ANNUAL REVIEWS

Annual Review of Vision Science

The Perceptual Science of Augmented Reality

Emily A. Cooper

Herbert Wertheim School of Optometry & Vision Science, Helen Wills Neuroscience Institute, University of California, Berkeley, California, USA; email: emilycooper@berkeley.edu

Annu. Rev. Vis. Sci. 2023. 9:455–78

First published as a Review in Advance on March 21, 2023

The *Annual Review of Vision Science* is online at vision.annualreviews.org

https://doi.org/10.1146/annurev-vision-111022-123758

Keywords

display systems, augmented reality, visual perception

Abstract

Augmented reality (AR) systems aim to alter our view of the world and enable us to see things that are not actually there. The resulting discrepancy between perception and reality can create compelling entertainment and can support innovative approaches to education, guidance, and assistive tools. However, building an AR system that effectively integrates with our natural visual experience is hard. AR systems often suffer from visual limitations and artifacts, and addressing these flaws requires basic knowledge of perception. At the same time, AR system development can serve as a catalyst that drives innovative new research in perceptual science. This review describes recent perceptual research pertinent to and driven by modern AR systems, with the goal of highlighting thought-provoking areas of inquiry and open questions.

INTRODUCTION

From their outset, electronic displays created new possibilities for visual experiences. These experiences, however, have largely been confined to flat surfaces that are separate from the rest of the world. For example, bulky displays attached to early computers allowed people to view the first dynamic computer graphics. With the rise of mobile computing, smartphones with compact and high-resolution touchscreen displays can now be held in the hand, providing a seemingly endless stream of palm-sized images controlled with a swipe of the finger. Now, emerging wearable near-eye display systems present the opportunity to immerse ourselves in novel visual experiences by wearing a device directly in front of our eyes. Wearable virtual reality (VR) systems, for example, leverage small displays similar to those in current smartphones and combine them with collimating optics in front of the eyes (e.g., magnifying lenses) to create a wholly virtual world. Wearable augmented reality (AR) systems can be achieved by adding optical combiners to create the perception that physical and virtual elements are intermingled in front of our eyes. Applications for modern AR systems are varied and include education, medical care, navigation assistance, vision enhancement, gaming, and entertainment (Dey et al. 2018).

The term augmented reality (or sometimes mixed reality) actually encompasses experiences enabled by a range of different technologies—from wearables, to hand-held devices, to room-scale projection systems (Milgram 1995). Some technologies aim to directly overlay virtual content onto the user's natural vision (optical see through), while others require the user to view a real-time video feed of the physical world with virtual content graphically integrated (video see through). Regardless of the specific hardware or approach, developing an effective AR system necessarily involves a deep understanding of the existing corpus of perceptual research. Indeed, AR system development is a compelling example of a virtuous cycle between basic and applied research: Research on human visual perception helps to guide innovative AR systems, while AR and related display system innovations spur research that advances our understanding of how we perceive the world. For example, research characterizing the focusing mechanism of the human eye has informed the design of AR displays that support more realistic and comfortable three-dimensional (3D) visuals (Lee et al. 2018, Liu et al. 2008, Wallach & Norris 1963, Watt et al. 2005a). Meanwhile, research into the best design of 3D display visuals has led to new understanding of how our eyes bring the world into focus (Hoffman & Banks 2010, MacKenzie et al. 2010). As another example, improvements in our understanding of the differences between central and peripheral vision have helped to drive advances in wide-field-of-view (FOV) rendering for wearable display systems (Brown et al. 2022, Freeman & Simoncelli 2011, Kim et al. 2019, Rosenholtz 2016, Walton et al. 2021). In turn, these advances are providing new insights into how we use peripheral vision during natural tasks (Duinkharjav et al. 2022).

This review primarily focuses on the perceptual science associated with a particular type of AR technology: wearable optical see-through systems that aim to convincingly merge 3D virtual objects with the physical world. These systems are the focus because they aim to achieve high perceptual fidelity, they can potentially support a broad set of applications in daily life, and commercial devices that use them are currently available. As such, their development represents a timely consideration in perceptual research. However, there is much overlap in the perceptual questions and concerns across the different approaches for AR, so much of the research covered in this review is relevant to other forms of AR, as well as to VR systems.

CREATING AUGMENTED REALITY

Three basic components of a wearable optical see-through display system are: light sources that generate imagery, optics that focus this imagery, and optical combiners that direct this imagery to

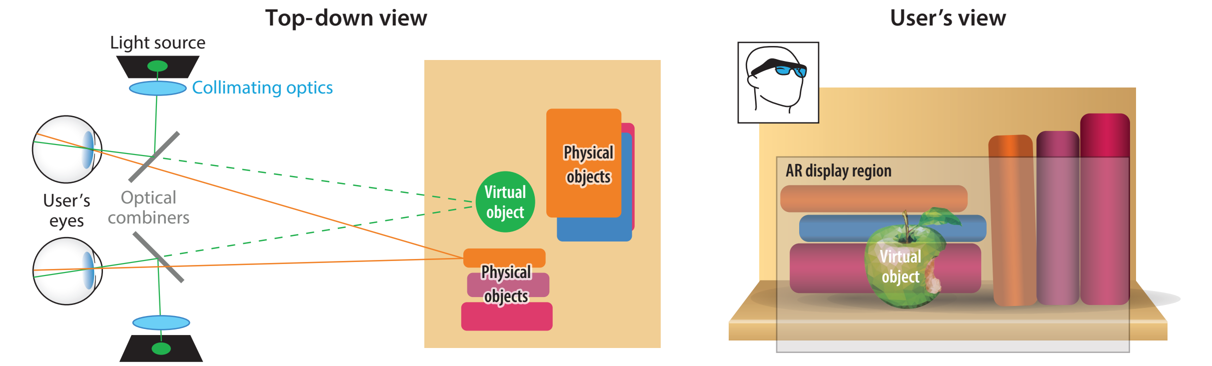

Figure 1

Creating augmented reality (AR) with a wearable optical see-through display system. While there are many potential platforms for AR, the figure focuses on wearable binocular systems that fit like a pair of bulky glasses or ski goggles (illustrated in the inset on the right). The user sees a combination of the real scene in front of them and imagery from displays that can simulate three-dimensional (3D) virtual objects. This experience can be accomplished with optical combiners that superimpose display imagery onto the user's natural view of the world. Light from the displays (*green lines*) is reflected to the eyes, while light from the world (*orange lines*) is also permitted to pass through. Additional optical elements, such as collimating lenses, can create a virtual image of the display at a comfortable focusing distance. The two displays present slightly different views to the two eyes, simulating the binocular disparities that users are used to seeing when viewing physical 3D objects. To produce convincing virtual 3D objects, the displayed images must also update as the user moves around a scene. The diagram of the system components is adapted with permission from Cholewiak et al. (2020), copyright 2020 The Optical Society, and the bookshelf and apple images are from Pixabay (**https://pixabay.com/service/license/**).

the user's pupils while also transmitting light from the world. **Figure 1** provides a diagram of these components that illustrates a simple AR system—commercial systems in practice are much more complex. In this diagram, the light sources are microdisplays, and the optical elements include magnifying lenses to create a virtual image at some comfortable focusing distance for the user and flat beam splitter combiners. With a system like this, virtual objects can be shown to the user through superimposition within a region of their natural visual field. Note that the system illustrated in **Figure 1** can only add patterns of light to the environment; it cannot selectively block incoming light. In addition to optics and display hardware, AR systems need other components to create convincing interactions between virtual objects and real objects. For example, a camera or other sensor that tracks the physical world and the user's motion is necessary to place virtual content sensibly in the world. High-speed graphical processing is required to update the displayed image with sufficient speed that virtual objects appear world-fixed as the user moves. If you move toward a table in front of you that contains both real and virtual objects, for example, then the virtual objects need to grow in size in your visual field just as the real objects do. The fidelity and functionality of these sensors can be just as important for the AR experience as the display system itself. Lastly, while presenting virtual imagery to a single eye is sufficient to create an AR experience, a binocular AR system can more accurately simulate 3D objects by showing slightly different views to the two eyes that contain the binocular disparities that we are used to seeing in real 3D scenes.

Laboratory-based AR systems were first developed in the 1960s, using miniature cathode-ray tubes and standard optics (magnifiers, half-silvered mirrors, and prisms) to relay 3D computer graphics (Sutherland 1968). While these systems were groundbreaking at the time, the graphics were primitive, and the hardware was cumbersome at best. Nowadays, a range of tiny light sources (e.g., organic light-emitting diode microdisplays, miniature projectors) and advanced optical elements (e.g., freeform optics, waveguides, exit pupil expanders) can be combined and codesigned to construct diverse AR systems. Computational advances allow for sophisticated corrections to optical aberrations and efficient rendering of virtual content in real time, while lightweight sensors can unobtrusively track the user's motion through the world. Some recent reviews have covered the impressive array of AR system hardware and software (see Kress 2020, Wetzstein 2020, Zhan et al. 2020).

While modern engineers have developed ingenious techniques for building and optimizing these AR systems, there is currently no commercial AR system that provides high-quality visual imagery across the full visual field in a glasses-like device that one might wear for hours on end. For example, one AR system may produce compelling 3D visuals but require a rather large, bulky headset and fill only a small region of the visual field. Another system may fit more like a normal pair of glasses but only be able to render simple imagery and require precise alignment with the user's pupil(s) for the graphics to be bright and visible. Determining the right balance between visual fidelity and practicality for modern AR systems remains a challenge, and perceptual research plays a key role in determining the specifications that a system must meet for a given application. The following sections examine the perceptual science that governs a selection of key factors for AR systems: getting light from the display into the user's pupils, filling the visual field with imagery, matching the resolution of the visual system, and producing realistic 3D virtual objects that integrate with the physical world.

GETTING LIGHT INTO THE PUPILS

Light from the natural world comes from many sources (the sun, starlight, man-made sources) and is reflected off of surfaces all around us. When we open our eyes, natural light floods our pupils from all directions (**Figure 2*a***). Light from an AR system, however, must be optically directed from an image source toward the user's pupil. This process results in a limited volume of space from which the complete displayed image can be seen. The term eyebox is often used to refer to a volume in space where the user's pupil must be located to receive an acceptable view of the content seen through an optical system (eye relief refers specifically to the ideal distance of the pupil from the optics). In the simple ray diagram in **Figure 2*b***, this eyebox is bounded by the diamond-shaped region where the emissions from all three pixels intersect (for clarity, the optical combiner from **Figure 1** has been omitted from this diagram). If the eyebox of an AR system is too small, then shifts in eye position or even just an eye rotation can result in an incomplete, vignetted, or distorted view of the imagery (**Figure 2*c***). At the same time, creating a larger eyebox is not always better because flooding the area around the pupil with more light requires using more power, which in wearable systems can add undesirable weight and bulk. As such, there is also pressure not to substantially overfill the pupil. An important perceptual question thus arises: How small can the eyebox be without creating noticeable perceptual artifacts during typical use?

