leaping tiger and a landing bird produce more complicated, nonlinear trajectories over a forest background. Shown in the supplement.

5.1.3 Discussion

In Figures 4, 6(i) and 7(i), we examine scenarios where the user’s eye tracks moving objects. A standard joint time-multiplexing approach produces speckled, blurry images where the fine details cannot be resolved thanks to sample-and-hold blur. An independent high-speed display approach mitigates these blurring effects, but remains speckled while losing overall image contrast. Our motion-aware approach can dramatically improve quality, but performance

drops if the eye moves unexpectedly. Without high-speed regularization, artifacts again manifest in the form of motion blur, while with high-speed regularization, slight halos appear near dark regions. Our stochastic approach mitigates these challenges, balancing performance over all potential eye motions with somewhat less speckle in the high-speed regularization case. Our kernel-based method mostly behaves similarly to the stochastic approach with high-speed regularization, but can manifest sharp halo artifacts without.

In Figures 5, 6(ii), and 7(ii), we examine stroboscopic artifacts that appear in traditional systems when eye motion does not match an object’s motion. All approaches that do not use high-speed regularization result in ghosting artifacts. In contrast, using high-speed regularization instead produces correct motion blur. Our motion-aware methodologies improve the contrast and speckle of this blur over an independent approach. Figure 8 evaluates our framework with only high-speed regularization and without motion-aware optimization. Increasing the rate of regularization reduces strobing to natural motion blur as shown in the text on the highway sign, but can decrease contrast. Using the full model with motion-aware optimization produces the best results.

Figure 7 evaluates our framework in the context of decreased compute. Performance follows the same trends as the previous examples. The stochastic method is somewhat

