

Factored Occlusion: Single Spatial Light Modulator Occlusion-capable Optical See-through Augmented Reality Display

Brooke Krajancich*, Nitish Padmanaban*, Gordon Wetzstein

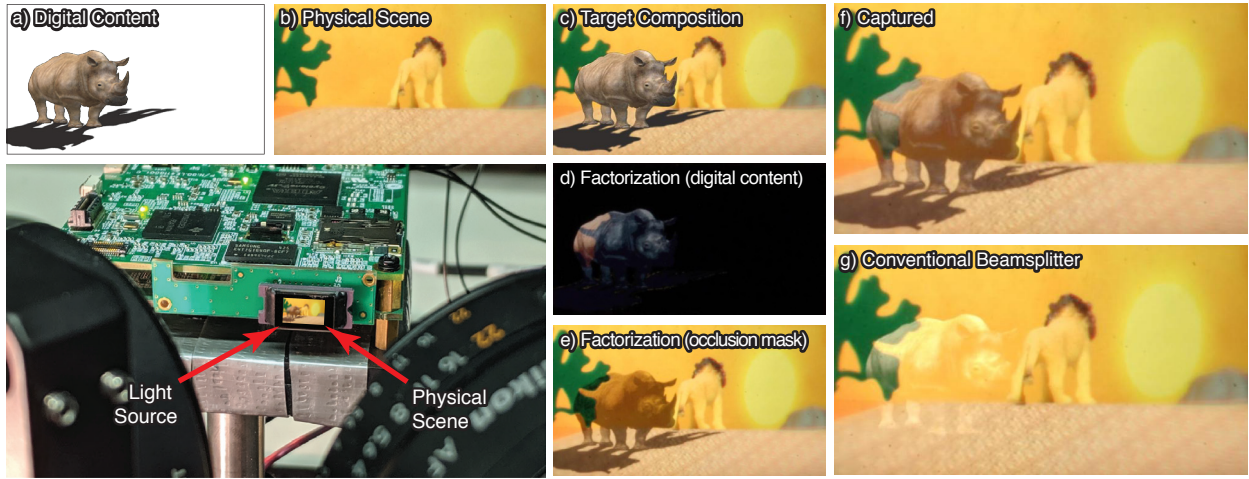


Fig. 1. Single spatial light modulator (SLM) occlusion-capable optical see-through augmented reality (OST-AR) display. The SLM is composed of micromirrors that reflect light from either (a) an illumination source or (b) a physical scene toward the user. (c) For AR applications, we aim to display a target composite of the physical scene (the lion and background) and some digital content (the rhino and shadows). The proposed factorization algorithm computes optimal mirror states for the (d) LED illumination and (e) light blocking of the physical scene, which combined form the best approximation to this target. Photographs of (f) this composition displayed on our prototype show significant improvement in color fidelity and light-blocking ability over (g) a conventional beamsplitter approach that additively combines digital and physical content.

Abstract—Occlusion is a powerful visual cue that is crucial for depth perception and realism in optical see-through augmented reality (OST-AR). However, existing OST-AR systems additively overlay physical and digital content with beam combiners – an approach that does not easily support mutual occlusion, resulting in virtual objects that appear semi-transparent and unrealistic. In this work, we propose a new type of occlusion-capable OST-AR system. Rather than additively combining the real and virtual worlds, we employ a single digital micromirror device (DMD) to merge the respective light paths in a multiplicative manner. This unique approach allows us to simultaneously block light incident from the physical scene on a pixel-by-pixel basis while also modulating the light emitted by a light-emitting diode (LED) to display digital content. Our technique builds on mixed binary/continuous factorization algorithms to optimize time-multiplexed binary DMD patterns and their corresponding LED colors to approximate a target augmented reality (AR) scene. In simulations and with a prototype benchtop display, we demonstrate hard-edge occlusions, plausible shadows, and also gaze-contingent optimization of this novel display mode, which only requires a single spatial light modulator.

Index Terms—Augmented reality, computational displays, mutual occlusion.

1 INTRODUCTION

Augmented reality (AR) is widely believed to become a next-generation computing platform, well-poised to fundamentally change how we consume information. By seamlessly augmenting digital content with a direct view of the physical world, optical see-through augmented reality (OST-AR) systems in particular offer unprecedented experiences in a wide range of applications, including human-computer interaction, communication, education, entertainment, and medicine [4, 55]. However, current OST-AR displays fall short of offering seamless experiences, primarily because they neglect one of the

most important depth cues of human vision: occlusion [8].

Occlusion describes a phenomenon where objects at some distance to the user partially or fully block the light from objects at farther distances. This cue provides the human visual system with important information about depth and failure to emulate occlusion not only detracts from perceptual realism [34, 41], but could be dangerous, inducing user error in tasks involving spatial judgment [11]. Current-generation AR displays use optical combiners that additively superimpose the digital content displayed by a spatial light modulator (SLM) on the physical world a user sees. These displays typically lack the capability to block physical light, leading to an inability to emulate occlusion. This in turn causes digital content to appear semi-transparent and unrealistic, especially in bright environments (see Fig. 1g).

Over the last few years, research has focused on developing occlusion-capable AR displays, though with varying approaches offering either soft-edge (i.e., out-of-focus) or hard-edge (i.e., pixel-precise) occlusion control [3, 12, 14, 30, 45, 54]. Often, occlusion is achieved using an additional SLM, such as a digital micromirror device (DMD) or liquid crystal on silicon (LCoS), in the optical path

• Brooke Krajancich, Nitish Padmanaban and Gordon Wetzstein are with Stanford University. E-mail: {brookek | nit | gordon.wetzstein}@stanford.edu. *indicates equal contribution.

Manuscript received xx.xxx. 201x; accepted xx.xxx. 201x. Date of Publication xx.xxx. 201x; date of current version xx.xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxx/TVCG.201x.xxxxxxx

between the physical world and the user's eye. However, this necessitates that the system has two SLMs: one for displaying the digital content and another for enabling occlusion control. Using multiple SLMs is impractical for AR systems, requiring significantly more power and increasing the device form factor through increased complexity of the optical and electronic systems. Furthermore, maintaining pixel-precise and robust calibration of the two SLMs may be difficult, especially for long-term use.

In this work, we propose a new approach to enabling hard-edge mutual occlusion in OST-AR displays. Our approach uses only a single SLM to generate both an occlusion mask that controls the physical light reaching the user's eye on a per-pixel basis and a digital image. This unprecedented capability is enabled by the unique properties of digital micromirror devices (DMDs), a reflective-type SLM consisting of an array of tiny mirrors that can flip between two different orientations at a rate of tens of thousands of times per second. Whereas this high-speed switching is typically used to create grayscale intensities by temporally dithering a constant light source, our system uses these mirrors to instead switch between a see-through state, which allows the user to observe a small part of the real world reflected in a micromirror, and a reflective state, where the same mirror blocks light from the scene in favor of light from an RGB light-emitting diode (LED). By performing a content-dependent optimization of the micromirror states and LED intensity, we demonstrate that it is possible to simultaneously block physical light in a pixel-precise manner and display a target digital image.

Our primary contributions are that we:

1. Introduce an occlusion-capable OST-AR display that reduces complexity by using only a single SLM for both the occlusion mask and virtual content generation,
2. Develop a mixed binary/continuous factorization approach that computes the DMD states and LED color values required to render both digital content and occlusion,
3. Demonstrate a gaze-contingent display mode that optimizes perceived image content for the foveal region of the display, and
4. Implement a proof-of-concept display prototype that demonstrates hard-edge occlusions for a range of scene compositions.

Overview of Limitations Although we demonstrate simultaneous occlusion and image display with a single SLM, the proposed algorithm optimizes a trade-off between accuracy of the desired occlusion mask and color fidelity of the digital image. Therefore, slight degradations of the display colors may be visible. Moreover, by shifting some of the design complexity from hardware into software, our system requires increased computational resources compared with conventional OST-AR displays.

