Varifocal Occlusion-Capable Optical See-through Augmented Reality Display based on Focus-tunable Optics

Kishore Rathinavel, Gordon Wetzstein, and Henry Fuchs

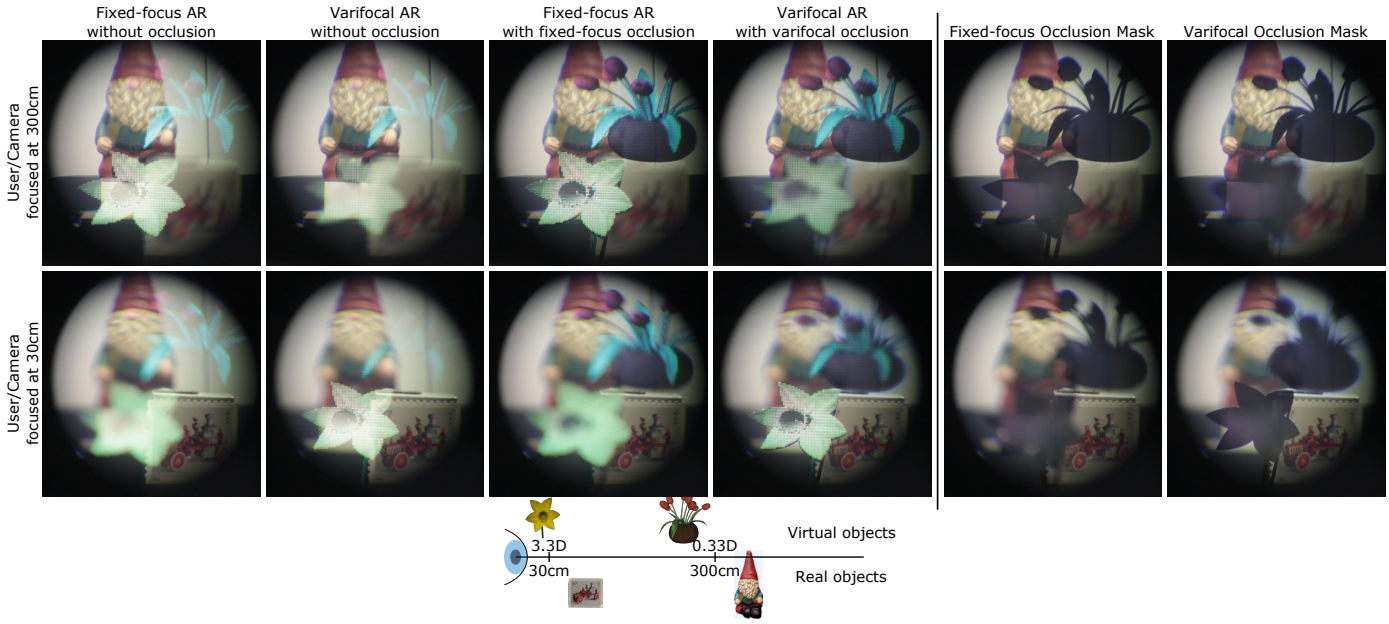

Fig. 1: **Left of the vertical line:** Photographs of an augmented reality (AR) display demonstrating various options for focusing the digital display as well as the occlusion layer. In this scene, the stamp and gnome are physical objects located at a close and far distance to the user, respectively. The AR display overlays two digital objects on this scene: a daffodil and a tulip flower pot, which are also located at a close and far distance. *Column 1:* conventional AR displays show the digital content at a fixed focal plane and do not support mutually consistent occlusions between digital and physical objects. For this reason, the digital content appears transparent and it is only focused at a single incorrect distance. *Column 2:* varifocal AR displays allow the digital content to appear at the correct distances, but without occlusion support this content still appears washed out. *Column 3:* previous proposals on fixed-focus occlusion-capable AR displays allow for consistent occlusions but only at a single distance, severely limiting image quality at out-of-focus distances. *Column 4:* the proposed varifocal occlusion-capable display technology supports hard-edge occlusion at various distances and thus creates perceptually more realistic experiences. **Right of the vertical line:** Comparison of occlusion masks between fixed-focus and varifocal occlusion-capable displays.

Abstract— Optical see-through augmented reality (AR) systems are a next-generation computing platform that offer unprecedented user experiences by seamlessly combining physical and digital content. Many of the traditional challenges of these displays have been significantly improved over the last few years, but AR experiences offered by today's systems are far from seamless and perceptually realistic. Mutually consistent occlusions between physical and digital objects are typically not supported. When mutual occlusion is supported, it is only supported for a fixed depth. We propose a new optical see-through AR display system that renders mutual occlusion in a depth-dependent, perceptually realistic manner. To this end, we introduce varifocal occlusion displays based on focus-tunable optics, which comprise a varifocal lens system and spatial light modulators that enable depth-corrected hard-edge occlusions for AR experiences. We derive formal optimization methods and closed-form solutions for driving this tunable lens system and demonstrate a monocular varifocal occlusion-capable optical see-through AR display capable of perceptually realistic occlusion across a large depth range.

Index Terms—Augmented Reality, Computational Displays, Varifocal Display, Occlusion

1 Introduction

Augmented Reality (AR) systems offer unprecedented experiences and are considered a next-generation computing platform. These wearable displays promise to seamlessly augment the physical world around us with digital content, such as information displays or user interfaces. Providing a seamless, perceptually realistic experience, however, requires the display to accurately support all depth cues of the human visual system [16, 38]. While current AR displays offer impressive capabilities, they typically do not support the most important depth cue: occlusion [11].

- *Kishore Rathinavel and Henry Fuchs are with UNC Chapel Hill. E-mail: {kishore | fuchs}@cs.unc.edu.*
- *Gordon Wetzstein is with Stanford University. E-mail: gordon.wetzstein@stanford.edu.*

Manuscript received 22 Mar. 2019; accepted 8 July. 2019.
Date of publication 13 Aug. 2019; date of current version 25 Oct. 2019.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TVCG.2019.2933120

Providing accurate, i.e., mutually consistent and hard-edge, occlusion between digital and physical objects with optical see-through AR displays is a major challenge. When digital content is located in front of physical objects, the former usually appear semi-transparent and unrealistic (see Fig. 1, columns 1 and 2). To adequately render these objects, the light reflected off of the physical object toward the user has to be blocked by the display before impinging on their retina. This occlusion mechanism needs to be programmable to support dynamic scenes and it needs to be perceptually realistic to be effective. The latter implies that occlusion layers are correctly rendered at the distances of the physical objects (see Fig. 2), allowing for pixel-precise, or hard-edge, control of the transmitted light rays.

Recent proposals on occlusion-capable optical see-through (OST) displays have only partially addressed this challenge. Global dimming [35], for example, is successful in controlling the light transmission of the display but without spatial control. Image-forming systems [7, 13, 21] enable consistent occlusions, but these are only correct at a single distance, severely limiting the image quality at other depths (see Fig. 1, column 3) and requiring bulky relay optics. Spatial light modulators (SLMs) for occlusion control can also be used without relay optics [18], but these will always be out of focus and require additional compensation techniques. Light field-based occlusion technology [32] offers somewhat sharper occlusion control without relay optics. Out-of-focus SLMs [18, 32] are usually based on liquid crystal displays (LCDs), which introduce diffraction artifacts of the physical world observed in OST displays, thus limiting the perceived image quality.

With this work, we introduce varifocal occlusion-capable optical see-through AR displays. These systems aim at providing a seamless and perceptually realistic experience by providing mutually consistent occlusions over a large depth range (see Fig. 1, column 4). Similar to varifocal near-eye displays, our approach uses focus-tunable lenses to dynamically shift the occlusion SLM to a single, but adaptive, optical distance. We envision this approach to operate in a gaze-contingent mode, where an eye tracker determines the distance of the fixated object and both the digital content and the occlusion system are dynamically focused at this distance.

A unique challenge of varifocal occlusion implemented with focus-tunable optics is precise control of the optical distortion of the physical light. As lenses change their focal power to align the occlusion SLM with different distance of the physical scene, the latter may also be magnified and its perceived distance altered, because the light of the physical scene and the occlusion SLM must share the same optical path. We derive a formal optimization approach and real-time heuristics to drive the proposed system in a perceptually accurate manner, preventing optical distortions of the physical world.

Specifically, we make the following contributions:

1. We introduce varifocal occlusion as an AR display capability that adaptively changes the focal distance of an occlusion mask to enable hard-edge occlusion over a large depth range.
2. We develop an optimization-based optical design approach for our focus-tunable optical system to achieve varifocal occlusion in a perceptually realistic manner without optically distorting the observed scene.
3. Using insights gained from the optimization approach, we use a ray-transfer matrix approaches to derive closed-form solutions for optical designs that allow for varifocal occlusion in real time.
4. We implement a monocular varifocal occlusion-capable AR display and demonstrate improved realism through depth-dependent occlusion.

