# X-Mask: Improving Soft-Edge Occlusion in Optical See-Through Displays with Cross-Shaped Pinholes

Xiaodan Hu[1*] Christoph Ebner[1†] Yan Zhang[2‡] Kiyoshi Kiyokawa[3§] Alexander Plopski[1¶]

[1]Graz University of Technology [2]Shanghai Jiao Tong University [3]Nara Institute of Science and Technology

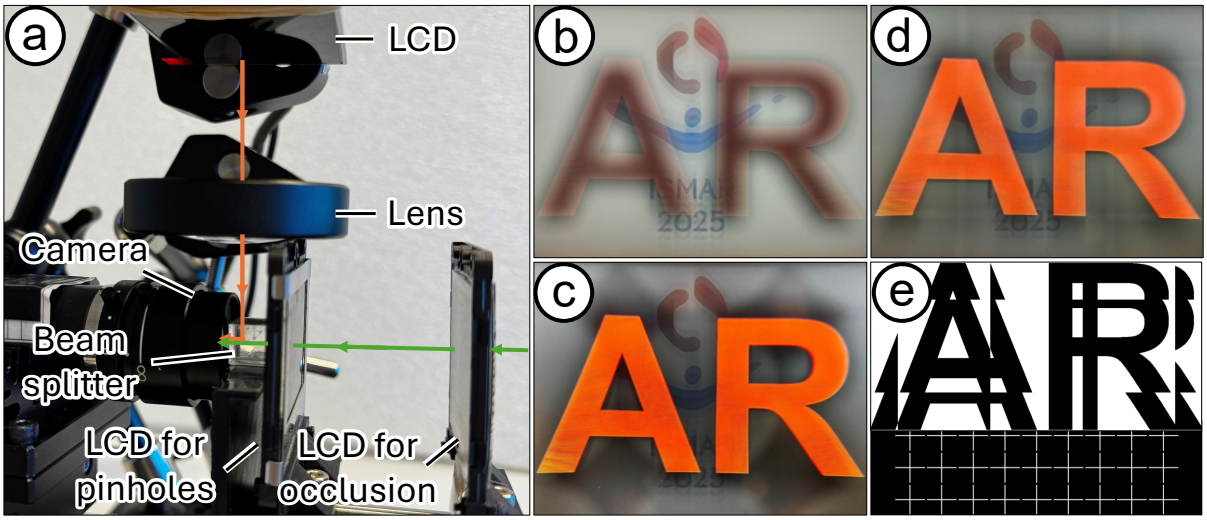


Figure 1: Our pinhole-array-based occlusion display (a) is capable of occluding the real world with minimal bleeding. Compared to conventional displays with a single-layer occlusion mask (b), and circular pinholes (c), our system generates a delinated occlusion mask around the virtual content, leading to high contrast and minimal bleeding of the occluded area (d). By modulating the shape of the pinholes (e, bottom) and the occlusion pattern (e, top), we balance occlusion bleeding with retinal tiling resulting in a more uniform mask with higher contrast.

## Abstract

Placing a transparent liquid crystal display (LCD) into the light path is a simple approach to create occlusion-capable optical see-through head-mounted displays (OST-HMDs) that suffers from defocused (soft-edge) occlusion where the mask leakage partially occludes surrounding content as well. Creating a focused (hard-edge) occlusion that does not suffer from mask leakage requires complicated, bulky optical setups. We present X-Mask, a pinhole-array-based OST-HMD that creates a sharp occlusion mask without the need for a bulky setup requiring only two transparent LCD layers. By rendering a pinhole array on the layer closer to the user's eye, our system functions as a programmable aperture layer that extends the effective depth of field and improves the sharpness of the occlusion mask rendered on the second LCD layer. Utilizing a conventional circular pinhole would result in non-uniform brightness and contrast. By changing the pinhole shape to a cross enables near-optimal retinal tiling with reduced overlaps and gaps. To accommodate pupil size variation, focus distance, and gaze direction, our system design allows for gaze-contingent adjustment of both LCD layers. We validate X-Mask in simulations and a physical prototype showing improved occlusion sharpness and visual uniformity.

**Index Terms:** Occlusion displays, pinhole array, light field displays, optical see-through head-mounted displays

*e-mail: xiaodan.hu@tugraz.at
†e-mail: christoph.ebner@tugraz.at
‡e-mail: yan-zh@sjtu.edu.cn
§e-mail: kiyo@is.naist.jp
¶e-mail: alexander.plopski@tugraz.at

## 1 Introduction

Augmented reality (AR) combines virtual information seamlessly with the real world. Optical see-through head-mounted displays (OST-HMDs) are considered a promising platform for integrating AR into daily life, as users can directly view the real environment through optical combiners while virtual content is typically reflected or diffracted into their eyes. However, the virtual content often blends with ambient real-world illumination, causing virtual objects to appear translucent, resulting in distorted colors and the loss of important depth cues [12].

Some OST-HMDs address this by creating a dark occlusion mask within the user's field of view (FOV) through physical modulation or blocking of incoming light with spatial light modulators (SLMs). Virtual content rendered against this occlusion mask can then avoid interference from ambient illumination, thus appearing more realistic and vivid [14, 28, 32].

Ideally, the occlusion mask should precisely match the size and shape of virtual objects within the user's FOV, thereby accurately blocking environmental illumination at the virtual object's position without affecting the user's view of other areas of the real world. This is known as *hard-edge occlusion* [14, 27, 28, 32, 30]. Hard-edge occlusion requires multiple optical elements (e.g., lenses) to ensure that users can simultaneously focus on both real-world scenes and virtual content. This inevitably leads to a bulky optical design, and the size of optical components like lenses often limits

the FOV of OST-HMDs. In contrast, if we eliminate the optical lenses and use only transmissive SLMs (e.g., transparent LCDs) to create the occlusion mask, we can significantly improve both the FOV and form factor. Nevertheless, the absence of focusing lenses results in defocus blur; specifically, when the user focuses on distant scenes, the occlusion mask positioned at close range appears blurred [26, 11, 8]. This effect is known as *soft-edge occlusion*. Despite recent research efforts [11, 8], it remains challenging to overcome the fundamental physical limitation causing blurred edges in the occlusion mask. Recently, Magic Leap 2 became the first OST-HMD to feature soft-edge occlusion [23], reflecting the growing interest in integrating occlusion functionality into real-world AR products.

In this paper, we present X-Mask, a pinhole-array-based occlusion display (shown in Fig. 1(a)) designed to generate an enhanced soft-edge occlusion mask without the need for complicated optical designs. While pinhole arrays have previously demonstrated effectiveness in constructing light-field displays [25, 24], their potential application for occlusion-capable OST-HMD remains largely unexplored. Our system consists of two layers of transmissive SLMs. The first SLM, closer to the user's eyes, generates a dynamic pinhole array. The pinhole array, thus, functions as a dynamically programmable aperture capable of extending the system's effective depth of field (DOF). Its spatial pattern can be adaptively adjusted in real time based on user-specific parameters, such as pupil size and focusing distance. The second SLM generates the occlusion mask so that it appears in focus. This mask is adapted to reflect changes in the pinhole layer.

Pinhole arrays used in light-field displays typically adopt circular or square apertures [1, 10, 25, 24]. However, because the pinholes are significantly defocused when placed close to the viewer's eye, conventional pinhole arrangements suffer from undesirable contrast distortion due to overlapping blur spots if the spacing is too small, or noticeable dark gaps if the spacing is too large [15, 29]. Such artifacts significantly degrade the quality of the final image.

The underlying reason for this limitation is that in typical OST-HMD scenarios, the user's focal distance ranges between approximately 1 and 2 meters, while the pinhole array resides very close to the eye, causing severe defocus. Circular or square pinhole shapes result in approximately circular blur spots. Geometrically, circles cannot tile a plane perfectly without overlaps or gaps, inevitably leading to contrast-reducing overlaps or image-degrading gaps [15].

To enhance image quality, we introduce a novel cross-shaped pinhole design. This leads to approximately square-shaped blur spots, enabling effective tiling. We validate the system through both simulation and a functional prototype. The results demonstrate that the cross-shaped pinhole configuration improves occlusion sharpness and visual uniformity compared to baseline designs, while preserving a wearable, lensless optical architecture suitable for integration into optical see-through displays.