Optical characterizations of image quality across a set of pupil positions are a useful first step in addressing this question. For example, a precisely controlled imaging device or an optical simulation can map out variations in the appearance of the display from different pupil positions. These analyses show that optical image quality in near-eye display systems can vary continuously around the sweet spot at the center of the eyebox, and that these variations incorporate multiple perceptually relevant degradations such as nonuniformity, reduced spatial resolution, spatial distortion, and color breakup (Austin et al. 2018, Cakmakci et al. 2019, Ratnam et al. 2019). One of the notable visual artifacts associated with a small eyebox is luminance nonuniformity, in which the visibility of the displayed image becomes reduced toward the edges due to a reduction in the light reaching the eye. Luminance nonuniformity is not unique to AR systems or even to eyebox optics. Indeed, any display panel or light source can suffer from similar nonuniformities or artifacts, and perceptually driven image quality models have been developed to predict their detectability that build

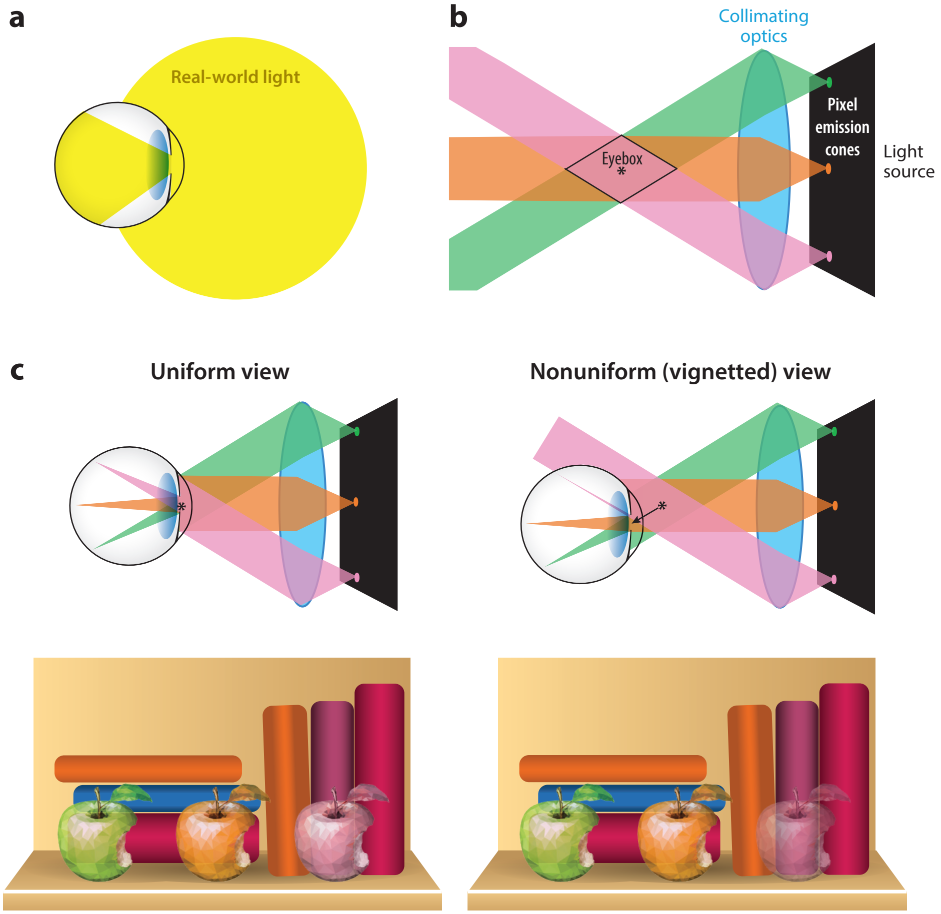

Figure 2

Augmented reality (AR) systems are subject to limitations in the visibility and quality of the displayed image. (*a*) Under most natural circumstances, light from the physical world comes from all directions, filling the pupil regardless of its location or orientation. (*b*) Light from a near-eye display is directed toward the expected location of the pupil, resulting in a limited eyebox volume in which the entire displayed image is visible. The asterisk indicates the center of the diamond-shaped eyebox. (*c*) A correctly positioned pupil, at the center of the eyebox, will be filled with the display light and will receive a uniform view of the displayed content (*left*), but a shifted or rotated pupil may be underfilled, creating a nonuniform or vignetted view (*right*; the line and arrow indicate the change in pupil position). In this example, the ray bundle from the pink apple is undersampled by the pupil, and the pink apple is vignetted because less light is added to the user's view from this portion of the display. The diagrams are adapted with permission from Cholewiak et al. (2020), copyright 2020 The Optical Society, and the bookshelf and apple images are from Pixabay (**https://pixabay.com/service/license/**).

on decades of psychophysical research characterizing the spatial sensitivity of the visual system (Mantiuk et al. 2021, Watson 2007, Watson & Ahumada 2005).

However, new perceptual elements emerge in AR systems that are not captured by existing models, making it challenging to predict whether visible nonuniformities will be problematic for an AR system user. For example, research on binocular vision suggests that binocular interactions will influence percepts if luminance nonuniformity differs between the two eyes (e.g., Ding & Levi 2017, Legge & Rubin 1981). If one eye in an AR system has a vignetted view, and the other does not, will the vignetting be masked or accentuated via binocular interactions? To address questions such as this one, recent applied perceptual studies have aimed to better understand binocular luminance perception in complex natural stimuli. In a recent study, Cholewiak and colleagues (2020) investigated these issues by implementing a realistic simulation of luminance nonuniformity (specifically vignetting) in AR, with simple semitransparent images overlaid on natural backgrounds. Perceptual judgments of nonuniformity increased systematically with increasing luminance vignetting. However, when the nonuniformity patterns differed between the eyes, sensitivity was much lower. This suppression of nonuniformity perception through binocular combination is consistent with existing models of binocular contrast summation, which predict that the eye seeing higher-contrast imagery—in this case, the eye with less vignetting—will dominate (Ding & Levi 2017, Legge & Rubin 1981). These empirical observations, and related rendering applications for binocular graphics, have led to renewed interest in extending models of binocular combination to apply to naturalistic stimuli that include complex spatial patterns and variable interocular contrast differences (Yang et al. 2012, Zhong et al. 2019). For example, while binocular combination may suppress the visibility of some artifacts, large interocular differences may lead to binocular rivalry, in which the percept alternates in time between content seen by either eye rather than being fused into a single percept (Levelt 1965). Zhong and colleagues (2019) recently showed that binocular contrast differences, but not binocular luminance differences, are a good predictor of binocular rivalry in natural images, and they used these findings to guide a new technique that aims to improve binocular image rendering without eliciting rivalry percepts.

In a practical sense, the combination of an eyebox volume that can comfortably accommodate typical variations in pupil size or location and a dynamic steerable element that can follow the pupil as people look around a scene may be necessary for removing all eyebox-related visual artifacts in AR systems. Perceptual studies that clearly delineate the acceptable spatial, temporal, and binocular characteristics for such a system will be essential and must have good external validity for actual AR experiences. For example, detection bounds for interocular differences in luminance, contrast, and color could be used to determine how similar the two eyes' positions need to be within their eyeboxes—and whether other artifacts, such as display panel color nonuniformities and optical distortions, are binocularly detectable. A range of pertinent psychophysical data on detection of binocular image differences already exists (e.g., Jennings & Kingdom 2016, Kingdom et al. 2019), but the applicability of models based on these data to naturalistic imagery similar to AR is yet to be fully tested.

FILLING THE VISUAL FIELD

Let us assume that a user's pupils are comfortably placed within the eyeboxes of an AR system, such that the user can see the entire displayed image for each eye. To create binocular imagery for a 3D AR experience, the displays need to present a pair of images consistent with virtual objects at specific locations in the world. Due to constraints on near-eye optics, however, most compact AR systems are only able to present this imagery over a subset of the natural visual field of each eye. Importantly, limited visual field coverage can limit the immersion of AR experiences and can impair performance of tasks that rely on information in the peripheral visual field (Lin et al. 2002, Ren et al. 2016, Trepkowski et al. 2019). Thus, quantifying the visual field coverage of AR systems is desirable, and techniques that can increase this coverage remain a high priority.

To understand the perceptual implications of this limitation, it is helpful to review how virtual imagery is generated and superimposed on the natural visual field. To start with, a pair of virtual cameras is used to render two side-by-side views of virtual objects, one view for each display (**Figure 3*a***). Each virtual camera captures content within a volume determined by the viewing

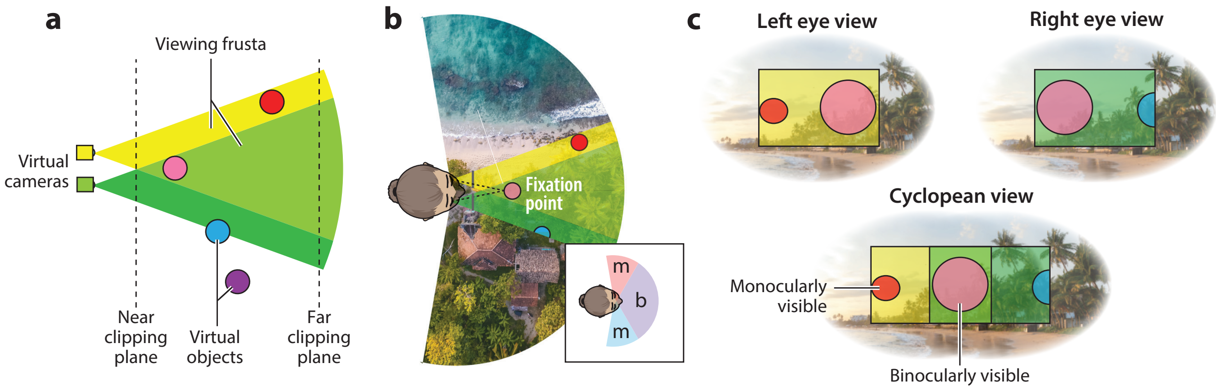

Figure 3

Visual field coverage in augmented reality (AR) systems. (*a*) To create stereoscopic AR imagery, a pair of virtual cameras captures two views of a virtual three-dimensional (3D) scene. Only the content that falls within the viewing frustum of each camera and the near and far clipping planes is rendered. (*b*) This imagery is superimposed on the natural visual field, using one display for each eye. The natural visual field has a central binocular region and two flanking monocular regions, but the displayed content in this example only appears in the binocular region. In the inset, the binocular region of the natural visual field is marked with "b," and the monocular regions are marked "m." (*c*) Depending on the virtual camera geometry, the alignment of the imagery in the left and right eyes, and the fixation point, there may be only partial overlap in the cyclopean (binocularly fused) visual field. In this example, the user is fixating the pink object, and therefore, they are converging their eyes near. Both cameras captured this object, and it is binocularly visible. However, the red and blue objects were only captured by one virtual camera: In the user's view, they are therefore monocularly visible. Note that the natural background in the cyclopean view contains unfused binocular disparities. Diagrams adapted from Wang & Cooper (2022) (CC BY 4.0); images obtained from Unsplash (Oliver Sjöström, Rowan Heuvel; **https://unsplash.com/license**).

frustum and near and far clipping planes, and the intercamera separation is (hopefully) equal to the user's interpupillary separation so as to create appropriate stereoscopic imagery. The resulting images are then presented to the user. In a properly fitted system, these images cover an angular region in each eye that matches the position and extent of the viewing frusta of the virtual cameras. For example, this region might be positioned straight ahead and subtend 40° horizontally and 30° vertically (**Figure 3*b***). The natural visual field is quite a bit larger than this. When the eyes fixate straight ahead, the full angular extent of visibility of each eye—the FOV—extends approximately 160° horizontally and 135° vertically (Spector 2011). These two monocular FOVs have a substantial portion of binocular overlap, such that both eyes see content that is approximately within ±60° of the midsagittal plane (see **Figure 3*b***).