2 Related Work

2.1 Varifocal Near-eye Displays

Varifocal displays are similar to conventional fixed-focus near-eye displays, but they dynamically adjust the distance of the magnified virtual image. This can be achieved using focus-tunable lenses [20, 24, 26, 31, 36, 39], deformable membranes [9, 12], or by mechanically actuating optical components [1, 36, 41]. Varifocal displays require eye tracking to determine the distance of the fixated object, to which the display is then focused in a gaze-contingent manner.

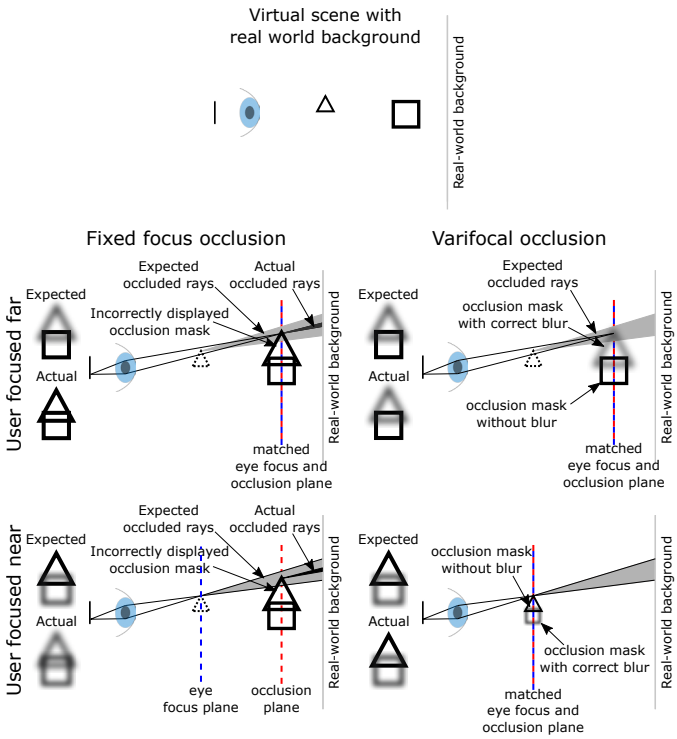

Fig. 2: *Topmost Row:* A virtual scene composed of one near and one far object placed in front of a real-world background. *Grid of figures:* Comparison of occlusion mechanism only (i.e. ignoring the digital or color image) for fixed-focus and varifocal occlusion displays for the above scene. Dashed blue and red lines indicate the user's focal plane and display's occlusion image plane respectively. Solid black lines indicate image formation for content placed in the user's focal plane. Images next to the eye show the "Expected" and "Actual" images seen by the user. Note that for fixed-focus occlusion, the occlusion plane is always at the far distance which causes the nearby object's occlusion mask to be seen incorrectly always and the far object's occlusion mask to be seen incorrectly when the eye is focused nearby. Varifocal occlusion-capable displays, on the other hand, move the occlusion plane to the user's focal plane and display an occlusion mask for in-focus objects as it is and a perceptually correct occlusion mask for out-of-focus objects by applying a computational blur.

Previous work on varifocal near-eye displays has primarily sought to adjust the virtual image of the digitally displayed content, primarily to mitigate the vergence-accommodation conflict [25, 27].

In this work, we extend the concept of varifocal displays to the problem of mutually consistent occlusion in AR, where the focus distance of an occlusion SLM is dynamically updated with the goal of improving perceptual realism. We discuss optical design strategies and demonstrate a varifocal occlusion-capable AR display that dynamically adjusts the focus of both digital image and occlusion SLM.

2.2 Occlusion-capable AR displays

2.2.1 Projection-based Lighting

Projection displays can be used to control the lighting of a scene in a spatially varying manner. Using such controlled illumination, mutually consistent occlusions, shading effects, and shadows in projector-based AR systems can be synthesized [2–4, 34]. The primary disadvantages of these systems are that projectors are required for the AR experience, which are not necessarily portable or wearable, and that they may not work in the presence of ambient illumination. We aim for a fully integrated occlusion-capable AR display that does not require additional

Products/Prototypes	AR focus mechanism	Occlusion focus mechanism
HoloLens, Meta2, MagicLeap, etc.	Fixed-focus	None
Itoh et al. [18]	Fixed-focus	Soft-edge
ELMO [21], Howlett and Smithwick [17], Cakmakci et al. [7]	Fixed-focus	Fixed-focus
Dunn et al. [12], Aksit et al. [1]	Varifocal	None
Hamasaki and Itoh [15], This work	**Varifocal**	**Varifocal**

Table 1: Summary of the type of focus cues that are supported for the virtual imagery and for occlusion by current AR products, previous research prototypes, and this work.

projectors.

2.2.2 Global Dimming

Commercial AR displays (e.g., Microsoft HoloLens, Magic Leap) often use a neutral density filter placed on the outside of the display module to reduce ambient light uniformly across the entire field of view. An adaptive version of global dimming was recently proposed by Mori et al. [35], where the amount of dimming is controlled by a single liquid crystal cell and responsive to its physical environment. While these approaches may be useful in some scenarios, they do not provide spatial control of the occlusion layer.

2.2.3 Fixed-focus Occlusion

The physical scene can be focused onto an occlusion SLM which selectively blocks its transmission in a spatially varying manner before it reaches the user's eye. This idea was first proposed by the seminal work of Kiyokawa et al. [21–23]. Improvements of related systems were later demonstrated [6, 7, 13, 14, 17, 43, 45].

Unfortunately, focusing a scene on an SLM usually requires a bulky optical system, first to focus it to the SLM, then to negate the effect of the first lens, and then to flip the resulting image the right way up. Moreover, as this approach only focuses a single distance of the scene on the occlusion SLM, hard-edge occlusion is only achieved at this fixed focus distance. This limitation is similar to the characteristics of fixed-focus near-eye displays, which has been alleviated by varifocal displays. In this work, we propose an extension of the concept of varifocal displays to occlusion.

Two key challenges for fixed-focus occlusion-capable displays are: (1) to ensure unit magnification of the see-through scene and (2) to ensure zero viewpoint offset between the see-through scene and the real-scene as seen without the display, so that the images of the real-world objects are at the correct distance. Both of these considerations are significantly more challenging for varifocal occlusion displays because unit magnification and zero viewpoint offset needs to be ensured while adjusting the focus of the SLM, which shares the optical path with the physical scene.

Kiyokawa et al. [21] derive optical design parameters that satisfy unit magnification for all real-world object distances and also propose an interesting geometric configuration of the optical components that make the offset between the real world objects and their images equal to zero. Cakmakci et al. [7] propose a compact optical design that satisfies the magnification requirements, but it does not achieve zero offset between the real viewpoint and the virtual viewpoint; however, the offset is small (5 cm). Howlett and Smithwick [17] propose an optical design approach based on ray-transfer matrices to achieve unit magnification and zero viewpoint offset, which is in turn inspired by optical cloaking [10]. We extend the optical design approach based on ray-transfer matrices to varifocal occlusion displays, and generalize the theory to asymmetrical optical designs.

2.2.4 Soft-edge Occlusion

To avoid a bulky optical system, a single LCD can be placed directly in front of the user's eyes [18, 43]. However, due to the fact that the occlusion LCD is out of focus, it always appears blurred. Itoh et al. [18] recently proposed to compensate for this blur by modifying the digitally displayed image. Such an approach could be interpreted as a hybrid optical see-through and video see-through AR display. Calibrating such a system requires extremely precise alignment and the mismatch in resolution (spatial and angular), latency, brightness, contrast, and color fidelity between digital display and physical world may contribute to perceived inconsistency and reduced perceptual realism in such a system [40]. Maimone et al. [33] also used an out-of-focus LCD, where the occlusion mask is calculated as the silhouette of the virtual object. None of these approaches achieves hard-edge occlusion, which severely limits perceptual realism.

2.2.5 Light Field Occlusion

Maimone and Fuchs [32] propose a 4D light field occlusion mask using stacked LCD layers placed out of focus in front of the eye, where the occluding patterns are calculated by light field factorization algorithms [30, 44]. The advantage of light field occlusion is that depth-dependent occlusion can be presented for virtual content at different depths simultaneously in a compact form factor. In practice, see-through LCDs mounted close to the eye are light inefficient and result in significant diffraction artifacts, which are due to the electronic components in each pixel as well as the wiring of the display panel. This effect significantly degrades the observed image quality of any soft-edge or light field occlusion system.

Another approach for light field occlusion is presented in [46] using concepts of integral imaging systems. This system has a very narrow field of view (4.3 degrees) and is fundamentally limited by the spatio-angular resolution tradeoff as well as diffraction.