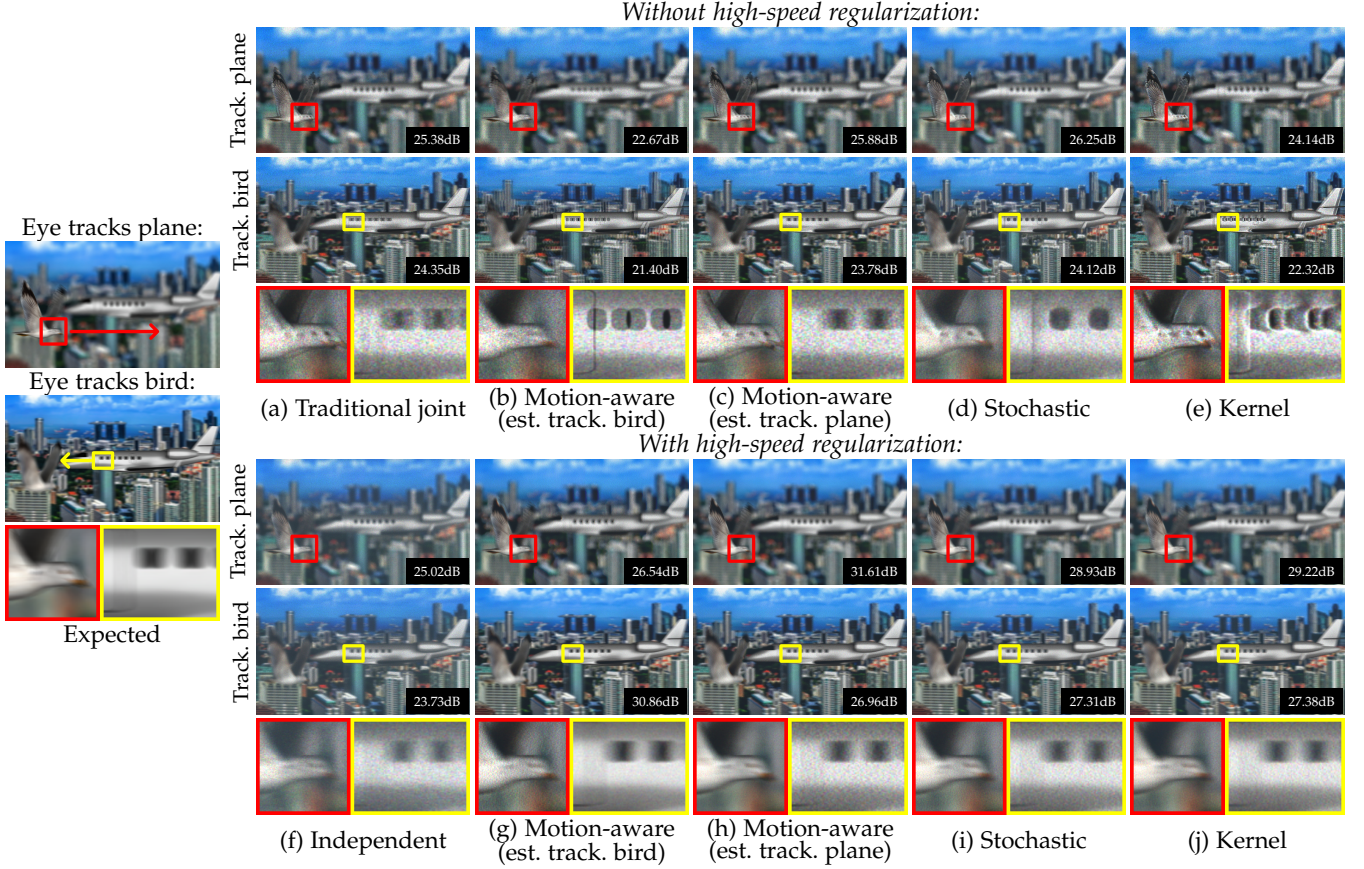


Fig. 5. **Tackling stroboscopic effects (simulation).** Traditional joint time-multiplexing results in strobing artifacts when object motion does not match eye motion, as can be seen in the repeated bird eye and plane windows (a). When used without high-speed regularization, our motion-aware framework results in similar ghosting (b), (c), (d), (e) — for instance, (b) is optimized for the eye tracking the bird, and thus results in strobing artifacts around the plane. By regularizing display outputs at high-speeds, these effects instead gracefully degrade into motion blur (f), (g), (h), (i), (j). Motion-aware optimization ensures that outputs are high contrast and speckle free (g), (h), (i), (j).

blurrier and contains more halos than the kernel-based approach as expected from Section 4.1.

For space reasons, Figures 6 and 7 contain abridged results. Please refer to the supplement for full versions, as well as evaluation on our Jungle scene with more complex motion trajectories, comparisons between sequential (Section 4.2) and full optimization, simulations on a binary SLM, experiments with different SLM framerates, and slow motion video visualizations.

5.1.4 Quantitative perceptual evaluation

To quantitatively evaluate the real perceptual quality of our methods, for each of our test scenes, we estimated possible eye trajectories according to the dominant object motion — the bird, plane and city motion in the *Skyline* scene, the ball, arm and hoop motion in the *Basketball* scene, the road sign, barrier and mountain motion in the *Highway* scene, and the tiger, bird and tree motion in the *Jungle* scene. With these trajectories and the associated scenes, we then ran our methods and estimated high-speed videos of perceived content via Equation (12) — to best capture the full gamut of potential visual effects, we render perceived output at the 1440 Hz framerate of the SLM. We then ran the state-of-the-art VR video metric FovVideoVDP [50] on this dataset of perceived videos.

As shown in Table 1, this metric yields similar conclusions to the previous qualitative discussion. A motion-aware method with correctly-estimated eye motion performs best, but performance decreases when the estimation is incorrect. Our stochastic and kernel-based approaches provide a middle ground. However, note that FovVideoVDP is designed for standard low-framerate content, and thus these numbers should be interpreted with caution (we are not aware of any perceptual video metrics for high-framerate content).

	Method	FovVideoVDP (\uparrow)
Without HSR	Traditional joint	4.12
	Mtn.-aware (correct est.)	3.92
	Mtn.-aware (incorrect est.)	3.13
	Stochastic	4.30
	Kernel-based	3.41
With HSR	Independent	5.10
	Mtn.-aware (correct est.)	6.84
	Mtn.-aware (incorrect est.)	5.32
	Stochastic	6.10
	Kernel-based	5.99

TABLE 1

Quantitative perceptual evaluation. We compute the average perceptual quality from FovVideoVDP [50] over the scenes shown in this paper, higher values are better. Using high-speed regularization (HSR) with correctly-estimated eye motion performs the best, but unexpected eye motion reduces performance. Our stochastic and kernel-based methods provide an alternative.