2 RELATED WORK

In this work, we present the first technique to utilize the SLM used to render virtual content to also provide a hard-edge occlusion mask for OST-AR displays. This is achieved using an optimization-based approach for programming the states of this SLM. As such, the following section surveys previously demonstrated occlusion-capable OST-AR displays and optimization techniques for computational near-eye displays.

2.1 Occlusion-capable OST-AR Displays

2.1.1 Global Dimming

Commercial OST-AR displays (e.g. Microsoft HoloLens, Magic Leap One) are not able to provide accurate (mutual and hard-edge) occlusion. Often a neutral density filter is used to uniformly reduce the brightness of the physical scene. Referred to as global dimming, this can reduce the ghost-like appearance of virtual content, particularly in bright environments. Mori et al. [48] recently proposed an adaptive version of this approach using a single liquid crystal cell, whereby the amount of dimming is dependent on the physical environment. However, this technique lacks spatial control and can detract from realism.

2.1.2 Soft-edge Occlusion

Placing an additional SLM in the optical path between the optical combiner and the user's eyes can be used to selectively block light from the physical scene [22, 45, 63]. However, this creates a blurred (or soft-edge) occlusion mask since the occlusion SLM is out of focus with respect to the virtual image. It is possible to compensate for this by extending the occlusion mask and computationally combining missing areas of the physical scene with images captured by an additional scene camera [22]. However, such a system requires complex calibration and can detract from the allure of OST-AR by augmenting the direct view of the physical scene [55].

Maimone and Fuchs [45] proposed a lensless multi-layer design using a system of three stacked LCD panels placed out of focus in front of the eye. By computing a factored light field, occlusion masks were realized by temporally switching the displayed image synchronously with an optical shutter. Although the multi-layer light field approach can in theory overcome some of the limitations of a single-layer blocking method and additionally demonstrate 3D occlusion in a relatively compact form factor, it is subject to several limitations: a stack of liquid crystal displays (LCDs) severely reduces the light transmission of such a system; diffraction through the LCD pixels fundamentally limits the optical resolution of a physical scene observed through the LCDs; and the depth of field of the light field is low, resulting in blurred occlusion masks at the distance of the physical objects. Therefore, none of these approaches is able to achieve truly hard-edge occlusion, limiting the perceptual realism of the display.

2.1.3 Hard-edge Occlusion

The out-of-focus nature of soft-edge occlusion-capable displays can be addressed through the use of additional relay optics that first focus the physical scene onto the SLM and then re-image it onto the user's retina. This approach was first proposed by the seminal work of Kiyokawa et al. [29, 30, 31]. Although effective at providing spatially controlled hard-edge occlusion, this approach requires a somewhat bulky optical system. Uchida et al. [62] proposed a similar system using a DMD, citing brightness advantages provided by the high contrast ratio of the reflective SLM. Recently, Kim et al. [28] prototyped this optical design, adding mirrors to redirect the optical path such that the user sees the physical scene directly behind the DMD. Yamaguchi and Takaki [66] recently extended this type of AR display using microlens array panels to produce an occlusion-capable integral light field display. This system was highly compact and facilitated 3D hard-edge occlusion, however it was limited to a very narrow field of view (4.3 degrees), diffraction artifacts, and as with all light-field displays, the spatio-angular resolution trade-off. Similar to some previous work, we also use a DMD with relay optics to achieve hard-edge occlusion. Unlike existing approaches, however, ours only requires a single DMD instead of multiple SLMs to simultaneously generate virtual content and occlusion masks, reducing form-factor and removing the need for careful SLM alignment.

Cakmakci et al. [3] improved upon Kiyokawa et al.'s work, using polarization-based optics and a reflective LCoS in conjunction with an organic light emitting diode (OLED) display to reduce form-factor, resolution, and switching-speed. Gao et al. [12] proposed the use of freeform optics, a two-layer folded optical architecture, along with a reflective LCoS to create a compact, high-resolution, occlusion-capable OST-AR display. This work showed great promise for both high optical performance and compact form factor, but it requires custom and expensive freeform lenses. Wilson and Hua [65] implemented a similar architecture using off-the-shelf components, using advances in SLM technology to demonstrate increased field of view, resolution, and optical performance. Recently, Hamasaki and Itoh [14] and Rathinavel et al. [54] extended this work to demonstrate varifocal occlusion-capable OST-AR displays using mechanical motion of the SLM and focus-tunable optics, respectively. Our work is complementary to all of these approaches and could help reduce the complexity of systems using either freeform lenses or varifocal AR displays.

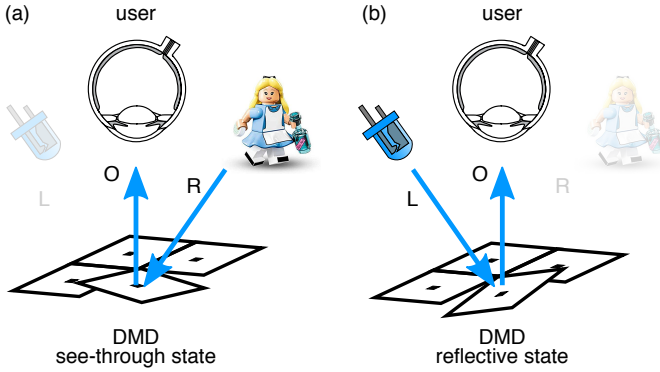


Fig. 2. Illustration of the principle of a single-SLM occlusion-capable AR display. Each pixel of the DMD can be flipped to one of two states. (a) In one state, the mirror reflects light of the physical scene, \mathbf{R} , toward the user. (b) In the other state, the mirror reflects the light of an LED, \mathbf{L} , toward the user. By quickly flipping the state of the mirror during the integration time of the user’s photoreceptors, this display is capable of optimizing a set of DMD pixel states and LED color values to display a target scene, \mathbf{O} , with mutual occlusions between physical and digital content.

2.2 Color Correction for OST-AR Displays

The additive superposition used by current generation OST-AR displays can also cause color distortion, since the user perceives the result of the displayed pixel color and the color of the current physical scene seen through the HMD. As such, several color correction approaches [9, 21, 36, 58, 61] have been developed, using knowledge of the physical scene to adjust rendered colors such that the final addition produces the intended color. Recently Itoh et al. [23] proposed a new paradigm that forms images by spatially subtracting colors of light. In this way, knowledge of the scene and a spatial light modulator, could be used to compute and display target color values. Our work can be viewed as an extreme extension of this concept, since we use the degrees of freedom offered by the DMD to effectively dim the physical scene component.

2.3 Factored Displays

Computational displays aim at shifting some of the design complexity of a display from hardware to software. Formal optimization, and non-negative matrix and tensor factorization in particular, have become some of the most powerful algorithms driving computational displays, for example to improve resolution [15, 16], dynamic range [6, 16, 59], light field capabilities [16, 37, 64], color gamut [26], and other aspects of monitor and projection-based displays [46]. The algorithmic approach proposed in this paper also builds on matrix factorization, but we apply it to a novel application of enabling mutual occlusion in OST-AR near-eye displays. Moreover, the inverse problem of optimizing binary DMD pixel states along with LED intensity values we face is a mixed binary/continuous problem, which is significantly more challenging than continuous problems proposed in these earlier works.

2.4 Computational Near-eye Displays

Computational near-eye displays are also an active area of research. In particular, foveated near-eye displays [13, 27, 52] and near-eye displays with focus cues that mitigate the vergence–accommodation conflict [33] have seen much interest in recent years. Focus-supporting near-eye displays include varifocal [1, 5, 10, 25, 32, 35, 40, 51, 60], multifocal [2, 7, 18, 39, 40, 42, 43, 47, 49, 53, 56], and light field displays [19, 20, 38, 44]. Our work aims at improving a complementary characteristic of near-eye displays (i.e., occlusion support) using a similar strategy of co-designing optics and image processing.

