

### Perceptual Guidelines for Optimizing Field of View in Stereoscopic Augmented Reality Displays

MINQI WANG and EMILY A. COOPER, University of California, Berkeley

Near-eye display systems for augmented reality (AR) aim to seamlessly merge virtual content with the user's view of the real-world. A substantial limitation of current systems is that they only present virtual content over a limited portion of the user's natural field of view (FOV). This limitation reduces the immersion and utility of these systems. Thus, it is essential to quantify FOV coverage in AR systems and understand how to maximize it. It is straightforward to determine the FOV coverage for monocular AR systems based on the system architecture. However, stereoscopic AR systems that present 3D virtual content create a more complicated scenario because the two eyes' views do not always completely overlap. The introduction of partial binocular overlap in stereoscopic systems can potentially expand the perceived horizontal FOV coverage, but it can also introduce perceptual nonuniformity artifacts. In this arrticle, we first review the principles of binocular FOV overlap for natural vision and for stereoscopic display systems. We report the results of a set of perceptual studies that examine how different amounts and types of horizontal binocular overlap in stereoscopic AR systems influence the perception of nonuniformity across the FOV. We then describe how to quantify the horizontal FOV in stereoscopic AR when taking 3D content into account. We show that all stereoscopic AR systems result in a variable horizontal FOV coverage and variable amounts of binocular overlap depending on fixation distance. Taken together, these results provide a framework for optimizing perceived FOV coverage and minimizing perceptual artifacts in stereoscopic AR systems for different use cases.

CCS Concepts: • Computing methodologies → Perception; Mixed/augmented reality;

Additional Key Words and Phrases: Augmented reality, binocular vision, near-eye display, field of view

#### **ACM Reference format:**

Minqi Wang and Emily A. Cooper. 2022. Perceptual Guidelines for Optimizing Field of View in Stereoscopic Augmented Reality Displays. *ACM Trans. Appl. Percept.* 19, 4, Article 19 (November 2022), 23 pages. https://doi.org/10.1145/3554921

#### 1 INTRODUCTION

Digital displays have become essential tools for education, work, healthcare, and entertainment. While conventional displays present imagery on an opaque panel, emerging **augmented reality** (**AR**) display systems aim at creating mixtures of real and virtual content that are visually immersive. These AR systems often rely on stereoscopic near-eye displays that optically combine the user's natural vision with 3D virtual content directed to the viewer's eyes from a pair of micro-displays or other light sources (Figure 1).

This work was supported by a grant from the National Science Foundation (2041726) and a fellowship from the Center for Innovation in Vision and Optics at University of California, Berkeley.

Authors' addresses: M. Wang, Herbert Wertheim School of Optometry and Vision Science, University of California, 381 Minor Hall Berkeley, CA 94720-2020; email: mwang67@berkeley.edu; E. A. Cooper, Herbert Wertheim School of Optometry and Vision Science, Helen Wills Neuroscience Institute, University of California, 381 Minor Hall Berkeley, CA 94720-2020; email: emilycooper@berkeley.edu.

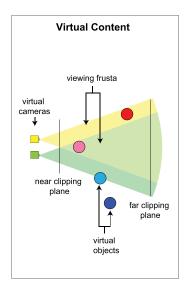


This work is licensed under a Creative Commons Attribution International 4.0 License.

© 2022 Copyright held by the owner/author(s). 1544-3558/2022/11-ART19 https://doi.org/10.1145/3554921

ACM Transactions on Applied Perception, Vol. 19, No. 4, Article 19. Publication date: November 2022.

Corrected Version of Record. V.1.1. Published February 01, 2023.



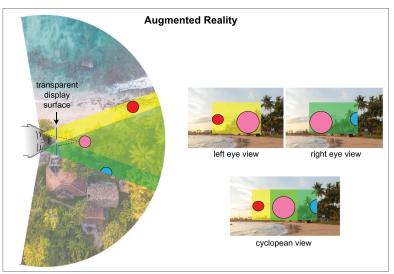


Fig. 1. Illustration of the virtual FOV and the natural FOV in augmented reality systems. Virtual cameras capture virtual content within their viewing frusta (left), and the resulting images are presented on a stereoscopic display that merges this content with a portion of each eye's natural FOV (right). When the eyes' images are fused, the "cyclopean view" can capture a greater horizontal extent than either monocular FOV, with some regions seen binocularly, and others seen monocularly. Image credits: Unsplash (Oliver Sjöström, Rowan Heuvel).

Many factors can influence the immersive nature of AR experiences. For example, physical realism can be limited by the resolution, contrast, and depth information provided by an AR system [31, 46]. Optical elements can interfere with display visibility and lead to distracting visual artifacts as users look around a scene [10, 11]. Importantly, the limited coverage of the user's natural **field of view** (**FOV**) by the display can impair the immersive experience [32] and can also affect performance on a variety of tasks [13, 36, 39, 42]. As such, consumer AR devices often aim at maximizing the FOV covered by the displays.

There are important engineering tradeoffs, however, between a near-eye display's FOV coverage and other factors such as device weight, display resolution, and eyebox size [28, 30]. For example, many current near-eye displays for AR use waveguides to optically combine virtual imagery with the natural FOV. Waveguides work by total-internal-reflection: The permissible direction of light that will be propagated through a waveguide and the coupling grating set limits on the achievable FOV [12, 31, 46]. In addition to optical factors, the physical requirements for covering a large region of the FOV with high spatial and temporal resolution are difficult to achieve without adding weight and size to the system, impacting the form factor [18]. New optical architectures are being developed to overcome these constraints (e.g., [8, 40, 45]), but are not yet mature or practical to manufacture. Currently, near-eye displays for AR still have quite limited FOV compared to the capacity of natural vision. Thus, finding ways to increase FOV coverage without sacrificing other important design factors is a high priority. Here, we describe in detail the concept of FOV for both natural human vision and stereoscopic AR display systems. We then report the results of user studies designed to develop updated perceptual guidelines for optimizing horizontal FOV in AR displays with minimal perceptual artifacts (specifically, visual nonuniformity). Our primary contributions are:

- (1) We clarify the importance of considering binocular overlap when quantifying FOV, both for natural vision and for stereoscopic AR systems.
- (2) We conduct user studies to evaluate two key design factors—the amount of binocular overlap and whether the overlap is convergent or divergent—that are thought to influence perceived FOV in AR. Our results

- suggest that increasing the amount of binocular overlap effectively reduces perceptual nonuniformity across the horizontal FOV. Contrary to prior work using simple stimuli, we find that divergent configurations are generally better than convergent configurations at the content distance.
- (3) While divergent configurations and large binocular overlap are preferred, we show that these properties of near-eye stereoscopic displays change when users look around a 3D scene. Using a simplified display model and combining this model with the user study results, we provide a guide to assist with determining the best display configuration for a given system.

#### 2 RELATED WORK

#### 2.1 The Natural Human Field of View

Knowing people's natural FOV, and how this FOV changes with eye movements, is important for creating technologies that aim at augmenting natural vision. The natural FOV limits the visual space that is available to the viewer at any given moment, and showing virtual content beyond this limit is excessive because the content will not be seen. On the other hand, if the FOV provided by an AR system is smaller than the natural FOV, this can compromise the immersive experience that these technologies seek to deliver.

The natural FOV for human vision is defined as the angular region of visibility for each eye. These angles, together with the direction of gaze, determine the volume of visual space that is visible at a given point in time. Different parts of this FOV are processed differently in the human visual system—for example, the fovea is a small region (about 5° [27]) of the retina with high resolution vision. When people look around a scene, they direct the foveas of both eyes to examine the object of interest (fixation point) and this changes the volume of visible space. Beyond the foveas, each eye sees a large monocular FOV with a shape determined by various factors that are fixed to the anatomy of the eye and the surrounding facial structures [43]. For most people, the upper and nasal sides of the monocular FOVs are head-fixed (that is, they are limited by facial anatomy such as the brow and nose bridge), while the lower and temporal sides are retina-fixed (that is, they are are limited by the edge of retina). For each eye, this monocular FOV extends approximately 60° upward, 75° downward, 60° nasally, and 100° temporally when the eyes look straight ahead [41].

The FOVs of the two eyes do not completely overlap each other. Certain regions of visual space are seen by both eyes (binocular) and other regions are seen by one eye only (monocular), as shown in the top-down views in Figure 2. The visual system takes these partially-overlapping retinal images from the two eyes and creates a *cyclopean* view of the world as if looking from a single viewpoint between the two eyes (as a cyclops in Greek mythology would). Despite having both monocular and binocular regions within this natural cyclopean FOV, the subjective experience is a seamless, singular view of the world. Together, both eyes provide a natural cyclopean FOV that extends to around  $\pm 100^\circ$  horizontally from the midsagittal plane when the eyes are gazing straight ahead, bounded by the temporal margin of each eye's visual field (Figure 2(A)). The binocular overlap region (purple), which extends to around  $\pm 60^\circ$ , plays an important role in depth perception, allowing for stereopsis and precise depth discrimination.