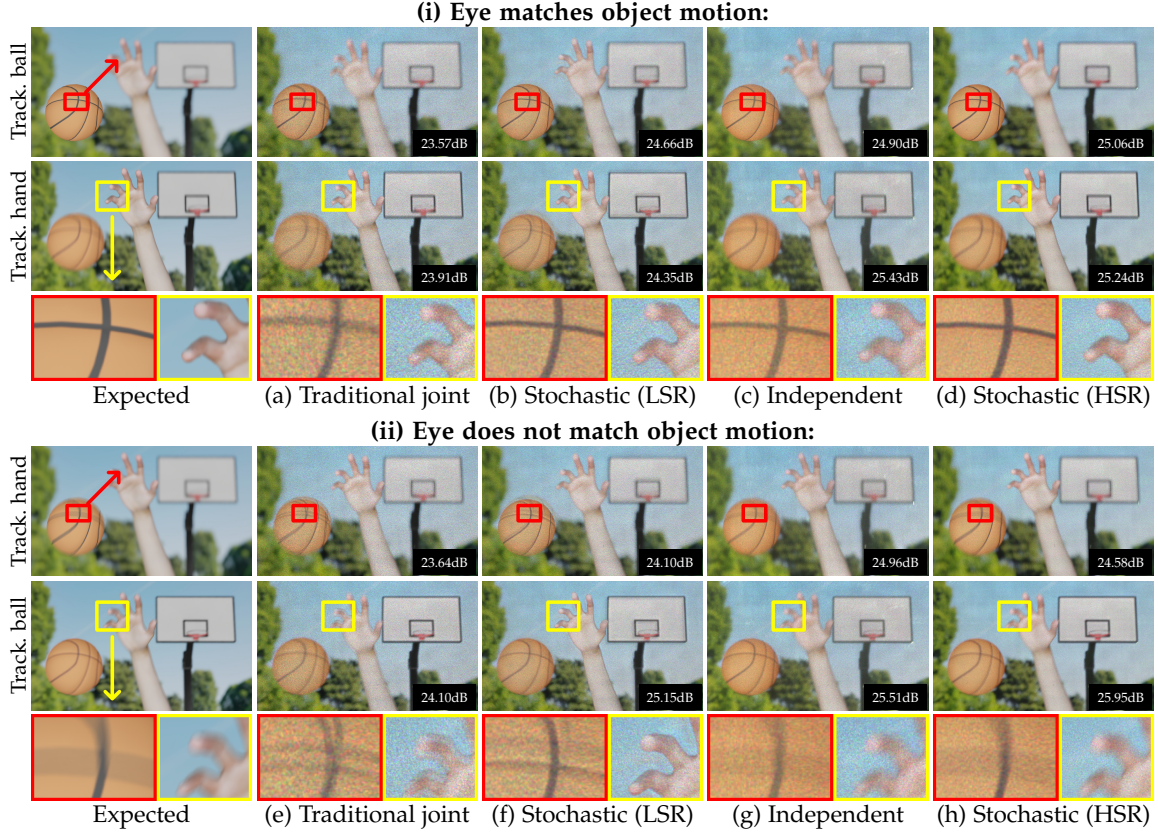


Fig. 6. **Results in a real system (abridged).** When the eye follows object motion (i), a traditional time-multiplexing approach results a blurry ball and hand, as well as increased speckle (a). Independent high-speed display (c) cleans up these artifacts, but results in loss of contrast while retaining speckle noise. Using motion-aware models that account for uncertainty can compensate for these effects (b), (d). Using high-speed regularization (HSR) reduces visible speckle (d) compared to without (LSR) (b). When the eye does not follow object motion (ii), traditional multiplexing (e) along with LSR (f) results in ghosting artifacts. Independent high-speed display produces natural motion blur (g), but reduced contrast. Our motion-aware model with HSR (h) mitigates these effects. To simplify calibration, the color results here are synthesized from captures at a single wavelength.

6 LIMITATIONS

Our high-speed regularization requires high-speed input content, which may not be directly available. However, virtual content could be readily rendered at higher rates, albeit with larger noise tolerances to avoid increasing overall computation. Alternatively, video interpolation could be used to synthesize high-speed content from low-speed input, *e.g.* RIFE [59] or FFMPEG’s *minterpolate*. For AR/VR world-locked content, affine transforms estimated from IMU data could be used to approximate high-speed frame changes akin to timewarp [60]. Given that many modern off-the-shelf displays already leverage motion interpolation to temporally upsample slow video content, we believe this is not a significant barrier to real-world use.

If eye tracking is not available, our proposed approach in Section 4.1 requires estimates of scene motion, which could add computational cost. But, if high-speed content is synthesized from low-speed input, the flow estimates that most motion-interpolation methods use could be reapplied. For rendered scenes, motion could be provided as part of the scene metadata. Our high-speed regularization also requires maintaining a history of N_{POV} previously displayed frames, resulting in increased memory usage over independent display. However, if the eye’s temporal persistence-of-vision is instead approximated with an exponential distribution, a single entry with an appropriate update rule would suffice.

7 CONCLUSION AND FUTURE WORK

In this work, we propose a paradigm of high-speed display on time-multiplexed holographic displays. Instead of repeatedly multiplexing the same static image, such systems can instead show unique, distinct content per frame. By modeling such a display with persistence-of-vision and eye motion, we can mitigate the sample-and-hold blur as well as stroboscopic artifacts of time-multiplexed holographic displays, while preserving contrast and reducing speckle. We test our approach in simulation and a real system.

Our current framework uses fairly simple perceptual models to represent human vision. More sophisticated models that account for foveation and other human visual effects could further improve visual quality [3], [4], [50]. During eye rotation, motion in the eye pupil may also need to be properly modeled following recent work in holographic displays [61], [62], [63], beyond just simple retinal translation.