3 IMAGE OPTIMIZATION

3.1 Image Formation

A digital micromirror device (DMD) is a reflective SLM that consists of a dense two-dimensional array of micro-electromechanical (MEMS) mirrors. These mirrors can be individually flipped to one of two states (i.e., $\pm 12^\circ$) at a rate of up to tens of thousands of times per second. DMDs are conventionally used in digital light projectors where each mirror reflects the light of some source, such as an LED, toward either a projection screen or alternatively into a light dump. Although each micromirror can only be set to a binary state of either ON or OFF, grayscale intensity values can be created by temporally dithering these binary patterns, and different colors can be displayed by time-sequentially varying the LED’s hue.

In this work, we propose using a DMD in an unconventional way specific to the needs of OST-AR displays. Again, the DMD reflects the light from an LED toward the user in one state, but instead of redirecting it into a dump for the other state, we repurpose this mirror state to reflect light from a physical scene toward the user instead. Therefore, each micromirror is either in a see-through state, allowing the user to directly observe the physical scene, or in a reflective mode where it directs whatever color and intensity the LED is currently showing to the user’s eye. Using fast temporal multiplexing, this display precisely controls how much light from both the scene and the LED reaches the user. This principle of operation is illustrated in Fig. 2.

Formally, we assume that the DMD contains a 2D array of $w \times h$ micromirrors and that these are capable of switching their states T times within the critical flicker fusion threshold of human vision, so it updates at about $60T$ Hz. We then define the set of binary micromirror states as $\mathbf{D} \in \{0, 1\}^{w \times h \times T}$, where the rows of this matrix indicate the vectorized spatial dimension and the columns the temporal variation of each mirror. We further define the continuous states of an LED with independent control of the red, green, and blue color channels as $\mathbf{L} \in [0, 1]^{T \times 3}$, where the rows indicate time and the columns refer to the three color channels. Finally, we define the target composite image and incident real world light as $\mathbf{O}, \mathbf{R} \in [0, 1]^{w \times h \times 3}$. The vectorized color image \mathbf{R} represents the physical scene in front of the user, as captured by a camera mounted on the user’s head. The target image \mathbf{O} contains both the physical scene and the digital content along with mutual occlusions, shadows, and shading effects that are rendered into \mathbf{O} . Despite the need for a captured representation of the physical scene for the optimization algorithm, the user directly views the scene at the end reflected off the mirrors, so our approach is a true optical see-through display.

Mathematically, we can express the image that the user observes, $\hat{\mathbf{O}}$, as a combination of the temporally varying LED intensity reflected by the DMD and the physical scene modulated by the inverse of the same DMD pattern:

$$\hat{\mathbf{O}} = \frac{\alpha}{T} \mathbf{D} \mathbf{L} + \mathbf{R} - \frac{1}{T} \mathbf{R} \circ (\mathbf{D} \mathbf{1}_{T \times 3}), \quad (1)$$

where the first, second, and third term model digital content as shown by LED and DMD, the direct-view physical scene, and the “crosstalk” between the two accounting for the limited degrees of freedom of the DMD in independently controlling LED and scene, respectively. The matrix $\mathbf{1}$ contains only ones and the subscript on the $\mathbf{1}$ matrix gives its dimensions. The \circ operator denotes the Hadamard product, i.e. element-wise multiplication. The scaling factor $1/T$ is required for appropriate normalization, and the factor α accounts for the total intensity of the LED, which can be brighter than the physical scene. We set this user-defined parameter to $\alpha = 3$.

3.2 DMD Factorization

The inverse problem at hand is that of determining the set of temporally varying continuous LED values \mathbf{L} and binary DMD mirror states \mathbf{D} that together with \mathbf{R} approximate the target image in the best way. To this end, we define the temporary variables $\mathbf{O}' = (\mathbf{O} - \mathbf{R}) \cdot T/\alpha$ and $\mathbf{R}' = \mathbf{R}/\alpha$ and aim at minimizing the difference between the observed image $\hat{\mathbf{O}}$ and the target image \mathbf{O} by optimizing the following

cost function:

$$J(\mathbf{D}, \mathbf{L}) = \frac{1}{2} \|\mathbf{O} - \hat{\mathbf{O}}\|_F^2 = \frac{1}{2} \|\mathbf{O}' - \mathbf{D}\mathbf{L} + \mathbf{R}' \circ (\mathbf{D}\mathbf{1}_{T \times 3})\|_F^2. \quad (2)$$

Since \mathbf{O}' is not nonnegative at all indices, commonly applied multiplicative update rules for nonnegative matrix factorization algorithms tend to perform poorly. We instead turn to the rank-one residue update proposed by Ho [17]. Using the following definition of the residue matrix,

$$\mathbf{O}'_t \triangleq \mathbf{O}' - \sum_{i \neq t} (\mathbf{D}_i \mathbf{L}_i^T - \mathbf{R}' \circ (\mathbf{D}_i \mathbf{1}_3^T)), \quad (3)$$

where \mathbf{D}_i and \mathbf{L}_i^T are the i^{th} column and row of \mathbf{D} and \mathbf{L} , respectively, we get the set of cost functions

$$J_t(\mathbf{D}, \mathbf{L}) = \|\mathbf{O}'_t - \mathbf{D}_t \mathbf{L}_t^T + \mathbf{R}' \circ (\mathbf{D}_t \mathbf{1}_3^T)\|_F^2. \quad (4)$$

The optimal update for \mathbf{D}_t is found to be

$$\mathbf{D}_{t,i}^* \leftarrow \begin{cases} 1 & \text{if } \mathbf{x}_{t,i} > 0 \\ 0 & \text{otherwise} \end{cases}, \quad (5)$$

$$\text{with } \mathbf{x}_t = 2(\mathbf{W} \circ \mathbf{O}'_t \circ (\mathbf{1}_{wh} \mathbf{L}_t^T - \mathbf{R}')) \mathbf{1}_3 - (\mathbf{W} \circ (\mathbf{1}_{wh} \mathbf{L}_t^T - \mathbf{R}')^2) \mathbf{1}_3, \quad (6)$$

where $\mathbf{W} \in [0, 1]^{wh \times T}$ is a weighting matrix that includes a relative weight for how important each pixel is for the optimization and $(\cdot)^2$ is defined element wise. In many cases, all elements of \mathbf{W} are one, but we could also weight the pixels in a gaze-contingent manner, setting the weights of pixels in the fovea to one and those for peripheral pixels to a lower value (see Sect. 6). Before we get to the update for \mathbf{L}_t , let us define two quantities,

$$\mathbf{y}_t = \mathbf{D}_t^T (\mathbf{W} \circ (\mathbf{O}'_t + (\mathbf{R}' \circ (\mathbf{D}_t \mathbf{1}_3^T))), \quad (7)$$

$$\mathbf{z}_t = \mathbf{D}_t^T (\mathbf{W} \circ \mathbf{D}_t \mathbf{1}_3^T). \quad (8)$$

The optimal update rule is then given by

$$\mathbf{L}_{t,i} \leftarrow \min \left\{ \frac{[y_{t,i}]_+}{z_{t,i}}, 1 \right\}. \quad (9)$$

With Equations 5 and 9 in hand, we have a set of simple update rules that allow us to optimize DMD and LED patterns of this mixed binary/continuous problem in an iterative manner. Pseudocode for this algorithm is outlined in Alg. 1 and source code will be made publicly available.