The natural FOV is also dynamic due to eye movements (Figure 2(B), (C)). For example, when people fixate on objects at near distances, their eyes rotate in opposite horizontal directions (called a vergence eye movement) and the horizontal sizes of each monocular FOV and the cyclopean FOV change. Specifically, when the eyes fixate a point that is near to the face, each eye rotates nasally (Figure 2(B)) to converge so that the near point falls on the foveas of both eyes. The nasal side of the monocular FOV of each eye shrinks (because the nose is head-fixed) and the cyclopean FOV also shrinks. On the other hand, the size of the binocular overlap region is the same because the visual angle taken up by the nose in the nasal field is compensated by the temporal field. The eyes also make conjugate horizontal and vertical movements to explore content within the same or different depth planes (Figure 2(C)). With horizontal eye movements, the binocular overlap region remains the same size and in the same position relative to the head, but the monocular and cyclopean FOVs change. Specifically, the

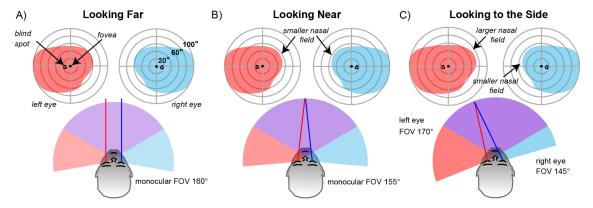


Fig. 2. The natural FOV of human vision is illustrated, including the top-down view of the horizontal FOV (bottom) and a direct depiction of the left and right eye's angular visual field (top). (A) When the eyes are looking far and straight ahead, each eye's temporal field subtends approximately 100° out from the fovea toward the temple, and the nasal field subtends approximately 60° toward the nose. (B) When the eyes converge to a near distance, more of the nasal fields are blocked by the nose. As a result, both the monocular and the cyclopean FOV decrease. However, the amount of binocular overlap remains unchanged. (C) When the eyes are looking to the side, the nasal limit of the FOV changes, expanding the FOV in one eye while shrinking in the other eye. The amount of expanding and shrinking is not necessarily equal between the two eyes, for example, if the eyes are also converged.

monocular portions of the cyclopean FOV change such that one side increases relative to the midsagittal plane, and the other side decreases.

#### 2.2 Field of View in Stereoscopic Near-Eye Displays

In AR, a portion of the natural FOV is augmented with content from digital displays. To avoid occluding natural vision with the display surface, AR devices typically redirect the virtual image of a display to each eye. The physical display is located away from the natural FOV, whereas the virtual images are presented near the center of each eye's natural FOV. The size, distance, and magnification of each virtual image determines the angular size of the FOV that can be stimulated for each eye (the virtual monocular FOV) and their position in each eye's field determines the virtual cyclopean FOV and the amount of binocular overlap.

An ideal AR system would allow the delivery of virtual content anywhere within the natural FOV of the viewer. Because the natural FOV is dynamic (it expands, shrinks, and reorients as described in the previous section), this would require each display to subtend a visual angle that is larger than the natural monocular FOV at any given point in time, or it would require a moving display that fills the instantaneous monocular FOV and moves with the eyes. However, existing systems cannot yet achieve this ideal. Commercially-available AR systems typically subtend around 30°–40° horizontally and vertically in each eye's FOV, creating a rectangular region in which content can be presented [29].

The development of wider FOV near-eye stereoscopic displays for AR has been a topic of research for several decades. In the 1980's, early near-eye displays were being developed for use in the military, and concerns were raised that a restricted FOV could be detrimental for certain operations such as target detection during flight (e.g., [44]). One consideration that emerged prominently during this period was the amount of full or partial binocular overlap in the virtual FOV [24]. In a full binocular overlap scenario, the cyclopean FOV coverage is the same as the monocular FOV coverage (Figure 3, left). However, in a partial overlap scenario, the total horizontal cyclopean FOV over which virtual content can be displayed is increased by horizontally displacing the physical displays. When both eyes view a virtual scene, a larger horizontal portion of the natural FOV is then filled with virtual content, with some regions of the virtual scene seen by both eyes, and other regions are seen by only

## 100% Binocular Overlap left eye FOV right eye FOV fixation point Cyclopean FOV

# Convergent Partial Overlap left eye FOV right eye FOV monocular region Cyclopean FOV

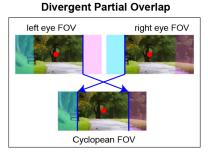


Fig. 3. Illustrations of the monocular and cyclopean FOV subtended when a user fixates at the same location in space (thus the same vergence eye position) for different binocular configurations. Examples are shown for complete binocular overlap (left) and partial overlap imagery: convergent (middle) and a divergent (right). The monocular regions at corresponding locations in each eye are shaded with cyan and magenta. Monocular edges of the display (blue lines) are present in one eye but not the other, creating monocular-binocular borders in the cyclopean view. Image credit: Unsplash (Gary Ellis).

one eye. This design yields a binocular overlap region flanked by two monocular regions, which is similar to the natural FOV except that the total coverage is still small enough to fit fully within the binocular overlap region of natural vision for most eye movements. Unlike the natural FOV, there are two possible configurations for partial overlap displays: convergent and divergent (Figure 3, middle and right). When looking at convergent displays, the monocular regions of the cyclopean FOV occur in the nasal field of each eye, while divergent displays present these monocular regions in the temporal field of each eye. This concept can be used for both **virtual and augmented reality technologies (VR and AR)**. In either modality, it is clear that the horizontal FOV coverage in stereoscopic near-eye systems has an important degree of freedom that can modify both the total cyclopean coverage, and the amount that is binocularly-visible.

#### 2.3 Nonuniformity Artifacts from Partial Field of View Overlap

The partially overlapping views illustrated in Figure 3 pose a problem for the visual system because the viewer sees monocular content within the binocular region of natural vision. Early work on partial overlap displays identified a range of perceptual artifacts associated with these designs [3, 15, 19–26, 33–35, 37, 38]. First, a perceptual fading (sometimes called *luning*) of the content around the monocular-binocular border (blue lines in Figure 3) was noted [26, 34, 35]. The monocular-binocular border was also found to be associated with elevated detection thresholds for presented targets [25]. Lastly, partial overlap was shown to lead to perceived fragmentation, in which the monocular regions appear to break up from the binocular region in some aspect of appearance (e.g., different depth, brightness) [23]. We refer to these collectively as nonuniformity artifacts.

These nonuniformity artifacts interfere with the percept of a continuous virtual FOV. The underlying cause of such artifacts is thought to be the fact that corresponding points in the two eyes are stimulated by highly discrepant stimuli. For example, in Figure 3, the cyan region of the left eye corresponds with the cyan region of the right eye (same for the magenta regions), but only one eye sees content in this region (in this illustration, the other eye simply sees nothing). This type of discrepancy in binocular inputs can lead to interocular suppression, in which the content seen by one eye is perceptually suppressed, or binocular rivalry, in which the content seen by the two eyes appears to alternate in time [7]. It is worth noting that monocularly-visible content does occur in the binocular region of natural vision as well, so their existence is not wholly unnatural [16]. For example, when a foreground object occludes a background, often one eye sees more of the background at the occlusion boundary (a so-called partial occlusion).

Extensive perceptual studies sought to understand and reduce the nonuniformity artifacts associated with partial binocular overlap, resulting in a set of guidelines for how to maximize the horizontal FOV in wearable

displays with minimal artifacts [23, 25, 26, 34]. Several different stimulus factors were found to reduce, but not eliminate, these artifacts and create a more coherent cyclopean view. These factors include adding a smooth luminance fall off toward the monocular-binocular border [34, 35], adjusting the relative luminance of the monocular regions [26], increasing the amount of binocular overlap [23], and adopting a convergent rather than a divergent display configuration (see [22, 24] for review). For example, one study measured how often people detected fading artifacts over the duration of half a minute trial, and showed that viewers reported seeing less fading over time with convergent overlap as compared to divergent overlap [26]. In another study, viewers chose convergent views to be better in terms of perceived uniformity across the display when both configurations were shown together at the same time on the screen to the viewer [23]. However, there are currently barriers to implementing these guidelines in modern stereoscopic AR displays. These prior studies largely used stimuli that do not reflect the visual appearance of current AR systems: They used simple gray scale images in which the entire FOV was limited to the virtual display with no other visual information outside of the display's FOV [23, 25, 26]. In AR, a smaller display FOV is superimposed over the larger natural FOV of human vision. In addition, most modern AR systems use additive light, making virtual content often semi-transparent. As such, both eyes will share more similar visual information in the monocular region, which may reduce the ability to detect nonuniformity artifacts. Thus, it is not clear that the strategies for mitigating artifacts will be similarly effective in AR.

#### 3 PERCEPTUAL STUDIES

We conducted two perceptual studies to examine how the amount of binocular overlap and the display configuration (convergent and divergent) influence the perceived quality of the FOV in AR. We focused on using visuals that are more similar to AR applications than those found in the existing literature on partial binocular overlap. We also chose one type of nonuniformity artifact to focus on: the fading of content near the binocular-monocular border. If the results of prior studies extend to AR visuals, we would expect users to experience less fading when viewing convergent configurations and configurations with more binocular overlap.

#### 3.1 Methods

- 3.1.1 Participants. In Experiment 1, 20 adults (ages 20–30 years, 1M 19F) participated. In Experiment 2, a different group of 20 adults (ages 19–27 years, 5M 15F) participated. One participant in Experiment 2 indicated that they could not fuse the fixation target, so their data were discarded and an additional participant was recruited. All participants had normal or corrected-to-normal visual acuity in both eyes, and normal stereo vision assessed with Randot Stereo Test. All participants were naïve to the study hypotheses, were compensated for their time, and gave informed consent for their participation. The experimental procedures were approved by the Institutional Review Board.
- 3.1.2 Display System. Stimuli were presented on a desk-mounted mirror stereoscope with two LCD displays (LG-32UD99-W, maximum luminance of 138 cd/m² measured by a PR-650 photometer) as shown in Figure 4. The viewing distance from the participant's eyes to each display was approximately 57 cm, resulting in a FOV for each display of 63° horizontally and 38° vertically. Each display was 3,840 by 2,160 pixels, resulting in a resolution of ~55 pixels per degree (~27.5 cycles per degree). During the study, the participants rested their chin on a chin rest while sitting in a dark room.
- 3.1.3 Stimuli. We used three types of visual stimuli: a simple oval, a simple rectangle, and simulated AR. The oval and rectangle stimuli consisted of uniform white shapes presented on a black background (Figure 4, simple (oval) and simple (rectangle)), emulating stimuli that have been used to study perceptual artifacts from partial overlap in prior literature [20, 21, 23, 25, 26]. Thus, we expected to replicate previous findings with these stimuli (i.e., fewer artifacts with larger binocular overlap regions and fewer artifacts with a convergent configuration compared to divergent).