The main contributions of this paper are:

- We propose X-Mask, a novel pinhole-array-based occlusion display to mitigate the severe defocus inherent in soft-edge occlusion displays, maintaining a wearable form factor and large FOV.
- We introduce a cross-shaped pinhole design that allows near-optimal retinal tiling with reduced gaps and overlaps,
- We analyze the imaging characteristics of cross-shaped pinholes and occlusion masks, and derive optimal layout parameters that balance visual uniformity and occlusion sharpness. These design insights are implemented and validated in a physical prototype through both simulation and see-through photographs.

## 2 Related Work

Our work combines aspects of occlusion-capable displays and pinhole-array displays, which we review in the following.

### 2.1 Occlusion-capable OST-HMDs

Tatham first proposed an AR display capable of mutual occlusion by placing a single-layer transparent LCD between real and virtual objects, representing the earliest implementation of soft-edge occlusion [26]. Subsequently, Kiyokawa et al. [14] employed a similar transparent LCD combined with multiple lenses and optical components to realize the first hard-edge occlusion OST-HMD, capable of achieving per-pixel occlusion. However, their ring-shaped optical design resulted in a large and bulky HMD system.

Subsequent research has focused on reducing the size of hard-edge occlusion displays [13, 28, 32], improving image quality [27, 3], increasing the FOV [30, 33], and adding new functionalities such as varifocal occlusion [6, 22, 7] and enabling stereo content in monoscopic displays [5]. Recently, Zhang et al. further compacted the original ring-shaped design by folding the optical path, resulting in a significantly smaller hard-edge add-on occlusion device (8.5×5.5×3 $cm^3$) for the HoloLens 1 [32]. Despite the substantial reduction in device size, the compact form factor limited the achievable FOV to H28.7°×V16.4°.

Similarly, other works have focused on improving the image quality of soft-edge occlusion displays [19, 11, 8, 17]. Typically, soft-edge occlusion displays do not employ focusing lenses, thereby substantially increasing the achievable FOV. However, because the occlusion mask is positioned far from the focal plane (usually at a viewing distance of about 1–2 m in OST-HMD scenarios), it becomes severely defocused. This defocus not only blurs the occlusion mask but also reduces its perceived size, reducing its effectiveness in blocking ambient illumination [11].

A straightforward method to enhance light blocking is to morphologically enlarge the mask [19, 11]. This approach is effective when the mask serves only to reduce excessive ambient light. For example, Hu et al. developed smart dimming sunglasses using the expanded occlusion mask specifically designed for individuals with light sensitivity [9]. They further conducted a user study, demonstrating that human perception of blurred soft-edge occlusion masks varies among individuals, with some perceiving sharper edge transitions than predicted by theoretical models [8]. However, it remains unclear whether users can still perceive the blurred edges distinctly when virtual objects are superimposed on morphologically expanded soft-edge occlusion masks.

Itoh et al. also addressed this issue by morphologically expanding the occlusion mask, but in addition to virtual objects, they rendered the real scene around the edges of the virtual content with color correction [11]. As a result, virtual objects appeared to blend seamlessly with the real environment, enabling the occlusion feature using only a single transparent LCD. However, the latency can impact the alignment between virtual and real-world content.

Maimone and Fuchs utilized stacked LCD panels to implement occlusion through a light-field approach [17]. Their compact AR eyeglasses leveraging light-field principles provided a wide FOV and multiple simultaneous focal depths. However, the lack of lens-based focusing resulted in blurred occlusion masks. Later, Maimone et al. further applied a light-field-based pinlight structure to achieve an occlusion-capable OST-HMD [18]. Nevertheless, generating the occlusion mask required deactivating the pinlights responsible for focusing, which also led to a blurred occlusion effect.

Zhang et al. incorporated a single pinhole into their large-FOV OST-HMD using paired ellipsoidal mirrors, aiming to extend the DOF and thereby sharpen the occlusion mask [31]. However, the use of the pinhole severely limited image brightness.

Building on these efforts, we propose an occlusion display based on a pinhole array. Compared to hard-edge occlusion systems with

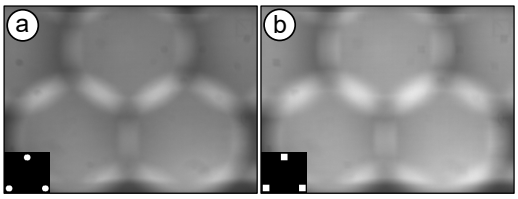


Figure 2: See-through views of (a) circular and (b) square pinhole array rendered on a transparent LCD, with the corresponding pinholes shown in the bottom-left corners. Due to defocus, both circular and square pinholes produce circular blur spots on the sensor. Overlapping regions appear brighter, leading to distorted contrast.

converging lenses, our design offers a wider FOV and a more compact form factor by using only two transparent LCDs. In contrast to lens-free soft-edge occlusion approaches, our method achieves sharper occlusion edges while maintaining system simplicity.

### 2.2 Pinhole-array-based Displays

To avoid introducing additional focusing lenses, which increase the form factor and limit the FOV, another viable approach is to employ a pinhole mask. However, using a single or a small number of pinholes significantly reduces image brightness [2, 31]. Thus, pinhole arrays are typically adopted to mitigate this issue.

Beyond improving brightness, pinhole arrays retain the ability to extend the DOF while also enabling light-field reconstruction. By limiting defocus blur and preserving directional light propagation, they help reconstruct light fields, making them useful for applications requiring depth representation. As a result, pinhole arrays have been widely explored in 3D imaging systems, light-field displays, and near-eye displays (NEDs).

Choi et al. implemented an active pinhole array using a transparent LCD, dynamically generating or eliminating pinholes to selectively control light transmission, enabling simultaneous 2D and 3D content presentation [4]. Huang et al. utilized pinhole arrays in a computational light-field vision-correcting display, allowing users to perceive clearly focused images without prescription eyewear [10]. Additionally, pinhole arrays have proven effective in NEDs. For instance, Song et al. combined a freeform prism with a pinhole array to implement a compact, light-field OST-HMD providing correct focus cues [25]. Akşit et al. demonstrated a slim, pinhole-array-based NED with a thickness of only 9 mm [1].

A significant challenge in pinhole-array-based NEDs is that the pinhole array itself becomes defocused. Due to the finite pupil size, each defocused pinhole produces an enlarged circular blur spot on the retina (shown in Fig. 2). Overlaps or gaps between these blur spots can cause contrast distortion or dark spots, respectively. Typically, an image is decomposed into multiple elemental images according to the pinhole array pattern, such that each elemental image passes through a corresponding pinhole and recombines on the retina to reconstruct the original image. However, overlapping blur spots lead to overlapping elemental images, resulting in regions of higher pixel intensity, severely degrading image quality.

To mitigate this issue, Yao et al. applied a nonlinear cosine-squared brightness gradient to the edges of elemental images [29]. This approach compensates for brightness horizontally and vertically during retinal overlap, thus reducing contrast distortion. However, since defocused pinholes form circular blur spots, overlap occurs in multiple directions. Recently, building upon this work, Lee et al. proposed a multiplexed retinal projection approach [15]. By dynamically rendering pinholes on a transparent LCD, they simultaneously adjusted the pinhole distribution and applied gradual intensity variations to elemental images across multiple frames. Consequently, four consecutive frames were accumulated to form a complete image. Although this method effectively reduces contrast distortion and dark spots, it inherently requires a high refresh rate

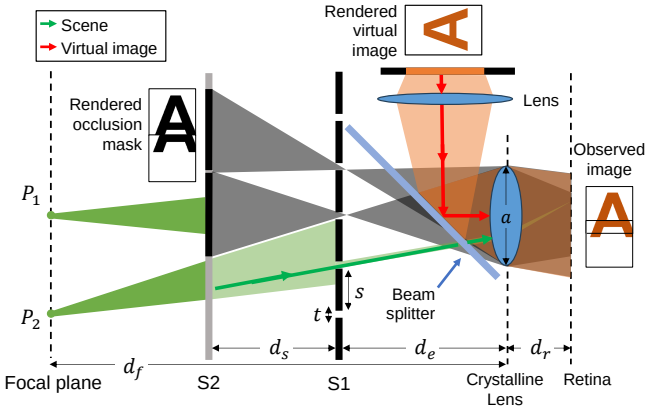


Figure 3: Optical Layout of our system. The front LCD (S1) renders a cross-shaped pinhole array, while the rear LCD (S2) displays the occlusion mask. When the eye focuses at the focal plane, light from point $P_1$ is blocked by the mask, whereas light from $P_2$ passes through S2 and is constrained into a narrow cone by the pinhole, allowing it to reach the retina. Each elemental mask is spatially modulated by a single pinhole, forming a complete occlusion pattern on the retina (note that although the occlusion mask blocks light, it is depicted in gray for clarity). The orange beams represent light from a virtual image. When aligned with the occlusion pattern, a solid virtual object is perceived.

display and may introduce flicker, potentially causing discomfort for hypersensitive users.