In the illustration in **Figure 3**, the AR imagery seen by each eye falls entirely within the natural binocular region. If we consider the cyclopean view (the fusion of the left and right eye's FOVs), then we can also see that the AR system presents both binocularly and monocularly visible content within the natural binocular region in the central visual field (**Figure 3*c***). The extent of this partial binocular overlap in AR systems depends on the specific system design (i.e., optics and display placement), as well as where people are looking in a 3D scene (i.e., the vergence angle of the eyes) (Aizenman et al. 2022, Wang & Cooper 2022). Regardless, basic psychophysical research can inform our understanding of perceptual consequences associated with seeing monocular stimuli in the central visual field (Levelt 1965, Weert & Levelt 1974, Wendt & Faul 2022). For example, in the illustration in **Figure 3*c***, if the user experiences binocular rivalry, then the result would be an alternation between seeing the virtual content visible to one eye and seeing only the real background visible to the other. That is, the red and blue balls would appear and disappear.

Nonetheless, early work on wearable display system design considered whether adjusting the display placement to increase monocularly visible content may be a viable strategy to increase horizontal FOV coverage, at the cost of losing some binocular (stereoscopic) depth information in the flanks of the image (McLean & Smith 1987, Melzer & Moffitt 1989). Applied perceptual studies have since shown that there are perceptual and performance deteriorations in the resulting monocularly visible regions, such as visual fading, image fragmentation, and reduced target detection (Klymenko et al. 1994, 1999; Wang & Cooper 2022). For example, Klymenko and colleagues (1999) tested people's ability to detect a target (a number from 2 to 9) that appeared and disappeared at a random location within a cluttered display of distractors. Reaction times were elevated, and targets were more likely to be mislocated, when targets were presented in monocularly visible regions of the display as compared to binocularly visible targets at the same eccentricity. However, the strength of these artifacts can be modulated by adjusting the amount and appearance of the monocularly visible content. Visual fading, for example, is consistently reduced if the size of the monocularly visible region is kept small relative to the binocular FOV (Klymenko et al. 1994, Wang & Cooper 2022). Indeed, small regions of monocularly visible content in the central visual field do occur during natural vision, such as when one eye can see more content behind a depth edge than the other eye. It has been suggested that systems leveraging regularities in our natural experience with depth edges hold promise for reducing artifacts and increasing effective FOV in AR and VR systems (Melzer & Moffitt 1991, Shimojo & Nakayama 1990). In this vein, basic vision research that aims to more deeply understand the constellation of percepts associated with depth edges in natural scenes may help to facilitate guidelines for rendering content in AR systems (Başgöze et al. 2020, Wilcox & Lakra 2007).

Given the perceptual challenges associated with small FOV coverage and monocularly visible content, the ability to robustly support wide FOV AR imagery remains a high priority in engineering. Ongoing applied research aims to create prototypes that incorporate novel design elements such as defocused point light sources and ellipsoidal mirrors to dramatically expand the FOV coverage of AR systems (e.g., Maimone et al. 2014, Zhang et al. 2020). As the capabilities of AR systems to cover the visual field advance, important additional questions emerge about the quality of visual field coverage; these questions are considered in the next section.

RENDERING THE DETAILS: RESOLUTION AND FOVEATION

The sensitivity of the human visual system is not uniform across the natural FOV of each eye. For example, acuity (the ability to visually resolve fine details) declines steeply with increasing eccentricity from the center into the periphery (Ludvigh 1941). Other aspects of vision, such as color sensitivity, also decline with increasing eccentricity (Noorlander et al. 1983). These changes occur due to eccentricity-dependent variations in the anatomical substrates of vision (e.g., imaging quality of the eye's optics, retinal sampling, and cortical representations) (Curcio & Allen 1990, Curcio et al. 1990, Sereno et al. 1995, Thibos 2020). For imagery that falls in the high-resolution central fovea of the retina, observers can resolve fine details (for example, gratings of up to 60 cycles per degree) under high-luminance and -contrast conditions. Moving just a few degrees off the fovea results in a decrease in acuity by more than a factor of 2 (Ludvigh 1941). Despite its low spatial resolution, peripheral vision is critical for monitoring the environment, detecting motion, and guiding eye movements and is thus essential for many potential AR applications, such as navigation and visual search (Vater et al. 2022).

In existing AR systems, and even VR systems, the resolution limit of the presented imagery tends to be well below the resolution of foveal vision (Spjut et al. 2020). The size of the individual display elements (i.e., pixels) in an AR system places a fundamental upper limit on the resolution of the imagery that can be displayed, and the viewing optics that combine the displayed image with the natural FOV also affect the resolution, brightness, and contrast of the image that gets to the retinas. Even if it were possible to create an AR system that displays in foveal resolution across the whole visual field, the graphical processing and power needed to drive those pixels in real time may be prohibitive, and much of the rendered content would be invisible in the periphery anyway. These observations have spurred a range of clever approaches to efficiently display and render visual detail for foveal and peripheral vision in near-eye and AR displays. One approach is to adopt foveated display hardware, in which the resolution varies from a high-resolution region designed to stimulate the fovea to a lower-resolution region designed to be sufficient for the peripheral visual field (e.g., Akşit et al. 2019, Kim et al. 2019, Rolland et al. 1998, Tan et al. 2018, Thomas et al. 1990). In a complementary line of research, foveated rendering algorithms are being developed to differentially render content in the foveal and peripheral regions of the visual field—this software-based approach could be used instead of or in addition to hardware-based foveation. With an eye tracker, the foveally matched region can be moved based on gaze direction such that, ideally, the displayed image is indistinguishable from an image with uniformly high visual resolution.

Based on psychophysical characterizations of peripheral vision, initial foveated rendering algorithms experimented with simply reducing the spatial resolution of peripheral graphical content before drawing to the display (e.g., Geisler & Perry 1998, Guenter et al. 2012, Swafford et al. 2016). This work confirmed that foveated imagery can be perceptually indistinguishable from full-resolution imagery, as expected from basic psychophysics (**Figure 4*a*,*b***). However, Patney and colleagues (2016) noted that pure resolution reduction (e.g., via a Gaussian low-pass filter) often produces a percept of reduced contrast in the periphery. They proposed that this artifact occurs because of a mismatch in the resolution and detection acuities in the peripheral visual field (Thibos et al. 1987). In particular, the ability of the visual system to simply detect visual patterns falls off less quickly in the peripheral visual field than the ability to resolve the specifics of the patterns. Patney and colleagues proposed that if the resolution threshold is used to drive foveated rendering, then peripheral stimuli will appear to have artificially low contrast, but that boosting peripheral contrast should restore expected levels of pattern detectability. To examine this possibility, they designed and tested a contrast-preserving low-pass filter and found that this approach substantially reduced the noticeability of foveated rendering in a small user study (**Figure 4*c***). It is notable that temporal sensitivity is also still quite high in the periphery, such that temporal instability in foveated rendering during user and scene motion can produce visual artifacts (Guenter et al. 2012, Krajancich et al. 2021, Patney et al. 2016).

More recently, foveated rendering techniques are being developed that aim to also exploit the perceptual crowding of visual features in peripheral vision that results from statistical pooling along the visual processing hierarchy (Brown et al. 2022, Freeman & Simoncelli 2011, Rosenholtz 2016, Walton et al. 2021) (**Figure 4*d***). These techniques create exotic-looking imagery that can nonetheless be indistinguishable from full resolution in the periphery because the images are matched in terms of their higher-order texture statistics. However, are these manipulations really innocuous? Recent behavioral experiments, motivated by a desire to understand whether and how foveated rendering may affect task performance, are leading to new insights into how peripheral vision is used during natural tasks. A recent study by Duinkharjav and colleagues (2022) suggests that sets of peripheral imagery that are perceptually the same may nonetheless result in different levels of behavioral performance, as measured via saccade latencies. This work raises new questions about the information processing operations supported by peripheral vision during natural tasks.