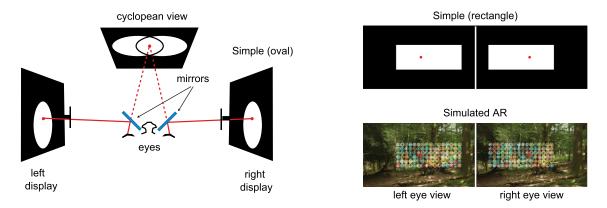


Fig. 4. Left: A top-down view of the stereoscope setup (not drawn to scale). Participants viewed a pair of desk-mounted displays through a pair of mirrors and fixated on a red dot to fuse the left and right eye's image. The cyclopean view illustrates the monocular-binocular borders where perceptual artifacts tend to occur, resulting in a nonuniform appearance of the white shape. Right: in addition to the simple oval shape shown on the left, two additional types of stimuli were used in the studies: simple rectangle and simulated AR. The monocular FOV could be either 30° or 40° wide and the monocular region could be either 4.5° or 9° wide. The red fixation dot is exaggerated for visibility. Natural image is © October 2016 SYNS Dataset [1], and icons are sourced from [14].

At typical use case of AR is showing application icons against the real environment. Thus, for the simulated AR stimuli, we used icon arrays for mobile applications [14] and tiled them on a virtual display that was superimposed over a stereoscopic natural background (Figure 4, simulated AR). For the natural backgrounds, we selected scenes from the SYNS natural stereo image dataset [1]. Adjacent views were stitched together to create a wider FOV image to fill our display [9]. These scenes included both outdoor and indoor environments, captured with stereo cameras with 6.3 cm separation. To combine the icons and the backgrounds, the background intensity was first normalized from 0 to 1 for all scenes, and then all pixel values were reduced by 66.6%. The icon's pixel intensities were normalized to range from 0% to 33.3% and then added to the background image. These percentages were selected in order to produce imagery in which the AR icons were clearly visible but also appeared semi-transparent as in most optical see-through AR systems. In reality, the amount of perceived contrast of AR content relative to the real background depends on both the luminance level of the environment and the settings of the display.

We varied several properties of the stimuli and examined the effect on nonuniformity artifacts. First, we examined whether the binocular overlap region size influenced nonuniformity artifacts for each stimulus type. To simulate binocular and monocular regions that may be typical of current AR systems, we tested two horizontal monocular FOV sizes for the virtual content: 30° and 40°, and two monocular region sizes: 4.5° and 9°. This resulted in binocular overlap region sizes that ranged from 21° to 35.5° horizontally, with horizontal cyclopean FOVs ranging from 34.5° to 49°. All stimuli were 15° tall vertically. We examined convergent versus divergent partial overlap for each of these binocular region sizes, resulting in eight conditions total for each stimulus type (oval/rectangle/AR).

The AR icons and the simple shapes were rendered at a fixed vergence distance of 1.5 m based on an **interpupillary distance** (**IPD**) of 6 cm, such that the icon arrays appeared to float in front of the background. An IPD of 6 cm was chosen for rendering since it allowed a majority of the viewers to fuse the stimuli without having to adjust the stereoscope. Individual IPDs were not measured due to social distancing protocols. To aid fusion, a red fixation dot was presented to match the vergence angle needed to fuse the binocular overlap regions in the two eyes' views.

#### 3.2 Experiment 1

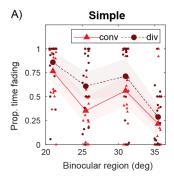
3.2.1 Procedure. In this experiment, participants were instructed to continuously indicate when they saw any fading of the stimuli (e.g., luning, fragmentation). They held down one keyboard key when they saw fading, and a different key when they did not see fading. On each trial, a particular combination of stimulus type (oval/rectangle/AR), binocular overlap region size, and convergent or divergent overlap was shown in pseudorandom order. Each trial was 30 s long. Participants were instructed to look at the fixation dot at the center of the screen for the duration of each trial so that within a trial, the amount of partial overlap was always fixed and determined by the stimulus design. Between trials, a uniform gray screen was shown for 5 s. Five unique AR scenes were used for this experiment. There were total of 56 trials.

3.2.2 Analysis. For each trial, we calculated the proportion of time that participants indicated that they saw fading, excluding the first 5 s to account for delays in starting to respond. Trials were included in analysis if at least 90% of the recorded key presses were valid responses (i.e., the participant pressed one of the two response keys). Based on this threshold, 2.6% of trials were omitted. When looking at the data with different thresholds, the results were very similar. For the valid trials, the proportion of fading time was calculated by dividing the duration that the participant indicated fading over the total duration of the responses.

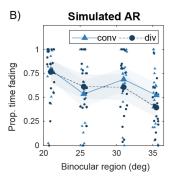
To examine the effects of the stimulus type, binocular region size, and convergent/divergence configuration on perceived fading, we fit the response data with a logistic regression model. Because the response data were bimodally distributed (many, but not all, proportions were near 0 or 1), prior to fitting the model we re-coded the responses into a binary variable indicating whether or not more than half of each trial had fading. The initial model included three main effects and all two-way interactions. In this initial model, we included the oval and rectangle shapes as separate categorical predictors; however, we did not observe an effect of the rectangle compared the oval so in the final analyses we combine these into a single variable that we call "simple stimuli", which was compared to the AR stimuli. Individual participants were modeled as random intercepts. Follow up logistic regression models were used to examine the significant interactions. For all statistical tests, a significance threshold of p < 0.05 was used.

3.2.3 Results. The results are shown in Figure 5(A), with the simple stimuli on the left and the AR stimuli on the right. The *x*-axis shows the size of the binocular region. The convergent and divergent configurations are plotted separately. Recall that previous studies found that a larger binocular region corresponded with fewer perceived artifacts [23]. Qualitatively, our data are consistent with prior work: There is a negative trend between the binocular region size and the proportion of time that participants reported fading. However, this effect seems to be stronger for simple stimuli compared to AR stimuli. Also consistent with prior literature, we see that the convergent configuration was associated with less fading compared to the divergent configuration for the simple stimuli. However, this did not appear to be the case for the AR stimuli, for which the convergent and divergent stimuli elicited similar responses.

The logistic regression model was consistent with this interpretation of the data. The model indicated that the coefficients for stimulus type, stimulus configuration, and binocular region size were all statistically significant. In addition, there were significant interactions between stimulus type and stimulus configuration, and stimulus type and binocular region size. The model accounted for 58% of the variation in the fitted data and the results are shown in Table 1. For the categorical predictors (stimulus type and configuration), we modeled the *simple* stimuli condition and the *convergent* condition as the intercept, so regression coefficients reflect the relative effects of the *AR* and *divergent* conditions. To better understand the interactions, we conducted a series of follow up analyses by fitting models to subsets of the data, based on the different categories of the main effects, and compared among them (Table 2). First, we examined the effect of configuration (convergent/divergent) separately for the simple stimuli and for the AR stimuli. For the simple stimuli, the divergent configuration was associated with an increase in fading compared to convergent configuration. Qualitatively, this increase is associated most strongly



Intercept



5.18

< 0.001

Fig. 5. Experiment 1 results for (A) simple stimuli and (B) simulated AR stimuli. The mean and 95% confidence intervals for the average proportion of time that participants reported fading during a trial are represented by the large symbols and the shaded regions. Different colors/symbols indicate different stimulus configurations. Individual subject's proportion data are plotted as small symbols. The binocular region size from small to large corresponds to the four different monocular FOV and monocular region combinations: 30° and 9°, 30° and 4.5°, 40° and 9°, and 40° and 4.5°.

Variable Coefficient (95% CI) t р Type (AR) -2.16(-3.91, -0.40)-2.420.02 Config. (Divergent) 2.05 (0.46, 3.64) 2.53 0.01\* Bino. region -0.17 (-0.23, -0.11)-5.78< 0.001 AR\*Divergent -0.97(-1.60, -0.35)-3.050.002\*AR\*Bino. region 0.10(0.04, 0.16)3.38 < 0.001 Divergent\*Bino. region -0.04(-0.09, 0.01)-1.650.10

Table 1. Logistic Regression Model Fit to Experiment 1 Data

For each predictor, the coefficient reflects an increase or decrease in the probability that fading was perceived for more than half of each trial. Positive values indicate more fading, and negative values indicate less fading. Coefficients that are significantly different from 0 based on the t-statistics are marked with asterisks (\*).

4.61 (2.86, 6.36)

with the binocular region size of 25°. There was no significant difference between the configurations for the AR stimuli. Next, we examined the effect of binocular region size separately for the simple stimuli and for the AR stimuli. For both the simple stimuli and AR stimuli, increasing binocular region size was associated with a significant decrease in fading, but the magnitude of the decrease was larger for the simple stimuli.