As opposed to any of these methods, the proposed varifocal occlusion approach achieves hard-edge occlusion at varying distances in the scene at high resolution, with better light efficiency, and using technology components that make it easily compatible with emerging varifocal near-eye display.

2.2.6 Varifocal Occlusion

Concurrently and independently of our work, Hamasaki and Itoh [15] also developed a strategy for a varifocal occlusion-capable AR display. Unlike our approach that builds on focus-tunable optics to dynamically adjust the depth of the occlusion layer, their approach requires mechanical motion of the occlusion SLM. Each approach has certain benefits and limitations. For example, robust calibration of the mechanically moving parts in their approach can be challenging, especially in a wearable display form factor. Our approach, on the other hand, requires focus-tunable optics, such as liquid lenses or Alvarez lenses (see Sec. 2.1).

2.3 Consistent Colors, Shading, and Shadows in AR

Spatial AR systems and optical see-through AR display often aim at providing radiometrically consistent, color-corrected or even color-stylized imagery [5, 19, 28, 29, 43]. All of these approaches are successful in enhancing the viewing experience in AR, but none of them tackles the problem of mutually consistent occlusions in optical see-through AR displays.

3 Optical Design

Our goal is to design a varifocal occlusion-capable OST AR display that satisfies several key requirements. These include

1. The virtual image of the occlusion SLM, i.e. the occlusion mask, and the digital image should be optically placed together in the scene and their distance be dynamically adjustable.
2. The lateral and longitudinal magnification of the physical scene seen through the display should be equal to one, such that the experience is similar to viewing the scene without any optical elements.
3. No mechanical motion should be introduced to any component (lenses, SLM, etc.) to adjust the distance of its virtual image. Instead, the virtual image should be moved by changing the focal powers of the employed lenses.

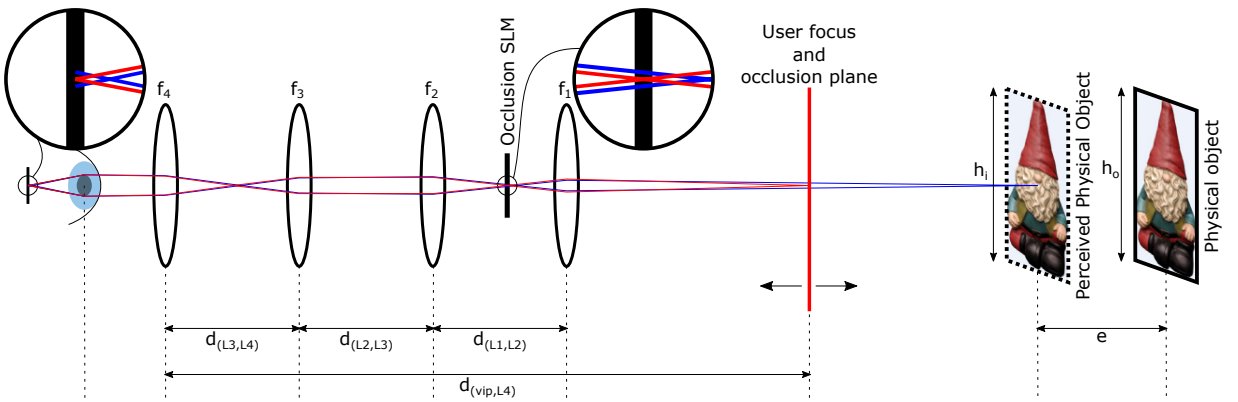

Fig. 3: Illustration of the unfolded optical path of a 4-lens system for image-forming occlusion-capable AR displays. With a varifocal display, the distance of virtual image and occlusion mask matches the user's focus distance, indicated by the thick vertical red line. Blue and red lines going from points in the scene through the optics onto the retina indicate ray diagrams for the image formation of the virtual/occlusion image and for physical objects, respectively. Enlarged inset at the occlusion SLM shows that the physical world at the user's focus plane is brought into focus at the SLM where portions of the real world can be occluded. Enlarged inset at the retina shows that the same rays (red) that are in-focus at the occlusion SLM are also in-focus at the retina – this property is utilized to also depict a perceptually correct occlusion mask for out-of-focus virtual objects by applying a computational blur. Finally, the image of real-world objects seen through the display should ideally have the same magnification and distances from the eye as compared to seeing the real world without the display, i.e. $\frac{h_i}{h_o} = 1$ and $e = 0$. In our implementation, we are able to match the magnification, but not the distance.

In the following, we first provide an overview of the optical design we consider, introduce a ray-transfer matrix analysis of prior work on fixed-focus occlusion-capable AR displays (see Sec. 2.2.3), and finally introduce our focus-tunable varifocal occlusion approach.

Overview of the optical design We consider an optical design composed of four lenses (see Fig. 3), whose respective functions are: The first lens brings the real world at a particular depth into focus at the SLM. This image is always flipped, similar to how the image of the real world that is formed on our retina inside our eyes is always flipped. The next two lenses re-invert the in-focus image at the SLM, similar to a 4f system. The last lens finally places the image back into the appropriate depth for comfortable viewing. Let us denote these lenses by L_1, L_2, L_3, L_4 (see Fig. 3).

The occlusion SLM can be placed in any of the image planes of the optical system. There are two locations for this, one is between L_1 and L_2 and the other is between L_3 and L_4. We place the occlusion SLM between L_1 and L_2 because it simplifies Eq. (13). The digital image SLM can also be placed in any of the image planes of the optical system. We choose to place it between L_1 and L_2 because in this case, we can treat both the occlusion SLM and the virtual SLM to be optically equivalent and derive just one set of conditions for both of them.

3.1 Modeling Fixed-focus Occlusion Masks

The light transport through optical components can be modeled using ray-transfer matrices. In this approach, a light ray is represented by a column vector composed of lateral distance (x) and angle of propagation (θ) with respect to the optical axis. The propagation of paraxial light rays through an optical component is modeled as the multiplication of the ray vector with a 2×2 ray-transfer matrix. Ray-transfer matrices are known for standard optical components, e.g. let us denote the ray-transfer matrix for a lens with focal length f by $\mathbf{M}$ and the ray-transfer matrix for free-space propagation with a distance d by $\mathbf{S}$. Then, $\mathbf{M}$ and $\mathbf{S}$ are given by:

$$\mathbf{M} = \begin{bmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{bmatrix}, \qquad \mathbf{S} = \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix}. \tag{1}$$

The composite ray transfer matrix that models the propagation of light rays through a series of optical components is simply the multiplication of the various individual ray transfer matrices of each optical component.

For our optical design (Fig. 3), the composite ray transfer matrix is represented as:

$$\mathbf{T} = \mathbf{M}_4\mathbf{S}_{(L_3,L_4)}\mathbf{M}_3\mathbf{S}_{(L_2,L_3)}\mathbf{M}_2\mathbf{S}_{(L_1,L_2)}\mathbf{M}_1, \tag{2}$$

where $\mathbf{M}_i$ is the ray-transfer matrix describing L_i and $\mathbf{S}_{(L_i,L_j)}$ describes the free-space propagation between lenses L_i and L_j.

The above linear system of equations is composed of four equations and seven unknowns (four unknown focal lengths and three unknown distances). This is an ill-posed inverse problem. Instead of attempting to solve it directly, previous works relied on symmetry constraints, such that $f_1 = f_4$, $f_2 = f_3$, and $d_{(L_1,L_2)} = d_{(L_3,L_4)}$.

Previous works have explored mainly two choices for the composite ray transfer matrix.

Shifted Perspective. In this configuration, the virtual viewpoint is shifted to the front of the optical system. In other words, the first lens and the last lens form conjugate aperture planes. Another way to think of it is that the light field entering the optical system and the light field exiting the optical system are equivalent. Mathematically, this condition represents $\mathbf{T} = \mathbf{I}$, where $\mathbf{I}$ is the 2×2 identity matrix. Some of the earlier prototypes of Kiyokawa et al. [22, 23] and Cakmakci et al. [7] had a shifted perspective.

Correct Perspective. In this configuration, a user looking through the optical system should see the exact same image of a physical scene behind it as if the optical system was absent. There is no shift in the viewpoint. Kiyokawa et al. [21] proposed a folded optics design for achieving correct perspective. This condition was analyzed formally with ray-transfer matrix equations for the first time in the context of optical cloaking [10] and later applied to the problem of occlusion in AR displays [17]. Mathematically, this is represented via the ray-transfer matrix $\mathbf{T} = \mathbf{S}_{(L_1,L_4)}$.

While an OST AR display should ideally be made to satisfy the correct perspective constraint, the disadvantage in doing so is that the field of view of the optical system is much smaller, being at most equal to the field of view seen through the first lens' aperture from a viewing distance of the length of the optical system. This limitation is exacerbated in our implementation by the small aperture (1 cm) of our focus-tunable lenses. For this reason, we design and implement an optical system that satisfies the shifted-perspective constraint.