Our methodology could potentially be applied to other variants of holographic systems. For instance, time multiplexing is often used in holographic displays beyond speckle reduction in tasks like color display [64] or eyebox expansion [65], and could produce related motion artifacts in these contexts. Beyond holographic displays, our framework could also be applied to other time-multiplexed displays such as DMD-based displays. The motion blur and stroboscopic effects we tackle in this work are ubiquitous.

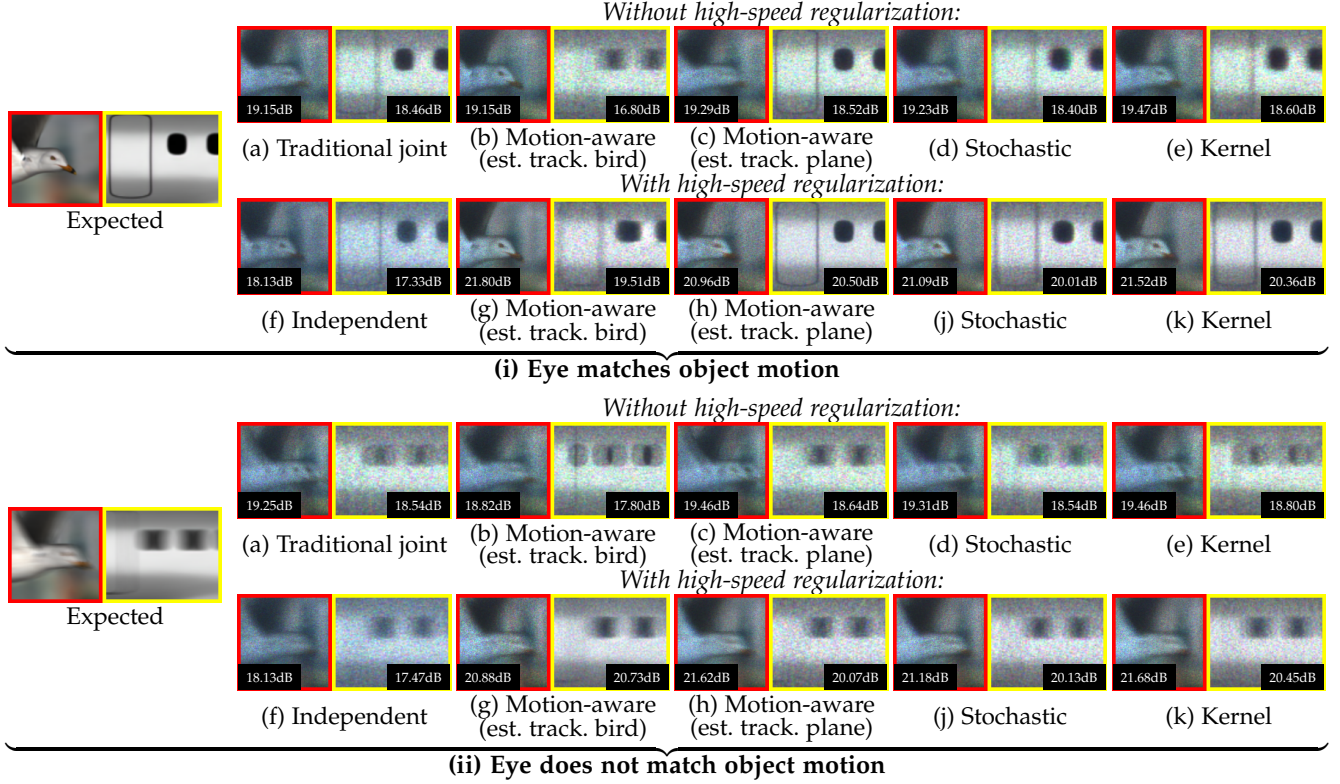


Fig. 7. **Low compute visualization (simulation, abridged).** We apply our methodology to the same simulated scene as Figure 4 and Figure 5, but with just 4 iterations of gradient descent instead of 500. Our proposed framework is again able to mitigate sample-and-hold blur as well as strobing effects. Unlike higher-compute settings, the kernel-based approach slightly outperforms the stochastic approach, as insufficient iterations are performed to cover the entire space of possible motions under the stochastic methodology.