3.2.4 Exploratory Analysis. In our main analysis, we chose to focus on binocular region size as a predictor because prior work suggested that this is a more reliable predictor of fading than monocular FOV, monocular region size, and total cyclopean FOV alone [23]. However, in the limit, binocular region size is unlikely to explain all of the variance in fading because it does not take the monocular region size into account at all. In an exploratory analysis, we looked at whether different ways of characterizing the virtual FOV correlate better with the proportion of fading time in our data. In addition to the binocular region size as suggested by previous work, we plotted our average AR results against the total cyclopean FOV and the proportion of each eye's FOV that is only monocularly-visible (i.e., the size of the monocular region divided by the size of FOV in one eye) (Figure 6). We fitted a simple linear regression line to these data. For these data, we found that using the monocular proportion explains more of the variance in responses ( $R^2 = 0.85$ ) compared to the original binocular region size metric ( $R^2 = 0.61$ ). The total cyclopean FOV metric was not a strong predictor of fading in these data ( $R^2 = 0.02$ ).

Variable	Coefficient (95% CI)	t	p			
Simple: Convergent vs. Divergent						
Divergent	Divergent 0.64 (0.17, 1.11)		0.01*			
Intercept	-0.12 (-0.62, 0.38) $-0.48$		0.63			
AR: Convergent vs. Divergent						
Divergent	-0.17 (-0.49, 0.14)	-1.06	0.29			
Intercept	0.62 (0.04, 1.20)	2.11	0.04*			
Simple: Bino. Region						
Bino. region	-0.19 (-0.24, -0.13)	-7.05	<0.001*			
Intercept	5.49 (3.90, 7.07)	6.82	<0.001*			
AR: Bino. Region						
Bino. region   -0.09 (-0.12, -0.06)		-5.84	<0.001*			
Intercept	3.11 (2.06, 4.16)	5.81	< 0.001*			

Table 2. Follow Up Tests Examining the Interaction Terms in the Main Model for Experiment 1

Data are split into subsets based on stimulus type. Statistically significant coefficients are marked with asterisks (\*).

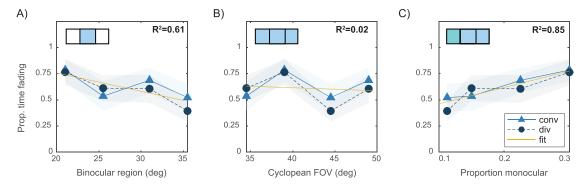
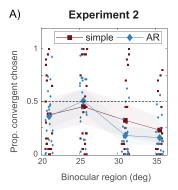


Fig. 6. Exploratory analysis looking at the proportion fading of the AR stimuli as a function of: (A) binocular region size, with monocular FOV and monocular region sizes of  $30^{\circ}$  and  $9^{\circ}$ ,  $30^{\circ}$  and  $4.5^{\circ}$ ,  $40^{\circ}$  and  $9^{\circ}$ , and  $40^{\circ}$  and  $4.5^{\circ}$  from left to right (B) cyclopean FOV, with monocular FOV and monocular region sizes of  $30^{\circ}$  and  $4.5^{\circ}$ ,  $30^{\circ}$  and  $9^{\circ}$ ,  $40^{\circ}$  and  $4.5^{\circ}$ ,  $40^{\circ}$  and  $9^{\circ}$  from left to right, and (C) proportion of each eye's FOV that is only monocularly-visible, with monocular FOV and monocular region sizes of  $40^{\circ}$  and  $4.5^{\circ}$ ,  $30^{\circ}$  and  $4.5^{\circ}$ ,  $40^{\circ}$  and  $4.5^{\circ}$ .

While by no means definitive, this strong trend suggests that as the ratio between the monocular region and the FOV increases, there may be a roughly linear increase in the fading time, at least for the combinations of parameters tested in our experiment. We will return to this measure as a possible predictor of perceptual artifacts in stereoscopic AR systems in the Discussion.

3.2.5 Summary. Experiment 1 replicated previous findings for simple stimuli, but suggests that the impact of binocular region size and convergent/divergent configuration differ for stimuli more closely approximating AR. Specifically, we did not find evidence that convergent configurations produce fewer perceptual nonuniformity artifacts in AR, suggesting that there is no need to favor systems that create convergent overlap over divergent overlap as previously thought—an important potential opportunity to relax design constraints when FOV is being optimized along with other factors. In the next experiment, we further examined the difference between convergent and divergent partial overlap.

ACM Transactions on Applied Perception, Vol. 19, No. 4, Article 19. Publication date: November 2022.



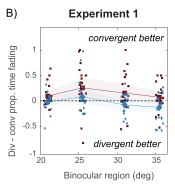


Fig. 7. (A) For Experiment 2, the proportion of trials that the convergent configuration was chosen over the divergent configuration. (B) For Experiment 1, the difference in proportion of fading time (divergent - convergent). In both panels, the dashed line represents the point of equality for convergent and divergent. Above the line means the convergent stimulus was preferred, and below the line means divergent was preferred. Data are otherwise plotted the same as Figure 5.

#### 3.3 Experiment 2

- 3.3.1 Procedure. In this experiment, participants were asked to directly compare two stimuli that were identical (the same stimulus type and binocular region size) except that one was in a convergent configuration and one was in a divergent configuration. Their task was to select the one that had "a wider field-of-view with minimum fading of the content". Because the FOV was actually identical for both stimuli on each trial, these instructions served to resolve ambiguous cases in which fading was so strong that the FOV actually appeared smaller for one stimulus. On each trial, participants could toggle back-and-forth between the two stimuli without any time limit. Twenty AR scenes were used for this experiment and each was repeated 4 times with a different icon set. The simple stimuli were each repeated 10 times. There were 400 trials total. As in Experiment 1, participants were instructed to fixate a point in the middle of the screen for the duration of the trial.
- 3.3.2 Analysis. For each unique condition, we calculated the proportion of trials that each participant chose the convergent stimulus over the divergent stimulus. We also performed a logistic regression on the trial-by-trial data, as described for Experiment 1. In this case, each response was coded with 0 if the divergent stimulus was chosen and 1 if the convergent stimulus was chosen.
- 3.3.3 Results. In this experiment, for both stimulus types (simple and AR), participants had an overall preference for divergent over convergent stimuli. In Figure 7(A), the results are plotted with the x-axis indicating the size of the binocular region and the y-axis indicating the proportion of trials for which the convergent stimulus was preferred. Values greater than 0.5 indicate a preference for convergent stimuli, and values less than 0.5 indicate a preference for divergent stimuli. For comparison, in Figure 7(B), we replot the data from Experiment 1, calculating the difference between convergent and divergent trials for the same stimulus, where a value greater than zero indicates that convergent trials had less fading, and less than zero indicates that divergent trials had less fading. From this comparison, it is clear that the forced-choice task in Experiment 2 resulted in a greater preference for the divergent configuration.

The results of the logistic regression for Experiment 2 are shown in Table 3. The main effects of stimulus type and binocular region size were both statistically significant, as was their interaction. In a follow up analysis, we performed two logistic regressions by categorizing the binocular region sizes into two levels: small ( $< 30^\circ$ ) and large ( $> 30^\circ$ ), and ran the model with each subset of the data to compare simple stimuli and AR stimuli. We grouped the binocular region sizes into two levels rather than running separate models for all four levels for ease of interpretation. These results are shown in Table 4. The results suggest that binocular region size modulated

Table 3. Logistic Regression Model Fit to Experiment 2 Data

Variable	Coefficient (95% CI)	t	р
Type (AR)	1.24 (0.56, 1.94)	3.54	<0.001*
Bino. region	-0.07 (-0.09, -0.05)	-6.07	<0.001*
AR*Bino. region	-0.05 (-0.08, -0.03)	-4.39	<0.001*
Intercept	1.02 (0.20, 1.83)	2.44	0.01*

For each variable, coefficients reflect the change in probability of choosing convergent over divergent stimuli (positive values indicate more likely to choose convergent, and negative values indicate less likely). Coefficients that are significantly different from zero based on the t-statistics are marked with asterisks (\*).

Table 4. Follow Up Analysis on the Interaction between Binocular Region Size and Stimulus Type

Variable	riable Coefficient (95% CI)		р	
Small Binocular Region: Simple vs. AR				
AR	0.11 (-0.07, 0.29)	1.17	0.24	
Intercept	-0.44 (-1.05, 0.17)	-1.43	0.15	
Large Binocular Region: Simple vs. AR				
AR	-0.83 (-1.04, -0.62)	-7.78	<0.001*	
Intercept	-1.46 (-2.23, -0.70)	-3.74	<0.001*	

Statistically significant coefficients are marked with asterisks (\*).

the effect of stimulus type on convergent preference. For small binocular region sizes, there was no difference between the AR and simple stimuli. For larger binocular regions, AR stimuli were associated with a stronger divergent preference than the simple stimuli. However, there was also a large amount of variation in the data across participants, and full model only accounted for 32% of the variance in the fitted data.

This regression analysis indicates how likely convergent stimuli were to be chosen across different conditions. However, it does not answer whether the tendency to select the convergent or divergent stimuli was significantly different from chance (i.e., that both are equally likely to be chosen). Thus, to investigate whether or not there was a significant preference for convergent or divergent stimuli in Experiment 2, we performed a chi-square goodness-of-fit test and asked whether the proportion of people preferring convergent and divergent deviated significantly from the expected values if half of the participants preferred convergent and half preferred divergent. For each condition, we coded each participant as having a preference for convergent if they chose convergent more than 50% of the time and as having a preference for divergent if they chose convergent less than or equal to 50% of the time. If there was equal preference for convergent and divergent stimuli, we would expect that 10 participants would prefer one over the other. The results are shown in Table 5. For larger binocular regions (>30°), we see that most people preferred divergent for both shape and AR stimuli. For smaller binocular region sizes, the results are more mixed, and the only significant effect was a divergent preference for the AR stimuli with the smallest binocular region. There were three participants who had no consistent preference (50/50), but these results did not change if we recoded the data such that they were included in the convergent preference group. Thus, contrary to previous work [23], we found no significant convergent preference across all conditions in this experiment.