3.2 Modeling Varifocal Occlusion Masks

Consider the general system of linear equations for image-forming occlusion optical designs (Eq. (2)). Recall that solving this is not

possible by simply analyzing the ray-transfer matrix equations because there are more unknowns than equations. Our approach is to apply an optimization approach to this problem. Gaining some insights from the optimization approach, we then revisit the ray-transfer matrices approach to derive closed-form solutions.

Both our approaches aim to satisfy these requirements:

1. the virtual image should be placed at a desired (but movable) distance.
2. the magnification of the see-through image of the real-world should be unity irrespective of the virtual image plane distance.

3.2.1 Optimization approach

The optimization approach needs to calculate the set of focal lengths that minimize the error in the magnification of the see-through view and the error in the virtual/occlusion image plane's depth.

To do this, we define an image formation model for OST occlusion-capable displays, a cost function for the errors, and apply known methods to minimize the error iteratively. We start off by assuming that all lenses are focus-tunable lenses.

Image Formation The image formation for the virtual and real-world is modeled by successive application of the Gaussian thin lens equations:

$$i = \frac{of}{o-f}, \tag{3}$$

where i is the image distance, o is the object distance, and f is the focal length of the lens.

For an optical system composed of multiple lenses, the object of the subsequent lens (L_{j+1}) is the image of the previous lens (L_j). So, the object distance for L_{j+1} is: $o_{j+1} = d_{(L_j,L_{j+1})} - i_j$.

Occlusion and Virtual Image Formation For the occlusion and virtual image, the objects are the occlusion and virtual SLMs which are optically placed together by design. Only the lenses between these SLMs and the eye (L_2, L_3, L_4) contribute to the virtual/occlusion image formation. So, the object distance ($d_{(\text{SLM},L_2)}$) is propagated through lenses L_2, L_3, and L_4, to obtain the distance to the perceived occlusion/virtual image plane from lens 4 ($d_{(\text{vip},L_4)}$). Let us denote this image formation function by:

$$[d_{(\text{vip},L_4)}] = I_V(f_2, f_3, f_4, d_{(\text{SLM},L_2)}, d_{(L_2,L_3)}, d_{(L_3,L_4)}), \tag{4}$$

where I_V is composed of the successive application of Eq. (3), beginning with

$$i_2 = \frac{d_{(\text{SLM},L_2)} f_2}{d_{(\text{SLM},L_2)} - f_2}, \qquad o_3 = d_{(L_2,L_3)} - i_2, \tag{5}$$

and ending with:

$$d_{(\text{vip},L_4)} = \frac{o_4 f_4}{o_4 - f_4}. \tag{6}$$

See-Through Image Formation For the real-world, we first discretize the real-world into N real-world depth planes, where the number N is chosen such that the system samples the real-world denser than the human eye's depth-of-field which has been measured to be 0.3 diopters [8, 42]. So, for a display whose nearest and farthest depth planes are at $D_{(R_{\text{near}},L_1)}$ diopters and $D_{(R_{\text{far}},L_1)}$ diopters respectively, the minimum number of discretized real-world depth planes should be:

$$N > \frac{D_{(R_{\text{near}},L_1)} - D_{(R_{\text{far}},L_1)}}{0.3}. \tag{7}$$

Each real-world depth ($d_{(R_j,L_1)}$) is propagated through lenses L_1, L_2, L_3, L_4 from which we get the see-through image depth from L_4 ($d_{(V_j,L_4)}$):

$$[d_{(V_1,L_4)}, d_{(V_2,L_4)}, ..., d_{(V_N,L_4)}] = I_R(f_1, f_2, f_3, f_4, d_{(L_1,L_2)}, d_{(L_2,L_3)}, d_{(L_3,L_4)}, d_{(R_1,L_1)}, d_{(R_2,L_1)}, ..., d_{(R_N,L_1)}), \tag{8}$$

where I_R is a successive application of Eq. (3) for each discretized real-world depth plane beginning with:

$$i_1 = \frac{d_{(R_j,L_1)} f_1}{d_{(R_j,L_1)} - f_1}, \qquad o_2 = d_{(L_1,L_2)} - i_1, \tag{9}$$

and ending with:

$$d_{(V_j,L_4)} = \frac{o_4 f_4}{o_4 - f_4}. \tag{10}$$

Error function The error associated with the occlusion/virtual image is the difference between desired occlusion/virtual image plane depth (d_{in}) and actual occlusion/virtual image plane depth ($d_{(\text{vip},L_4)}$) calculated as: $d_{in} - d_{(\text{vip},L_4)}$.

The error associated with the magnification of the physical scene is the difference between one and the magnification of the see-through image, where magnification is calculated as $m = -\frac{\text{image distance}}{\text{object distance}}$. However, in calculating the magnification, we need to be careful about what we consider as the object distance: Recall that in the see-through image formation function (Eq. (8)), we've defined the real-world object distances with respect to the first lens ($d_{(R_j,L_1)}$), whereas the final image distance is calculated with respect to the last lens ($d_{(V_j,L_4)}$). This discrepancy is alright when the optical system is designed to satisfy the shifted-perspective constraint. However, for the correct-perspective constraint, the object distance should be modified to $d_{(R_j,L_1)} + d_{(L_1,L_4)}$. For our display, where the correct object distance is $d_{(R_j,L_1)}$ and the magnification is given by $m_j = -\frac{d_{(V_j,L_4)}}{d_{(R_j,L_1)}}$

The combined error vector is given below:

$$E = \begin{bmatrix} d_{\text{in}} - d_{(\text{vip},L_4)} \\ 1 - m_1 \\ 1 - m_2 \\ ... \\ 1 - m_N \end{bmatrix}. \tag{11}$$

The optimization problem is to find a set of focal lengths (f_1, f_2, f_3, f_4) that minimize the above error:

$$\underset{f_1,f_2,f_3,f_4}{\text{argmin}} \; ||E||^2. \tag{12}$$

Our implementation of this indicates that the set of focal lengths that minimizes the above error function always has a fixed f_2 and f_3.

Unfortunately, the execution time of this optimization is not real-time. We could calculate the dynamic values of f_1 and f_4 for different occlusion mask distances and use the calculated values in a look-up table to get real-time performance. Alternatively, we could use the new information that a fixed f_2 and f_3 can satisfy all the requirements to calculate closed-form solutions, as discussed below.

3.2.2 Closed-form solutions

Consider the same 4-lens optical design for a varifocal occlusion-capable display composed of the following parameters: $f_1^{(t)}, f_2, f_3, f_4^{(t)}, d_{(L_1,L_2)}, d_{(L_2,L_3)}, d_{(L_3,L_4)}, d_{(\text{SLM},L_1)}$, where the superscript $\cdot^{(t)}$ indicates a dynamically changing parameter.

Using the Gaussian thin lens equation (Eq. (3)), $f_1^{(t)}$ is calculated based on the desired virtual image plane distance ($d_{(\text{vip},L_1)}^{(t)}$) and the distance between L_1 and the occluding SLM ($d_{(\text{SLM},L_1)}$):

$$f_1^{(t)} = \frac{d_{(\text{vip},L_1)}^{(t)} d_{(\text{SLM},L_1)}}{d_{(\text{vip},L_1)}^{(t)} + d_{(\text{SLM},L_1)}}. \tag{13}$$

Solving for the rest of the parameters needs an analysis of the ray-transfer matrix equation. To satisfy the shifted-perspective condition,

the ray-transfer matrix needs to satisfy:

$$\mathbf{I} = \mathbf{M}_4^{(t)} \mathbf{S}_{(L_3,L_4)} \mathbf{M}_3 \mathbf{S}_{(L_2,L_3)} \mathbf{M}_2 \mathbf{S}_{(L_1,L_2)} \mathbf{M}_1^{(t)}. \tag{14}$$

Finding optical parameter values that satisfy the above equation automatically ensures that the requirements listed in the beginning of Sec. 3.2 will be satisfied. Since we have learned from our optimization experiments that solutions exists where L_2 and L_3 are fixed-focal length lenses, we solve Eq. (14) for $\mathbf{M}_4^{(t)}$ and analyze the conditions that ensure that the constants of matrix $\mathbf{M}_4^{(t)}$ (i.e., the ones and zero of $\mathbf{M}_4^{(t)}$) are their appropriate values:

$$\mathbf{M}_4^{(t)} \overset{a}{=} \begin{bmatrix} \frac{1+BC}{\frac{C}{f_1^{(t)}}+A} & C \\ B & \frac{C}{f_1^{(t)}}+A \end{bmatrix} \overset{b}{=} \begin{bmatrix} 1 & 0 \\ -\frac{1}{f_4^{(t)}} & 1 \end{bmatrix}, \tag{15}$$

where $\overset{a}{=}$ is obtained by solving Eq. (14) for $\mathbf{M}_4^{(t)}$ and $\overset{b}{=}$ is obtained because $\mathbf{M}_4^{(t)}$ should have the ray-trasfer matrix for a lens, and where A, B, C are the following:

$$A = 1 - \frac{d_{(L_3,L_4)} + d_{(L_2,L_3)} \left(1 - \frac{d_{(L_3,L_4)}}{f_3}\right)}{f_2} - \frac{d_{(L_3,L_4)}}{f_3}, \tag{16}$$

$$B = \frac{1 - \frac{d_{(L_2,L_3)}}{f_3} - d_{(L_1,L_2)} \left(\frac{1 - \frac{d_{(L_2,L_3)}}{f_3}}{f_2} + \frac{1}{f_3} \right)}{f_1^{(t)}} + \frac{1 - \frac{d_{(L_2,L_3)}}{f_3}}{f_2} + \frac{1}{f_3}, \tag{17}$$

$$C = d_{(L_2,L_3)} \left(1 - \frac{d_{(L_3,L_4)}}{f_3}\right) + d_{(L_3,L_4)} + d_{(L_1,L_2)} A. \tag{18}$$