Foveated techniques may ultimately be an essential part of AR systems, and a compelling range of research has emerged in recent years on this topic. While perceptual assessments of foveated display techniques have not yet focused on AR imagery in particular, the principles that this work is establishing for visibility and performance in peripheral vision seem likely to translate to AR

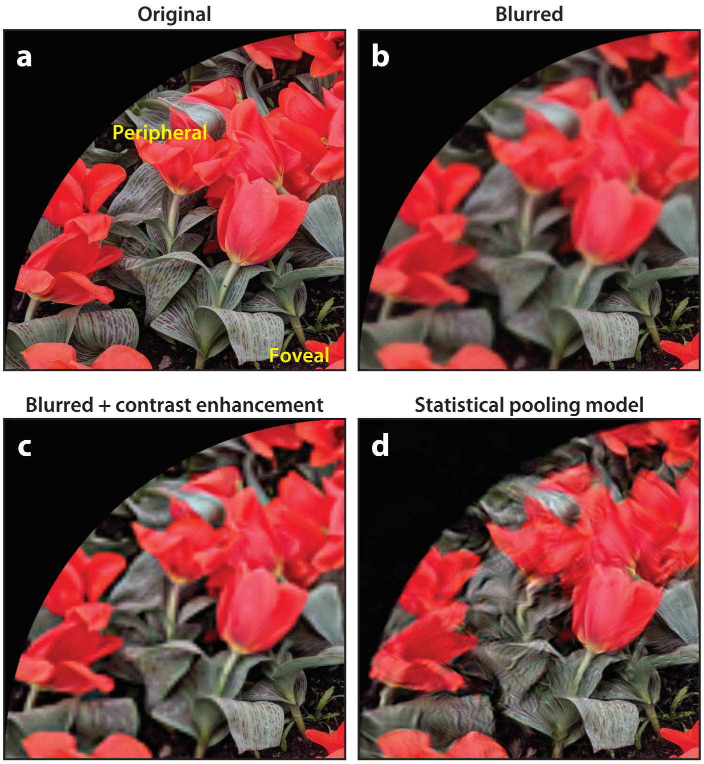

Figure 4

Examples of foveated rendering techniques that could be applied in augmented reality (AR) systems. (*a*) A full-resolution image that will be subjected to foveated rendering, with the fovea modeled in the lower right corner. (*b*) Resolution is reduced from the foveal region into the periphery by applying Gaussian blur. (*c*) A customized contrast-preserving low-pass filter is used to counteract the perception of reduced contrast in panel *b* (Patney et al. 2016). (*d*) Pooling of visual features is used to model cortical processing of peripheral visual information and produce a lower-bandwidth representation of the original scene (Brown et al. 2022). Note that the peripheral flower petals in this image still appear relatively sharp, but their specific contours and texture are altered. To reproduce the viewing conditions described by Brown et al., this panel should be viewed from approximately 35 cm while fixating the lower right corner. Images courtesy of Anjul Patney and Vasha DuTell.

systems. In addition to the display of fine details, the luminance range and color gamut of any display also limit the appearance of the simulated scenes and objects that the display can produce. These limitations become all the more notable in AR systems that aim to merge virtual objects into the real world. As such, the remainder of this review focuses on the appearance of virtual objects in AR.

PRESENTING VIRTUAL OBJECTS IN REAL ENVIRONMENTS

Surfaces: Lightness, Contrast, Color, and Transparency

When we encounter physical objects, we perceive their surfaces to have a constellation of reflectance properties (e.g., lightness, contrast, color, transparency). However, these percepts are not a direct read-out of the luminance levels and chromatic spectrum of light reflected from objects into the eyes. For example, the wavelength spectrum emanating from a green apple will differ depending on how it is illuminated, but most people would agree that an apple does not appear to change its color if you carry it from your kitchen to your back porch. Perceptual research indicates that the visual system can compensate for such environmental lighting effects by estimating and discounting the properties of the illumination (i.e., it can achieve color and lightness constancy), but the robustness of these inferences can vary depending on the available contextual information (Brainard & Radonjić 2014).

In optical see-through AR experiences, the light reaching the user's eye from virtual surfaces is a blend of the light emitted from the display and the light in the physical environment. As a result, one of the most notable surface percepts of virtual objects is that they appear semitransparent: You see the object, but you also see the background through it. While this appearance is similar to the appearance of real-world semitransparent objects, like a colored glass vase, it is not the same. The AR display is itself an emissive light source, and the emitted light is not affected by changes in lighting in the surrounding environment, as the reflected light from any real object would be (Huang et al. 2021, Murdoch 2020, Zhang & Murdoch 2021). Another key feature of these AR displays is that fluctuations in the amount of ambient illumination in the world modulate the mean intensity of virtual content. This modulation can reduce visibility and perceived contrast because our ability to detect luminance variations approximately follows Weber's law. That is, the just noticeable difference in luminance increases as a function of the mean luminance, causing virtual content with a fixed display luminance to become less visible in brighter environments (e.g., Moffitt & Browne 2019, Van Nes & Bouman 1967). As such, the demands on AR display system luminance can differ substantially, for example, depending on whether the system is being used in an indoor environment or outside on a sunny day.

Initial work on surface appearance in AR focused on characterizing the distortions caused by this blending and designing compensatory adjustments to preserve the spectral composition of the display imagery (e.g., Gabbard et al. 2010, Hincapie-Ramos et al. 2015, Weiland et al. 2009). In the simplest case, one might use a forward-facing camera to capture a three-channel color RGB image of the world in front of the user (RGB_{world}), subtract these RGB values from the desired RGB values for the virtual content ($RGB_{desired}$), and then render the compensated values to the display ($RGB_{display} = RGB_{desired} - RGB_{world}$). While this technique is simple in theory, in practice, the limited dynamic range of displays makes it infeasible. For example, if one wants to show a bright green virtual apple ($RGB_{desired} = 0{,}255{,}0$), but the background is bright red ($RGB_{world} = 255{,}0{,}0$), then it is not possible to display the required negative intensities ($RGB_{display} = -255{,}255{,}0$). Some emerging AR systems incorporate spatial light modulation (SLM) to circumvent these optical blending issues (e.g., Cakmakci et al. 2004, Kiyokawa et al. 2003, Wilson & Hua 2017). SLMs can selectively block (subtract) patterns of light from the physical world and allow true occlusion by virtual objects. However, precise light subtraction is still not commonplace. In a software-based approach, Hincapie-Ramos and colleagues (2015) developed a real-time algorithm to adjust the displayed RGB values so as to minimize the difference between the desired spectrum and the actual spectrum reaching the user's eye, subject to the limitations of the display. However, perceptual improvements in color fidelity associated with this method were modest.

A clue to why such RGB-based compensation approaches might not achieve the desired results lies in our broader understanding of surface perception and how it relies on perceptual inferences about the physical scene (Hassani & Murdoch 2019, Murdoch 2020). For example, one study asked how a set of AR manipulations affected the perceived brightness (equivalent to lightness in the experiment) of a cube (Murdoch 2020). The author of this study considered multiple AR scenarios that supported different inferences about the relationship between the virtual surface and the real surface. These scenarios included a physical cube with a virtual surface precisely aligned to create a skin and others in which the real and virtual imagery were not aligned. In the first scenario, people tended to respond as if they merged the physical object and virtual skin into one surface, and lightness percepts were generally consistent with a blend between the light from the world and display. However, when the edges of the virtual and real surfaces were not aligned, people tended to discount the light added by the overlay, as if they were separately estimating the reflectance properties of two distinct objects (a semitransparent patch and an opaque cube). The findings in the latter condition are consistent with the basic literature on perception of transparency in 3D scenes and the associated surface percepts. For example, observers can estimate the lightness of real surfaces even when surfaces are obstructed by a semitransparent occluder that distorts the retinal image, so long as sufficient contextual information is provided (Anderson 2003, Gilchrist & Jacobsen 1983, Kingdom 2011). Foundational work on the perception of transparency describes this as a process of scission—that is, the perceptual interpretation of a visual stimulus (e.g., a particular luminance level) as two different overlapping layers (Metelli 1974). If the perceptual interpretation of AR surface properties can vary depending on the context, then the most effective way to compensate for lightness and color distortions caused by optical blending will likely also need to be context dependent.

Because AR systems appear capable of eliciting a variety of inferences about surface relationships and reflectance properties, they present a potential testbed for investigating how higher-level inferences affect our perception of surface properties. For example, recent experiments characterizing color and lightness constancy with realistic graphical rendering might be extended into AR, enabling assessments with even more natural 3D information (Radonjić et al. 2015, Singh et al. 2022). To date, however, the research on perceived surface properties in AR has been dominated by simple stimuli with limited 3D information. Ultimately, properly rendering surface properties in AR will likely require an understanding of the inferences that users make about the underlying geometry, the patterns of environmental illumination, and the relationships between physical and virtual surfaces.

Accommodation and Focus

Light reflected from objects in our physical environment enters the pupils with different amounts of ray divergence, depending how far the object is away from us. For example, light rays reflected from a distant mountain range are effectively parallel as they enter the pupil, whereas the rays reflected from a magazine held in the hands are diverging. The cornea and the crystalline lens of our eyes together converge these light rays to form the retinal image. The required amount of light convergence (optical power) to create a clear retinal image depends on how divergent the incoming rays are—that is, from how far away they are being reflected. If the eye had a fixed optical power, then we would only be able to clearly see objects at one distance. To achieve clear vision of the 3D world, our visual system actively changes the optical power of the lens to bring objects at different distances into focus (this process is called accommodation). Our visual system also has a relatively permissive depth of field, such that objects slightly closer or farther than the focus distance can appear equally sharp. Psychophysical studies indicate that the size of the depth of field is approximately ±0.2–0.4 diopters (inverse meters) from the focus distance, but this range depends on the pupil size and visual content (Campbell 1957, Sebastian et al. 2015).

Display systems using conventional optics present imagery at a single optical distance. The eyes need to accommodate to this distance (or close to it) to see the displayed content sharply. In AR systems, a unique challenge emerges. If the display's optical distance is not well-matched to the surrounding aspects of the physical 3D environment, then when the eyes focus on real objects, virtual objects or content that are supposed to be in the same depth plane will become blurry, and vice versa. For example, **Figure 5*a*** illustrates an AR system that aims to add virtual annotations (a label and an arrow) to a real object (a door). The optical distance of the AR system

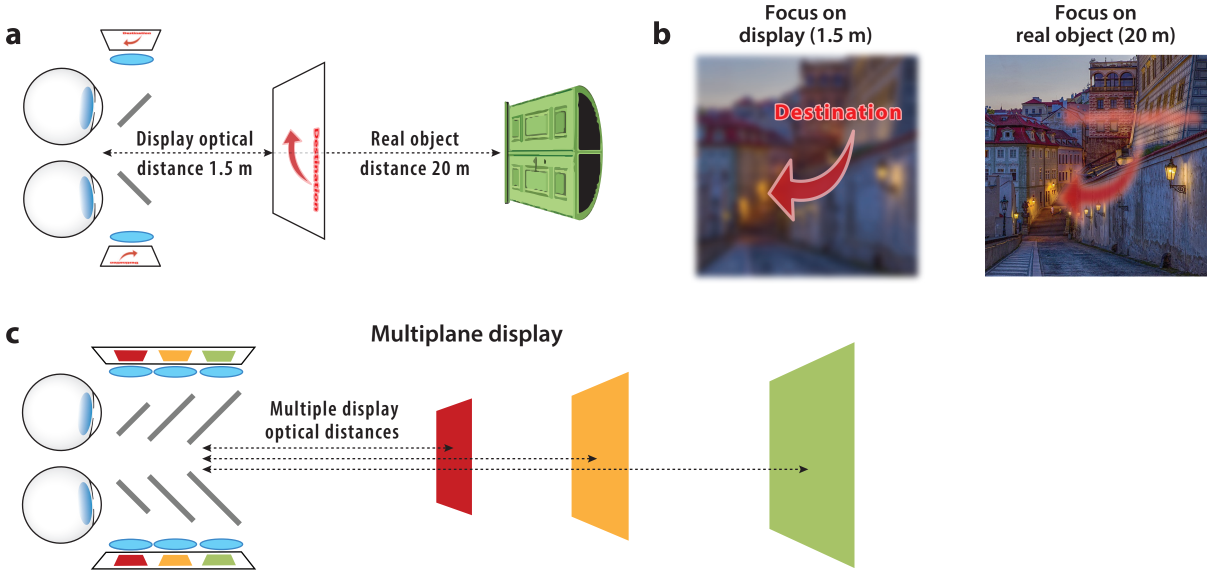

Figure 5

Different focus distances present a challenge for visual clarity in augmented reality (AR). (*a*) When viewing physical objects that are far or near, the eye actively accommodates to bend light rays the appropriate amount to achieve a sharp retinal image. If the optical distance of the AR system display is inconsistent with the nearby physical objects, then this inconsistency can lead to undesirable defocus. In this example, the display optical distance is 1.5 m, and the virtual annotations on the display are being used to label a door that is 20 m away. (*b*) When the eyes focus on the virtual content, the real object is out of focus. When the eyes focus on the real object, the virtual content is out of focus. (*c*) The use of multiplane displays is one approach to expanding the optical distances at which virtual content can be presented. In this approach, virtual content can be blended across a set of semitransparent display surfaces, each with a different optical distance. Door and city street images obtained from Pixabay (**https://pixabay.com/service/license/**).