3.3.4 Summary. When participants were asked to make a direct comparison of convergent and divergent configurations in Experiment 2, we again observed no consistent preference for convergent AR stimuli. However, unlike Experiment 1, we observed a tendency for participants to prefer the divergent configuration for both

Bino. Region	Num. Conv.	Num. Div.	$\chi^2$	Ð	
Simple					
21°	7	13	1.8	0.18	
25.5°	8	12	0.8	0.37	
31°	5	15	5	0.03*	
35.5°	4	16	7.2	0.007*	
AR					
21°	5	15	5	0.03*	
25.5°	10	10	0	1	
31°	2	18	12.8	<0.001*	
35.5°	1	19	16.2	<0.001*	

Table 5. Chi-square Test Results for Experiment 2

Statistically significant results are marked with asterisks (\*).

stimulus types, especially for large binocular region sizes. These results suggest that there may be task-dependent or timing-depending differences in perceived nonuniformity, which we will take up in the Discussion.

#### 4 3D FIELD OF VIEW IN AUGMENTED REALITY

In the previous section, we treated binocular region size and convergent/divergent overlap as static properties of an AR system, similar to prior literature. However, this is not accurate for AR systems in which people view and interact with content at both near and far distances. When eyes converge and diverge, the display's monocular FOV limits stay fixed in the world, but they are not fixed on the retina, which can alter the sign and amount of partial overlap. The binocular overlap, the total FOV, and configuration can change quite substantially depending on where the viewer is looking. In this section, we provide a simplified model for determining the cyclopean FOV and binocular overlap when different fixation distances are taken into account (see [2] for a similar analysis based on VR systems). This model, combined with the perceptual study results, allows for maximizing the cyclopean FOV over particular distances depending on the use case and display configuration.

#### 4.1 Viewing Geometry and Camera Frusta

To create 3D-AR experiences on a near-eye display, a pair of stereoscopic images needs to be generated and displayed according to the appropriate 3D viewing geometry. Specifically, two images should be constructed such that the directions of the light rays entering each eye's pupil from the displays match the intended locations of virtual objects in the 3D space. Typically, virtual cameras that are horizontally offset from each other are used to project a 3D virtual world into a pair of images. Each virtual camera will capture a left or a right view, which are then presented to the corresponding eye of the user. To achieve the correct 3D viewing, the horizontal offset of the cameras needs to match the user's IPD [17], the dimensions of each camera's viewing frustum need to match the visual angle subtended by each display, and the frusta positions need to align with the visual direction of the displays relative to the viewer's eyes (see Figure 1). In the following analyses, we assume that this viewing geometry is followed, but see [5] for a more systematic description of possible configurations to achieve correct stereoscopic viewing geometry. Because we assume correct viewing geometry, the camera's FOV is equivalent to the FOV subtended by the AR content for the user, and we will visualize the virtual camera frusta in front of the eyes of the observer, corresponding to the region of the virtual world that can be presented and merged with the natural FOV.

In this setup, each virtual camera's horizontal frustum determines the horizontal monocular FOV angle ( $\theta$ ) covered by the display for each eye. A standard camera frustum is symmetrical about the camera's optical axis—that is, the nasal field ( $\theta_n$ ) and the temporal field ( $\theta_t$ ) are equal (Figure 8(A)). But other arrangements are possible

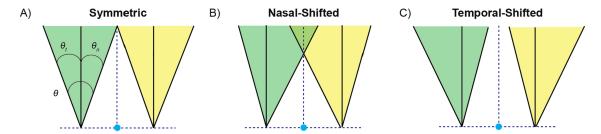


Fig. 8. Examples of three types of camera frusta, showing the horizontal monocular FOV  $(\theta)$ , nasal field  $(\theta_n)$ , and temporal field  $(\theta_t)$ . Each camera has a  $\theta$  of 40° horizontally. (A) Symmetric frusta,  $|\theta_n| = |\theta_t|$ . (B) Nasal-shifted frusta,  $|\theta_n| > |\theta_t|$ . (C) Temporal-shifted frusta,  $|\theta_n| < |\theta_t|$ . The dashed lines indicate a cyclopean coordinate system, with the blue dot indicating the origin.

in which the nasal field and the temporal field are not equal (Figure 8(B), (C)). These arrangements are called asymmetric frusta. There are two types of asymmetric frusta, which we will call nasal-shifted and temporal-shifted to avoid confusion with the convergent and divergent terminology that has been used historically for partial overlap displays. With nasal-shifted frusta, both cameras' nasal fields are extended and the temporal fields are reduced (Figure 8(B)). The left camera captures more content in the right field, and the right camera captures more content in the left field. For temporal-shifted frusta, the temporal fields are extended and the nasal fields are reduced, so the left camera captures more of the left field and vice versa for the right camera (Figure 8(C)).

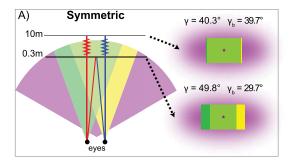
#### 4.2 Qualitative Horizontal Field-of-View Analysis

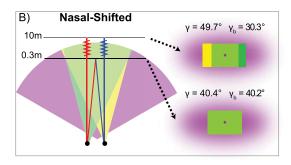
In 3D-AR, the angular cyclopean FOV (which we will refer to as  $\gamma$ ) and the angular size of the binocular overlap region ( $\gamma_b$ ) measured from the midpoint between the two eyes depend on the size of the individual camera frusta (monocular FOV,  $\theta$ ), the frustum configuration (symmetric, nasal-shifted, or temporal-shifted), the camera separation (a — in this case equal to the user's IPD), and the distance to the content that is being fixated (d). Nasal-shifted and temporal-shifted frusta are generally associated with the convergent and divergent partial overlap described in previous work, respectively. But we will show that there is not a one-to-one relationship between the binocular overlap considered in previous work and frustum asymmetry when considering 3D-AR content.

We start here with a geometric demonstration of how two variables—fixation distance and frustum asymmetry—interact. A visual comparison of the different camera frusta is shown in Figure 9 when the eyes are looking at a far object (10 m) and a near object (0.3 m). As illustrated, each camera configuration does not have a fixed cyclopean FOV ( $\gamma$ ) and binocular overlap ( $\gamma_b$ ), instead, these values change dynamically depending on the distance that is being fixated.

For far fixation distances with a symmetric configuration (Figure 9(A), top), the two monocular frusta fully overlap each other, so the cyclopean FOV is roughly equal to each camera's monocular FOV and the cyclopean FOV is (almost) fully binocular. At far fixation distances, nasal-shifted and temporal-shifted frusta result in a larger cyclopean FOV but smaller binocular region (Figure 9(B), (C) top). The nasal-shifted frusta create convergent partial overlap, and the temporal-shifted frusta create divergent overlap at far distances.

At nearer fixation distances, the eyes converge and fixate on points that are shifted rightward on the left display and leftward on the right display. For symmetric and temporal-shifted frusta, this results in a view that now has divergent partial overlap (Figure 9(A), (C) bottom). For nasal-shifted frusta, the overlap direction and amount are both distance-dependent: eye convergence may result in convergent partial overlap, divergent partial overlap, or full overlap of content (Figure 9(B) bottom, see next section). Thus, asymmetric frusta achieve a wider FOV compared to symmetric frusta for most, but not for all, distances. In short, the preferred frustum configuration to achieve a certain FOV varies with fixation distance.





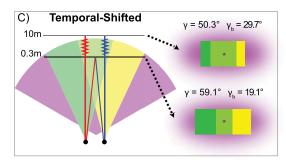


Fig. 9. Comparison of the cyclopean FOV  $(\gamma)$  and binocular overlap region  $(\gamma_b)$  for (A) symmetric camera frusta, (B) asymmetric nasal-shifted camera frusta, and (C) asymmetric temporal-shifted camera frusta at two different planes of fixation. Fixation distance is not drawn to scale. For each panel, monocular FOV  $\theta=40^\circ$ , asymmetry  $s=0^\circ,+10^\circ$  or  $-10^\circ$ , and interpupilary distance a=6 cm. Purple region indicates the natural binocular FOV and green/yellow rectangles indicate the regions covered by the display in each eye.

#### 4.3 Quantitative Horizontal Field-of-View Analysis

Here, we describe how to calculate the horizontal cyclopean FOV and binocular overlap for a given configuration. We adopt a 2D Cartesian coordinate system with the cyclopean eye at the origin (midpoint between the two eyes), the x-axis co-linear to the interocular axis (positive rightward), and the z-axis co-linear to the midsagittal plane (positive forward). Clockwise angles in this system are positive.

The angular horizontal extent (in radians) for each camera frustum ( $\theta$ ) is defined as the sum of the magnitude of the nasal field,  $\theta_n$  and the temporal field,  $\theta_t$ :

$$\theta = |\theta_n| + |\theta_t|. \tag{1}$$

We define the amount of asymmetry (s) as the difference in magnitude of these two angles:

$$s = |\theta_n| - |\theta_t|. \tag{2}$$

When  $|\theta_n| = |\theta_t|$ , then s = 0 and the cameras have symmetric frusta. To make asymmetric frusta but maintain the same monocular FOV, an equal angle is added to  $|\theta_n|$  and subtracted from  $|\theta_t|$  for nasal-shifted (s > 0), and vice versa for temporal-shifted (s < 0). We assume the shift is small such that the angular size,  $\theta$ , is relatively constant for a given display size.