From Eq. (15), we can infer that $C = 0$, and thereby, we can derive that $A = 1$ by substituting $C = 0$ in:

$$1 = \frac{C}{f_1^{(t)}} + A. \tag{19}$$

Re-arranging Eq. (16) by substituting $A = 1$:

$$-\frac{f_2}{f_3} = 1 + \frac{d_{(L_2,L_3)}}{d_{(L_3,L_4)}} \left(1 - \frac{d_{(L_3,L_4)}}{f_3}\right). \tag{20}$$

Re-arranging Eq. (18) by substituting $C = 0$ and $A = 1$:

$$-\frac{d_{(L_1,L_2)}}{d_{(L_3,L_4)}} = 1 + \frac{d_{(L_2,L_3)}}{d_{(L_3,L_4)}} \left(1 - \frac{d_{(L_3,L_4)}}{f_3}\right). \tag{21}$$

This gives us the condition that:

$$\frac{d_{(L_1,L_2)}}{d_{(L_3,L_4)}} = \frac{f_2}{f_3}. \tag{22}$$

$d_{(L_2,L_3)}$ can be derived by re-arranging Eq. (18) and substituting $d_{(L_1,L_2)} = \frac{d_{(L_3,L_4)} f_2}{f_3}$, $A = 1$, and $C = 0$:

$$\begin{aligned} -\frac{f_2 d_{(L_3,L_4)}}{f_3} &= d_{(L_3,L_4)} + d_{(L_2,L_3)} \left(1 - \frac{d_{(L_3,L_4)}}{f_3}\right) \\ \implies d_{(L_2,L_3)} &= \frac{d_{(L_3,L_4)} \left(1 + \frac{f_2}{f_3}\right)}{\frac{d_{(L_3,L_4)}}{f_3} - 1}. \end{aligned} \tag{23}$$

$d_{(L_2,L_3)}$ has to be positive. This gives us an improved version of the condition in Eq. (22):

$$\frac{d_{(L_3,L_4)}}{f_3} = \frac{d_{(L_1,L_2)}}{f_2} > 1. \tag{24}$$

$f_4^{(t)}$ is primary calculated from Equations (15) and (17):

$$f_4{}^{(t)} = -\frac{1}{B}. \tag{25}$$

Summary Here are steps that can be taken to arrive at the static parameters of the optical design:

1. Using Eq. (24), choose any three among $d_{(L_1,L_2)}, d_{(L_3,L_4)}, f_2, f_3$ and calculate for the fourth parameter. This choice can be based on the available fixed-focus lenses for f_2 and f_3 or based on constraints placed upon $d_{(L_1,L_2)}$ and $d_{(L_3,L_4)}$ by the hardware prototype. Although $d_{(\mathrm{SLM},L_1)}$ doesn't feature in any of the conditions that we've derived, it should also be considered carefully in this step because it influences $d_{(L_1,L_2)}$ in that $d_{(L_1,L_2)} > d_{(\mathrm{SLM},L_1)}$.
2. $d_{(L_2,L_3)}$ can now be calculated using Eq. (23).

During the operation of the display, the dynamic parameters ($f_1^{(t)}$ and $f_4^{(t)}$) are calculated using Equations (13) and (25) which are in turn dependent on only one dynamic value which is the virtual image distance ($d_{(\mathrm{vip},L_1)}^{(t)}$).

Again, these equations ensure Eq. (14) which means that the see-through image of the real world would have unit magnification, although with a longitudinal shift which is equal to the length of the optical system from L_1 to L_4.

4 Implementation

We demonstrate varifocal occlusion with a monocular benchtop prototype (see Fig. 4 (A)). Optical design details and components details are discussed in the following.

Optical Design. To minimize distortion and chromatic aberrations in the prototype, all fixed-focus lenses (L_2, L_3) in our prototype are Nikon Nikkor 35-mm f/2 camera lenses. We use a 30-mm cage polarizing beamsplitter cube (ThorLabs CCM1-PBS251) to combine the real-world view after occlusion and the digital image. This design choice and the bulkiness of the Nikon imaging lenses constrains $d_{(L_1,L_2)}$ to a minimum of 10 cm. With this choice of parameters, and for an augmented scene whose minimum and maximum occlusion/virtual image plane depths are 30 cm and 300 cm, respectively, we obtain $f_1^{(t)}$ to lie in the range 25–28.5 diopters and $f_4^{(t)}$ in the range 2.64–5.67 diopters by using our closed-form solutions (Sec. 3.2.2). However, neither of these ranges of optical powers is directly supported by the focus-tunable lenses. The focus-tunable lenses in our prototype are Optotune EL-10-30-TC whose focal range is 8.3–20 diopters and Optotune EL-10-30-C whose focal range is 5–10 diopters.

Additional offset lenses are necessary to bring the operating range of optical powers into the supported range. The combined lens power (D_{combined}) of a focus-tunable lens ($D_{\mathrm{tunable}}^{(t)}$) and an offset lens ($D_{\mathrm{offset}}$) is theoretically $D_{\mathrm{combined}} = D_{\mathrm{offset}} + D_{\mathrm{tunable}}^{(t)}$. In practice, however, we cannot place the offset lens exactly on top of the focus-tunable lens, so it is necessary to modify the composite ray-transfer matrix equations to additionally model the free-space propagation between offset and focus-tunable lenses.

Adding offset lenses changes the composite ray-transfer matrix and solving the equations analytically is tedious. Instead, we used the optimization based method (Sec. 3.2.1) because it is easy to introduce additional offset lenses in Eqs. 4 and 8 run it through the optimization. The resulting optical design is shown in Fig. 4.

Optimization. Our display's nearest depth plane is $D_{R_{\mathrm{near}},L_1} = 3.33$ diopters and the farthest distance is $D_{R_{\mathrm{far}},L_1} = 0.33$ diopters. The number of discretized real-world depth planes (N) considered for optimization can be calculated using Eq. (7) to be at-least 11 planes. The software for our optimization framework is implemented in Python using the package SciPy and the optimization function used is *differential_evolution*. The best optimization result out of 10 trials is chosen as the final optimization result. Tables 2 shows the focal lengths of the focus-tunable lenses calculated using our optimization approach. The

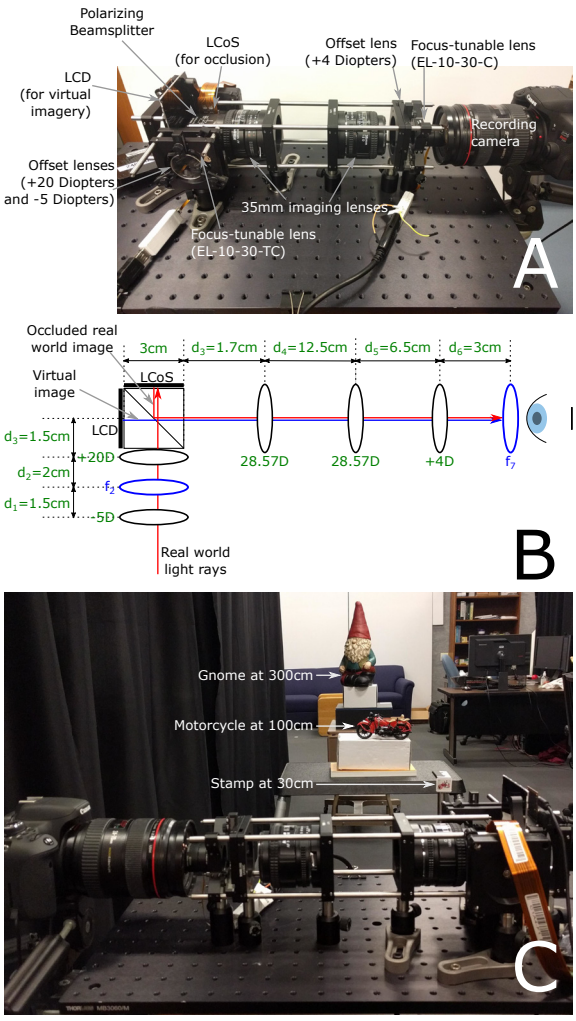

Fig. 4: *(A)* Photo of our varifocal occlusion-capable AR display *(B)* Optical design of the prototype. Static design parameters are denoted in green. Propagation of real-world light through the system is depicted with red arrows. Propagation of the virtual image is depicted with blue arrows. The arrows are only representative of the general direction of propagation and do not depict the exact path taken by the light rays. *(C)* Photo of lab set-up which shows the prototype and the three real objects: stamp at 30cm, motorcycle at 100cm, and gnome at 300cm.

optimization for each virtual image plane distance (i.e. each column of Table 2) takes about 4 seconds.