and therefore the annotations is fixed at 1.5 m (0.667 diopters), but the door is 20 m in front of the user (0.05 diopters). Therefore, when the user accommodates to the door, the text is outside of their depth of field, and vice versa (**Figure 5*b***). In addition to the issue of clarity, the accommodative response and the associated patterns of focus and defocus provide useful information that affects depth percepts (Fisher & Ciuffreda 1988, Hoffman & Banks 2010), and discomfort can result when accommodative responses are mismatched with other oculomotor responses (so called vergence–accommodation conflicts) (Kooi & Toet 2004, Lambooij et al. 2009, Shibata et al. 2011). To address these issues, a range of technological approaches exist that expand the optical distances over which virtual imagery can be presented, which could minimize or eliminate issues related to accommodation in AR (Banks et al. 2016, Kramida 2016). For instance, multiplane display designs have been proposed that employ an ensemble of transparent display planes positioned at different distances from the eye (Akeley et al. 2004, Chen et al. 2019, Love et al. 2009, Rolland et al. 2000) (**Figure 5*c***). However, significant engineering is required to miniaturize and integrate this design into a wearable AR system. Indeed, these and similar approaches typically incur substantial additional computational cost and optical hardware and thus need to be optimized for the specific requirements of natural accommodation.

With these advanced system designs, perceptual questions naturally arise about how accurate and precise the focusing information needs to be. For example, with multiplane displays, it is important to know how many display planes are sufficient, what their maximum depth separation can be, and whether it is possible to accurately simulate content at depths between the display planes. A series of studies combining optical modeling and perceptual tasks have examined these questions in detail, yielding perceptually motivated guidelines for display focus cues, as well as insights into the basic focusing properties of the eye (Hoffman et al. 2008; MacKenzie et al. 2010, 2011; Watt et al. 2005b). MacKenzie and colleagues (2010, 2011), for example, asked whether it is possible to stimulate natural accommodative responses to virtual objects at depths between two display planes. That is, if there is a display at 1 diopter and another at 2 diopters, is it possible to drive accommodation to 1.5 diopters by drawing the image at half intensity on each display? They considered two hypotheses, each with some support in the existing perceptual literature: People may simply accommodate to the display plane with higher contrast, or they may accommodate to an intermediate point between the display planes (Mandelbaum 1960, Owens 1979, Rosenfield & Ciuffreda 1991). MacKenzie and colleagues found evidence for the latter hypothesis for small display plane differences (<~1 diopter) and for the former for larger differences. Using this paradigm, they were also able to resolve an apparent conflict in our basic understanding of accommodation and stimulus spatial frequency. Empirically, the accommodative system appears to rely solely on low-to-mid-range spatial frequencies (below approximately 8 cycles per degree) in the stimulus (Mathews & Kruger 1994, Owens 1980). This observation is surprising because, presumably, high-spatial-frequency patterns (fine details) are informative for fine-tuning the accommodative distance when a target is only slightly out of focus (Charman & Tucker 1978). By analyzing the responses to blended imagery and comparing them to an optical model of the eye, MacKenzie & Watt (2010, 2011) showed that high spatial frequencies simply do not provide reliable accommodative information in natural viewing situations (e.g., when optical aberrations in the eye are considered). They thus provided a new explanation for why different stimuli can or cannot drive accommodation.

The use of multiplane displays is just one example of the approaches that are being considered for expanding the focus distances of an AR system. For example, a complementary set of resolution- and quality-related questions arises for lightfield and 3D holographic displays, which similarly aim to represent natural focus cues with sufficient fidelity (e.g., Choi et al. 2021, Lanman & Luebke 2013, Maimone et al. 2017). In general, the desire for ever more nuanced predictions and control of the accommodative response of the eye has spurred a renewed interest in this topic. New research in this domain has revealed, for instance, that accommodation is more accurate than previously appreciated (Labhishetty et al. 2021) and that chromatic aberrations in the eye can drive accommodative responses with surprising vigor (Cholewiak et al. 2018). Lastly, it is essential to consider that accommodation degrades significantly and systematically with age (resulting in presbyopia), which has been shown to have key implications for both the approach to and importance of focus cues for the broader population (Padmanaban et al. 2017). Ultimately, understanding how factors such as transparency, optical aberrations, and user age affect the focusing response of the eyes is essential for providing guidelines for AR systems that can effectively integrate digital information with the surrounding 3D environment for a broad set of situations and users.

Distance and Shape

As highlighted in the previous sections, inferences about the 3D shapes of objects and the relationships between them play a key role in our perceptual experience of AR. A spate of basic research has shown that our perception of the distances and shapes of objects around us results from a rich combination of many sources of visual information (Howard & Rogers 2002). We can, for example, triangulate the distances to objects using the different viewpoints provided by our two eyes and using our change in viewpoint as we move around a scene. We can use the properties of perspective projection to determine relative distances based on reasonable assumptions about our environmental geometry, allowing us to infer depth relationships even from content shown on flat pictures (or flat displays). The curvature of 3D shapes can be inferred from patterns of shading across the surface, as well as from the distortions of surface texture. At this point, our best understanding of how we perceive the distances and shapes of physical objects is that we combine information across multiple sources and leverage prior assumptions about the most likely properties of the physical world around us (Knill & Richards 1996). This approach leads to robust and useful, although certainly imperfect, perception of 3D space.

In the best-case scenario, perception of the distances and shapes of virtual objects in AR should only be expected to be as good as our perception of these properties for physical objects. Perceptual studies of the accuracy and precision of 3D vision in natural environments demonstrate that our percepts can be surprisingly good, but they are of course imperfect. In a classic study, for example, Smith & Smith (1961) asked people to throw balls to a set of targets a few meters away and found high accuracy and precision under full-cue conditions. This performance is even more impressive when considering that the ball-throwing paradigm incorporates errors in motor responses. In practice, the accuracy of real-world distance perception likely depends on both the depth information available in the particular environment (e.g., whether there are textured surfaces with rich perspective cues) and the distance regime under investigation (e.g., peripersonal space versus long distances).

Distance perception for virtual objects has been more extensively studied in VR as compared to AR, and this body of work has identified a range of system-level factors, such as FOV and realism, that appear to modulate the accuracy of distance perception to virtual objects in virtual environments (Kelly 2023, Renner et al. 2013). In general, distances in fully virtual environments tend to be underestimated in comparison to estimates in similar real environments. Based on our normative understanding of depth perception, this underestimation likely results from a variety of factors, including the lack of some naturally occurring depth cues in the graphics used in experiments (e.g., less rich perspective cues) and the presence of conflicting cues that indicate a flat surface (e.g., the pixel array that makes up the VR display).

Studies of distance perception in AR have the advantage of being able to directly compare a real and virtual object presented in the same (real) context. Studies of egocentric AR distance perception suggest that systematic underestimation and overestimation errors can be made for virtual objects, but that these errors are generally of a similar size to errors made for real objects and smaller than those made in fully virtual environments (Jones et al. 2008, Rolland et al. 2002, Stefanucci et al. 2021, Swan et al. 2007). A recent study by Gagnon and colleagues (2021a), however, examined egocentric distance perception to real and virtual people over long distances (10–35 m) and consistently found that distances were underestimated to virtual people as compared to real people across different environments and different response paradigms. This result suggests that cues related to the objects themselves may be less informative in AR than in the real world. It has been hypothesized that increasing contextual cues—such as shadows for virtual objects on the ground—can increase accuracy; however, rendering shadows is challenging in optical see-through AR systems that do not block light (Adams et al. 2022).

Additional studies have explored the unique distance estimation problems that can be posed by depth cue conflicts between virtual objects and physical ones in AR. Ellis & Menges (1998), for instance, asked participants to adjust the distance of a physical pointer to match the distance of a virtual object. Sometimes, a second physical object was placed along the line of sight such that the virtual object was visually superimposed on it. While observers were quite accurate at matching the distance of the virtual object in isolation, the introduction of superposition resulted in systematic changes. When the physical object was placed closer in depth than the virtual object, observers at times perceived the physical object to become transparent, as if the observer had X-ray vision (for descriptions of similar phenomena, see Singh et al. 2010). **Figure 6** contains a stereopair illustrating this observation. Conflicting occlusion cues are important in AR because they can

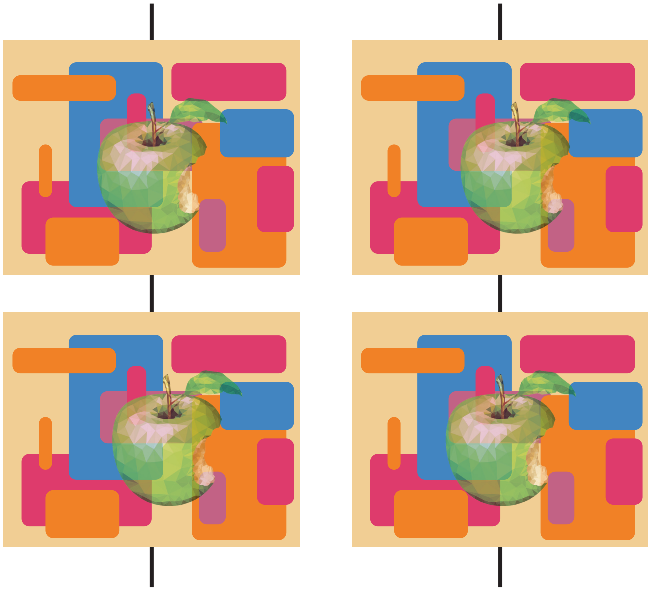

Figure 6

Cue conflicts in augmented reality (AR) can result in unusual depth percepts. Each row is a cross- or divergently fusible stereopair. If cross-fused, the top row shows an example in which the occlusion and binocular disparity cues both indicate that the apple is in the foreground, and the bottom row shows an example in which the occlusion indicates that the apple is in the foreground, but binocular disparity indicates that it is in the background. The cue conflict can create a percept of X-ray vision. If divergently fused, the top and bottom configurations are reversed. Apple image obtained from Pixabay (**https://pixabay.com/service/license/**).