Algorithm 1 DMD Factorization Algorithm

```

1: function FACTORIZE( $\mathbf{O}, \mathbf{R}$ )
2:    $\mathbf{O}' = (\mathbf{O} - \mathbf{R}) \cdot T/\alpha$ ,  $\mathbf{R}' = \mathbf{R}/\alpha$ 
3:   for each NMF iteration do
4:     for  $t = 1 \dots T$  do
5:       compute  $\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t$  ▷ see Equations 6–8
6:        $\mathbf{D}_t \leftarrow \mathbf{x}_t > 0$ 
7:        $\mathbf{L}_t \leftarrow (\mathbf{y}_t / \mathbf{z}_t) \cdot \text{clip}(0, 1)$ 
8:     end for
9:   end for
10:  return  $\mathbf{D}, \mathbf{L}$ 
11: end function

```

3.3 Analysis

Although the proposed DMD factorization algorithm is iterative in nature, it usually converges in a few tens of iterations. For example, Fig. 3 (left) shows the peak signal-to-noise ratio (PSNR) plotted for an increasing number of iterations for the teaser scene. Although the binary nature of this inverse problem makes this plot non-smooth, the convergence is generally well-behaved.

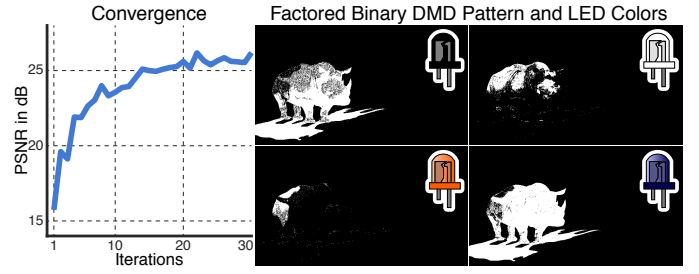


Fig. 3. (left) The convergence plot shows, here for the example of the teaser scene, that the proposed iterative algorithm converges in about 20–30 iterations. (right) Four of the factored binary DMD patterns and corresponding RGB LED values are shown.

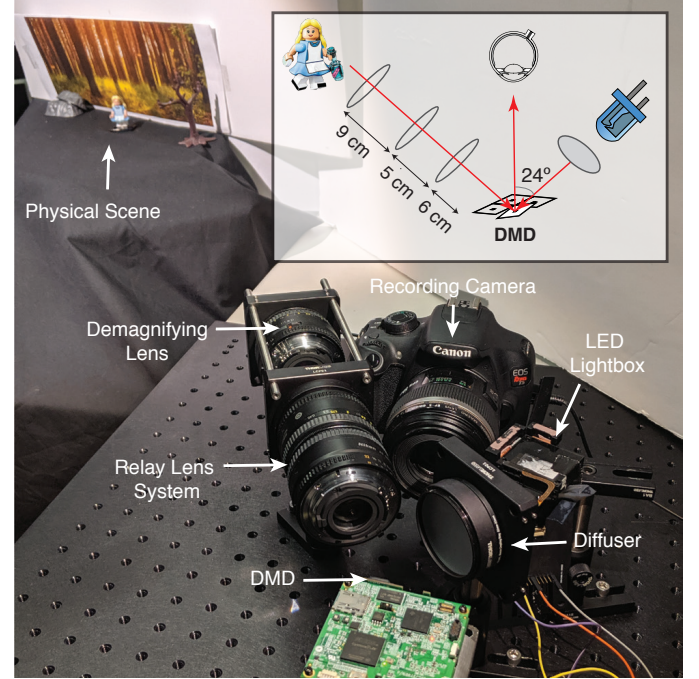


Fig. 4. Photograph of our factored occlusion display and the *Alice* physical scene. (inset) Optical schematic of the prototype. Light from the LED and physical scene are incident on the DMD from opposing but equal angles, then reflected toward the viewer depending on the state of each pixel.

The algorithm factors a target image into a set of binary DMD patterns and corresponding LED color values. We show 4 out of a total of 48 DMD patterns and LED values for the teaser scene in Fig. 3 (right). When the LED is black (upper left pattern), the DMD selectively blocks incident physical light without adding any additional light from the DMD. However, the color of the LED can take on arbitrary values for each of the DMD subframes.

4 IMPLEMENTATION

To evaluate the proposed approach to occlusion-capable OST-AR using a single SLM, we built a benchtop prototype (see Fig. 4). Component details, optical design, and system design are discussed in the following.

Display. We use a Texas Instruments LightCrafter Digital Light Projector unit, detaching the included lightbox module such that the conventional OFF state of the DMD is redirected toward a physical scene. To achieve even illumination of the DMD with the detached lightbox, the RGB-LED was placed behind a Thorlabs N-BK7 ground glass diffuser (1500 Grit) at the required angle of 24° from the normal.



Fig. 5. Demonstration of factored occlusion for rendering hard-edge occlusion and mutually consistent shadows. We refer to the scenes, in order from top to bottom, as *armchair*, *elephant*, and *tea party*. (left column) We combine each physical scene with a digital image to form a target composition. (center left column) Unable to block light from the scene, a conventional beamsplitter configuration produces largely transparent renderings. (center right column) In comparison, the composition captured with our approach shows significant improvements to both light blocking and color fidelity. Our factored occlusion algorithm computes (right column) the occlusion mask and digital content components, which multiplicatively combine with the physical scene on the single SLM to best approximate the target composition.

We connect an Arduino Uno to obtain 8-bit control of each of the RGB-LED color channels. The DMD itself has a resolution of 608×684 , a screen diagonal of $0.3''$, and a maximum frame rate of 4000 Hz.

Optical Design. To minimize distortion and chromatic aberrations in the prototype, we use three Nikon Nikkor 50-mm f/2 camera lenses to focus the physical scene on the DMD: one to demagnify the scene and two to relay and invert the image. Due to the bulkiness of the Nikon imaging lenses and the incident angle required for the DMD, the relay system is placed slightly further away (60 mm) from the DMD than ideal (50 mm). This gives a field of view of approximately 8.7° . We do not correct for the sheared focal plane created by the DMD (see Sect. 7 for discussion).

Recording Setup. A Canon EOS T5 Rebel camera with a Canon EF-S 60 mm f/2.8 lens is used to capture photographs of the display. For each result presented in this paper the camera settings were: 23 mm, f/18, ISO-200, and 2 sec exposure time. Occlusion mask and digital content factorizations are captured by blocking the light path from the RGB-LED or physical scene, respectively.

System Control. We use the graphical user interface (GUI) provided by Texas Instruments to control the mirror states of the DMD. A photograph with all mirrors turned to reflect the physical scene provides the knowledge of the background needed for our factorization algorithm (see Sect. 3.2). For images presented in this work we implement our method in Matlab, running 30 iterations to ensure convergence. The resulting set of binary micromirror states (**D**) is uploaded to the GUI as a series of 1-bit BMP files and the RGB-LED colors (**L**) as 8-bit color values to a simple Arduino LED control program. We also connect this Arduino to the DMD, facilitating synchronous triggering of the micromirror and RGB-LED states. The GUI sets a maximum number of patterns that can be loaded into memory, lim-

iting us to using 48 subframes per image. Although the DMD can switch at up to 4000 Hz, we use a 5-ms-long subframe (200 Hz) to minimize the effect of small errors in trigger synchronization on the displayed image, giving a frame rate of approximately 4.2 Hz.

System Calibration. Before capturing results, we calibrate the colors and intensities of light incident on the DMD from the physical scene and RGB-LED. By positioning a small color calibration target (Gretag Macbeth) in the physical scene, we first calibrate white balance of the recording camera. We then calibrate the intensities of each color of the RGB-LED by adding different resistors to each LED until it best matches the white target square in the scene. Finally, we adjust the lighting of the physical scene such that the intensities of white LED color and target square match.

Emulating Conventional Beamsplitter Displays. In addition to demonstrating occlusion-capable OST-AR, we use our display to emulate the image composition that would be obtained without any occlusion capability. This is achieved by capturing the digital image displayed on the DMD with the physical scene path blocked, representative of how it would conventionally be displayed on an SLM. We then add the digital image to the image of the physical scene (linearized using $\gamma = 2.2$).