Displays. For the occlusion SLM, we use a reflection mode liquid crystal on silicon (LCoS) modulator (Silicon Micro Display ST1080) with a resolution of 1,920 × 1,080 and a screen diagonal of 0.74". For the digitally superimposed imagery, we use a liquid crystal display (LCD, Topfoison TF60010A) with a resolution of 2,560 × 1,440 pixels and a screen diagonal of 5.98". Both of these displays are placed at the same optical distance with respect to the user/camera. The pixel density of the LCD is much lower than that of the LCoS panel, which results in pixelated virtual imagery, observed in Figures 1, 5. An additional polarizer was placed on top of the virtual image's LCD panel and manually adjusted to reduce its brightness enough to match with the real world's brightness.

Real-time System. The software for real-time rendering of the occlusion and virtual images is implemented in C++ using OpenGL/GLSL. Multi-pass shaders implement rendering of the RGB image and linearized depth map of the scene, which is used to calculate the depth-dependent computational blur for the occlusion and virtual image. The PC controlling the displays and the focus-tunable

d_{om}	3.33	3.03	2.73	2.43	2.13	1.83	1.53	1.23	0.93	0.63	0.33
f_2	17.9	17.7	17.5	17.3	17.0	16.8	16.5	16.3	16.0	15.8	15.5
f_7	6.47	6.77	7.06	7.36	7.66	7.96	8.26	8.55	8.85	9.15	9.45

Table 2: Focus settings of the focus-tunable lenses for each setting of the occlusion mask distance (d_{om}) modeled in our optimization routine for the prototype display shown in Fig. 4. All values are in units of diopters.

lenses uses an Intel Xeon E5-2630 2.4 GHz processor with an NVIDIA GeForce GTX 980 running Windows 7.

Recording Setup. An augmented reality scene was set up as shown in Figure 4 (C) and it is composed of three real objects: a stamp placed at 30 cm, a toy motorcycle placed at 100 cm, and a garden gnome placed at 300 cm. The scene seen through the display includes several digitally superimposed objects, i.e. one virtual object placed adjacent to each physical object. A Canon T6i Rebel camera with a Canon 24-70 mm f/2.8 lens is used to capture photographs through the display. For each see-through view presented in this paper (Figs. 1, 5, 6), the camera settings were: 70 mm, f/14, ISO-1600, 0.6 s exposure time.

Emulating different AR and occlusion displays. In addition to demonstrating varifocal occlusion, our display is capable of emulating previous AR display technologies which differ from each other in terms of whether or not they provide accommodation support or occlusion support. We utilize this to compare different AR technologies. Here are the four major types of previous AR displays we compare, and the method by which these technologies are emulated:

- **Fixed-focus AR without occlusion**: Current commercially available AR displays present a fixed-focus virtual image without support for occlusion. These displays are emulated by setting our prototype to always present an image at the farthest virtual image plane distance and by setting the occlusion image to full white (reflects as much of the incident light as possible).
- **Varifocal AR without occlusion**: These displays are emulated by dynamically adjusting the focal lengths of the focus-tunable lenses for the given virtual image plane distance, and by applying a computational blur that mimics the perceived retinal blur to the virtual objects that are supposed to be defocused. Occlusion support is turned off by setting the occlusion image to full white.
- **Fixed-focus AR with fixed-focus occlusion**: Previous prototypes of hard-edge occlusion always present the occlusion and virtual imagery at a far distance. These displays are emulated by setting our prototype to always present the image at a far distance while displaying a silhouette of the virtual objects as the occlusion mask.
- **Varifocal AR with varifocal occlusion**: Our proposed display technology dynamically adjusts the focal lengths of the focus-tunable lenses for the given virtual image plane distance, and by applying a computational blur to the virtual objects that are supposed to be defocused. The varifocal occlusion mask is computed by applying a similar computational blur to the silhouette of the virtual image.

5 Results

5.1 See-through images

Figures 1 and 5 show a comparison of the see-through view of different AR and occlusion technologies. In each of these figures, the augmented scene is composed of real-world objects and virtual placed at different distances. At each distance, one virtual object is placed slightly in front of the real world object to demonstrate our display's ability to occlude real world objects. The mechanism by which the different occlusion and AR displays are emulated is explained in Sec. 4. The see-through view for the different AR and occlusion technologies are shown column-wise:

- **Column One:** Emulates commercially available AR displays. In these displays, the virtual imagery looks transparent and is