In our work, we also adopt a transparent LCD to implement a dynamic pinhole array. While previous studies used transparent LCDs to control the presence or absence of pinholes for combining 2D and 3D content or for time-multiplexed displays [4, 15], our goal is to address the limitations introduced by defocus. Specifically, changes in pupil size and focusing distance affect the PSF at the pinhole plane, which alters how pinholes are projected onto the retina. Our use of a transparent LCD enables real-time control of the pinhole layout, supporting flexible spatial modulation and content alignment.

## 3 System Design

In this section, we describe the system architecture, the design rationale behind the pinhole shape based on defocus blur, the impact of pinhole parameters on retinal image formation, and the generation of elemental occlusion masks through the pinhole layer.

### 3.1 System Overview

Our occlusion display prototype consists of two transmissive SLMs, implemented using transparent LCD panels, positioned in front of the user's eye. As illustrated in Fig. 3, the front LCD (S1) renders a pinhole array, while the rear LCD (S2) displays the occlusion mask. Both LCD panels are placed parallel to each other. When the observer focuses on a scene at a distance $d_f$, light originating from source $P_1$ is blocked by the occlusion mask on S2, whereas light from source $P_2$ is partially attenuated due to polarization effects induced by the liquid crystal and filtered by the polarizers of S2. Nonetheless, a portion of this light can still pass through the pinholes, forming a clear image on the retina.

Without the pinhole array, the occlusion mask on S2 would appear severely blurred on the retina due to its distance from the focal plane. However, with the pinhole array in place, although the image plane of S2 remains behind the retina when the eye is focused at $d_f$, the pinhole array reduces the angular spread of light rays reaching the eye, resulting in a sharper occlusion mask on the retina.

By inserting a beam splitter between S1 and the observer and adding a color display combined with a lens for virtual image pro-

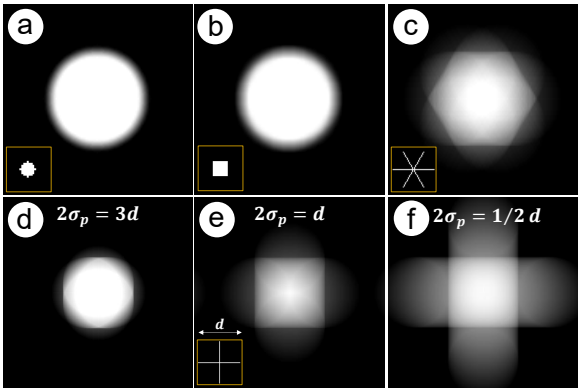


Figure 4: (a-c) Simulated defocus blur of different pinhole shapes: (a) circular, (b) square, (c) star-shaped. Insets show the original pinhole shapes. (d-f) Defocus simulations of cross-shaped pinholes with varying line lengths. As the line length $d$ increases while the PSF radius $\sigma_p$ remains fixed, the resulting blur pattern transitions from circular (PSF-dominated) to shapes that closely resemble uniformly tileable patterns when $d \approx 2\sigma_p$, and eventually approximates the original pinhole geometry when $d \gg 2\sigma_p$.

jection, the occlusion mask effectively blocks ambient illumination at the position of the virtual content. As a result, users perceive a clear and solid virtual image superimposed onto a sharply visible real-world scene.

## 3.2 Pinhole Shape and Defocus Blur

### 3.2.1 Filled-Shape Pinhole Designs

Conventional pinhole-array-based displays commonly use circular or square pinholes [1, 29, 15, 25]. These shapes, however, produce circular blur spots under defocus due to the round pupil aperture of the human eye, as illustrated in Fig. 2 and Fig. 4(a-b).

The optimal pinhole diameter is usually determined by balancing diffraction-induced blur and geometric blur [1]. Smaller pinholes introduce stronger diffraction, while larger ones increase defocus blur. This trade-off yields an approximate optimal diameter $t \approx \sqrt{2\lambda d_s}$, where $\lambda$ is the wavelength of the incident light, and $d_s$ represents the distance between the two LCD panels. Practically, when $d_s$ ranges from approximately 1 cm to 4 cm, the optimal pinhole diameter typically falls within 0.1 mm to 0.2 mm.

Consequently, the retinal projection of the pinhole pattern, that is, the image formed by light transmitted through the pinhole, can be described mathematically as a convolution:

$$I_1(x,y) = P_o(x,y) \otimes H_{p1}(x,y), \tag{1}$$

where $P_o(x,y)$ denotes the transmittance function of the pinhole and $H_{p1}(x,y)$ is the PSF approximated by a circular averaging filter. The transmittance function of a circular pinhole is represented as:

$$P_o(x,y) = \begin{cases} 1, & x^2+y^2 \le (\frac{t}{2})^2 \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

and the circular averaging filter is written as:

$$H_{p1}(x,y) = \begin{cases} \frac{1}{\pi\sigma_p^2}, & x^2+y^2 \le \sigma_p^2, \\ 0, & \text{otherwise}. \end{cases} \tag{3}$$

The radius of the PSF, $\sigma_p$, is governed by the defocus between the focal plane and the pinhole layer. It can be estimated by combining the thin lens formula $\frac{1}{d_f} + \frac{1}{d_r} = \frac{1}{f}$, and triangle similarity as:

$$\sigma_p = \frac{a d_r}{2}\left(\frac{1}{d_e} - \frac{1}{d_f}\right) \cdot m_{rp}, \quad \text{where } m_{rp} = \frac{d_e}{d_r}, \tag{4}$$

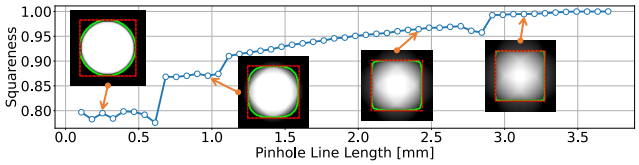


Figure 5: Relationship between pinhole line length from 0 to $2\sigma_p$ and blur spot squareness for PSF radius $\sigma_p = 1.86$ mm, with sample blur images. The squareness ratio is computed from the extracted blur contour (green line) and its minimum bounding rectangle (red dotted line), capturing how closely the defocused blur resembles a square.

where $f$ is the focal length, $d_r$ is the distance from the eye lens to the retina, $a$ is the pupil diameter, $d_e$ is the distance from the front LCD (S1) to the eye lens, and $d_f$ is the focusing distance of the eye. To align with the coordinate system of $P_o(x,y)$, we express the PSF radius on the pinhole plane by applying a magnification factor $m_{rp}$ from the retinal plane.

Geometrically, the convolution can be interpreted as expanding the pinhole shape with a blur kernel, conceptually similar to a Minkowski sum with a disk of radius $\sigma_p$. Under significant defocus, the resulting blur spots, regardless of the original pinhole shape, tend to become approximately circular, as shown in Fig. 4(a–b). Such circular blur spots inherently cannot tile the retinal plane efficiently, inevitably causing overlaps or gaps that lead to reduced contrast and visual quality degradation.

### 3.2.2 Line-Based Pinhole Designs: Cross and Star Shapes

Conventional circular or squared pinholes produce circular blur spots under defocus, which limit retinal tiling. In contrast, we explore line-based pinhole designs, such as cross-shaped and star-shaped patterns, which consist of intersecting strokes and produce square- or hexagon-like blur spots that better support uniform tiling.