5 RESULTS

The *rhino* scene (Fig. 1) shows a result demonstrating hard-edge occlusion and the rendering of shadows using our benchtop prototype (see supplementary material for simulation results). Comparing the digital component of the factorization (Fig. 1d) to the original digital image (Fig. 1a) clearly shows the factorization working to create a smooth rendering, despite the discontinuity of color and intensity in the background. More red light must be added to compensate for the areas where the rhino must occlude the green in the physical scene.

Similarly, other areas of the rhino require less occlusion. For example, the algorithm takes advantage of the yellow color in the background to help construct the digital image. The captured result (Fig. 1f) indeed shows an improvement in occlusion ability and rendered color of the rhino compared to that produced by a conventional beamsplitter configuration (Fig. 1g). In particular, the ability to block light from the physical scene means that realistic shadow effects can be rendered. However, as one might expect, our method is not without its trade-offs. It can be seen that the leaves of the tree are still visible through the rhino rendering, with the system being unable to recreate the complex shading of the rhino while simultaneously subtracting out the background.

In Fig. 5, we show several additional scenes demonstrating mutual occlusion and rendering of shadows. Much like the *rhino* scene, the conventional beamsplitter approach for these scenes results in digital images that appear largely transparent and lacking in color. In particular, the chair in the *armchair* scene and table in the *tea party* scene are barely visible. The tree in the *elephant* scene also creates noticeable artifacts in the rendering of the elephant. In comparison, our proposed factorization algorithm enables hard-edge occlusion, noticeably improving the rendering of these digital images, including the ability to render shadows. The factored occlusion mask and digital-only components show the concurrent addition and subtraction of light that makes this possible. For example, on the left side of the *elephant* rendering, patches that are more pink in color are present, corresponding to regions where color compensation is required to properly occlude the dark green color of the tree. While the *armchair* and *elephant* scenes are almost indistinguishable from their target compositions, the *tea party* scene shows noticeable deviation in the rendering of the table. Again, as with the *rhino* scene, this illustrates the trade-off between addition and subtraction of light within the optimization problem. In this case, the algorithm favors correctly rendering the brightly colored parts of the image (the teapot and cup), leaving the table overly blue and transparent. The perceptual effects of this trade-off can be mitigated by a locally weighted factorization, as discussed in the following.

6 TOWARD GAZE-CONTINGENT DISPLAY

Using the weighting matrix \mathbf{W} (Sect. 3.2), we can directly influence the relative importance of each scene part when the factorization algorithm has to make trade-offs. This enables a gaze-contingent display mode, where the weights of the region that the user fixates, i.e. the foveal region, are higher than those in the periphery of the user’s visual field. In this way, we demonstrate gaze-contingent occlusion display by optimizing the perceived image quality primarily for the foveal display region. As proof of concept, we select the gaze position manually, but we envision this being continuously updated using eye-tracking technology, as demonstrated in previous work for other near-eye display techniques [51].

The *bird* scene (Fig. 6) shows an example of the improvement attainable with gaze-contingent optimization, as captured using our prototype. The target composition contains both red and blue bird images on a green and orange background, respectively. This forces a rendering trade-off between adding red while reducing green or adding blue while reducing orange. The naïve global optimization lacks in color intensity and fine details, as expected. Including a weighting matrix (refer to the supplement for further details) to favor the region of the blue bird noticeably improves the intensity of the blue and the finer details of the bird’s wings are now discernable. On the other hand, the red bird fades in color and occludes the tree in the scene less; however, we do not expect this to be noticeable since it would be in the periphery of the visual field. Weighting the red bird instead reverses this, with the red bird exhibiting stronger color and occlusion and the blue bird losing intensity and detail. However, the birds are separated such that we do not expect them to be fixated simultaneously when using a near-eye display, and thus eye-tracked rendering in this way is viable.

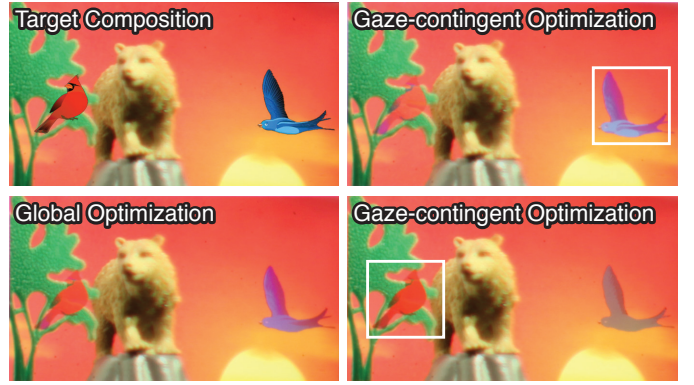


Fig. 6. Photographs of the gaze-contingent factored occlusion display mode with the *bird* scene. (top left) We create a target scene containing digital images of two different birds. With starkly different colors and physical backgrounds to occlude, (bottom left) applying our approach to form a best global approximation produces a composition that resembles, but cannot fully replicate this target. However, in a near-eye display system, the birds are likely to be viewed separately, allowing gaze-contingent rendering. We demonstrate that including a gaze-contingent weighting in our approach improves the rendering of the fixated object. (top right) Weighting the region containing the blue bird improves its color fidelity and increases the level of detail. (bottom right) Similarly, weighting the red bird improves its color fidelity and occlusion of the tree.

7 DISCUSSION

In summary, we introduce a new type of occlusion-capable OST-AR system, employing a single DMD to multiplicatively merge physical scene and digital content light paths on a pixel-by-pixel basis. We derive a factorization algorithm to optimize time-multiplexed binary DMD patterns and LED colors to approximate a target AR scene. This approach reduces the transparency and color artifacts produced by a conventional beamsplitter configuration and removes the additional SLM required by previously demonstrated occlusion-capable systems. Here we discuss the current limitations of the technique and current prototype, and suggest possible directions for future work.

Dynamic Range Limitations Our approach is unique in that it facilitates the use of the same SLM to both render digital content and hard-edge occlusion. This reduction in hardware complexity comes at the cost of a trade-off between proper occlusion and accurate color rendering. Although we derive a formal factorization algorithm to optimize this problem, the system is inherently limited in dynamic range. In many scenes, the rendering may necessarily be less accurate than in a system with multiple SLMs. We demonstrate this in Fig. 7, simulating the rendering of a complex target, forcing the rendering and occlusion of multiple different color combinations concurrently. Although significantly better than that produced by a conventional additive beamsplitter configuration, it is clear that the algorithm struggles to add all the light required to maintain color fidelity while also subtracting light for full occlusion. This can be locally improved by using the gaze-contingent optimization approach, but at the cost of further degrading other areas, as demonstrated by the selected region crops.

Perceptually-driven Optimization. The factorization algorithm we present is driven by a least squares loss function applied to linear intensities. While this works well for many scenes in practice (e.g., Figs. 1 and Fig. 5), reformulating the objective to minimize the error in a perceptually more uniform space, such as CIE Lab, could further improve the perceived quality of the results – albeit at the cost of increased computation. Kauvar et al. [26], for example, implemented a perceptually uniform loss for factored spectral displays which can be readily applied to our method. We leave this effort for future work.