To indicate a distance at which to calculate the FOV, we define a line within this coordinate system at distance d in front of the eyes, parallel to the interocular axis. The bounds of the frusta intersect this line at four points:  $\ell_L$ ,  $\ell_R$ ,  $r_L$ ,  $r_R$ , where  $\ell$  and r denote the left and right cameras, and r and r subscripts denote the left and right

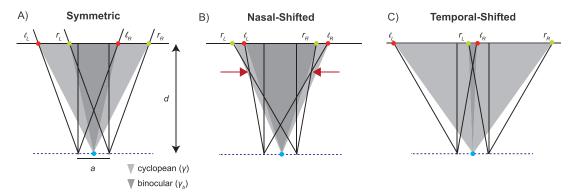


Fig. 10. Examples of three types of camera frusta with the same distance (d) to the depth plane of interest. The points of intersection bounding the binocular region ( $\gamma_b$ ) and the cyclopean FOV ( $\gamma$ ) are illustrated on this depth plane. (A) Symmetric frusta. (B) Nasal-shifted frusta. The red arrows indicate the plane at which the two frusta completely overlap each other. (C) Temporal-shifted frusta. For these frusta, there is no distance with complete overlap.

bounds of the frusta (Figure 10). The x-coordinates of these points are:

$$\ell_L = d \tan(-|\theta_t|) - \frac{a}{2},$$

$$r_L = d \tan(-|\theta_n|) + \frac{a}{2},$$

$$\ell_R = d \tan(|\theta_n|) - \frac{a}{2},$$

$$r_R = d \tan(|\theta_t|) + \frac{a}{2}.$$
(3)

Note, the sign of the angles differs depending on the reference eye because the temporal field is clockwise from the right eye, but counterclockwise from the left eye. Because the two cameras' frusta are mirror-symmetric about the origin, all calculations can be done by considering only the right-side edges of the frusta ( $\ell_R$ ,  $r_R$ ) and multiplying by a factor of 2. These points will define the angular cyclopean FOV ( $\gamma$ ) and binocular region size ( $\gamma_b$ ) for that distance. The point with the smaller x-value ( $\min(\ell_R, r_R)$ ) will bound the binocular region ( $\gamma_b$ ) on the right side, and the point with the larger x-value ( $\max(\ell_R, r_R)$ ) will bound the cyclopean FOV ( $\gamma$ ):

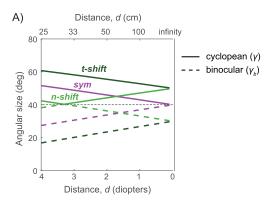
$$\gamma_b = 2\tan^{-1}\left(\frac{\min(\ell_R, r_R)}{d}\right),\tag{4}$$

$$\gamma = 2\tan^{-1}\left(\frac{\max(\ell_R, r_R)}{d}\right). \tag{5}$$

Note that when  $\gamma_b$  is negative, there is no binocular overlap for content at that distance.

We can now quantitatively examine  $\gamma$  and  $\gamma_b$  for different display configurations and viewing situations. Figure 11(A) shows examples of the horizontal cyclopean FOV ( $\gamma$ , solid lines) and binocular region ( $\gamma_b$ , dotted lines) for different camera configurations at different depth planes for a fixed monocular FOV ( $\theta$ ) of 40° and IPD (a) of 6 cm. All angles have been converted to degrees. For symmetric frusta (purple lines), the horizontal offset created by a is negligible at large distances and the two frusta have essentially 100% binocular overlap with  $\gamma$  equal to  $\theta$ . At close distances,  $\gamma$  increases, but  $\gamma_b$  decreases, due to increasing partial (divergent) overlap of the two eyes' views.

For asymmetric frusta, we plot the results for an asymmetry (s) of 10° in either direction. The cyclopean FOV ( $\gamma$ ) at far distances is larger than for symmetric frusta, but the binocular region ( $\gamma_b$ ) is smaller. With temporal-shifted frusta (dark green lines),  $\gamma$  increases at nearer distances and  $\gamma_b$  decreases (Figure 9), becoming quite small



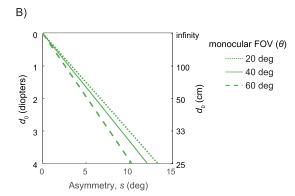


Fig. 11. (A) The cyclopean FOV and binocular overlap region size for the three types of frusta (sym = symmetric; n-shift = nasal-shifted, t-shift = temporal shifted). Distances are reported in both diopters and centimeters for fixed parameters of:  $\theta$ =40°, a = 6 cm, and s =  $\pm$ 10°. (B) For different values of  $\theta$ , the content distance that contains complete overlap ( $d_0$ ) varies by the amount of asymmetry (s) of the nasal-shifted frustum.

for fixation distances less than 50 cm. At all distances, the binocular overlap is divergent. For nasal-shifted frusta (light green lines), the effect of fixating nearer is quite different: fixating at nearer distances decreases  $\gamma$  and increases  $\gamma_b$  (Figure 9). There is a distance where 100% overlap occurs, shown here at 30 cm. For distances nearer than this, the line slope reverses. This 100% overlap distance reflects a transition from convergent to divergent overlap. Thus, nasal-shifted and temporal-shifted frusta will result in quite different patterns of binocular overlap as a user looks around a 3D scene. Increasing the amount of nasal or temporal asymmetry (s) effectively shifts the lines defining  $\gamma$  upwards, and shifts the lines defining  $\gamma_b$  downwards. By eye, it appears that nasal-shifted frusta provide a better compromise between expanding the cyclopean FOV across a range of distances, without producing overly small binocular overlap regions at any distance.

For nasal-shifted frusta, the distance that results in 100% binocular overlap ( $d_0$ ), can be calculated as follows:

$$d_0 = \frac{a}{\tan(|\theta_n|) - \tan(|\theta_t|)}.$$
(6)

In Figure 11(B), we plot  $d_0$  as a function of the nasal-shifted asymmetry angles (s) for three different monocular FOV  $(\theta)$  values. As s increases,  $d_0$  moves to closer and closer depth planes. However, the slope of these lines depends on the frustum size, such that larger values of  $\theta$  also result in complete binocular overlap at closer distances.

Lastly, it is notable that for one distance, the nasal-shifted frusta have the same  $\gamma$  and  $\gamma_b$  as the symmetric frusta (intersections of solid purple and light green lines in Figure 11(A) at 59 cm). When this happens, the symmetric frusta have produced divergent overlap for the user, while the nasal-shifted frusta have produced convergent overlap. For temporal-shifted frusta, there is no distance at which the frusta have 100% overlap and no distance that has the same  $\gamma$  and  $\gamma_b$  as the symmetric frusta.

The provided equations can be used to determine and customize the cyclopean FOV and binocular overlap region for different AR systems and use cases (e.g., near viewing versus far viewing). For example, a system could be optimized to have 90% convergent binocular overlap at a working distance of 1 m using a nasal-shifted design, and the implications for FOV at other distances could then be examined.

#### 5 DISCUSSION

The question of FOV coverage in stereoscopic AR systems is deceptively complex. All stereoscopic AR systems can have variable horizontal FOV coverage and variable binocular overlap, whether they are explicitly designed

to be "partial overlap systems" or not. The variability emerges from the impact of fixation distance on the overlap in the FOV coverage of the two eyes. In situations where binocular overlap is partial, the implications for perceptual nonuniformity across the visual field must be considered. A geometric analysis of system design can predict the amount of FOV covered for an AR system at different distances, but cannot determine the visibility of these artifacts. By combining perceptual results with a geometric analysis, we hope to provide a new toolkit for optimizing the FOV of stereoscopic near-eye display systems for AR. Here, we discuss the implications of the perceptual study as well as how to incorporate these insights into designing AR FOV.

#### 5.1 Convergent and Divergent Configurations

A key question that arises when considering FOV in stereoscopic AR is whether convergent or divergent partial overlap is preferable. Prior work using simple stimuli strongly suggested that convergent partial overlap produces fewer nonuniformity artifacts than divergent partial overlap [23, 26]. Our results using simulated AR, however, suggest that this finding does not generalize well to modern stereoscopic AR systems. In Experiment 1, our results showed no strong difference between the two configurations for AR stimuli. In Experiment 2, we asked participants to directly compare convergent and divergent configurations, and we found that both simple and AR stimuli had better visual quality (i.e., a wider FOV with less fading of content) with a divergent configuration than a convergent configuration, and the difference was more pronounced as binocular region size increased.

For AR stimuli, our results suggest no need to prioritize convergent overlap to minimize nonuniformity artifacts per se. However, we are left with a challenge to understand why our results with simple shapes in Experiment 2 differ from the canon of prior work. One potential explanation is that we used larger binocular region sizes. For example, in one prior study that used a similar comparison paradigm [23], convergent configurations were found to have a more uniform FOV compared to divergent configurations 93.3% of the time. However, that study used stimuli with a much smaller binocular overlap region (15.6° monocular FOV, 7.8° monocular region, and 7.8° binocular region). If we look at the trend in our data from Experiment 2 (Figure 7), we see that smaller binocular region sizes were associated with a weaker preference for divergent configurations. Indeed, if we use the logistic regression model fit to these data and predict the probability that the convergent configurations was preferred for a binocular region of this size (7.8°), for simple stimuli, we obtain 0.62 for simple stimuli and 0.79 for AR stimuli. That is, our model would indeed predict a convergent preference for this smaller binocular region. Importantly, the range of FOVs used in the current studies is more reflective of current AR systems than previous work.

While the different FOV regime might explain differences in our results from previous work, it does not explain why Experiments 1 and 2 suggested slightly different biases. One important difference between these paradigms is the timing. Experiment 1 involved a continuous response over 30 s, while viewing times for Experiment 2 were self-controlled and presumably much shorter. While Experiment 2 was designed to directly compare the two configurations, we suggest that the results from Experiment 1 might be more reflective of artifact visibility during typical use of AR systems, when the amount of overlap would not rapidly alternate unless someone was possibly alternating their fixation depth back-and-forth. In future work, it would be fruitful to examine artifact detection using more naturalistic paradigms, such as performing target detection or reading tasks in AR.