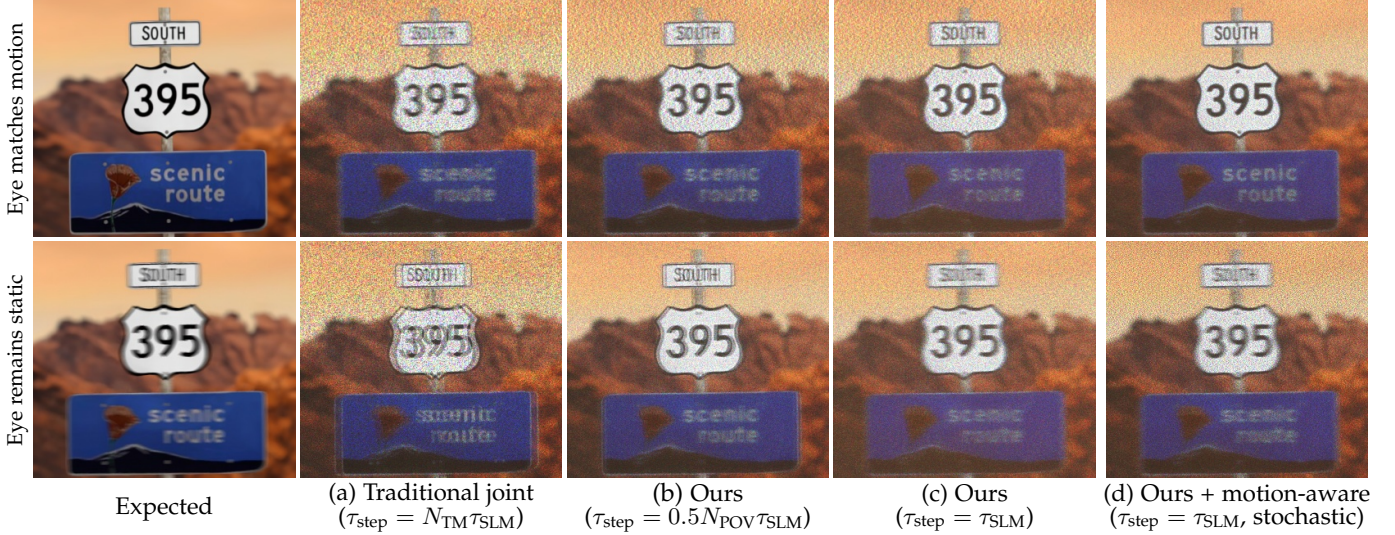


Fig. 8. **With high-speed regularization, without motion-aware optimization (simulation).** In this figure, we experiment with our framework with just high-speed regularization without motion-aware optimization, *i.e.*, assuming the eye is static. As shown by (a), (b) and (c), increasing the rate of regularization by decreasing τ_{step} turns stroboscopic effects into natural motion blur, but contrast slightly decreases. Simultaneously, higher regularization speeds sharpen perceived details when the eye does match object motion. Incorporating motion-aware optimization provides the best results (d).