Computational Requirements. The proposed system reduces hardware complexity at the cost of increased computational require-

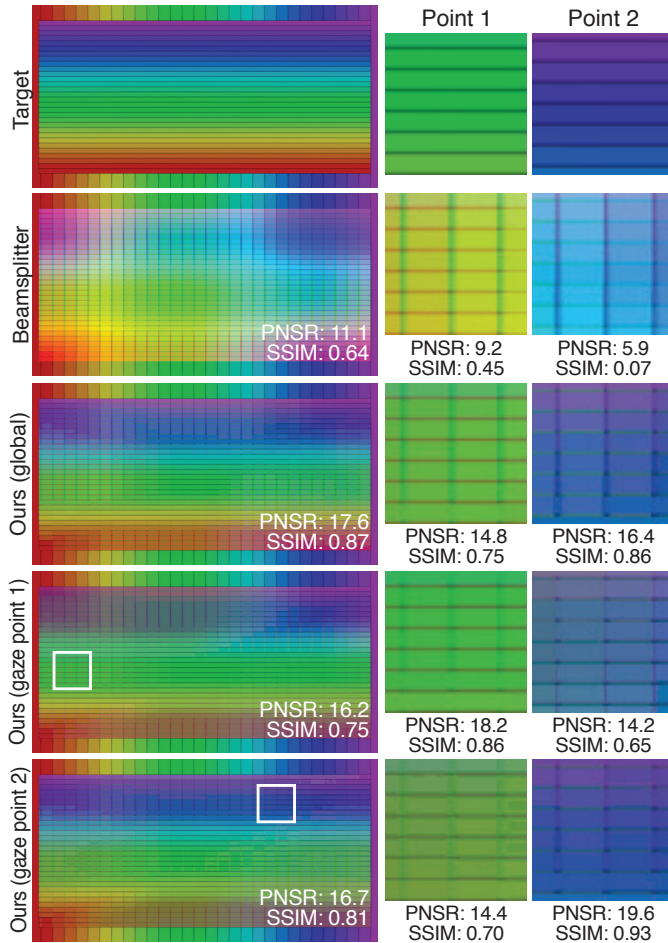


Fig. 7. *Rainbow scene simulations demonstrating dynamic range limitations.* We simulate a complex target composition: vertical rainbow patterns represent the physical scene and horizontal ones represent the digital content. This forces the algorithm to attempt to render and occlude many different color combinations concurrently. A conventional beamsplitter configuration results in large intensity and color artifacts. Our approach shows significant improvement in replicating the target, as demonstrated by the increased peak signal-to-noise ratio (PNSR) and structural similarity (SSIM) index values, although there are still some color artifacts and regions where the virtual overlay fails to completely occlude the background. We select two such regions and demonstrate that although the overall correspondence to the target decreases, gaze-contingent optimization improves the rendering accuracy of these regions. For direct comparison, we show enlarged views and calculate PNSR and SSIM for these regions across all discussed rendering modes.

ments. We obtain runtimes of about 15 seconds with an unoptimized Matlab implementation of the algorithm. However, although not immediately straightforward to implement, the proposed optimization consists almost entirely of matrix-based calculations and converges in only a few iterations, thus GPU implementation could enable near real-time frame rates, as demonstrated for other types of factored displays [64]. Moreover, computing resources on wearable devices quickly advance with custom processors being commercially deployed to optimize AR-specific task execution. Algorithms like the proposed could become a part of future application-specific integrated circuits (ASICs).

Alignment Challenges. Factored occlusion is based on a pixel-precise optimization given a known physical scene, and as such, requires an RGB image of this scene as it would be seen by the user. Since an RGB-depth camera would likely be needed to guide spa-

tial mapping, including determining boundaries needed for mutual occlusion rendering, this would not mandate any additional hardware. With both the camera and the DMD rigidly fixed, and eye-tracking functionality also seeing increased inclusion in HMDs, a homography transformation could be used to map the camera image to a perspective aligned to the user’s viewpoint. Inaccuracy in this transformation, though likely only a few pixels if implemented well, would result sub-optimal virtual and physical scene combinations, manifesting as color or occlusion artifacts (refer to the supplement for an example).

Furthermore, factored occlusion is currently effective only for static images, since scene motion within the rendering time would result in more significant misalignment. However, this is only limited by current computation speed, and thus real-time rendering with a GPU implementation could enable extension to moving scenes.

Optical Design Improvements. In this work we present a proof-of-concept implementation to demonstrate the operation of this new type of occlusion-capable OST-AR system. As such, our optical stack leaves out a few refinements. First, we do not correct for the shearing of the focal plane induced by the DMD micromirror tilt. As demonstrated previously [50], this is straightforward to overcome with transmissive or reflective diffraction optics. Second, the small field of view we observe is limited by the size of the DMD and the relatively long focal length of the relay optics. The use of a larger DMD and higher optical power, or even custom, optics could greatly improve this. Finally, although we demonstrate a monocular, fixed-focus prototype, our method is complementary to and can be combined with efforts to develop optics for varifocal [54, 65] and binocular displays [29]. As recently demonstrated by Kim et al. [28], employing mirrors to redirect the optical path of the physical scene could also be used to allow a user to see-through a real scene in the original direction.

System Miniaturization. Significant progress has also been made in reducing the device form factor of occlusion-capable systems [3, 65]. Although we reduce the number of SLMs required and thus enable shrinking the optical engine footprint, our current prototype uses rather bulky focusing optics. Despite having bulky optics, we place no limits on the type of optical elements that could be used, allowing a multitude of paths forward toward miniaturization.

Display Improvements. The DMD in our prototype is well short of the current state of the art. Off-the-shelf displays with higher pixel densities (over 1920×1080) and switching speeds (over 22 kHz) are readily purchasable. Simply swapping out our DMD with a new one could therefore both markedly improve the resolution of the scenes we demonstrate and also enable faster overall frame rates beyond the human flicker fusion threshold. Further evidencing the potential of DMD-based display systems, the technology has also recently seen adoption in commercial display devices such as Avegant’s VR and AR headsets.

Advancing electrochromatic mirror technology could also see the development of new display technologies [24, 57]. One such technology could be an array of switchable mirrors: the functional equivalent of a transmissive DMD. With these emerging SLM designs, we would no longer require a folded optical path for the physical scene, and the virtual image can be projected from the side as is typical in modern near-eye displays. Co-design of this type of transmissive DMD with our factored occlusion approach would represent a significant step forward in OST-AR system design.

8 CONCLUSION

Achieving accurate hard-edge occlusion is critical in realizing seamless, perceptually realistic OST-AR experiences. With this work, we present a computational approach to achieving such occlusion, forgoing the traditional additive approach and instead computing a multiplicative blend of real and digital content on a single SLM. We demonstrate in simulations and on a prototype benchtop display the ability to render hard-edge occlusions and plausible shadows, and further demonstrate a gaze-contingent optimization of this novel display mode. Going forward, accurate rendering of hard-edge occlusion in

OST-AR systems will require advances in optical assembly, such as those from the myriad efforts complementary to our own, but also novel computational strategies to overcome the physical impossibility of creating negative light. Factored occlusion represents the first in this line of computational approaches.