Our design is inspired in part by the pinlight display, which used a dense array of point light sources (pinlight) to project light through a modulation plane toward the eye [18]. In that system, the image on the modulation plane is shaped into hexagonal patches. When illuminated by pinlights, these patches tile the view and reduce artifacts. In contrast, our system deals with incoming ambient light from the real environment, which is modulated or blocked by a pinhole array and occlusion layer near the eye. Since the occlusion plane is significantly defocused when the observer views distant scenes, we cannot rely purely on geometric projection. Ideally, we should determine the required pinhole shape by taking the deconvolution of optimal tiling shapes (e.g., squares and hexagons). However, given that the PSF is large under typical viewing conditions, a stable deconvolution is impractical due to the resulting ill-posed inverse problem. Thus, as an alternative, we select cross-shaped and star-shaped pinholes, whose convolution with the large circular PSF naturally yields blurred shapes that closely approximate uniform-tileable patterns, as shown in Fig. 4(c) and (e).

To determine the optimal line length $d$ for the cross-shaped and star-shaped pinholes, we analyzed how $d$ relates to the effective blur shape under defocus. The relationship can be approximated as follows:

$$\text{Blur shape} \approx \begin{cases} \text{Determined by pinhole shape}, & d \gg 2\sigma_p \\ \text{Approximately square or hexagonal}, & d \approx 2\sigma_p \\ \text{Dominated by the circular PSF}, & d \ll 2\sigma_p \end{cases}$$

as illustrated in Fig. 4(d–f).

To analyze how the blur shape evolves with line length, we define a squareness metric (shown in Fig. 5). We first extract the outer contour of the blur spot by applying the Sobel operator to the normalized intensity image to compute the gradient magnitude, followed

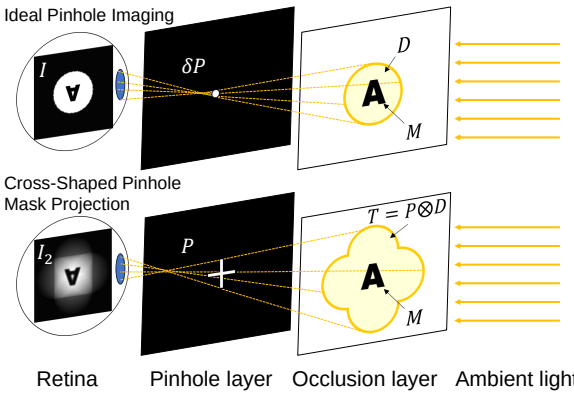


Figure 6: Comparison between conventional pinhole imaging (top) and our occlusion-capable system using a cross-shaped pinhole (bottom). Top row: A finite pinhole $\delta P$ transmits a narrow cone of light, forming a sharp retinal projection $I$. Bottom row: The shaped pinhole $P$, combined with the angular filtering imposed by the pupil $D$, defines an effective transmittance $T = P \otimes D$. The observed image $I_2$ results from modulating the ambient light field with the occlusion pattern $M$ and convolving with the system PSF.

by thresholding and contour detection. Among the detected contours, we select the one with the largest area, denoted as $A_{\text{contour}}$. We then fit a minimum-area bounding rectangle to this contour and denote its area as $A_{\text{rect}}$, with side lengths $w$ and $h$. The squareness ratio is defined as:

$$\text{Squareness} = \frac{\min(w,h)}{\max(w,h)} \cdot \frac{A_{\text{contour}}}{A_{\text{rect}}}.$$

The trend indicates that the system exhibits tolerance in the selection of line length: even when $d < 2\sigma_p$ , the resulting blur spot maintains square-like characteristics and corner sharpness. This flexibility allows a broader range of design choices that balance visual uniformity and crosstalk. However, the final choice of these parameters must also consider the combined effect of the pinhole array and the occlusion mask, which we analyze in Sec. 5.1.

### 3.3 Occlusion Mask Formation

#### 3.3.1 Ideal Pinhole Imaging

To clearly illustrate how our occlusion mask forms on the retina, we first consider an idealized scenario of pinhole imaging. As depicted in Fig. 6 (top), an infinitely small pinhole transmits only a narrow cone of ambient light. Before reaching the pinhole, the ambient illumination is spatially modulated by the mask pattern $M(x,y)$. Subsequently, the eye's pupil further limits which rays can reach the retina, effectively defining a circular region $D(x,y)$ on the occlusion layer from which light can enter the pupil. We define this pupil-projected region $D(x,y)$ as a binary disk function:

$$D(x,y) = \begin{cases} 1, & x^2 + y^2 \le \sigma_D^2 \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

This disk results from the projection of the pupil through the pinhole onto the occlusion layer. The radius $\sigma_D$ of this projected region can be computed using triangle similarity as:

$$\sigma_D = \frac{a}{2} \cdot \frac{d_s}{d_e}, \tag{6}$$

where $d_s$ is the spacing between the pinhole and the occlusion layer. Thus, the retinal intensity distribution for an ideal pinhole system is given by:

$$I(x,y) = [M(x,y) \cdot D(x,y)] \otimes \delta P(x,y), \tag{7}$$

where $\delta P(x,y)$ represents the ideal point pinhole.

In theory, the finite aperture of the pupil introduces a defocus blur, which can be modeled as a point spread function $H_{p2}$. However, because the angular cone of light is already tightly constrained by the pinhole, the peripheral portions of $H_{p2}$ are effectively clipped. This truncation makes the contribution of defocus blur negligible in practice, especially when the pinhole is placed close to the eye.

#### 3.3.2 Cross-shaped Pinhole Imaging

In our system (shown in Fig. 6 (bottom)), the pinhole aperture is no longer infinitely small. Instead, it has a finite, cross-shaped aperture described by the transmittance function $P_+(x,y)$ as

$$P_+(x,y) = \begin{cases} 1, & \text{if } \left(|x| \le \frac{d}{2}, |y| \le \frac{t}{2}\right) \text{ or } \left(|y| \le \frac{d}{2}, |x| \le \frac{t}{2}\right) \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

Here, $t$ is no longer the pinhole diameter but the line thickness. Unlike the ideal point pinhole, this shaped aperture allows ambient light modulated by the occlusion mask to transmit through multiple directions.

Similar to the ideal pinhole imaging, we computed the pupil-projected region through the cross-shaped pinhole as $T(x,y)$, which can be derived by:

$$T(x,y) = D(x,y) \otimes P_+(x,y). \tag{9}$$

Eq. (9) defines the effective transmittance function of $T(x,y)$ on the occlusion layer. Thus, an effective occlusion region given by the product of $T(x,y)$ and the modulation of mask pattern $M(x,y)$ is formed on the occlusion layer. As a consequence, the final retinal intensity distribution $I_2(x,y)$ is formulated as:

$$I_2(x,y) = [M(x,y) \cdot T(x,y)] \otimes H_{p2}(x,y), \tag{10}$$

where $H_{p2}(x,y)$ is the defocus-induced PSF, originating from the mismatch between the occlusion layer and the eye's focal plane. Although $H_{p2}$ is inherently large due to the significant defocus distance, the shaped pinhole also significantly restricts the angular spread of transmitted rays, leading to a sharper retinal image than conventional soft-edge occlusion approaches.

This convolutional formulation is conceptually related to coded aperture imaging [16, 34], where a scene is modulated by a spatially structured aperture, and the resulting image can be interpreted as the convolution of the aperture pattern with the incoming light field.

However, the convolution in Eq. (10) introduces spatial crosstalk across the occlusion mask. Theoretically, similar to the deblurring approach used in coded aperture systems, one could apply deconvolution to generate a pre-filtered occlusion mask $M'(x,y)$ such that, after convolution with the system's effective PSF, the retinal result approximates the intended binary mask:

$$M'(x,y) \cdot T(x,y)] \otimes H_{p2}(x,y) \approx M(x,y). \tag{11}$$

Moreover, due to the interaction between the pinhole array and the pupil, the effective transmittance region $T(x,y)$ takes on a soft-edged, rounded shape, as illustrated in Fig. 6. This shape complicates uniform tiling across the occlusion layer. To simplify analysis and enable regular tiling, we approximate each $T(x,y)$ by a square region $\tilde{T}(x,y)$ centered at the corresponding $T(x,y)$, as shown in Fig. 7. Since the peripheral regions of $T(x,y)$ contribute less intensity, this square approximation retains the majority of the energy while facilitating an easier layout.