#### 5.2 Predicting Fading in 3D-AR Experiences

There is an inherent tradeoff between the amount of binocular overlap and the cyclopean FOV for a fixed monocular FOV, as demonstrated by the geometric analysis. With the perceptual studies, we can now also add predicted nonuniformity artifacts across various fixation distances into the mix to better understand the perceptual implications of specific design decisions.

By way of example, we focus here on the AR stimulus from Experiment 1. We combine the data from the divergent and convergent configurations since they were highly similar. We can then fit the average proportion fading time (F) data as a function of the proportion of each eye's FOV that was only monocularly visible as

mentioned earlier in Section 3.2.4. Here, we use  $(p_m)$  to denote this monocular proportion. We chose this variable because it was the strongest FOV-based predictor of the fading time. In order to extrapolate beyond the ratios included in the experiment, we assert that the amount of fading should be 0 when this variable is 0 (i.e., complete binocular overlap). When adding the point (0,0) to the fitting, a straight line can no longer capture the data. We found the best fitting non-linear function of the form  $b\sqrt{p_m}$ , where b denotes the only free parameter. Because  $p_m$  should not exceed 1, we obtain the following piece-wise non-linear function:

$$F = \begin{cases} 1.405\sqrt{p_m}, & 0 \le p_m \le 0.50\\ 1, & p_m > 0.50. \end{cases}$$
 (7)

This fit is plotted along with the average data in Figure 12(A). The first case in this equation can be solved for  $p_m$  as a function of F to calculate the largest acceptable ratio for a given threshold on fading

$$p_m = \left(\frac{F}{1.405}\right)^2. \tag{8}$$

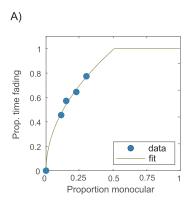
For example, if a fading proportion of 0.50 is the maximum acceptable level ( $F \le 0.5$ ), that would suggest the proportion of each eye's FOV that is monocularly-visible should be 0.13 or less.

Of course, it is unlikely for any 3D-AR experience to have a fixed monocular proportion if users are looking around a scene at different depths. To add 3D viewing into consideration, we can combine this analysis with the geometric analysis. Specifically, we can apply Equation (7) to calculate the predicted amount of fading for an arbitrary frustum and display setup. For example, given the binocular region size and the total cyclopean FOV in Figure 11(A) for a monocular FOV of 40°, we can determine the monocular region size at various distances for the three types of frusta and input the ratio into Equation (7). The predicted proportion fading times as a function of fixation distance are shown in Figure 12(B). The important thing to note is that across all distances the nasal-shifted frusta produce less fading than the temporal-shifted frusta (although they also produce a smaller cyclopean FOV). Recall that the nasal shift produces a mixture of convergent and divergent overlap. We also see that the symmetric frusta may have fewer artifacts over time than nasal-shifted frusta for use cases with a relatively far fixation distance (about 2 m and farther). However, more perceptual data using stimuli with different FOV sizes are needed to validate the function plotted in Figure 12(A). Validating and expanding on these guidelines with more diverse visual stimuli will be an important direction for future work. It is also important to keep in mind that a large monocular region not only induces nonuniformity artifacts, but also lacks stereo cues for depth. Indeed, a recent analysis of virtual reality screen placement and typical fixation distances during virtual reality tasks produced guidelines advising a small nasal shift so as to achieve near 100% binocular overlap and maximize stereo cues [2]. Given the current differences in monocular FOV in VR and AR systems and their different use cases, it is likely that different tradeoffs should be prioritized for these systems.

#### 5.3 Monocular Regions in Natural Vision

At the core of these nonuniformity artifacts is the issue of AR systems that create monocularly-visible regions within the natural binocular FOV. However, we know that monocularly-visible regions are not inherently problematic. For example, the blind spots in the two eyes create two monocularly-visible regions, which are not detected during most of daily life. In addition to the blind spots, monocularly-visible regions often result at occlusion boundaries in 3D scenes. However, these regions tend to serve as a source of depth information instead of causing nonuniformity artifacts [16], possibly because the size of naturally occurring monocularly-visible regions tend to be small compared to what is generated by partial overlap displays. These naturally-occurring monocular regions also have statistical regularities that may aid the visual system in detecting and integrating them into the natural binocular percept without producing interocular competition [6].

Indeed, previous work argued that convergent partial overlap configurations are better because they simulate a more natural occlusion geometry. With the convergent configuration, the location of the monocular regions



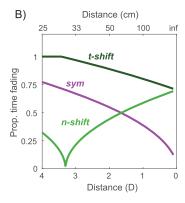


Fig. 12. (A) A piece-wise function was fit to the AR data from Experiment 1, including the (0,0) theoretical point. The *x*-axis is the ratio between monocular region and monocular FOV shown as a proportion. The *y*-axis is the predicted fading proportion and is capped at 1. (B) Predictions for the proportion fading time across various fixation distances for the three types of camera frusta described in the geometric analysis (parameters are the same as in Figure 11(A)).

simulates situations in which the viewer looks through an aperture, whereas the divergent configuration simulates a situation when the eyes are fixating on a near surface that occludes a background [4, 23, 26, 35]. Importantly, because the AR stimuli are spatially contiguous, for divergent stimuli this means the visual pattern on the near surface (binocular region) and the occluded background (monocular region) would match, which may be unlikely in natural scenes. However, it is worth remembering that the natural FOV is also divergent, with temporal flanking monocular regions in the far periphery[15]. Perhaps a deeper understanding of monocular regions in natural vision can yield new insights into how these features can be designed to mitigate artifacts in AR.

#### 5.4 Vertical Partial Overlap

In this work, we focused on the horizontal FOV and horizontal partial overlap. Of course, horizontal partial overlap has no effect on extending the vertical FOV. In principle, a vertical partial overlap could be created with one frustum/display shifted up and the other one down. However, unless there is 100% horizontal binocular overlap for all fixation distances, the resulting vertical monocular regions would be offset to the left and right as well. It seems unlikely that the this strategy would produce a large increase in vertical FOV without introducing other issues, but small vertical offsets may be interesting to explore.

#### 5.5 Limitations and Future Work

While we made an effort to design stimuli that replicate modern AR use cases, there are still limitations that should be taken into consideration when translating our results into more complex viewing situations. For example, during real AR experiences, there may be different motion in the AR content and the natural background, which could influence the strength of nonuniformity artifacts. In addition, the presence of blur in the background due to different focal distances of the AR content and the background may affect the visibility of virtual content and detection of fading artifacts.

Indeed, integrating natural eye movements will be an important direction for future work. In the current perceptual experiments, the eyes were always fixating at the center of the binocular overlap region, but when users look directly at the monocular region there may be differences in how well they can detect the artifacts. In addition, changes in fixation distance will result in monocular regions that shrink and expand dynamically. These movements could suppress or enhance fading and warrant further investigation. If eye movements make

fading more visible, it would be interesting to consider integrating partial overlap with an actuated display and eye tracking that could maintain binocular overlap at the fixation point as users look around a scene. In addition, it has been speculated that the binocular region need not to be greater than 40° relative to fixation since binocular interactions are diminished beyond this eccentricity [15]. Our desk-mounted experiment setup could not achieve a wide enough FOV to test this prediction, but it could be explored with a VR headset.

#### 6 CONCLUSION

AR systems aim to create immersive mixtures of real and virtual content, but limited FOV coverage remains a persistent barrier to realizing this goal. In this work, we reviewed the importance of considering partial binocular overlap when analyzing FOV coverage in AR. We highlighted the need for a set of perceptual guidelines to optimize the horizontal FOV in AR systems, and conducted two perceptual studies to facilitate the development of these guidelines. Our results suggest that a large binocular overlap region and divergent configuration may reduce perceptual nonuniformity artifacts when viewing partial overlap imagery in AR, allowing more effective FOV expansion. We provide a model that can be used for optimizing FOV in AR headsets while taking into account the dynamic FOV during 3D-AR viewing. By better understanding the factors that affect perceived FOV size and quality during dynamic 3D interactions, we hope to facilitate the development of systems with sufficient FOV coverage that FOV is no longer a limiting factor in AR experiences.

#### **ACKNOWLEDGMENTS**

The authors would like to thank Marty Banks, Joohwan Kim, and Jacob Hadnett-Hunter for helpful feedback and discussions.