commonly occur if software-based occlusion handling is unreliable (Macedo & Apolinario 2023). Occlusion handling refers to computational strategies used to determine whether a physical object is occluding a virtual object and then to render only the unoccluded portion of the virtual object. Effectively, the stimulus in **Figure 6** puts binocular disparity cues in conflict with occlusion cues (the virtual object is still occluding the physical object, even though binocular disparity indicates that it is farther away). This apparent conflict results in multiple possible interpretations. It is possible that the occlusion cue is correct, and other cues are incorrect (the virtual object is nearer than the other cues suggest), but it is also possible that the physical surface is not opaque and has become transparent. Basic research suggests that, when depth cues from occlusion geometry and binocular disparity are put in conflict, occlusion is typically the stronger cue, but the cue that ultimately wins can still depend on the spatial layout of the stimulus (Braunstein et al. 1986, Chen et al. 2018). Extreme cue conflicts like those illustrated in **Figure 6** push the limits of perceptual inference and show how the visual system works to resolve conflicting information in a naturalistic context, even if the ultimate perceptual interpretation is still rather bizarre.

Lastly, emerging lines of research are examining 3D spatial perception by focusing on judgments of action capabilities in AR spaces (Gagnon et al. 2021b, Pointon et al. 2018). For example, participants may be asked to judge whether they can step over a virtual gap on the ground or pass through the space between two virtual poles. Such studies are expanding our understanding of the fidelity of spatial percepts elicited by current AR systems and the consequences of these percepts for action judgments. However, if AR systems get to the point of eliciting affordance judgments that are identical to fully physical spaces, then these lines of research will create compelling opportunities to study perception-in-action across a range of diverse scenarios that are not possible when one is limited to wholly physical environments. Future studies, for instance, could explore the rules that govern depth cue combination during natural behavior by systematically varying the 3D information available about virtual objects and examining the effects on people's actions, in addition to their perception, within an AR space.

RELATED TOPICS

Augmented reality has not been around for nearly as long as physical reality. Nonetheless, the body of perceptual research pertinent to AR is impressively diverse. In this section, I briefly point the reader to a range of other research areas that are not addressed in this review. Research that helps to specify the necessary accuracy, speed, and robustness of AR system calibration and user motion tracking is key for developing experiences with consistent visual cues and for avoiding motion artifacts. For example, recent work has shown that the visual system is impressively sensitive to motion artifacts termed visual jitter in AR—in which virtual objects appear unanchored from the real world—which can result from user motion tracking errors and latencies (Wilmott et al. 2022). Beyond the perception of 3D objects in AR, researchers are examining other considerations for AR applications, such as the ability to reliably read text (e.g., Gabbard et al. 2006) and the factors that contribute more generally to discomfort in AR devices and similar wearables (e.g., Kaufeld et al. 2022, Stanney et al. 2020). This review focuses on the visual sense, but a fully immersive AR system would include the ability to touch, hear, and possibly smell the virtual content (e.g., Maisto et al. 2017, Sodnik et al. 2006). Indeed, multisensory AR relates to the issues of physical realism and sense of presence, which are core to some AR experiences (Marto & Gonçalves 2022). AR systems that incorporate eye and hand tracking also afford opportunities for new forms of interaction, such as moving virtual objects just by looking at them (Plopski et al. 2022). These topics are just a few examples of the breadth of AR-related research that is ongoing in fields from vision science, to human factors, to human–computer interaction.

CONCLUSION

The desire to understand, model, and predict perception of AR imagery has pushed scientists to ask whether foundational perceptual findings based on simple stimuli can reliably generalize to real-world scenarios. It has also driven a new wave of research using more naturalistic stimuli, grappling with the complexities that these stimuli introduce for developing quantitative models of visual perception. Importantly, while the enabling technologies for AR have evolved over time, the perceptual principles that govern AR experiences remain the same, allowing the impact of this work to accumulate and to support the next generation of technologies.

SUMMARY POINTS

1. Augmented reality (AR) systems that aim to merge virtual imagery with our view of the physical world can suffer from visual limitations and perceptual artifacts. Addressing these flaws requires a comprehensive understanding of visual perception.
2. Light from an optical see-through AR system must be directed from an image source to the user's pupil, which places constraints on (*a*) the pupil locations from which the displayed image can be seen, (*b*) the amount of the user's visual field that the displayed image

covers, and (*c*) the spatial and temporal resolution of the imagery. Perceptual research can be used to guide design specifications for each of these factors, but this research must have good external validity for AR experiences.

3. AR systems that aim to create high-fidelity three-dimensional (3D) experiences often present imagery to both eyes, consistent with our natural stereoscopic vision. Considerations of the advantages and challenges of binocular vision are thus key across a range of perceptual factors in AR.
4. Ultimately, presenting virtual objects in AR with realistic 3D surface properties, focus cues, and depth information relies on leveraging our scientific understanding of the multifaceted inferences that people make about materials, lighting, and geometry in their environment. Current perceptual research on AR is testing our understanding of these inferences and highlighting key areas for future inquiry.
5. To date, the development of AR systems and related technologies has both leveraged our existing knowledge of perception and driven perceptual research in new directions.

FUTURE ISSUES

1. Computational models that can reliability predict visual quality and artifact visibility in AR will be useful for anticipating and avoiding potential perceptual issues in these systems and for driving innovation. A range of psychophysical data already exists that is relevant to AR system development; however, the applicability of models based on these data to naturalistic visual imagery, such as that encountered with AR systems, is yet to be fully tested.
2. While many perceptual issues are general across different display systems, emerging hardware will also continue to raise new questions. For example, holographic displays that use phase-only holograms produce distinct visual noise patterns that should spur new lines of inquiry in perceptual research.
3. As AR systems become more sophisticated, they will continue to present new opportunities for advancing areas of perceptual science, such as 3D surface perception and perception-in-action, that aim to understand how our perceptual systems operate in natural contexts.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The author would like to thank Minqi Wang, Rachel Brown, Vasha DuTell, and Anjul Patney for assistance with graphics and Marty Banks, Hany Farid, Dennis Levi, Emma Alexander, Minqi Wang, Dylan Fox, Iona McLean, Joohwan Kim, Michael Murdoch, and David Hoffman for helpful feedback on earlier versions of the article. The author's work on this review was supported by the National Science Foundation (grant 2041726).

LITERATURE CITED

Adams H, Stefanucci J, Creem-Regehr S, Bodenheimer B. 2022. Depth perception in augmented reality: the effects of display, shadow, and position. In *Proceedings of the 2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 792–801. Piscataway, NJ: IEEE

Aizenman AM, Koulieris GA, Gibaldi A, Sehgal V, Levi DM, Banks MS. 2022. The statistics of eye movements and binocular disparities during VR gaming: implications for headset design. *ACM Trans. Graph.* 42(1):7

Akeley K, Watt SJ, Girshick AR, Banks MS. 2004. A stereo display prototype with multiple focal distances. *ACM Trans. Graph.* 23(3):804–13

Akşit K, Chakravarthula P, Rathinavel K, Jeong Y, Albert R, et al. 2019. Manufacturing application-driven foveated near-eye displays. *IEEE Trans. Vis. Comput. Graph.* 25(5):1928–39

Anderson BL. 2003. The role of occlusion in the perception of depth, lightness, and opacity. *Psychol. Rev.* 110(4):785–801

Austin RL, Denning BS, Drews BC, Fedoriouk VB, Calpito RC. 2018. Qualified viewing space determination of near-eye and head-up displays. *J. Soc. Inf. Disp.* 26(9):567–75

Banks MS, Hoffman DM, Kim J, Wetzstein G. 2016. 3D displays. *Annu. Rev. Vis. Sci.* 2:397–435

Basgöze Z, White DN, Burge J, Cooper EA. 2020. Natural statistics of depth edges modulate perceptual stability. *J. Vis.* 20(8):10

Brainard DH, Radonjić A. 2014. Color constancy. In *The New Visual Neurosciences*, ed. JS Werner, LM Chalupa, pp. 545–56. Cambridge, MA: MIT Press

Braunstein ML, Andersen GJ, Rouse MW, Tittle JS. 1986. Recovering viewer-centered depth from disparity, occlusion, and velocity gradients. *Percept. Psychophys.* 40(4):216–24

Brown R, DuTell V, Walter B, Rosenholtz R, Shirley P, et al. 2022. Efficient dataflow modeling of peripheral encoding in the human visual system. *ACM Trans. Appl. Percept.* 20(1):1

Cakmakci O, Ha Y, Rolland JP. 2004. A compact optical see-through head-worn display with occlusion support. In *Proceedings of the Third IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 16–25. Piscataway, NJ/New York: IEEE/ACM

Cakmakci O, Hoffman DM, Balram N. 2019. 3D eyebox in augmented and virtual reality optics. *SID Symp. Dig. Tech. Pap.* 50(1):438–41

Campbell FW. 1957. The depth of field of the human eye. *Opt. Acta Int. J. Opt.* 4(4):157–64

Charman WN, Tucker J. 1978. Accommodation as a function of object form. *Am. J. Optom. Physiol. Opt.* 55(2):84–92

Chen Q, Peng Z, Li Y, Liu S, Zhou P, et al. 2019. Multi-plane augmented reality display based on cholesteric liquid crystal reflective films. *Opt. Express* 27(9):12039–47

Chen Z, Denison RN, Whitney D, Maus GW. 2018. Illusory occlusion affects stereoscopic depth perception. *Sci. Rep.* 8:5297

Choi S, Gopakumar M, Peng Y, Kim J, Wetzstein G. 2021. Neural 3D holography: learning accurate wave propagation models for 3D holographic virtual and augmented reality displays. *ACM Trans. Graphics* 40(6):240

Cholewiak SA, Başgöze Z, Cakmakci O, Hoffman D, Cooper EA. 2020. A perceptual eyebox for near-eye displays. *Opt. Express* 28(25):38008–28

Cholewiak SA, Love GD, Banks MS. 2018. Creating correct blur and its effect on accommodation. *J. Vision* 18(9):1

Curcio CA, Allen KA. 1990. Topography of ganglion cells in human retina. *J. Comp. Neurol.* 300(1):5–25