As shown in Fig. 7, if the pinhole pitch on the pinhole layer is $s_1$, the center-to-center spacing and side length of the corresponding $T(x,y)$ on the S2 become:

$$s_2 = \sigma_D \cdot s_1, \tag{12}$$

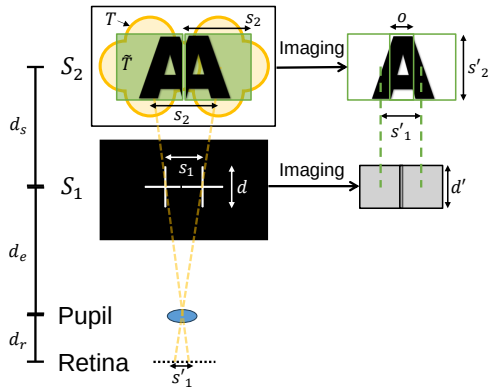


Figure 7: Elemental mask generation geometry. Square regions $\tilde{T}$ on the occlusion layer approximate the pupil-constrained transmittance of each pinhole. Their projections partially overlap on the retina, forming a complete occlusion pattern.

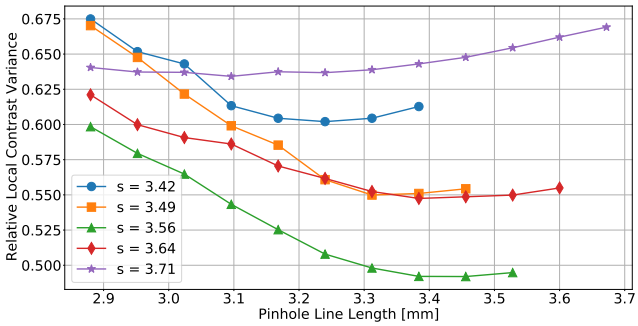


Figure 8: Relative local contrast variance for different pinhole pitches and line lengths. Each data point corresponds to a simulated retinal image under specific pinhole parameters, evaluated using the gradient-based local contrast metric described in Sec. 5.1. Lower values indicate more uniform appearance.

where $\sigma_D$ is the pupil projection radius onto the occlusion layer (defined in Eq. (6)), which can be interpreted as the geometric magnification from the pinhole plane to the occlusion plane. Each elemental mask $\tilde{T}$ thus occupies a square region of size $s_2 \times s_2$.

Each $\tilde{T}$ passes through a corresponding pinhole and projects onto the retina. The projected center-to-center spacing on the retina is denoted $s'_1$, and the projected size of each mask is $s'_2$. Based on geometric relationships, we derive the retinal overlap ratio $R_o$ as:

$$R_o = \frac{o}{s'_2} = 1 - \frac{2(d_e + d_s)}{a d_s}, \tag{13}$$

where $o$ is the overlap between adjacent projected masks.

We adopt the content generation procedure proposed by Akşit et al. [1] to synthesize the final occlusion mask. Specifically, once the overlap ratio is determined, we crop and map regions from a target image onto each elemental mask according to the indexing and offset scheme.

## 4 Implementation

In this section, we describe the implementation of our prototype system from both hardware and software perspectives.

### 4.1 Hardware Setup

To evaluate the retinal appearance of our occlusion mask, we constructed a benchtop prototype, as shown in Fig. 1(a). To ensure parallel alignment of the two LCD panels and maintain coaxial optical geometry, the LCDs are mounted within a semi-cage optical structure. A camera is attached to an XYZ translation stage for flexible positioning and precise alignment along the optical axis. While pixel-level alignment improves image sharpness, our system does not require subpixel precision, as each elemental mask spans multiple pixels, allowing for some tolerance in misalignment

We use two Sony LCX017 transmissive LCD panels (1024×768 pixels, pixel pitch=0.036 mm, 36.9×27.6 mm active area, 60 Hz, monochrome), placed $d_s = 40$ mm apart. The 40 mm gap limits its geometric-defocus blur to $\approx 0.1^\circ$ (full width) while preventing overlap of the square regions $\tilde{T}$ on S2 [1]. Such overlap would let multiple S2 pixels pass through neighbouring pinholes and greatly increase crosstalk The inner polarizers between the two LCDs were removed to improve light transmission without affecting display functionality when stacked. Since the FOV is based on the size of the LCDs and the separation between them, the prototype has a maximum FOV of H37°×V28°. We render virtual content on a 640×480 LCD panel. A 50 mm focal length convex lens is used to focus the image onto a 10×10 mm T90:R10 cube beam splitter, forming a see-through virtual image overlay. The LCD panels are controlled by external driver boards connected by VGA and HDMI links to an Intel Core i7-12700K PC with an NVIDIA GeForce RTX 4080 SUPER GPU running Windows 11.

The camera is configured with a fixed focusing distance of 1 m and an aperture diameter of approximately 3.78 mm, which reflects typical AR usage scenarios. As discussed in Equation 4, the PSF stabilizes beyond 1 m, and while closer focusing could reduce the PSF size and improve sharpness, our choice of 1 m demonstrates that the occlusion remains effective even under practical conditions. Importantly, the camera aperture is kept at its maximum circular setting, as any deviation from a circular pupil would distort the PSF shape, causing the defocused blur of the cross-shaped pinholes to deviate from the desired square-like form. The distance from the camera lens to the front LCD (S1) is approximately 20 mm.

### 4.2 Software

The rendering of both the pinhole pattern and the occlusion mask is implemented using OpenCV with CUDA acceleration.

Once the pinhole parameters are determined, we first create a canvas where white seed pixels are placed at regular intervals according to the specified pinhole pitch. A cross-shaped morphological kernel is then constructed based on the desired pinhole line width and length. Applying an erosion operation using this kernel produces the final cross-shaped pinhole layout. Since all operations are executed on the GPU, real-time updates such as scaling or translating the pinhole grid are fully supported, with a measured latency of approximately 0.03 s.

The rendering of the elemental occlusion masks follows the procedure illustrated in Fig. 7, where the original mask is partitioned into elemental regions based on the current pinhole layout.

## 5 Results and Evaluation

In this section, we present three types of evaluation: a defocus-based simulation model to analyze how pinhole layout parameters affect retinal image quality (Sec. 5.1), simulated user views to evaluate the perceived quality of the occlusion and the virtual image (Sec. 5.2), and real-world see-through observations using our physical prototype (Sec. 5.3).

### 5.1 Parameters Selection

#### 5.1.1 Pinhole Parameters

Given the fixed hardware parameters described above, we compute the PSF radius $\sigma_p$ using Eq. (4), which yields approximately 1.85 mm. Thus, when the center-to-center pitch of the pinholes is set to $s = 1.85 \times 2 = 3.7$ mm, the defocused pinhole images just barely avoid overlapping on the retina.

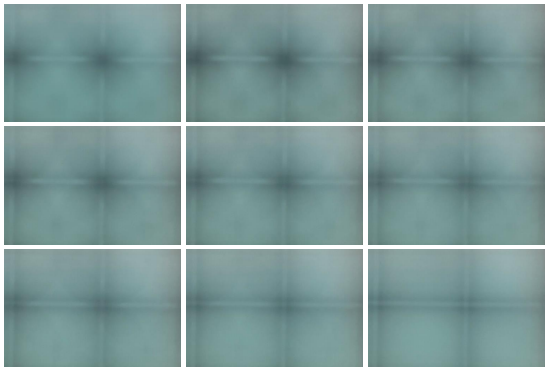

Figure 9: Photographs of real see-through views captured using our prototype, with a fixed pinhole pitch of 3.56 mm and varying pinhole line lengths from 2.91 mm to 3.56 mm. Images are arranged left to right, top to bottom, in order of increasing line length.

However, as shown in Fig. 4(e), the square blur spots exhibit slightly lower intensity near their edges. Therefore, introducing a small degree of overlap can help balance the overall brightness and produce a more uniform appearance. Moreover, from the curvature analysis in Fig. 5, we observe that even when the pinhole line length $d$ is slightly smaller than the PSF diameter $2\sigma_p$, the curvature of the corners remains relatively stable. This indicates a certain tolerance in line length selection.