REFERENCES

- [1] A. Maimone, A. Georgiou, and J. S. Kollin, "Holographic near-eye displays for virtual and augmented reality," *ACM Transactions on Graphics*, vol. 36, no. 4, pp. 1–16, 2017.
- [2] S. Choi, M. Gopakumar, Y. Peng, J. Kim, and G. Wetzstein, "Neural 3d holography: Learning accurate wave propagation models for 3d holographic virtual and augmented reality displays," *ACM Transactions on Graphics*, vol. 40, no. 6, pp. 1–12, 2021.
- [3] P. Chakravarthula, Z. Zhang, O. Tursun, P. Didyk, Q. Sun, and H. Fuchs, "Gaze-contingent retinal speckle suppression for perceptually-matched foveated holographic displays," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 11, pp. 4194–4203, 2021.
- [4] A. Georgiou, J. Kollin, C. Hewitt, P. Chakravarthula, and B. Guenter, "Visual perception of noise in a simulated holographic display—a user study," *Displays*, vol. 76, p. 102333, 2023.
- [5] G. Kuo, F. Schiffers, D. Lanman, O. Cossairt, and N. Matsuda, "Multisource holography," *ACM Transactions on Graphics*, vol. 42, no. 6, pp. 1–14, 2023.
- [6] Y. Deng and D. Chu, "Coherence properties of different light sources and their effect on the image sharpness and speckle of holographic displays," *Scientific Reports*, vol. 7, no. 1, p. 5893, 2017.
- [7] Y. Peng, S. Choi, J. Kim, and G. Wetzstein, "Speckle-free holography with partially coherent light sources and camera-in-the-loop calibration," *Science advances*, vol. 7, no. 46, p. eabg5040, 2021.
- [8] F. A. Schiffers, G. Kuo, N. Matsuda, D. Lanman, and O. Cossairt, "Holochrome: Polychromatic illumination for speckle reduction in holographic near-eye displays," pp. 1–18, 2025.
- [9] 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*, 2022, pp. 1–9.
- [10] D. Kim, S.-W. Nam, S. Choi, J.-M. Seo, G. Wetzstein, and Y. Jeong, "Holographic parallax improves 3d perceptual realism," *ACM Transactions on Graphics*, vol. 43, no. 4, pp. 1–13, 2024.
- [11] V. R. Curtis, N. W. Cairra, J. Xu, A. G. Sata, and N. C. Pégard, "Dcgh: dynamic computer generated holography for speckle-free, high fidelity 3d displays," in *2021 IEEE Virtual Reality and 3D User Interfaces*. IEEE, 2021, pp. 1–9.
- [12] B. Lee, D. Kim, S. Lee, C. Chen, and B. Lee, "High-contrast, speckle-free, true 3d holography via binary cgh optimization," *Scientific Reports*, vol. 12, no. 1, p. 2811, 2022.
- [13] D. Kim, S.-W. Nam, B. Lee, J.-M. Seo, and B. Lee, "Accommodative holography: improving accommodation response for perceptually realistic holographic displays," *ACM Transactions on Graphics*, vol. 41, no. 4, pp. 1–15, 2022.
- [14] M. Abrash, "Down the vr rabbit hole: Fixing judder," <http://blogs.valvesoftware.com/abrash/down-the-vr-rabbit-hole-fixing-judder/>, 2013.
- [15] M. Rejhon, "The stroboscopic effect of finite frame rate displays," <https://blurbusters.com/the-stroboscopic-effect-of-finite-framerate-displays/>, 2019.
- [16] E. Cuervo, K. Chintalapudi, and M. Kotaru, "Creating the perfect illusion: What will it take to create life-like virtual reality headsets?" in *Proceedings of the 19th international workshop on mobile computing systems & applications*, 2018, pp. 7–12.
- [17] G. Kuo, L. Waller, R. Ng, and A. Maimone, "High resolution étendue expansion for holographic displays," *ACM Transactions on Graphics*, vol. 39, no. 4, pp. 66–1, 2020.
- [18] S. Monin, A. C. Sankaranarayanan, and A. Levin, "Analyzing phase masks for wide étendue holographic displays," in *2022 IEEE International Conference on Computational Photography*. IEEE, 2022, pp. 1–12.
- [19] —, "Exponentially-wide étendue displays using a tilting cascade," in *2022 IEEE International Conference on Computational Photography*. IEEE, 2022, pp. 1–12.
- [20] L. Shi, B. Li, C. Kim, P. Kellnhöfer, and W. Matusik, "Towards real-time photorealistic 3d holography with deep neural networks," *Nature*, vol. 591, no. 7849, pp. 234–239, 2021.
- [21] Y. Peng, S. Choi, N. Padmanaban, and G. Wetzstein, "Neural holography with camera-in-the-loop training," *ACM Transactions on Graphics*, vol. 39, no. 6, pp. 1–14, 2020.
- [22] J. Kim, M. Gopakumar, S. Choi, Y. Peng, W. Lopes, and G. Wetzstein, "Holographic glasses for virtual reality," in *ACM SIGGRAPH 2022 Conference Proceedings*, 2022, pp. 1–9.
- [23] T. Kozacki, M. Chlipala, J. Martinez-Carranza, R. Kukołowicz, and M. S. Idicula, "Led near-eye holographic display with a large non-paraxial hologram generation," *Optics Express*, vol. 30, no. 24, pp. 43 551–43 565, 2022.
- [24] Y. Jo, D. Yoo, D. Lee, M. Kim, and B. Lee, "Multi-illumination 3d holographic display using a binary mask," *Optics Letters*, vol. 47, no. 10, pp. 2482–2485, 2022.
- [25] B. Chao, M. Gopakumar, S. Choi, J. Kim, L. Shi, and G. Wetzstein, "Large étendue 3d holographic display with content-adaptive dynamic fourier modulation," in *SIGGRAPH Asia 2024 Conference Papers*, 2024, pp. 1–12.
- [26] P. Chakravarthula, Y. Peng, J. Kollin, H. Fuchs, and F. Heide, "Wirtinger holography for near-eye displays," *ACM Transactions on Graphics*, vol. 38, no. 6, pp. 1–13, 2019.
- [27] D. Yang, W. Seo, H. Yu, S. I. Kim, B. Shin, C.-K. Lee, S. Moon, J. An, J.-Y. Hong, G. Sung *et al.*, "Diffraction-engineered holography: Beyond the depth representation limit of holographic displays," *Nature Communications*, vol. 13, no. 1, p. 6012, 2022.
- [28] A. Cable, E. Buckley, P. Mash, N. Lawrence, T. Wilkinson, and W. Crossland, "53.1: Real-time binary hologram generation for high-quality video projection applications," in *SID Symposium Digest of Technical Papers*, vol. 35, no. 1. Wiley Online Library, 2004, pp. 1431–1433.
- [29] K. Masuda, Y. Saita, R. Toritani, P. Xia, K. Nitta, and O. Matoba, "Improvement of image quality of 3d display by using optimized binary phase modulation and intensity accumulation," *Journal of Display Technology*, vol. 12, no. 5, pp. 472–477, 2016.
- [30] E. Lee, Y. Jo, S.-W. Nam, M. Chae, C. Chun, Y. Kim, Y. Jeong, and B. Lee, "Speckle reduced holographic display system with a jointly optimized rotating phase mask," *Optics Letters*, vol. 49, no. 19, pp. 5659–5662, 2024.
- [31] Y. Takaki and M. Yokouchi, "Speckle-free and grayscale hologram reconstruction using time-multiplexing technique," *Optics express*, vol. 19, no. 8, pp. 7567–7579, 2011.
- [32] J.-P. Liu, M.-H. Wu, and P. W. Tsang, "3d display by binary computer-generated holograms with localized random down-sampling and adaptive intensity accumulation," *Optics Express*, vol. 28, no. 17, pp. 24 526–24 537, 2020.
- [33] D. Chan, S. G. Narasimhan, and M. O'Toole, "Holocurtains: Programming light curtains via binary holography," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17 886–17 895.
- [34] J. Cheng, C. Gu, D. Zhang, and S.-C. Chen, "High-speed femtosecond laser beam shaping based on binary holography using a digital micromirror device," *Optics letters*, vol. 40, no. 21, pp. 4875–4878, 2015.
- [35] K. J. Mitchell, S. Turtaev, M. J. Padgett, T. Čížmár, and D. B. Phillips, "High-speed spatial control of the intensity, phase and polarisation of vector beams using a digital micro-mirror device," *Optics express*, vol. 24, no. 25, pp. 29 269–29 282, 2016.
- [36] S. Turtaev, I. T. Leite, K. J. Mitchell, M. J. Padgett, D. B. Phillips, and T. Čížmár, "Comparison of nematic liquid-crystal and dmd based spatial light modulation in complex photonics," *Optics express*, vol. 25, no. 24, pp. 29 874–29 884, 2017.
- [37] D. Stuart, O. Barter, and A. Kuhn, "Fast algorithms for generating binary holograms," *arXiv preprint arXiv:1409.1841*, 2014.
- [38] D. Stuart and A. Kuhn, "Single-atom trapping and transport in dmd-controlled optical tweezers," *New Journal of Physics*, vol. 20, no. 2, p. 023013, 2018.
- [39] J. Cheng, C. Gu, D. Zhang, D. Wang, and S.-C. Chen, "Ultrafast axial scanning for two-photon microscopy via a digital micromirror device and binary holography," *Optics letters*, vol. 41, no. 7, pp. 1451–1454, 2016.
- [40] Q. Geng, D. Wang, P. Chen, and S.-C. Chen, "Ultrafast multi-focus 3-d nano-fabrication based on two-photon polymerization," *Nature communications*, vol. 10, no. 1, pp. 1–7, 2019.
- [41] P. Pozzi, D. Wilding, O. Soloviev, H. Verstraete, L. Bliet, G. Vdovin, and M. Verhaegen, "High speed wavefront sensorless aberration correction in digital micromirror based confocal microscopy," *Optics Express*, vol. 25, no. 2, pp. 949–959, 2017.
- [42] A. Jones, I. McDowall, H. Yamada, M. Bolas, and P. Debevec, "Rendering for an interactive 360 light field display," in *ACM SIGGRAPH 2007 papers*, 2007, pp. 40–es.
- [43] O. Cossairt, A. R. Travis, C. Moller, and S. A. Benton, "Novel view sequential display based on dmd technology," in *Stereoscopic Displays and Virtual Reality Systems XI*, vol. 5291. SPIE, 2004, pp. 273–278.