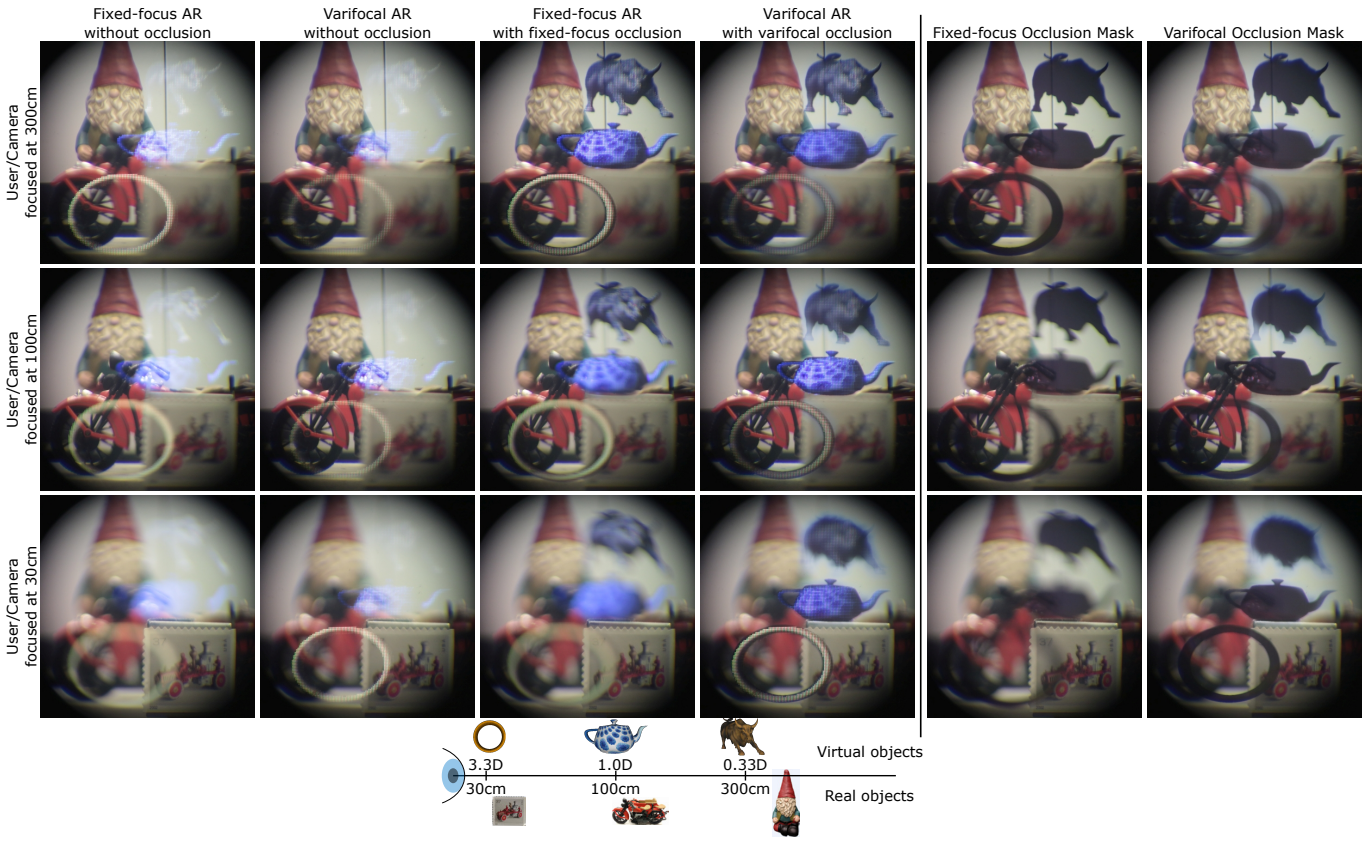

Fig. 5: **Left of the vertical line:** views through our prototype AR display, which is emulating different AR display technologies for each column. The augmented scene is composed of real-world objects (stamp, motorcycle, and gnome) and virtual objects (ring, teapot, and bull). Objects are distributed at different depths: stamp and ring at 30cm, motorcycle and teapot at 100cm, and gnome and bull at 300cm. *(Column 1)* Commercially available AR displays: a transparent virtual image is presented at a fixed distance. Important depth cues such as occlusion and accommodation are absent. *(Column 2)* Varifocal AR displays: virtual image can be moved to different depths, but images are still transparent. *(Column 3)* Fixed-focus occlusion-capable AR display: Occlusion and virtual image is fixed at a single depth, limiting realism when the user is focused to other depths. Note how all virtual objects, including the nearby ones, are in focus when the camera is focused far, and all virtual objects are focused when the camera is focused near. *(Column 4)* Varifocal occlusion-capable AR displays: virtual and occlusion image plane can be moved to different depths enabling perceptually correct depth cues for occlusion and accommodation. Note how objects at the same depth, e.g., near objects (stamp and ring) or far objects (gnome and bull), are correctly in focus or defocused depending on the focus state of the user/camera. **Right of the vertical line:** Comparison of occlusion masks between fixed-focus and varifocal occlusion-capable displays.

placed at a fixed distance, which does not provides the user with important depth cues like occlusion or accommodation.

- **Column Two:** Emulates varifocal AR displays. The virtual image plane is movable in these displays and should be designed to match the user's focal distance. A computational blur can be applied optionally to virtual content that is out-of-focus with the focal distance. The improvement over commercially available AR displays is that accommodation cues are provided in a perceptually correct manner, but these displays still lack the ability to show the most important depth cue, namely occlusion [11].
- **Column Three:** Emulates fixed-focus occlusion-capable AR displays. In these displays, occlusion of real objects by virtual objects can be displayed, but the occlusion mask and virtual image are always displayed at a fixed depth, which reduces the realism for virtual objects located at other depths. Note how in Fig. 5, all three virtual objects, namely ring, teapot, and bull are in-focus when the camera is focused far and all three objects are defocused when the camera is focused at other distances.
- **Column Four:** Demonstrates our varifocal occlusion-capable AR display. Our display is able to move the occlusion and virtual image planes to different distances, and hence, is able to provide depth-dependent occlusion and accommodation depth cues. Note how in Fig. 5, the camera correctly records only one virtual and one real object in-focus at each focus setting.
- **Columns Five and Six:** Comparison of only the occlusion masks for fixed-focus and varifocal occlusion displays.

5.2 Quality of real world magnification

For any AR display, whether occlusion-capable or not, the magnification of see-through images of the real world should be unity irrespective of the virtual image plane distance. Ensuring this property is particularly challenging for a varifocal occlusion-capable AR display. Section 3.2.1 and 3.2.2 discuss complementary strategies to ensure this. Our prototype display shown in Fig. 4 was designed using the optimization approach (Sec. 3.2.1) and Tables 2 and 3 show the focal length settings of the focus-tunable lenses and the magnification of the see-through image for different settings of the occlusion or virtual image plane distance.

Note that the optimization approach (Sec. 3.2.1) requires a discretization of only the real-world distances, but accepts continuously changing values for the occlusion mask. Tables 2 and 3 are calculated for a finite set of occlusion mask distances only to indicate the performance of the display for different occlusion mask distance settings.

Table 3 shows that the optimization predicts that the see-through image magnification values are close to unity, but not exactly equal to unity. Using the closed-form approach would have ensured exact unit magnification for all combinations of real world distance and virtual image plane distance, however, as discussed in Sec. 4, due to some

d_{om} / d_{rw}	3.33	3.03	2.73	2.43	2.13	1.83	1.53	1.23	0.93	0.63	0.33
3.33	0.93	0.94	0.95	0.95	0.96	0.97	0.98	0.99	1	1.01	1.02
3.03	0.93	0.94	0.95	0.96	0.96	0.97	0.98	0.99	1	1.01	1.02
2.73	0.93	0.94	0.95	0.96	0.96	0.97	0.98	0.99	1	1.01	1.02
2.43	0.93	0.94	0.95	0.96	0.97	0.97	0.98	0.99	1	1.01	1.02
2.13	0.94	0.94	0.95	0.96	0.97	0.98	0.98	0.99	1	1.01	1.01
1.83	0.94	0.95	0.96	0.96	0.97	0.98	0.98	0.99	1	1.01	1.01
1.53	0.95	0.95	0.96	0.97	0.97	0.98	0.99	0.99	1	1.01	1.01
1.23	0.95	0.96	0.97	0.97	0.98	0.98	0.99	0.99	1	1	1.01
0.93	0.97	0.97	0.98	0.98	0.98	0.99	0.99	1	1	1	1.01
0.63	0.99	0.99	1	1	1	1	1	1	1	1	1
0.33	1.07	1.07	1.06	1.05	1.04	1.03	1.02	1.01	1	0.99	0.98

Table 3: Magnification predicted by our optimization routine for each real world distance (d_{rw}) propagated through the optical system for each setting of the occlusion mask distance (d_{om}) for the prototype display shown in Fig. 4. Distances (d_{om} and d_{rw}) are in diopters. Note that all magnification values are close to 1.0, indicating good optimization quality.

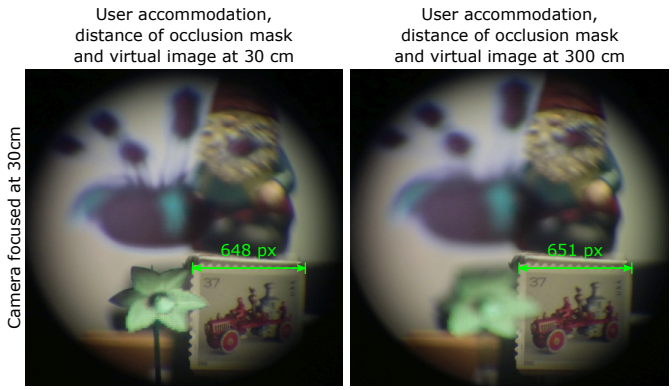

Fig. 6: View through our prototype occlusion-capable AR display for different settings of occlusion/virtual image plane depth with camera focus fixed on foreground. The user fixates the foreground objects (left) and background objects (right) and the virtual image distance and occlusion mask are following their fixation distance. The camera remains focused on the foreground object, demonstrating that changing optical settings of the display do not change the magnification of the physical scene, as indicated by the stamp's size.

hardware constraints, the focal range predicted by the closed-form solutions is unattainable with the focus-tunable lenses at our disposal. Hence, the best we can do currently is the solution predicted by the optimization routine. A similar table could be shown for the other error considered in the optimization approach, i.e. the error in the occlusion or virtual image plane distance (see Eq. 11), however we omit this because these errors are negligible (always less than one centimeter).

We verify the quality of real-world magnification of our prototype by capturing see-through images of our display for different display focus settings for a fixed camera focus distance (see Fig. 6). In the left subfigure, the user is assumed to fixate the daffodil in the foreground. In this setting, the other flower pot is blurred due to the computational blur that emulates perceived retinal blur. The camera is also focused on the foreground objects. In the right subfigure, the user now fixates at an object at the farther distance and the virtual image distance along with the occlusion mask are updated to the farther distance. We intentionally keep the camera focus on the foreground object to highlight the fact that refocusing the virtual image and the occlusion mask does not change the magnification of the physical scene in a noticeable manner. This is highlighted by the size of the stamp being roughly constant. Note that the user would never see the camera image shown in the right subfigure, because in a varifocal display, the distance of the object they fixate is the same as the virtual image distance. Nevertheless, this experiment demonstrates our prototype display's capability to maintain constant magnification of the real-world independent of the virtual image distance.

5.3 Display specifications

The display's field of view is 15.3°. The supported occlusion/virtual image plane depth is from optical infinity to 30 cm. In our results, we do not include real or virtual objects beyond 300 cm because 300 cm seems equivalent to optical infinity for the display. The eyebox is about 1 cm, equal to the aperture of the last lens in the system.

6 Discussion

In summary, we introduce varifocal-occlusion capable AR displays based on focus-tunable optics. This approach improves the realism of optical see-through displays by enabling mutually consistent occlusions between digital and physical objects over a large depth range. We derive a formal optimization approach and real-time heuristics to tune the optical settings of our system to avoid distortions of the physical scene and demonstrate improved realism with a prototype AR display.