Based on these observations, we explore a range of pinhole pitches $s \in [3.42, 3.71]$ and line lengths $d \in [2.88, 2s]$, and simulate the resulting retinal images. Each simulated retinal image is divided into 8px $\times$ 8px subregions, and we calculate the average gradient magnitude within each subregion using the Sobel operator, as local gradients effectively reflect perceptual non-uniformities and edge saliency in the image. Specifically, we compute horizontal and vertical gradients using the Sobel kernels, denoted by $G_x(x,y)$ and $G_y(x,y)$, and define local contrast as:

$$\text{LocalContrast}_{i,j} = \frac{1}{|R_{i,j}|} \sum_{(x,y)\in R_{i,j}} \sqrt{G_x^2(x,y) + G_y^2(x,y)},$$

where $R_{i,j}$ denotes the $i,j$-th local patch. A lower value indicates less perceptual edge strength. We then compute the variance-to-mean ratio of these local contrasts to quantify visual uniformity:

$$\text{RelativeVariance} = \frac{\text{Var}\left(\sum \text{LocalContrast}\right)}{\text{Mean}\left(\sum \text{LocalContrast}\right)}$$

The results are summarized in Fig. 8. We observe that when the pinhole pitch is set to $s = 3.56$ mm, the relative local contrast variance reaches its minimum. This corresponds to a slight overlap as shown in Fig. 10(c). Compared to the circular pinhole layout shown in Fig. 10(a) (replicating the configuration in [1]), we visually reduce global contrast. Furthermore, we find that when the pinhole line length lies within the range of 3.35 mm to 3.56 mm, the relative variance remains relatively stable. This suggests that a smaller cross shape can be used without significantly compromising perceptual uniformity. Reducing the line length in this manner also helps minimize spatial crosstalk between adjacent pinholes. These findings are also qualitatively supported by subjective visual assessments conducted in real-world observations, as shown in Fig. 9.

In addition to pitch and line length, the line width $t$ also plays a critical role in shaping the retinal blur spots. Narrower lines produce sharper square-shaped spots. However, reducing the line width too much can severely limit transmittance. To balance spot sharpness with light throughput, we choose $t = 2$ pixels, which corresponds to 0.072 mm.


Figure 10: See-through views under different pinhole configurations (without occlusion mask) with the camera focused on a monitor 1m away. (a) Circular pinhole layout replicating the configuration in [1], (b) Star-shaped pinholes, (c) Cross-shaped pinholes ($s = 3.56$ mm).

#### 5.1.2 Occlusion Mask Generation

In our implementation, we initially explored deconvolution-based approaches to generate occlusion masks that would counteract the spatial blurring introduced by the pinhole PSF. However, we found that stable deconvolution was challenging in practice, primarily due to the non-Gaussian and anisotropic shape of the cross-shaped pinhole kernel, which introduces high-frequency components and ill-conditioned inverses in the Fourier domain. This often resulted in noise amplification and instability.

As a practical compromise, we adopted a simpler approach by directly partitioning the original mask into elemental regions. Although this approach introduces some degree of spatial crosstalk, our results suggest that the level of interference for cross-shaped pinholes remains visually acceptable and perceptually tolerable for most regions.

While star-shaped pinholes produce more uniform brightness distributions across the retina due to their better tiling properties (Fig. 10(b)), they also exhibit more severe crosstalk because of the wider angular support of each element. In light of these trade-offs, we selected the cross shape for our final prototype, balancing simplicity of content generation with acceptable visual quality.

### 5.2 Simulation Results

We evaluated the quality of our system using simulated user views. To this end, we simulated our display in software and rendered a light field across the user's pupil (a = 3.78mm) and accumulated the light field views to generate the retinal image of the user. All light field views share a common image plane where objects have zero disparity at the focus distance of the user which we set to 1 m. Using this method, we virtually placed the display in front of a rendered scene, which we augmented with virtual content. For evaluation, we first generated reference images, in which the virtual content completely occludes the real world. To better reflect real-world conditions, we set the brightness of the virtual content to be only a third of the real-world brightness. We further normalized the perceived brightness of the real world across all different display configurations. We compare the ground truth images to three different AR displays: one without any occlusion capability, one with an occlusion layer but without the pinhole layer, and our system. The simulated system configuration is consistent with our physical prototype described in Sec. 4.1. To quantitatively assess the image quality in all of these cases, we used the Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index Measure (SSIM).

The quantitative results indicate that our display yields significantly higher PSNR and SSIM values compared to a system without occlusion capabilities. Furthermore, our system yields slightly higher values than the occlusion-capable display without pinholes as well, indicating that the image quality is on par with a non-pinhole system. However, the metrics do not entirely capture the sharper occlusion capability of our system, as can be seen in Fig. 11 with $d = s = 3.56$ mm.

### 5.3 See-through Results

To evaluate the observed occlusion effects under real-world viewing conditions, we captured a series of see-through photographs using our prototype with $s = 3.56$ mm, $d = 3.35$ mm. The camera was

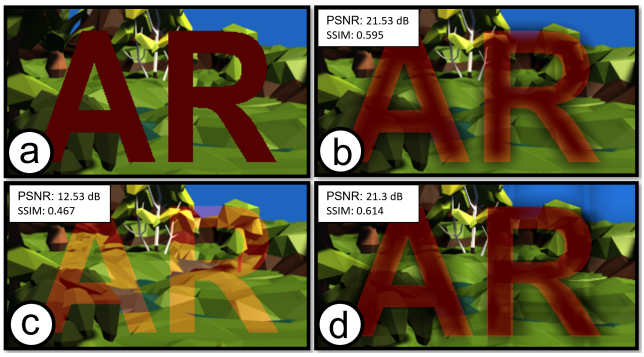


Figure 11: Simulated results using a light field placed at the location of the eyebox. We render virtual content on top of a real scene. (a) the ground truth (real world fully blocked with perfect edges). (b) an occlusion-capable system without a pinhole layer. (c) conventional system without occlusion-capabilities. (d) our system ($d = s = 3.56$ mm). The comparison indicates, that our system leads to less light bleeding than the other systems.

focused on a physical background located 1 m behind the display, simulating a typical use-case scenario for OST-HMDs.

Fig. 12 presents two sets of comparison images. The top row (a–c) shows the occlusion masks under a uniform white background. Figure (a) depicts the case without the pinholes, while (b) and (c) correspond to circular and cross-shaped pinholes, respectively. Both pinhole configurations share the same spatial layout, and thus use identical elemental occlusion masks. Compared to the defocused mask in (a), (b) and (c) produce sharper edges due to the increased DOF from the pinhole layer. However, (b) shows prominent diamond-shaped gaps caused by the limited tiling ability of circular pinholes. In contrast, the cross-shaped pinholes in (c) enable better tiling and a more continuous mask, though with slightly more blur due to increased angular support and diagonal crosstalk.

The bottom row (d–f) of Fig. 12 displays real see-through views of the "ISMAR 2025" background. In (d), without the pinhole layer, the background text appears sharp and unobstructed. In (e) and (f), the occlusion masks generated by the circular and cross-shaped pinholes partially block the background, demonstrating the system's ability to modulate light transmission in space. Notably, the cross-shaped configuration results in slightly more diffused edges, but still maintains effective occlusion. These observations align with the simulated PSF behavior and support the practical viability of the proposed occlusion-capable display system.

Finally, we demonstrate virtual content rendering using the same optical setup. As shown in Fig. 1(b–d), (b) shows a defocused mask causing the virtual image to appear semi-transparent. (c) uses circular pinholes to produce a solid virtual image, but visible diamond-shaped gaps remain. (d) shows our cross-shaped pinholes achieving both solid virtual content and a clear background view.

To further assess the visual performance of the occlusion display under real-world conditions, we computed the modulation transfer function (MTF) curves of the displayed occlusion mask from the captured see-through images. Fig. 13 shows the MTF extracted from edge profiles in the occlusion mask region, for three configurations: without pinholes (Fig. 12(a)), with circular pinholes (Fig. 12(b)), and with cross-shaped pinholes (Fig. 12(c)). In the no-pinhole configuration, the MTF declines most rapidly, approaching MTF20 at the spatial frequency of 0.4 cycle/degree, confirming significant blur and poor spatial detail preservation. The circular pinhole configuration achieves the highest sharpness among the tested setups, with its MTF decreasing more gradually and maintaining upon MTF20 until 1.6 cycle/degree. The cross-shaped pinhole configuration exhibits performance intermediate between these two extremes, where MTF20 is achieved at 1.5 cycle/degree, demonstrating that our method provides a notable improvement in preserving the sharpness of the occlusion mask edges while enabling better tiling in the see-through view.

## 6 Discussion and Limitation

This section outlines the key limitations observed in our current prototype and simulation, including crosstalk, transmittance, diffraction artifacts, perceptual factors, and the potential integration of eye tracking. We also suggest potential solutions and future directions to overcome these challenges.

