

# Towards Indistinguishable Augmented Reality: A Survey on Optical See-Through Head-Mounted Displays

YUTA ITOH, Tokyo Institute of Technology, Japan  
TOBIAS LANGLOTZ, University of Otago, New Zealand  
JONATHAN SUTTON, University of Otago, New Zealand  
ALEXANDER PLOPSKI, University of Otago, New Zealand

Adding virtual information that is indistinguishable from reality has been a long-awaited goal in Augmented Reality (AR). While already demonstrated in the 1960s, only recently have Optical See-Through Head-Mounted Displays (OST-HMDs) seen a reemergence, partially thanks to large investments from industry, and are now considered to be the ultimate hardware for augmenting our visual perception. In this paper, we provide a thorough review of state-of-the-art OST-HMD-related techniques that are relevant to realize the aim of an AR interface almost indistinguishable from reality. In this work, we have an initial look at human perception to define requirements and goals for implementing such an interface. We follow up by identifying three key challenges for building an OST-HMD based AR interface that is indistinguishable from reality: spatial realism, temporal realism, and visual realism. We discuss existing works that aim to overcome these challenges while also reflecting against the goal set by human perception. Finally, we give an outlook into promising research directions and expectations for the years to come.

CCS Concepts: • **Human-centered computing** → **Mixed / augmented reality**; **Graphical user interfaces**;

Additional Key Words and Phrases: augmented reality, mixed reality, head-mounted display, near eye display, optical see through, wearable computing, visual coherence, review

## ACM Reference Format:

Yuta Itoh, Tobias Langlotz, Jonathan Sutton, and Alexander Plopski. 2021. Towards Indistinguishable Augmented Reality: A Survey on Optical See-Through Head-Mounted Displays. *ACM Comput. Surv.* 0, 0, Article 00 (2021), 36 pages. <https://doi.org/00000001.0000001>

## 1 INTRODUCTION

Providing users with additional information generated by computer systems is essential to many research fields in *human-computer interaction* (HCI). The fields of *mixed realities* (MR) and *augmented realities* (AR) particularly emphasise this idea of integrating virtual and real information in a single common space called a mixed or augmented reality [4, 121].

For some applications, the visual coherence of this integration in a common space (e.g. for displaying mostly context-based textual information or notifications) is not important; for many applications of MR and AR, however, it is key to integrate virtual information so that it closely resembles reality in its appearance, making it visually indistinguishable. Examples include applications that try to trick the brain for therapeutic reasons (e.g. stroke rehabilitation [141] and phobia

---

Authors' addresses: Yuta Itoh, Tokyo Institute of Technology, Japan, [yuta.ito@c.titech.ac.jp](mailto:yuta.ito@c.titech.ac.jp); Tobias Langlotz, University of Otago, New Zealand, [tobias.langlotz@otago.ac.nz](mailto:tobias.langlotz@otago.ac.nz); Jonathan Sutton, University of Otago, New Zealand, [sutjo752@student.otago.ac.nz](mailto:sutjo752@student.otago.ac.nz); Alexander Plopski, University of Otago, New Zealand, [alexander.plopski@otago.ac.nz](mailto:alexander.plopski@otago.ac.nz).

---

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

© 2021 Copyright held by the owner/author(s).

0360-0300/2021/0-ART00

<https://doi.org/00000001.0000001>

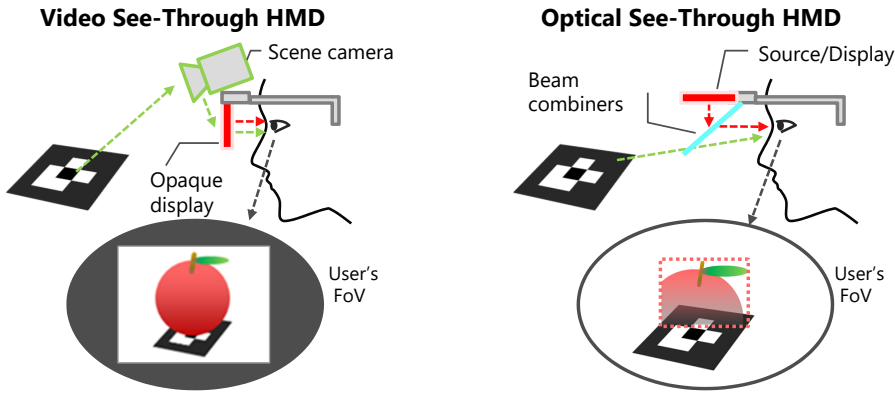


Fig. 1. Comparison of VST and OST displays under a typical scenario of marker-based AR.

treatment [70]), architectural AR applications and computer games. For all these applications, virtual and real information must be difficult to distinguish—a system must make the physical environment and the superimposed information visually coherent. Coherence, as a general aim, also often applies to other modalities used in AR (e.g. for audio augmentations [99] and haptic augmentations [67]). In this work, however, we focus on the coherence of visual AR as one of the most explored research directions, particularly the technical difficulties of achieving this visual coherence in *optical see-through head-mounted displays* (OST-HMDs).