6.1 Limitations

Similar to other varifocal-type displays, ours would require eye tracking to determine where to focus the display. Our current prototype does not include an eye tracker, although this capability has been demonstrated with previous varifocal VR displays [36]. The field of view of our prototype is limited by the size of commercially available focus-tunable lenses, although these are steadily increasing [37]. Finally, our prototype shares limitations of other, fixed-focus occlusion-capable AR displays in being implemented as a benchtop system. Although it does not seem straightforward how to miniaturize the proposed optical design, we believe that the capabilities offered by our system are unique and important; we hope to inspire others to address some of the remaining questions on optimizing device form factors for occlusion-capable displays in general.

6.2 Future Work

First and foremost, the device form factor of this and other occlusion-capable displays should be reduced to enable wearable occlusion-capable displays. This is a major optical design challenge, beyond the scope of this paper. Eye tracking should be incorporated into such a wearable system. While most occlusion-capable displays aim at computing a binary occlusion mask, one could also envision the attenuation pattern to be optimized to enable consistent illumination, shading, and shadows of digital and physical objects along with consistent occlusion [4] or enable other types of optical image processing capabilities [43].

6.3 Conclusion

To enable seamless experiences with AR displays, hard-edge occlusion control is critical. With this work, we take steps towards improving the realism of optical see-through displays with varifocal occlusion capabilities. Yet, many challenges in this area remain to design and build small, light-weight AR glasses that offer perceptually realistic and seamless experiences.

Acknowledgments

K.R. and H.F.'s research was supported in part by NSF grant 1405847 ("II-New: Seeing the Future: Ubiquitous Computing in EyeGlasses"), and by a research gift from Intel Corporation, and by the BeingTogether Center, a collaboration between Nanyang Technological University (NTU) Singapore and University of North Carolina (UNC) at Chapel Hill. The BeingTogether Center is supported by the Singapore National Research Foundation, Prime Minister's Office, under its International Research Centers in Singapore Funding Initiative.

G.W. was supported by an Okawa Research Grant and a Sloan Fellowship. Other funding for the project was provided by NSF (award numbers 1553333 and 1839974) and by Intel.

REFERENCES

[1] K. Akşit, 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] G. Avveduto, F. Tecchia, and H. Fuchs. Real-world occlusion in optical see-through ar displays. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, p. 29. ACM, 2017.

[3] O. Bimber and B. Fröhlich. Occlusion shadows: Using projected light to generate realistic occlusion effects for view-dependent optical see-through displays. In *Proc. IEEE ISMAR*, 2002.

[4] O. Bimber, A. Grundhöfer, G. Wetzstein, and S. Knödel. Consistent illumination within optical see-through augmented environments. In *Proc. IEEE ISMAR*, pp. 198–207, 2003.

[5] O. Bimber, D. Iwai, G. Wetzstein, and A. Grundhoefer. The Visual Computing of Projector-Camera Systems. *Computer Graphics Forum*, 2008.

[6] O. Cakmakci, Y. Ha, and J. Rolland. Design of a compact optical see-through head-worn display with mutual occlusion capability. In *Proc. SPIE 5875*, 2005.

[7] O. Cakmakci, Y. Ha, and J. P. Rolland. A compact optical see-through head-worn display with occlusion support. In *Proc. IEEE ISMAR*, pp. 16–25, 2004.

[8] F. W. Campbell. The depth of field of the human eye. *Optica Acta: International Journal of Optics*, 4(4):157–164, 1957.

[9] P. Chakravarthula, D. Dunn, K. Akşit, 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.

[10] J. S. Choi and J. C. Howell. Paraxial ray optics cloaking. *OSA Opt. Express*, 22(24):29465–29478, 2014.

[11] 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.

[12] D. Dunn, C. Tippets, K. Torell, P. Kellnhofer, K. Akşit, 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.

[13] C. Gao, Y. Lin, and H. Hua. Occlusion capable optical see-through head-mounted display using freeform optics. In *Proc. IEEE ISMAR*, pp. 281–282, 2012.

[14] C. Gao, Y. Lin, and H. Hua. Optical see-through head-mounted display with occlusion capability. In *Proc. SPIE 8735*, 2013.

[15] T. Hamasaki and Y. Itoh. Varifocal occlusion for optical see-through head-mounted displays using a slide occlusion mask. *IEEE TVCG*, 25(5):1961–1969, 2019.

[16] I. P. Howard and B. J. Rogers. *Seeing in Depth*. Oxford University Press, 2002.

[17] I. D. Howlett and Q. Smithwick. Perspective correct occlusion-capable augmented reality displays using cloaking optics constraints. *Journal of the Society for Information Display*, 25(3):185–193, 2017.

[18] Y. Itoh, T. Hamasaki, and M. Sugimoto. Occlusion leak compensation for optical see-through displays using a single-layer transmissive spatial light modulator. *IEEE TVCG*, 23(11):2463–2473, 2017.

[19] 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 TVCG*, 25(5):1951–1960, 2019.

[20] 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.

[21] K. Kiyokawa, M. Billinghurst, B. Campbell, and E. Woods. An occlusion-capable optical see-through head mount display for supporting co-located collaboration. In *Proc. IEEE ISMAR*, 2003.

[22] K. Kiyokawa, Y. Kurata, and H. Ohno. An optical see-through display for mutual occlusion of real and virtual environments. In *Proc. ISAR*, pp. 60–67, 2000.

[23] 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, 2001.

[24] 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.

[25] F. L. Kooi and A. Toet. Visual comfort of binocular and 3d displays. *Displays*, 25(2-3):99–108, 2004.

[26] P.-Y. Laffont, A. Hasnain, P.-Y. Guillemet, S. Wirajaya, J. Khoo, D. Teng, and J.-C. Bazin. Verifocal: A platform for vision correction and accommodation in head-mounted displays. In *ACM SIGGRAPH 2018 Emerging Technologies*, pp. 21:1–21:2, 2018.

[27] M. Lambooij, M. Fortuin, I. Heynderickx, and W. IJsselsteijn. Visual discomfort and visual fatigue of stereoscopic displays: A review. *Journal of Imaging Science and Technology*, 53(3):30201–1, 2009.

[28] T. Langlotz, M. Cook, and H. Regenbrecht. Real-time radiometric compensation for optical see-through head-mounted displays. *IEEE TVCG*, 22(11):2385–2394, 2016.

[29] T. Langlotz, J. Sutton, S. Zollmann, Y. Itoh, and H. Regenbrecht. Chromaglasses: Computational glasses for compensating colour blindness. In *Proc. SIGCHI*, pp. 390:1–390:12, 2018.

[30] D. Lanman, M. Hirsch, Y. Kim, and R. Raskar. Content-adaptive parallax barriers: Optimizing dual-layer 3d displays using low-rank light field factorization. In *ACM SIGGRAPH Asia*, pp. 163:1–163:10, 2010.

[31] 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.

[32] A. Maimone and H. Fuchs. Computational augmented reality eyeglasses. In *Proc. IEEE ISMAR*, pp. 29–38, 2013.

[33] 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. (SIGGRAPH)*, 33(4):89:1–89:11, 2014.

[34] A. Maimone, X. Yang, N. Dierk, A. State, M. Dou, and H. Fuchs. General-purpose telepresence with head-worn optical see-through displays and projector-based lighting. In *2013 IEEE Virtual Reality (VR)*, pp. 23–26. IEEE, 2013.

[35] 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 *Proc. IEEE VR*, pp. 251–258, 2018.

[36] 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.

[37] N. Padmanaban, R. Konrad, and G. Wetzstein. Autofocals: Evaluating gaze-contingent eyeglasses for presbyopes. *Science Advances*, 2019.

[38] S. E. Palmer. *Vision Science - Photons to Phenomenology*. MIT Press, 1999.

[39] 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.

[40] J. P. Rolland and H. Fuchs. Optical versus video see-through head-mounted displays in medical visualization. *Presence: Teleoperators and Virtual Environments*, 9(3):287–309, 2000. doi: 10.1162/105474600566808

[41] 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.

[42] S. J. Watt, K. Akeley, M. O. Ernst, and M. S. Banks. Focus cues affect perceived depth. *Journal of vision*, 5(10):7–7, 2005.

[43] G. Wetzstein, W. Heidrich, and D. Luebke. Optical image processing using light modulation displays. *Computer Graphics Forum*, 29(6):1934–1944, 2010.

[44] G. Wetzstein, D. Lanman, M. Hirsch, and R. Raskar. Tensor displays: Compressive light field synthesis using multilayer displays with directional backlighting. *ACM Trans. Graph. (SIGGRAPH)*, 31(4):80:1–80:11, 2012.

[45] A. Wilson and H. Hua. Design and prototype of an augmented reality display with per-pixel mutual occlusion capability. *OSA Opt. Express*, 25(24):30539–30549, 2017.

[46] Y. Yamaguchi and Y. Takaki. See-through integral imaging display with background occlusion capability. *OSA Appl. Opt.*, 55(3):A144–A149, 2016.