- [44] O. Cossairt, J. Napoli, S. L. Hill, R. K. Dorval, and G. E. Favalora, "Occlusion-capable multiview volumetric three-dimensional display," *Applied Optics*, vol. 46, no. 8, pp. 1244–1250, 2007.
- [45] J.-H. R. Chang, B. V. Kumar, and A. C. Sankaranarayanan, "Towards multifocal displays with dense focal stacks," *ACM Transactions on Graphics*, vol. 37, no. 6, pp. 1–13, 2018.
- [46] I. McDowall and M. Bolas, "Fast light for display, sensing and control applications," in *IEEE VR Workshop on Emerging Display Technologies*, vol. 8, 2005.
- [47] S. J. Koppal, S. Yamazaki, and S. G. Narasimhan, "Exploiting dlp illumination dithering for reconstruction and photography of high-speed scenes," *International Journal of Computer Vision*, vol. 96, pp. 125–144, 2012.
- [48] R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, and H. Fuchs, "The office of the future: A unified approach to image-based modeling and spatially immersive displays," in *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, 1998, pp. 179–188.
- [49] D. Cotting, M. Naef, M. Gross, and H. Fuchs, "Embedding imperceptible patterns into projected images for simultaneous acquisition and display," in *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*. IEEE, 2004, pp. 100–109.
- [50] R. K. Mantiuk, G. Denes, A. Chapiro, A. Kaplanyan, G. Rufo, R. Bachy, T. Lian, and A. Patney, "Fovvideovdp: A visible difference predictor for wide field-of-view video," *ACM Transactions on Graphics*, vol. 40, no. 4, pp. 1–19, 2021.
- [51] C. A. Burbeck and D. Kelly, "Spatiotemporal characteristics of visual mechanisms: excitatory-inhibitory model," *Journal of the Optical Society of America*, vol. 70, no. 9, pp. 1121–1126, 1980.
- [52] J. Laird, M. Rosen, J. Pelz, E. Montag, and S. Daly, "Spatio-velocity csf as a function of retinal velocity using unstabilized stimuli," in *Human Vision and Electronic Imaging XI*, vol. 6057. SPIE, 2006, pp. 32–43.
- [53] S. T. Hammett and A. Smith, "Two temporal channels or three? a re-evaluation," *Vision Research*, vol. 32, no. 2, pp. 285–291, 1992.
- [54] J. W. Goodman, *Introduction to Fourier Optics*, ser. McGraw-Hill physical and quantum electronics series. W. H. Freeman, 2005.
- [55] A. M. Feit, S. Williams, A. Toledo, A. Paradiso, H. Kulkarni, S. Kane, and M. R. Morris, "Toward everyday gaze input: Accuracy and precision of eye tracking and implications for design," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 1118–1130.
- [56] W. Wang and J. Shen, "Deep visual attention prediction," *IEEE Transactions on Image Processing*, vol. 27, no. 5, pp. 2368–2378, 2018.
- [57] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proceedings of the 3rd International Conference for Learning Representations*, 2015.
- [58] G. Farneback, "Two-frame motion estimation based on polynomial expansion," in *Image Analysis: 13th Scandinavian Conference*. Springer, 2003, pp. 363–370.
- [59] Z. Huang, T. Zhang, W. Heng, B. Shi, and S. Zhou, "Real-time intermediate flow estimation for video frame interpolation," in *European Conference on Computer Vision*. Springer, 2022, pp. 624–642.
- [60] J. M. P. Van Waveren, "The asynchronous time warp for virtual reality on consumer hardware," in *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, 2016, pp. 37–46.
- [61] F. Schiffers, P. Chakravarthula, N. Matsuda, G. Kuo, E. Tseng, D. Lanman, F. Heide, and O. Cossairt, "Stochastic light field holography," in *2023 IEEE International Conference on Computational Photography*. IEEE, 2023, pp. 1–12.
- [62] P. Chakravarthula, S.-H. Baek, F. Schiffers, E. Tseng, G. Kuo, A. Maimone, N. Matsuda, O. Cossairt, D. Lanman, and F. Heide, "Pupil-aware holography," *ACM Transactions on Graphics*, vol. 41, no. 6, pp. 1–15, 2022.
- [63] L. Shi, D. Ryu, and W. Matusik, "Ergonomic-centric holography: Optimizing realism, immersion, and comfort for holographic display," *Laser & Photonics Reviews*, vol. 18, no. 4, p. 2300651, 2024.
- [64] K. Kavaklı, L. Shi, H. Urey, W. Matusik, and K. Akşit, "Multi-color holograms improve brightness in holographic displays," in *SIGGRAPH Asia 2023 Conference Papers*, 2023, pp. 1–11.
- [65] B. Lee, D. Yoo, J. Jeong, S. Lee, D. Lee, and B. Lee, "Wide-angle speckleless dmd holographic display using structured illumination with temporal multiplexing," *Optics Letters*, vol. 45, no. 8, pp. 2148–2151, 2020.