Curcio CA, Sloan KR, Kalina RE, Hendrickson AE. 1990. Human photoreceptor topography. *J. Comp. Neurol.* 292(4):497–523

Dey A, Billinghurst M, Lindeman RW, Swan JE. 2018. A systematic review of 10 years of augmented reality usability studies: 2005 to 2014. *Front. Robot. AI* 5:37

Ding J, Levi DM. 2017. Binocular combination of luminance profiles. *J. Vis.* 17(13):4

Duinkharjav B, Chakravarthula P, Brown R, Patney A, Sun Q. 2022. Image features influence reaction time: a learned probabilistic perceptual model for saccade latency. *ACM Trans. Graph.* 41(4):144

Ellis SR, Menges BM. 1998. Localization of virtual objects in the near visual field. *Hum. Factors* 40(3):415–31

Fisher SK, Ciuffreda KJ. 1988. Accommodation and apparent distance. *Perception* 17(5):609–21

Freeman J, Simoncelli EP. 2011. Metamers of the ventral stream. *Nat. Neurosci.* 14(9):1195–201

Gabbard JL, Swan JE, Hix D. 2006. The effects of text drawing styles, background textures, and natural lighting on text legibility in outdoor augmented reality. *Presence* 15(1):16–32

Gabbard JL, Swan JE, Zedlitz J, Winchester WW. 2010. More than meets the eye: an engineering study to empirically examine the blending of real and virtual color spaces. In *Proceedings of the 2010 IEEE Virtual Reality Conference (VR)*, pp. 79–86. Piscataway, NJ: IEEE

Gagnon HC, Rosales CS, Mileris R, Stefanucci JK, Creem-Regehr SH, Bodenheimer RE. 2021a. Estimating distances in action space in augmented reality. *ACM Trans. Appl. Percept.* 18(2):7

Gagnon HC, Zhao Y, Richardson M, Pointon GD, Stefanucci JK, et al. 2021b. Gap affordance judgments in mixed reality: testing the role of display weight and field of view. *Front. Virtual Real.* 2. **https://doi.org/10.3389/frvir.2021.654656**

Geisler WS, Perry JS. 1998. Real-time foveated multiresolution system for low-bandwidth video communication. *Proc. SPIE* 3299:294–305

Gilchrist AL, Jacobsen A. 1983. Lightness constancy through a veiling luminance. *J. Exp. Psychol. Hum. Percept. Perform.* 9(6):936–44

Guenter B, Finch M, Drucker S, Tan D, Snyder J. 2012. Foveated 3D graphics. *ACM Trans. Graph.* 31(6):164

Hassani N, Murdoch MJ. 2019. Investigating color appearance in optical see-through augmented reality. *Color Res. Appl.* 44(4):492–507

Hincapie-Ramos JD, Ivanchuk L, Sridharan SK, Irani PP. 2015. SmartColor: real-time color and contrast correction for optical see-through head-mounted displays. *IEEE Trans. Vis. Comput. Graphics* 21(12):1336–48

Hoffman DM, Banks MS. 2010. Focus information is used to interpret binocular images. *J. Vis.* 10(5):13

Hoffman DM, Girshick AR, Akeley K, Banks MS. 2008. Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *J. Vis.* 8(3):33

Howard IP, Rogers BJ. 2002. *Seeing in Depth*, Vol. 2: *Depth Perception*. Toronto: Univ. Toronto Press

Huang H-P, Wei M, Chen S. 2021. White appearance of virtual stimuli produced by augmented reality. *Color Res. Appl.* 46(2):294–302

Jennings BJ, Kingdom FAA. 2016. Detection of between-eye differences in color: interactions with luminance. *J. Vis.* 16(3):23

Jones JA, Swan JE, Singh G, Kolstad E, Ellis SR. 2008. The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception. In *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization*, pp. 9–14. New York: ACM

Kaufeld M, Mundt M, Forst S, Hecht H. 2022. Optical see-through augmented reality can induce severe motion sickness. *Displays* 74:102283

Kelly JW. 2023. Distance perception in virtual reality: a meta-analysis of the effect of head-mounted display characteristics. *IEEE Trans. Visual Comput. Graphics*. In press

Kim J, Jeong Y, Stengel M, Akşit K, Albert R, et al. 2019. Foveated AR: dynamically-foveated augmented reality display. *ACM Trans. Graph.* 38(4):99

Kingdom FAA. 2011. Lightness, brightness and transparency: a quarter century of new ideas, captivating demonstrations and unrelenting controversy. *Vis. Res.* 51(7):652–73

Kingdom FAA, Seulami NM, Jennings BJ, Georgeson MA. 2019. Interocular difference thresholds are mediated by binocular differencing, not summing, channels. *J. Vis.* 19(14):18

Kiyokawa K, Billinghurst M, Campbell B, Woods E. 2003. An occlusion capable optical see-through head mount display for supporting co-located collaboration. In *Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 133–41. Piscataway, NJ/New York: IEEE/ACM

Klymenko V, Harding TH, Beasley HH, Martin JS. 1999. *The effect of helmet mounted display field-of-view configurations on target acquisition*. Rep. 99-19, U. S. Army Aeromed. Res. Lab., Fort Rucker, AL. **https://apps.dtic.mil/sti/citations/ADA368601**

Klymenko V, Verona RW, Beasley HH, Martin JS, McLean WE. 1994. *Factors affecting the visual fragmentation of the field-of-view in partial binocular overlap displays*. Rep. 94–29, U. S. Army Aeromed. Res. Lab., Fort Rucker, AL. **https://apps.dtic.mil/sti/citations/ADA283081**

Knill DC, Richards W. 1996. *Perception as Bayesian Inference*. Cambridge, UK: Cambridge Univ. Press
Kooi FL, Toet A. 2004. Visual comfort of binocular and 3D displays. *Displays* 25(2–3):99–108
Krajancich B, Kellnhofer P, Wetzstein G. 2021. A perceptual model for eccentricity-dependent spatio-temporal flicker fusion and its applications to foveated graphics. *ACM Trans. Graph.* 40(4):47
Kramida G. 2016. Resolving the vergence-accommodation conflict in head-mounted displays. *IEEE Trans. Vis. Comput. Graph.* 22(7):1912–31
Kress BC. 2020. *Optical Architectures for Augmented-, Virtual-, and Mixed-Reality Headsets*. Bellingham, WA: SPIE
Labhishetty V, Cholewiak SA, Roorda A, Banks MS. 2021. Lags and leads of accommodation in humans: fact or fiction? *J. Vis.* 21(3):21
Lambooij M, Fortuin M, Heynderickx I, IJsselsteijn W. 2009. Visual discomfort and visual fatigue of stereoscopic displays: a review. *J. Imag. Sci. Technol.* 53(3):30201
Lanman D, Luebke D. 2013. Near-eye light field displays. *ACM Trans. Graph.* 32(6):220
Lee S, Cho J, Lee B, Jo Y, Jang C, et al. 2018. Foveated retinal optimization for see-through near-eye multi-layer displays. *IEEE Access* 6:2170–80
Legge GE, Rubin GS. 1981. Binocular interactions in suprathreshold contrast perception. *Percept. Psychophys.* 30(1):49–61
Levelt WJM. 1965. *On binocular rivalry*. PhD Diss., Inst. Percept. RVO–TNO, Soesterberg, Neth.
Lin JJ-W, Duh HBL, Parker DE, Abi-Rached H, Furness TA. 2002. Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. In *Proceedings of the 2002 IEEE Virtual Reality Conference (VR)*, pp. 164–71. Piscataway, NJ: IEEE
Liu S, Cheng D, Hua H. 2008. An optical see-through head mounted display with addressable focal planes. In *Proceedings of the 2008 IEEE Virtual Reality Conference (VR)*, pp. 164–71. Piscataway, NJ: IEEE
Love GD, Hoffman DM, Hands PJW, Gao J, Kirby AK, Banks MS. 2009. High-speed switchable lens enables the development of a volumetric stereoscopic display. *Opt. Express* 17(18):15716–25
Ludvigh E. 1941. Extrafoveal visual acuity as measured with Snellen test-letters. *Am. J. Ophthalmol.* 24(3):302–10
Macedo MCF, Apolinario AL. 2023. Occlusion handling in augmented reality: past, present and future. *IEEE Trans. Vis. Comput. Graph.* 29(2):1590–609
MacKenzie KJ, Dickson RA, Watt SJ. 2011. Vergence and accommodation to multiple-image-plane stereoscopic displays: "real world" responses with practical image-plane separations? *Proc. SPIE* 7863:786315
MacKenzie KJ, Hoffman DM, Watt SJ. 2010. Accommodation to multiple-focal-plane displays: implications for improving stereoscopic displays and for accommodation control. *J. Vis.* 10:22
MacKenzie KJ, Watt SJ. 2010. Eliminating accommodation-convergence conflicts in stereoscopic displays: Can multiple-focal-plane displays elicit continuous and consistent vergence and accommodation responses? *Proc. SPIE* 7524:752417
MacKenzie KJ, Watt SJ. 2011. The stimulus to accommodation: changes in retinal contrast matter, not the spatial frequency content. *J. Vis.* 11:516
Maimone A, Georgiou A, Kollin JS. 2017. Holographic near-eye displays for virtual and augmented reality. *ACM Trans. Graph.* 36(4):85
Maimone A, Lanman D, Rathinavel K, Keller K, Luebke D, Fuchs H. 2014. Pinlight displays: wide field of view augmented reality eyeglasses using defocused point light sources. *ACM Trans. Graph.* 33(4):89
Maisto M, Pacchierotti C, Chinello F, Salvietti G, De Luca A, Prattichizzo D. 2017. Evaluation of wearable haptic systems for the fingers in augmented reality applications. *IEEE Trans. Haptics* 10(4):511–22
Mandelbaum J. 1960. An accommodation phenomenon. *Arch. Ophthalmol.* 63(6):923–26
Mantiuk RK, Denes G, Chapiro A, Kaplanyan A, Rufo G, et al. 2021. FovVideoVDP: a visible difference predictor for wide field-of-view video. *ACM Trans. Graph.* 40(4):49
Marto A, Gonçalves A. 2022. Augmented reality games and presence: a systematic review. *J. Imag. Sci. Technol.* 8(4):91
Mathews S, Kruger PB. 1994. Spatiotemporal transfer function of human accommodation. *Vis. Res.* 34(15):1965–80