#### REFERENCES

- [1] Wendy J. Adams, James H. Elder, Erich W. Graf, Julian Leyland, Arthur J. Lugtigheid, and Alexander Muryy. 2016. The southamptonyork natural scenes (SYNS) dataset: Statistics of surface attitude. *Scientific Reports* 6, 1 (2016), 35805.
- [2] Avi M. Aizenman, George A. Koulieris, Agostino Gibaldi, Vibhor Sehgal, Dennis M. Levi, and Martin S. Banks. 2022. The statistics of eye movements and binocular disparities during VR gaming: Implications for headset design. ACM Transactions on Graphics (2022). https://dl.acm.org/doi/10.1145/3549529.
- [3] Mohammad S. Alam, S. H. Zheng, Khan Iftekharuddin, and Mohammad A. Karim. 1992. Study of field-of-view overlap for night vision applications. In *Proceedings of the IEEE 1992 National Aerospace and Electronics Conference*. IEEE, 1249–1255.
- [4] Barton L. Anderson and Ken Nakayama. 1994. Toward a general theory of stereopsis: Binocular matching, occluding contours, and fusion. *Psychological Review* 101, 3 (1994), 414–445.
- [5] Martin S. Banks, David M. Hoffman, Joohwan Kim, and Gordon Wetzstein. 2016. 3D displays. *Annual Review of Vision Science* 2, 1 (2016), 397–435.
- [6] Zeynep Başgöze, David N. White, Johannes Burge, and Emily A. Cooper. 2020. Natural statistics of depth edges modulate perceptual stability. *Journal of Vision* 20, 8 (2020), 10.
- [7] Randolph Blake and Hugh Wilson. 2011. Binocular vision. Vision Research 51, 7 (2011), 754-770.
- [8] Vladimir N. Borisov, Nikolay V. Muravyev, Roman A. Okun, Aleksandr E. Angervaks, Gavril N. Vostrikov, and Mikhail V. Popov. 2021. A DOE-based waveguide architecture of wide field of view display for augmented reality eyewear. In Proceedings of the Optical Architectures for Displays and Sensing in Augmented, Virtual, and Mixed Reality. SPIE, 94–102.
- [9] Gary Bradski. 2000. The OpenCV library. Dr. Dobb's Journal of Software Tools 25, 11 (2000), 120-123.
- [10] Ozan Cakmakci, David M. Hoffman, and Nikhil Balram. 2019. 31-4: Invited paper: 3D eyebox in augmented and virtual reality optics. SID Symposium Digest of Technical Papers 50, 1 (2019), 438–441.
- [11] Steven A. Cholewiak, Zeynep Başgöze, Ozan Cakmakci, David M. Hoffman, and Emily A. Cooper. 2020. A perceptual eyebox for near-eye displays. *Optics Express* 28, 25 (2020), 38008–38028.
- [12] Edward DeHoog, Jason Holmstedt, and Tin Aye. 2016. Field of view limitations in see-through HMD using geometric waveguides. Applied Optics 55, 22 (2016), 5924–5930.
- [13] Stephen R. Ellis, Bernard D. Adelstein, Ronald J. Reisman, Joelle R. Schmidt-Ott, Jonathan Gips, Jimmy Krozel, and Malcolm Cohen. 2002. Augmented reality in a simulated tower environment: Effect of field of view on aircraft detection. NASA/TM-2002-211853 (2002).
- [14] Freepik. 2017. Retrieved May 9th, 2019 from https://www.flaticon.com/authors/freepik.

- [15] Scott S. Grigsby and Brian H. Tsou. 1994. Visual processing and partial-overlap head-mounted displays. Journal of the Society for Information Display 2, 2 (1994), 69–74.
- [16] Julie M. Harris and Laurie M. Wilcox. 2009. The role of monocularly visible regions in depth and surface perception. *Vision Research* 49, 22 (2009), 2666–2685.
- [17] Robert T. Held and Martin S. Banks. 2008. Misperceptions in stereoscopic displays: A vision science perspective. In *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization*. 23–32.
- [18] Qichao Hou, Qiwei Wang, Dewen Cheng, and Yongtian Wang. 2016. Geometrical waveguide in see-through head-mounted display: A review. In *Proceedings of the Optical Design and Testing*. SPIE, 52–59.
- [19] Sion Jennings and Manfred Dion. 1997. An investigation of helmet-mounted display field-of-view and overlap tradeoffs. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. 32–36.
- [20] Victor Klymenko, Thomas H. Harding, Howard H. Beasley, John S. Martin, and Clarence E. Rash. 1999. The effect of helmet mounted dsiplay field-of-view configurations on target acquisition. USAARL Report No. 99-19 (1999).
- [21] Victor Klymenko, Thomas H. Harding, Howard H. Beasley, and Clarence E. Rash. 2001. Visual search performance in HMDs with partial overlapped binocular fields-of-view. *USAARL Report* No. 2001-05 (2001).
- [22] Victor Klymenko, Robert W. Verona, Howard H. Beasley, and John S. Martin. 1994. Convergent and divergent viewing affect luning, visual thresholds, and field-of-view fragmentation in partial binocular overlap helmet-mounted displays. In *Proceedings of the Helmet-and Head-Mounted Displays and Symbology Design Requirements*. SPIE, 82–96.
- [23] Victor Klymenko, Robert W. Verona, Howard H. Beasley, John S. Martin, and William E. McLean. 1994. Factors affecting the visual fragmentation of the field-of-view in partial binocular overlap displays. *USAARL Report* No. 94-29 (1994).
- [24] Victor Klymenko, Robert W. Verona, Howard H. Beasley, John S. Martin, and William E. McLean. 1994. Visual perception in the field-of-view of partial binocular overlap helmet-mounted displays. USAARL Report No. 94-40 (1994).
- [25] Victor Klymenko, Robert W. Verona, John S. Martin, Howard H. Beasley, and William E. McLean. 1994. The effect of binocular overlap mode on contrast thresholds across the field-of-view as a function of spatial and temporal frequency. USAARL Report No. 94-49 (1994).
- [26] Victor Klymenko, Robert W. Verona, John S. Martin, Howard H. Beasley, and William E. McLean. 1994. Factors affecting the perception of luning in monocular regions of partial binocular overlap displays. USAARL Report No. 94-47 (1994).
- [27] Helga Kolb. 2007. Facts and figures concerning the human retina. Webvision. Retrieved Dec. 1, 2021 from https://webvision.med.utah.edu/book/part-xiii-facts-and-figures-concerning-the-human-retina/.
- [28] George A. Koulieris, Kaan Akşit, Michael Stengel, Rafal K. Mantiuk, Katerina Mania, and Christian Richardt. 2019. Near-eye display and tracking technologies for virtual and augmented reality. Computer Graphics Forum 38, 2 (2019), 493–519.
- [29] Bernard C. Kress. 2019. Digital optical elements and technologies (EDO19): Applications to AR/VR/MR. In *Proceedings of the Digital Optical Technologies 2019*. SPIE, 343–355.
- [30] Bernard C. Kress. 2019. Optical waveguide combiners for AR headsets: Features and limitations. In *Proceedings of the Digital Optical Technologies 2019*. SPIE, 75–100.
- [31] Bernard C. Kress and Ishan Chatterjee. 2021. Waveguide combiners for mixed reality headsets: A nanophotonics design perspective. Nanophotonics 10, 1 (2021), 41–74.
- [32] James J. W. Lin, Henry B. L. Duh, Don E. Parker, Habib Abi-Rached, and Thomas A. Furness. 2002. Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. In *Proceedings of the IEEE Virtual Reality* 2002. 164–171.
- [33] Bill McLean and Steve Smith. 1987. Developing a wide field of view HMD for simulators. In *Proceedings of the Display System Optics*. 0778. 79–82.
- [34] James E. Melzer and Kirk Moffitt. 1989. Partial binocular-overlap in helmet-mounted displays. In *Proceedings of the Display System Optics II*. SPIE, 56–62.
- [35] James E. Melzer and Kirk W. Moffitt. 1991. Ecological approach to partial binocular overlap. In Proceedings of the Large Screen Projection, Avionic, and Helmet-Mounted Displays. SPIE, 124–131.
- [36] Tao Ni, Doug A. Bowman, and Jian Chen. 2006. Increased display size and resolution improve task performance in information-rich virtual environments. In *Proceedings of the Graphics Interface 2006*. Canadian Information Processing Society, 139–146.
- [37] Robert Patterson, Marc D. Winterbottom, and Byron J. Pierce. 2006. Perceptual issues in the use of head-mounted visual displays. Human Factors 48, 3 (2006), 555–573.
- [38] Clarence E. Rash, William E. McLean, Ben T. Mozo, Joseph R. Licina, and Joseph B. McEntire. 1999. Human factors and performance concerns for the design of helmet-mounted displays. *USAARL Report* No. 99-08 (1999).
- [39] Donghao Ren, Tibor Goldschwendt, YunSuk Chang, and Tobias Höllerer. 2016. Evaluating wide-field-of-view augmented reality with mixed reality simulation. In *Proceedings of the 2016 IEEE Virtual Reality*. IEEE, 93–102.
- [40] Zhujun Shi, Wei Ting Chen, and Federico Capasso. 2018. Wide field-of-view waveguide displays enabled by polarization-dependent metagratings. In Proceedings of the Digital Optics for Immersive Displays. SPIE, 272–277.
- [41] Robert H. Spector. 1990. Visual fields. In Proceedings of the Visual Fields—Clinical Methods: The History, Physical, and Laboratory Examinations. Butterworth Publishers, Chapter 116.

- [42] Christina Trepkowski, David Eibich, Jens Maiero, Alexander Marquardt, Ernst Kruijff, and Steven Feiner. 2019. The effect of narrow field of view and information density on visual search performance in augmented reality. In *Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces*. 575–584.
- [43] Andrew B. Watson. 2018. The field of view, the field of resolution, and the field of contrast sensitivity. *Electronic Imaging* 2018, 14 (2018), 1–11.
- [44] Maxwell J. Wells, Michael Venturino, and Robert K. Osgood. 1989. The effect of field-of-view size on performance at a simple simulated air-to-air mission. In *Proceedings of the Helmet-Mounted Displays*. SPIE, 126–137.
- [45] Jianghao Xiong, Guanjun Tan, Tao Zhan, and Shin-Tson Wu. 2020. Breaking the field-of-view limit in augmented reality with a scanning waveguide display. OSA Continuum 3, 10 (2020), 2730–2740.
- [46] Tao Zhan, Kun Yin, Jianghao Xiong, Ziqian He, and Shin-Tson Wu. 2020. Augmented reality and virtual reality displays: Perspectives and challenges. iScience 23, 8 (2020), 101397.

Received 5 April 2022; revised 3 July 2022; accepted 20 July 2022