Dorian Chan received his PhD from the Computer Science Department at Carnegie Mellon University, advised by Matthew O'Toole. He also spent time at Reality Labs Research, Meta and Snap Research. His work has been recognized with the Best Paper Honorable Mention Award at CVPR 2022, and 1st place at the SIGGRAPH 2020 Student Research Competition. Today, he works broadly across computational photography for both scientific and creative applications.

Oliver Cossairt is a research scientist at Reality Labs Research, Meta and an Associate Professor in the Electrical Engineering and Computer Science Department at Northwestern University. He holds an M.S. from the MIT Media Lab, and a Ph.D. in Computer Science from Columbia University. With a background as an Optical and Software Engineer at Actuality Systems, Oliver has received several accolades, including the NSF Graduate Research Fellowship, ICCP Best Paper and Honorable Mention, and an NSF CAREER Award.

Nathan Matsuda is a research scientist at Meta developing computational camera and display systems that support immersive visual experiences in VR. He completed his PhD at Northwestern University while supported by the NSF Graduate Research Fellowship. Prior to grad school, Nathan worked in visual effects for film and television.

Grace Kuo is a research scientist at Reality Labs Research, Meta working on the joint design of optical hardware and algorithms for imaging and display systems. She earned her PhD from the Department of Electrical Engineering and Computer Sciences at UC Berkeley, advised by Dr. Laura Waller and Dr. Ren Ng.