### 6.1 Crosstalk

A key limitation of our approach lies in the presence of spatial crosstalk between adjacent pinholes. In the cross-shaped configuration, each pinhole has extended horizontal and vertical arms that increase the angular support of the transmitted light. Small spacing between pinholes can lead to overlapping light fields at the retina, thereby reducing edge sharpness and introducing a mild glow or leakage around the mask edges. This effect is particularly evident when occlusion masks contain high-frequency details.

While our results suggest that the level of crosstalk remains within perceptually tolerable bounds, further reduction may be possible by optimizing the pinhole design. Potential strategies include implementing non-uniform transmittance profiles or applying tailored deconvolution methods to compensate for angular leakage. Specifically, we consider deep-learning-based inverse-diffraction design approaches to compute grayscale pinhole patterns with optimized PSF behavior.

In fact, we initially explored a star-shaped pinhole configuration as shown in Fig. 10(b) which provides higher tiling efficiency. However, this design also introduces more severe crosstalk, as light is allowed to spread along three angular axes rather than two, resulting in significant degradation of occlusion quality. For this reason, we did not pursue the star-shaped design further in this work. Nevertheless, we believe that with an appropriate crosstalk compensation mechanism, this configuration could become a promising direction for future exploration.

### 6.2 Scene Brightness

Another limitation of our system lies in the scene brightness. The transparent LCD panel we use (Sony LCX017) has a light transmittance of only 21%, which already imposes a baseline reduction in brightness. Although the use of a pinhole array improves transmittance compared to a single pinhole, it further restricts the amount of light that reaches the eye.

It is worth noting, however, that the cross-shaped pinholes in our prototype can achieve higher transmittance than circular pinholes under comparable sharpness conditions. While increasing the diameter of a circular pinhole compromises occlusion sharpness, cross-shaped pinholes can extend their line lengths while keeping the line width narrow. For instance, in the second column of Fig. 12, the circular pinhole occupies approximately 113 pixels in area, whereas the cross-shaped pinhole spans 194 pixels in total. This design allows more light to pass through while still maintaining adequate edge definition.

A potential solution to the brightness limitation is to render pinholes only within the viewer's FOV. By concentrating pinhole placement in central vision, it may be possible to preserve perceived sharpness while improving overall light transmittance. Moreover, because perceived brightness compresses physical luminance nonlinearly [21], our informal observations suggest that visibility remains acceptable, which warrants further investigation.

### 6.3 Diffraction Issue of Transparent LCDs

In our prototype, we use two identical transparent LCD panels as SLMs. As shown in the second row of Fig. 12, we observe no-

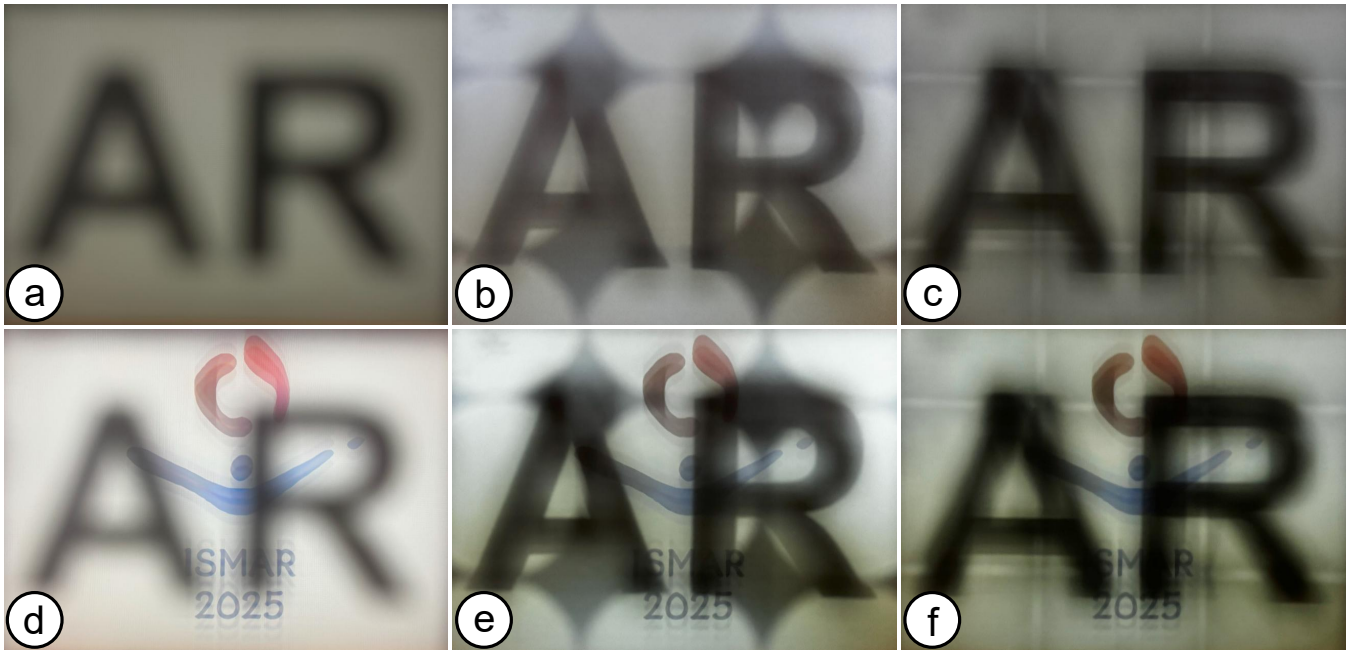


Figure 12: Comparison of see-through occlusion performance under different pinhole configurations. The camera focus is set to 1 m. (a–c) show occlusion mask visibility under uniform illumination with no background content: (a) no pinhole, (b) circular pinholes with 0.22 mm radius, and (c) cross-shaped pinholes ($s = 3.56$ mm, $d = 3.35$ mm). Both (b) and (c) use identical spatial layouts and thus share the same elemental occlusion mask. (d–f) show real-world see-through views with a physical background placed at 1 m: (d) no pinhole, (e) circular pinholes, and (f) cross-shaped pinholes. Exposure times were adjusted per condition to ensure visibility: (a,d) 1/60 s, (b,e) 1/30 s, and (c.f) 1/40 s.

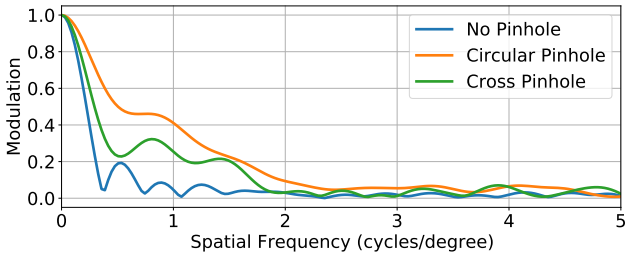


Figure 13: MTF curves of the formed occlusion masks for three pinhole configurations: no pinhole, circular pinhole, and our cross-shaped pinhole, measured from real-world see-through views.

ticeable ghosting in the background text, which is caused by the diffraction due to the gaps between liquid crystal cells. This effect is a known limitation of transparent LCDs and becomes particularly visible under high-contrast see-through conditions. To make transparent LCDs a viable platform for occlusion-capable displays, it is essential to mitigate the impact of diffraction. Recent research has explored alternative approaches, such as using photochromic materials to implement dynamic occlusion layers without active pixel structures [3, 20]. Additionally, devices like Magic Leap 2 have replaced traditional square pixel cells with curved geometries to suppress diffraction artifacts [23]. These strategies highlight promising directions for improving optical clarity in future implementations.

### 6.4 Human Visual Perception

Although our evaluation includes both simulations and real-world photographs, certain visual effects, such as crosstalk or edge softness, are not always accurately reflected by image metrics alone [8]. Human perception is influenced by factors such as background content, attention, and contrast sensitivity. Slight artifacts that are measurable may go unnoticed, while others may appear more prominent than expected. In future work, we plan to conduct user studies to better understand how different pinhole designs and rendering strategies affect the perceived quality of occlusion, particularly under varying lighting and background conditions.

### 6.5 Integration with Eye Tracking

One advantage of rendering the pinhole array on a transparent LCD is the ability to make the pinholes dynamic. This allows the pinhole parameters to adapt in real time to changes in pupil size. With the addition of an eye-tracking feature, the system can further adjust the spatial layout of the pinholes based on the user's gaze direction.