REFERENCES

- [1] K. Afsit, W. Lopes, J. Kim, P. Shirley, and D. Luebke. Near-eye Varifocal Augmented Reality Display Using See-through Screens. *ACM Trans. Graph.*, 36(6):189:1–189:13, 2017.
- [2] K. Akeley, S. J. Watt, A. R. Girshick, and M. S. Banks. A stereo display prototype with multiple focal distances. *ACM Trans. Graph. (SIGGRAPH)*, 23(3):804–813, 2004.
- [3] O. Cakmakci, Yonggang Ha, and J. P. Rolland. A compact optical see-through head-worn display with occlusion support. In *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 16–25, Nov. 2004. doi: 10.1109/ISMAR.2004.2
- [4] J. Chakraborty, B. Furht, M. Anisetti, P. Ceravolo, E. Damiani, and M. Irvovic. Augmented reality technologies, systems and applications. *Multimedia Tools and Applications*, 51(1):341–377, Jan. 2011. doi: 10.1007/s11042-010-0660-6
- [5] P. Chakravarthula, D. Dunn, K. Afsit, and H. Fuchs. Focusar: Auto-focus augmented reality eyeglasses for both real world and virtual imagery. *IEEE transactions on visualization and computer graphics*, 24(11):2906–2916, 2018.
- [6] J.-H. R. Chang, B. V. K. V. Kumar, and A. C. Sankaranarayanan. 2¹⁶ shades of gray: high bit-depth projection using light intensity control. *Opt. Express*, 24(24):27937–27950, 2016.
- [7] J.-H. R. Chang, B. V. K. V. Kumar, and A. C. Sankaranarayanan. Towards multifocal displays with dense focal stacks. *ACM Trans. Graph. (SIGGRAPH Asia)*, 37(6):198:1–198:13, 2018.
- [8] J. Cutting and P. Vishton. Perceiving layout and knowing distances: The interaction, relative potency, and contextual use of different information about depth. In W. Epstein and S. Rogers, eds., *Perception of Space and Motion*, chap. 3, pp. 69–117. Academic Press, 1995.
- [9] J. David Hincapi-Ramos, L. Ivanchuk, S. K. Sridharan, and P. P. Irani. SmartColor: Real-Time Color and Contrast Correction for Optical See-Through Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics*, 21(12):1336–1348, Dec. 2015. doi: 10.1109/TVCG.2015.2450745
- [10] D. Dunn, C. Tippets, K. Torell, P. Kellnhofer, K. Akit, P. Didyk, K. Myszkowski, D. Luebke, and H. Fuchs. Wide Field Of View Varifocal Near-Eye Display Using See-Through Deformable Membrane Mirrors. *IEEE TVCG*, 23(4):1322–1331, 2017.
- [11] H. Fuchs, M. A. Livingston, R. Raskar, D. Colucci, K. Keller, A. State, J. R. Crawford, P. Rademacher, S. H. Drake, and A. A. Meyer. Augmented reality visualization for laparoscopic surgery. In W. M. Wells, A. Colchester, and S. Delp, eds., *Medical Image Computing and Computer-Assisted Intervention MICCAI98*, Lecture Notes in Computer Science, pp. 934–943. Springer Berlin Heidelberg, 1998.
- [12] C. Gao, Y. Lin, and H. Hua. Occlusion capable optical see-through head-mounted display using freeform optics. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 281–282, Nov. 2012. doi: 10.1109/ISMAR.2012.6402574
- [13] B. Guenter, M. Finch, S. Drucker, D. Tan, and J. Snyder. Foveated 3d graphics. *ACM Trans. Graph. (SIGGRAPH Asia)*, 31(6):164:1–164:10, 2012.
- [14] T. Hamasaki and Y. Itoh. Varifocal Occlusion for Optical See-Through Head-Mounted Displays using a Slide Occlusion Mask. *IEEE Transactions on Visualization and Computer Graphics*, 25(5):1961–1969, May 2019. doi: 10.1109/TVCG.2019.2899249
- [15] F. Heide, D. Lanman, D. Reddy, J. Kautz, K. Pulli, and D. Luebke. Cascaded displays: spatiotemporal superresolution using offset pixel layers. *ACM Trans. Graph. (SIGGRAPH)*, 33(4):60, 2014.
- [16] M. Hirsch, G. Wetzstein, and R. Raskar. A compressive light field projection system. *ACM Trans. Graph. (SIGGRAPH)*, 33(4):58, 2014.
- [17] N.-D. Ho. *Nonnegative matrix factorization algorithms and applications*. PhD thesis, Université catholique de Louvain, 2008.
- [18] X. Hu and H. Hua. Design and assessment of a depth-fused multi-focal plane display prototype. *J. Disp. Technol.*, 10(4):308–316, 2014.
- [19] H. Hua and B. Javidi. A 3D integral imaging optical see-through head-mounted display. *Opt. Express*, 22(11):13484–13491, 2014.
- [20] F.-C. Huang, K. Chen, and G. Wetzstein. The light field stereoscope: Immersive computer graphics via factored near-eye light field display with focus cues. *ACM Trans. Graph. (SIGGRAPH)*, 34(4), 2015.
- [21] Y. Itoh, M. Dzitsiuk, T. Amano, and G. Klinker. Semi-Parametric Color Reproduction Method for Optical See-Through Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics*, 21(11):1269–1278, Nov. 2015. doi: 10.1109/TVCG.2015.2459892
- [22] Y. Itoh, T. Hamasaki, and M. Sugimoto. Occlusion Leak Compensation for Optical See-Through Displays Using a Single-Layer Transmissive Spatial Light Modulator. *IEEE Transactions on Visualization and Computer Graphics*, 23(11):2463–2473, Nov. 2017. doi: 10.1109/TVCG.2017.2734427
- [23] Y. Itoh, T. Langlotz, D. Iwai, K. Kiyokawa, and T. Amano. Light Attenuation Display: Subtractive See-Through Near-Eye Display via Spatial Color Filtering. *IEEE Transactions on Visualization and Computer Graphics*, 25(5):1951–1960, May 2019. doi: 10.1109/TVCG.2019.2899229
- [24] K. R. Jeong, I. Lee, J. Y. Park, C. S. Choi, S.-H. Cho, and J.-L. Lee. Enhanced black state induced by spatial silver nanoparticles in an electrochromic device. *NPG Asia Materials*, 9(3):e362, Mar. 2017. doi: 10.1038/am.2017.25
- [25] P. V. Johnson, J. A. Parnell, J. Kim, C. D. Saunter, G. D. Love, and M. S. Banks. Dynamic lens and monovision 3d displays to improve viewer comfort. *OSA Opt. Express*, 24(11):11808–11827, 2016.
- [26] I. Kauvar, S. J. Yang, L. Shi, I. McDowall, and G. Wetzstein. Adaptive color display via perceptually-driven factored spectral projection. *ACM Trans. Graph. (SIGGRAPH Asia)*, 34(6):165–1, 2015.
- [27] J. Kim, Y. Jeong, M. Stengel, K. Afsit, R. Albert, B. Boudaoud, T. Greer, J. Kim, W. Lopes, Z. Majercik, P. Shirley, J. Spjut, M. McGuire, and D. Luebke. Foveated ar: Dynamically-foveated augmented reality display. *ACM Trans. Graph. (SIGGRAPH)*, 38(4):99:1–99:15, 2019.
- [28] K. Kim, D. Heo, and J. Hahn. Occlusion-capable Head-mounted Display. In *PHOTOPTICS*, 2019.
- [29] K. Kiyokawa, M. Billingham, B. Campbell, and E. Woods. An occlusion capable optical see-through head mount display for supporting co-located collaboration. In *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings.*, pp. 133–141, Oct. 2003. doi: 10.1109/ISMAR.2003.1240696
- [30] K. Kiyokawa, Y. Kurata, and H. Ohno. An optical see-through display for mutual occlusion of real and virtual environments. In *Proceedings IEEE and ACM International Symposium on Augmented Reality (ISAR 2000)*, pp. 60–67, Oct. 2000. doi: 10.1109/ISAR.2000.880924
- [31] K. Kiyokawa, Y. Kurata, and H. Ohno. An optical see-through display for mutual occlusion with a real-time stereovision system. *Computers & Graphics*, 25(5):765–779, Oct. 2001. doi: 10.1016/S0097-8493(01)00119-4
- [32] R. Konrad, E. A. Cooper, and G. Wetzstein. Novel optical configurations for virtual reality: Evaluating user preference and performance with focus-tunable and monovision near-eye displays. In *Proc. ACM SIGCHI*, pp. 1211–1220, 2016.
- [33] F. L. Kooi and A. Toet. Visual comfort of binocular and 3d displays. *Displays*, 25(2):99–108, Aug. 2004. doi: 10.1016/j.displa.2004.07.004
- [34] E. Kruijff, J. E. Swan, and S. Feiner. Perceptual issues in augmented reality revisited. In *2010 IEEE International Symposium on Mixed and Augmented Reality*, pp. 3–12, Oct. 2010. doi: 10.1109/ISMAR.2010.5643530
- [35] P.-Y. Laffont, A. Hasnain, P.-Y. Guillemet, S. Wirajaya, J. Khoo, D. Teng, and J.-C. Bazin. Varifocal: A platform for vision correction and accommodation in head-mounted displays. In *ACM SIGGRAPH 2018 Emerging Technologies*, pp. 21:1–21:2, 2018.
- [36] T. Langlotz, M. Cook, and H. Regenbrecht. Real-Time Radiometric Compensation for Optical See-Through Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics*, 22(11):2385–2394, Nov. 2016. doi: 10.1109/TVCG.2016.2593781
- [37] D. Lanman, M. Hirsch, Y. Kim, and R. Raskar. Content-adaptive parallax barriers: optimizing dual-layer 3d displays using low-rank light field factorization. *ACM Transactions on Graphics (SIGGRAPH Asia)*, 29(6):163, 2010.
- [38] D. Lanman and D. Luebke. Near-eye light field displays. *ACM Trans. Graph. (SIGGRAPH Asia)*, 32(6):220:1–220:10, 2013.
- [39] S. Lee, J. Cho, B. Lee, Y. Jo, C. Jang, D. Kim, and B. Lee. Foveated Retinal Optimization for See-Through Near-Eye Multi-Layer Displays. *IEEE Access*, 6:2170–2180, 2018. doi: 10.1109/ACCESS.2017.2782219