McLean B, Smith S. 1987. Developing a wide field of view Hmd for simulators. *Proc. SPIE* 0778:79–82
Melzer JE, Moffitt K. 1989. Partial binocular-overlap in helmet-mounted displays. *Proc. SPIE* 1117:56–62
Melzer JE, Moffitt KW. 1991. Ecological approach to partial binocular overlap. *Proc. SPIE* 1456:124–31
Metelli F. 1974. The perception of transparency. *Sci. Am.* 230(4):90–98
Milgram P, Takemura H, Utsumi A, Kishino F. 1995. Augmented reality: a class of displays on the reality-virtuality continuum. *Proc. SPIE* 2351:282–92
Moffitt K, Browne MP. 2019. Visibility of color symbology in head-up and head-mounted displays in daylight environments. *Opt. Eng.* 58(5):051809
Murdoch MJ. 2020. Brightness matching in optical see-through augmented reality. *J. Opt. Soc. Am. A* 37(12):1927–36
Noorlander C, Koenderink JJ, den Ouden RJ, Edens BW. 1983. Sensitivity to spatiotemporal colour contrast in the peripheral visual field. *Vis. Res.* 23(1):1–11
Owens DA. 1979. The Mandelbaum effect: evidence for an accommodative bias toward intermediate viewing distances. *J. Opt. Soc. Am. B* 69(5):646–52
Owens DA. 1980. A comparison of accommodative responsiveness and contrast sensitivity for sinusoidal gratings. *Vis. Res.* 20(2):159–67
Padmanaban N, Konrad R, Stramer T, Cooper EA, Wetzstein G. 2017. Optimizing virtual reality for all users through gaze-contingent and adaptive focus displays. *PNAS* 114(9):2183–88
Patney A, Salvi M, Kim J, Kaplanyan A, Wyman C, et al. 2016. Towards foveated rendering for gaze-tracked virtual reality. *ACM Trans. Graph.* 35(6):179
Plopski A, Hirzle T, Norouzi N, Qian L, Bruder G, Langlotz T. 2022. The eye in extended reality: a survey on gaze interaction and eye tracking in head-worn extended reality. *ACM Comput. Surv.* 55(3):53
Pointon G, Thompson C, Creem-Regehr S, Stefanucci J, Joshi M, et al. 2018. Judging action capabilities in augmented reality. In *Proceedings of the 15th ACM Symposium on Applied Perception*, Art. 6. New York: ACM
Radonjić A, Cottaris NP, Brainard DH. 2015. Color constancy in a naturalistic, goal-directed task. *J. Vis.* 15(13):3
Ratnam K, Konrad R, Lanman D, Zannoli M. 2019. Retinal image quality in near-eye pupil-steered systems. *Opt. Express* 27(26):38289–311
Ren D, Goldschwendt T, Chang Y, Hollerer T. 2016. Evaluating wide-field-of-view augmented reality with mixed reality simulation. In *Proceedings of the 2016 IEEE Virtual Reality Conference (VR)*, pp. 93–102. Piscataway, NJ: IEEE
Renner RS, Velichkovsky BM, Helmert JR. 2013. The perception of egocentric distances in virtual environments—a review. *ACM Comput. Surv.* 46(2):23
Rolland JP, Krueger MW, Goon A. 2000. Multifocal planes head-mounted displays. *Appl. Opt.* 39(19):3209–15
Rolland JP, Meyer C, Arthur K, Rinalducci E. 2002. Method of adjustments versus method of constant stimuli in the quantification of accuracy and precision of rendered depth in head-mounted displays. *Presence* 11(6):610–25
Rolland JP, Yoshida A, Davis LD, Reif JH. 1998. High-resolution inset head-mounted display. *Appl. Opt.* 37(19):4183–93
Rosenfield M, Ciuffreda KJ. 1991. Accommodative responses to conflicting stimuli. *J. Opt. Soc. Am. A* 8(2):422–27
Rosenholtz R. 2016. Capabilities and limitations of peripheral vision. *Annu. Rev. Vis. Sci.* 2:437–57
Sebastian S, Burge J, Geisler WS. 2015. Defocus blur discrimination in natural images with natural optics. *J. Vis.* 15(5):16
Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, et al. 1995. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* 268(5212):889–93
Shibata T, Kim J, Hoffman DM, Banks MS. 2011. The zone of comfort: predicting visual discomfort with stereo displays. *J. Vis.* 11(8):11
Shimojo S, Nakayama K. 1990. Real world occlusion constraints and binocular rivalry. *Vis. Res.* 30(1):69–80
Singh G, Swan JE, Jones JA, Ellis SR. 2010. Depth judgment measures and occluding surfaces in near-field augmented reality. In *APGV '10: Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*, pp. 149–56. New York: ACM

Singh V, Burge J, Brainard DH. 2022. Equivalent noise characterization of human lightness constancy. *J. Vis.* 22(2):2

Smith PC, Smith OW. 1961. Ball throwing responses to photographically portrayed targets. *J. Exp. Psychol.* 62:223–33

Sodnik J, Tomazic S, Grasset R, Duenser A, Billinghurst M. 2006. Spatial sound localization in an augmented reality environment. In *Proceedings of the 18th Australia Conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments*, pp. 111–18. New York: ACM

Spector RH. 2011. Visual fields. In *Clinical Methods: The History, Physical, and Laboratory Examinations*, ed. HK Walker, WD Hall, JW Hurst, pp. 565–672. Boston: Butterworths

Spjut J, Boudaoud B, Kim J, Greer T, Albert R, et al. 2020. Toward standardized classification of foveated displays. *IEEE Trans. Vis. Comput. Graph.* 26(5):2126–34

Stanney K, Lawson BD, Rokers B, Dennison M, Fidopiastis C, et al. 2020. Identifying causes of and solutions for cybersickness in immersive technology: reformulation of a research and development agenda. *Int. J. Hum.-Comput. Interface* 36(19):1783–803

Stefanucci JK, Creem-Regehr S, Bodenheimer B. 2021. Comparing distance judgments in real and augmented reality. In *Proceedings of the 2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 82–86. Piscataway, NJ: IEEE

Sutherland IE. 1968. A head-mounted three dimensional display. In *Proceedings of the December 9–11, 1968, Fall Joint Computer Conference*, Part I, pp. 757–64. New York: ACM

Swafford NT, Iglesias-Guitian JA, Koniaris C, Moon B, Cosker D, Mitchell K. 2016. User, metric, and computational evaluation of foveated rendering methods. In *Proceedings of the ACM Symposium on Applied Perception*, pp. 7–14. New York: ACM

Swan JE II, Jones A, Kolstad E, Livingston MA, Smallman HS. 2007. Egocentric depth judgments in optical, see-through augmented reality. *IEEE Trans. Vis. Comput. Graph.* 13(3):429–42

Tan G, Lee Y-H, Zhan T, Yang J, Liu S, et al. 2018. Foveated imaging for near-eye displays. *Opt. Express* 26(19):25076–85

Thibos LN. 2020. Retinal image formation and sampling in a three-dimensional world. *Annu. Rev. Vis. Sci.* 6:469–89

Thibos LN, Cheney FE, Walsh DJ. 1987. Retinal limits to the detection and resolution of gratings. *J. Opt. Soc. Am. A* 4(8):1524–29

Thomas ML, Siegmund WP, Antos SE, Robinson RM. 1990. Fiber optic development for use on the fiber optic helmet mounted display. *Opt. Eng.* 29(8):855–61

Trepkowski C, Eibich D, Maiero J, Marquardt A, Kruijff E, Feiner S. 2019. The effect of narrow field of view and information density on visual search performance in augmented reality. In *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 575–84. Piscataway, NJ: IEEE

Van Nes FL, Bouman MA. 1967. Spatial modulation transfer in the human eye. *J. Opt. Soc. Am. B* 57(3):401–6

Vater C, Wolfe B, Rosenholtz R. 2022. Peripheral vision in real-world tasks: a systematic review. *Psychon. Bull. Rev.* 29:1531–57

Wallach H, Norris CM. 1963. Accommodation as a distance cue. *Am. J. Psychol.* 76:659–64

Walton DR, Anjos RKD, Friston S, Swapp D, Akşit K, et al. 2021. Beyond blur: real-time ventral metamers for foveated rendering. *ACM Trans. Graph.* 40(4):48

Wang M, Cooper EA. 2022. Perceptual guidelines for optimizing field of view in stereoscopic augmented reality displays. *ACM Trans. Appl. Percept.* 19(4):19

Watson AB. 2007. The spatial standard observer: a new tool for display metrology. *Inf. Disp.* 23(1):12

Watson AB, Ahumada AJ Jr. 2005. A standard model for foveal detection of spatial contrast. *J. Vis.* 5(9):717–40

Watt SJ, Akeley K, Ernst MO, Banks MS. 2005a. Focus cues affect perceived depth. *J. Vis.* 5(10):834–62

Watt SJ, Akeley K, Girshick AR, Banks MS. 2005b. Achieving near-correct focus cues in a 3D display using multiple image planes. *Proc. SPIE* 5666:393–401

Weert CMM, Levelt WJM. 1974. Binocular brightness combinations: additive and non-additive aspects. *Percept. Psychophys.* 15:551–62

Weiland C, Braun A-K, Heiden W. 2009. Colorimetric and photometric compensation for see-through displays. In *Universal Access in Human-Computer Interaction: Intelligent and Ubiquitous Interaction Environments*, ed. C Stephanidis, pp. 603–12. Lect. Notes Comput. Sci. 5615. Berlin: Springer

Wendt G, Faul F. 2022. Binocular luster—a review. *Vis. Res.* 194:108008

Wetzstein G. 2020. Augmented and virtual reality. In *NANO-CHIPS 2030: On-Chip AI for an Efficient Data-Driven World*, ed. B Murmann, B Hoefflinger, pp. 467–99. Berlin: Springer

Wilcox LM, Lakra DC. 2007. Depth from binocular half-occlusions in stereoscopic images of natural scenes. *Perception* 36(6):830–39

Wilmott JP, Erkelens IM, Murdison ST, Rio KW. 2022. Perceptibility of jitter in augmented reality head-mounted displays. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, pp. 470–78. Piscataway, NJ: IEEE

Wilson A, Hua H. 2017. Design and prototype of an augmented reality display with per-pixel mutual occlusion capability. *Opt. Express* 25(24):30539–49

Yang X, Zhang L, Wong T-T, Heng P-A. 2012. Binocular tone mapping. *ACM Trans. Graph.* 31(4):93

Zhan T, Yin K, Xiong J, He Z, Wu S-T. 2020. Augmented reality and virtual reality displays: perspectives and challenges. *iScience* 23(8):101397

Zhang L, Murdoch MJ. 2021. Perceived transparency in optical see-through augmented reality. In *Proceedings of the 2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 115–20. Piscataway, NJ: IEEE

Zhang Y, Isoyama N, Sakata N, Kiyokawa K, Hua H. 2020. 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)*, pp. 301–11. Piscataway, NJ: IEEE

Zhong F, Koulieris GA, Drettakis G, Banks MS, Chambe M, et al. 2019. DiCE: dichoptic contrast enhancement for VR and stereo displays. *ACM Trans. Graph.* 38(6):211