In particular, it is desirable to ensure that at least one pinhole is positioned precisely along the gaze direction. This alignment helps prevent repeated zone artifacts, which may otherwise occur when the eye rotates and no pinhole is centered along the new viewing direction [24], and effectively makes the eyebox primarily limited by the LCD size rather than by physical pinhole placement. Additionally, the approach described in Sec. 6.2, rendering the pinhole array only in the central FOV, can also be realized through gaze-contingent rendering.

## 7 Conclusion

We proposed X-Mask, a wearable occlusion-capable OST-HMD based on a cross-shaped pinhole array. Our system improves occlusion sharpness and visual uniformity using only two transparent LCD panels, which render a pinhole array and an occlusion mask, respectively. Simulation and prototype evaluations confirm its effectiveness under defocus. These results highlight the potential of cross-shaped pinhole arrays as a practical solution for lightweight, lensless occlusion displays.

## Acknowledgments

This work was supported by Snap Inc. and JSPS KAKENHI (22H00539), Japan.

## REFERENCES


[1] K. Akşit, J. Kautz, and D. Luebke. Slim near-eye display using pinhole aperture arrays. *Applied optics*, 54(11):3422–3427, 2015. 2, 3, 4, 6, 7

[2] K. Akşit, A. H. G. Niaki, E. Ulusoy, and H. Urey. Super stereoscopy technique for comfortable and realistic 3d displays. *Optics letters*, 39(24):6903–6906, 2014. 3

[3] M. Chae, K. Bang, Y. Jo, C. Yoo, and B. Lee. Occlusion-capable see-through display without the screen-door effect using a photochromic mask. *Optics Letters*, 46(18):4554–4557, 2021. 2, 9

[4] H. Choi, J. Kim, S.-W. Cho, Y. Kim, J. B. Park, and B. Lee. Three-dimensional–two-dimensional mixed display system using integral imaging with an active pinhole array on a liquid crystal panel. *Applied optics*, 47(13):2207–2214, 2008. 3

[5] C. Ebner, P. Mohr, T. Langlotz, Y. Peng, D. Schmalstieg, G. Wetzstein, and D. Kalkofen. Off-axis layered displays: Hybrid direct-view/near-eye mixed reality with focus cues. 29(5):2816–2825, 2023. doi: 10.1109/TVCG.2023.3247077 2

[6] 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, 2019. 2

[7] W. Han, J.-W. Lee, J.-Y. Shin, M.-H. Choi, H.-R. Kim, and J.-H. Park. Varifocal occlusion in an optical see-through near-eye display with a single phase-only liquid crystal on silicon. *Photonics Research*, 12(4):833–853, 2024. 2

[8] X. Hu, Y. Zhang, A. Plopski, Y. Itoh, M. Perusquía-Hernández, N. Isoyama, H. Uchiyama, and K. Kiyokawa. Perception-driven soft-edge occlusion for optical see-through head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics*, 2024. 2, 9

[9] X. Hu, Y. Zhang, H. Uchiyama, N. Isoyama, N. Sakata, and K. Kiyokawa. Smart dimming sunglasses for photophobia using spatial light modulator. *Displays*, 81:102611, 2024. 2

[10] F.-C. Huang, G. Wetzstein, B. A. Barsky, and R. Raskar. Eyeglasses-free display: towards correcting visual aberrations with computational light field displays. *ACM transactions on graphics (TOG)*, 33(4):1–12, 2014. 2, 3

[11] 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, 2017. doi: 10.1109/TVCG.2017.2734427 2

[12] 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, 2019. 1

[13] Y.-G. Ju, M.-H. Choi, P. Liu, B. Hellman, T. L. Lee, Y. Takashima, and J.-H. Park. Occlusion-capable optical-see-through near-eye display using a single digital micromirror device. *Optics letters*, 45(13):3361–3364, 2020. 2

[14] 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. IEEE, 2000. 1, 2

[15] H. Lee, J.-H. Jung, S. Hong, and H.-J. Choi. Near eye display based on multiplexed retinal projections for robust compensation of eye pupil variance. *Optics Express*, 32(2):2631–2643, 2024. 2, 3, 4

[16] A. Levin, R. Fergus, F. Durand, and W. T. Freeman. Image and depth from a conventional camera with a coded aperture. *ACM transactions on graphics (TOG)*, 26(3):70–es, 2007. 5

[17] A. Maimone and H. Fuchs. Computational augmented reality eyeglasses. In *2013 IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2013*, pp. 29–38, 2013. doi: 10.1109/ISMAR.2013.6671761 2

[18] 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. In *ACM SIGGRAPH 2014 emerging technologies*, pp. 1–1. 2014. 2, 4

[19] S. K. Nayar and V. Branzoi. Adaptive dynamic range imaging: Optical control of pixel exposures over space and time. In *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2, pp. 1168–1175, 2003. doi: 10.1109/iccv.2003.1238624 2

[20] C.-W. Ooi, Y. Hiroi, and Y. Itoh. A compact photochromic occlusion capable see-through display with holographic lenses. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 237–242. IEEE, 2023. 9

[21] A. Radonjić, S. R. Allred, A. L. Gilchrist, and D. H. Brainard. The dynamic range of human lightness perception. *Current Biology*, 21(22):1931–1936, 2011. 8

[22] K. Rathinavel, G. Wetzstein, and H. Fuchs. Varifocal occlusion-capable optical see-through augmented reality display based on focus-tunable optics. *IEEE transactions on visualization and computer graphics*, 25(11):3125–3134, 2019. 2

[23] A. I. Russell, J. I. Trisnadi, V. Mathur, M. R. Johnson, and C. Carlisle. Geometries for mitigating artifacts in see-through pixel arrays, U.S. Patent 00,038,72 A1, Jan. 7, 2021. 2, 9

[24] W. Song, Q. Cheng, P. Surman, Y. Liu, Y. Zheng, Z. Lin, and Y. Wang. Design of a light-field near-eye display using random pinholes. *Optics Express*, 27(17):23763–23774, 2019. 2, 9

[25] W. Song, Y. Wang, D. Cheng, and Y. Liu. Light field head-mounted display with correct focus cue using micro structure array. *Chinese Optics Letters*, 12(6):060010, 2014. 2, 3, 4

[26] E. W. Tatham. Optical occlusion and shadows in a 'see-through' augmented reality display. In *1999 IEEE International Conference on Information Visualization (Cat. No. PR00210)*, pp. 128–131. IEEE, 1999. 2

[27] 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, 2017. 1, 2

[28] A. Wilson and H. Hua. Design of a pupil-matched occlusion-capable optical see-through wearable display. *IEEE Transactions on Visualization and Computer Graphics*, 28(12):4113–4126, 2021. 1, 2

[29] C. Yao, D. Cheng, and Y. Wang. Uniform luminance light field near eye display using pinhole arrays and gradual virtual aperture. In *2016 International Conference on Virtual Reality and Visualization (ICVRV)*, pp. 401–406. IEEE, 2016. 2, 3, 4

[30] Y. Zhang, X. Hu, K. Kiyokawa, N. Isoyama, N. Sakata, and H. Hua. Optical see-through augmented reality displays with wide field of view and hard-edge occlusion by using paired conical reflectors. *Optics letters*, 46(17):4208–4211, 2021. 1, 2

[31] Y. Zhang, X. Hu, K. Kiyokawa, N. Isoyama, H. Uchiyama, and H. Hua. Realizing mutual occlusion in a wide field-of-view for optical see-through augmented reality displays based on a paired-ellipsoidal-mirror structure. *Optics Express*, 29(26):42751–42761, 2021. 2, 3

[32] Y. Zhang, X. Hu, K. Kiyokawa, and X. Yang. Add-on occlusion: Turning off-the-shelf optical see-through head-mounted displays occlusion-capable. *IEEE Transactions on Visualization and Computer Graphics*, 2023. 1, 2

[33] Y. Zhang, N. Isoyama, N. Sakata, K. Kiyokawa, and H. Hua. Super Wide-view Optical See-through Head Mounted Displays with Per-pixel Occlusion Capability. In *Proceedings of the 2020 IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2020*, pp. 301–311, 2020. doi: 10.1109/ISMAR50242.2020.00056 2

[34] C. Zhou, S. Lin, and S. Nayar. Coded aperture pairs for depth from defocus. In *2009 IEEE 12th international conference on computer vision*, pp. 325–332. IEEE, 2009. 5