- [40] S. Liu, D. Cheng, and H. Hua. An optical see-through head mounted display with addressable focal planes. In *Proc. IEEE ISMAR*, pp. 33–42, 2008.
- [41] M. A. Livingston, J. E. Swan, J. L. Gabbard, T. H. Hollerer, D. Hix, S. J. Julier, Y. Baillet, and D. Brown. Resolving multiple occluded layers in augmented reality. In *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings.*, pp. 56–65, Oct. 2003. doi: 10.1109/ISMAR.2003.1240688
- [42] P. Llull, N. Bedard, W. Wu, I. Tosic, K. Berkner, and N. Balram. Design and optimization of a near-eye multifocal display system for augmented reality. In *OSA Imaging Appl. Opt.*, 2015.
- [43] G. D. Love, D. M. Hoffman, P. J. W. Hands, J. Gao, A. K. Kirby, and M. S. Banks. High-speed switchable lens enables the development of a volumetric stereoscopic display. *Opt. Express*, 17(18):15716–25, 2009.
- [44] A. Maimone and H. Fuchs. Computational augmented reality eyeglasses. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 29–38, Oct. 2013. doi: 10.1109/ISMAR.2013.6671761
- [45] A. Maimone, D. Lanman, K. Rathinavel, K. Keller, D. Luebke, and H. Fuchs. Pinlight Displays: Wide Field of View Augmented Reality Eyeglasses Using Defocused Point Light Sources. *ACM Trans. Graph.*, 33(4):89:1–89:11, July 2014. doi: 10.1145/2601097.2601141
- [46] B. Masia, G. Wetzstein, P. Didyk, and D. Gutierrez. A survey on computational displays: Pushing the boundaries of optics, computation, and perception. *Computers & Graphics*, 37(8):1012–1038, 2013.
- [47] O. Mercier, Y. Sulai, K. Mackenzie, M. Zannoli, J. Hillis, D. Nowrouzezahrai, and D. Lanman. Fast Gaze-contingent Optimal Decompositions for Multifocal Displays. *ACM Trans. Graph.*, 36(6):237:1–237:15, Nov. 2017. doi: 10.1145/3130800.3130846
- [48] S. Mori, S. Ikeda, A. Plopski, and C. Sandor. BrightView: Increasing Perceived Brightness of Optical See-Through Head-Mounted Displays Through Unnoticeable Incident Light Reduction. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 251–258, Mar. 2018. doi: 10.1109/VR.2018.8446441
- [49] R. Narain, R. A. Albert, A. Bulbul, G. J. Ward, M. S. Banks, and J. F. O’Brien. Optimal Presentation of Imagery with Focus Cues on Multi-plane Displays. *ACM Trans. Graph.*, 34(4):59:1–59:12, July 2015. doi: 10.1145/2766909
- [50] M. O’Toole, J. Mather, and K. N. Kutulakos. 3d shape and indirect appearance by structured light transport. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(7):1298–1312, July 2016. doi: 10.1109/TPAMI.2016.2545662
- [51] N. Padmanaban, R. Konrad, T. Stramer, E. A. Cooper, and G. Wetzstein. Optimizing virtual reality for all users through gaze-contingent and adaptive focus displays. *PNAS*, 114:2183–2188, 2017.
- [52] A. Patney, M. Salvi, J. Kim, A. Kaplanyan, C. Wyman, N. Bentley, D. Luebke, and A. Lefohn. Towards foveated rendering for gaze-tracked virtual reality. *ACM Trans. Graph. (SIGGRAPH Asia)*, 35(6):179:1–179:12, 2016.
- [53] K. Rathinavel, H. Wang, A. Blate, and H. Fuchs. An extended depth-at-field volumetric near-eye augmented reality display. *IEEE TVCG*, 24(11):2857–2866, 2018.
- [54] K. Rathinavel, G. Wetzstein, and H. Fuchs. Varifocal occlusion-capable optical see-through augmented reality display based on focus-tunable optics. *IEEE TVCG (Proc. ISMAR)*, 2019.
- [55] J. P. Rolland and H. Fuchs. Optical Versus Video See-Through Head-Mounted Displays in Medical Visualization. *Presence*, 9(3):287–309, June 2000. doi: 10.1162/105474600566808
- [56] J. P. Rolland, M. W. Krueger, and A. Goon. Multifocal planes head-mounted displays. *Appl. Opt.*, 39(19):3209–3215, 2000.
- [57] D. R. Rosseinsky and R. J. Mortimer. Electrochromic Systems and the Prospects for Devices. *Advanced Materials*, 13(11):783–793, 2001. doi: 10.1002/1521-4095(200106)13:11<783::AID-ADMA783>3.0.CO;2-D
- [58] J.-H. Ryu, J.-W. Kim, K.-K. Lee, and J.-O. Kim. Colorimetric background estimation for color blending reduction of OST-HMD. In *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, pp. 1–4, Dec. 2016. doi: 10.1109/APSIPA.2016.7820764
- [59] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs. High dynamic range display systems. *ACM Trans. Graph. (SIGGRAPH)*, 23(3):760–768, 2004.
- [60] S. Shiwa, K. Omura, and F. Kishino. Proposal for a 3-d display with accommodative compensation: 3ddac. *Journal of the Society for Information Display*, 4(4):255–261, 1996.
- [61] S. K. Sridharan, J. D. Hincapi-Ramos, D. R. Flatla, and P. Irani. Color Correction for Optical See-through Displays Using Display Color Profiles. In *Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology, VRST ’13*, pp. 231–240. ACM, New York, NY, USA, 2013. event-place: Singapore. doi: 10.1145/2503713.2503716
- [62] T. Uchida, K. Sato, and S. Inokuchi. An Optical See-through MR Display with Digital Micro-mirror Device. *Transactions of the Virtual Reality Society of Japan*, 7(2), 2002.
- [63] G. Wetzstein, W. Heidrich, and D. Luebke. Optical Image Processing Using Light Modulation Displays. *Computer Graphics Forum*, 29(6):1934–1944, 2010. doi: 10.1111/j.1467-8659.2010.01660.x
- [64] G. Wetzstein, D. Lanman, M. Hirsch, and R. Raskar. Tensor displays: compressive light field synthesis using multilayer displays with directional backlighting. *ACM Transactions on Graphics (SIGGRAPH)*, 2012.
- [65] A. Wilson and H. Hua. Design and prototype of an augmented reality display with per-pixel mutual occlusion capability. *Optics Express*, 25(24):30539–30549, Nov. 2017. doi: 10.1364/OE.25.030539
- [66] Y. Yamaguchi and Y. Takaki. See-through integral imaging display with background occlusion capability. *Applied Optics*, 55(3):A144–A149, Jan. 2016. doi: 10.1364/AO.55.00A144