### 1.1 Motivation and Contribution

Azuma gave three criteria for his widely used definition of AR: combining the virtual and real, being registered in 3D (with six *degrees of freedom* (DoF)), and being interactive in real time [4]. However, these criteria do not automatically lead to an AR experience, where the real and virtual become indistinguishable; they depend on the technical approach taken to creating an AR experience [22, 93].

One such technical approach is *video see-through augmented reality* (VST), which is based on the concept of seeing the world through cameras. Typical examples are mobile phone-based AR systems [98] and *video see-through head-mounted displays* (VST-HMD), which add external cameras to closed, HMDs typically used in *virtual reality* (VR) [147] (Fig. 1 left). When using video see-through AR displays, the user never directly sees the physical world but only sees it as represented in the camera feed. This has inherent practical issues, as it decouples the user from reality, and social issues, such as disabling mutual eye contact. The advantage, however, is that creating a visually coherent representation is easier because the information is visually merged in camera-image space [22, 84].

Besides video see-through AR, research has also investigated the use of projectors to directly augment the real environment with virtual information, which is commonly referred to as *spatial augmented reality* or *projection mapping* [11, 45]. To achieve a realistic, visually coherent augmentation, spatial AR usually depends on geometrically and radiometrically aligning projectors and their surroundings [10, 11, 45].

Finally, an AR experience can also be created using *optical see-through displays*. In fact, the first approach demonstrating an AR display was an optical see-through display [160]. Optical see-through displays are semi-transparent displays in the user's view. These semi-transparent displays can be used to superimpose computer-generated graphics on the user's view (Fig. 1 right). Common

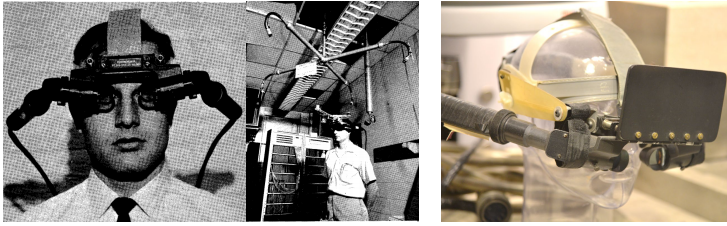


Fig. 2. (Left) The Sword of Damocles [160]. (Right) The Sword of Damocles exhibited at the Computer History Museum, Mountain View (photo taken by the authors at the Computer History Museum, CA, USA).

examples are *head-up displays* (HUDs) [32] and OST-HMD. While the original idea dates to the 1960s [160], optical see-through displays—and OST-HMDs in particular—have recently received more attention from industry and research. In comparison to VST-HMDs, optical see-through displays are closer to resembling normal glasses, which allow us to see the real environment at its highest fidelity (e.g. unaffected by technical limitations), making them more applicable to many scenarios. Unfortunately, building a convincing, visually coherent AR experience with OST-HMDs is generally considered more challenging.

This work gives an overview of the state-of-the-art techniques that allow realistic representations in OST-HMDs. We identify challenges in three key areas in particular: spatial realism, temporal realism and visual realism. The contributions of this paper are a) identifying key challenges to providing visually indistinguishable augmentations, b) clearly analysing the requirements for each key area to fulfil the goal of indistinguishability, c) reviewing the current solutions for the identified key areas, and finally d) discussing promising research directions and key future techniques.

## 1.2 Background

There are several reviews and taxonomies on AR, but they do not focus on OST-HMDs and their issues. More specific works cover topics presented here. Kruijff et al. and Collins et al. discuss perceptual challenges for AR [22, 93], the latter particularly for HMDs. Both focus on issues, however, not solutions or the current state of the technology.

Besides several papers that introduce HMD technologies common in AR and MR [75, 77, 78, 87, 90], some review papers and books have focused on OST-HMD but mainly on their optical design [15, 47, 87, 90, 91, 145, 149]. Some works go further to discuss display units or virtual image creation [47, 87, 145], how occlusions can be achieved in optical see-through systems [79], or how HMDs are commonly tracked and interacted with [87]. However, these do not cover important issues, such as geometric image distortion, latency and colour representation, and most do not cover issues such as accommodation, focus problems and spatial consistency. Finally, nearly all of these works miss the last 5–10 years of scientific progress apart from recent optical designs.

Research suggests that even a low-fidelity display can create a convincing AR experience and that high visual fidelity, which was an aim for most of the work presented in this paper, is not necessary [34, 142, 143]. Fischer et al. and Regenbrecht et al. achieved this by simplifying the appearance of the physical world; in contrast, we propose matching the visual fidelity of the real world.

## 2 OPTICAL SEE-THROUGH HEAD-MOUNTED DISPLAYS

The history of HMDs originated in the 19th century. In 1838, Wheatstone developed the stereoscope [169], a device that realises so-called stereograms by providing parallax images for both

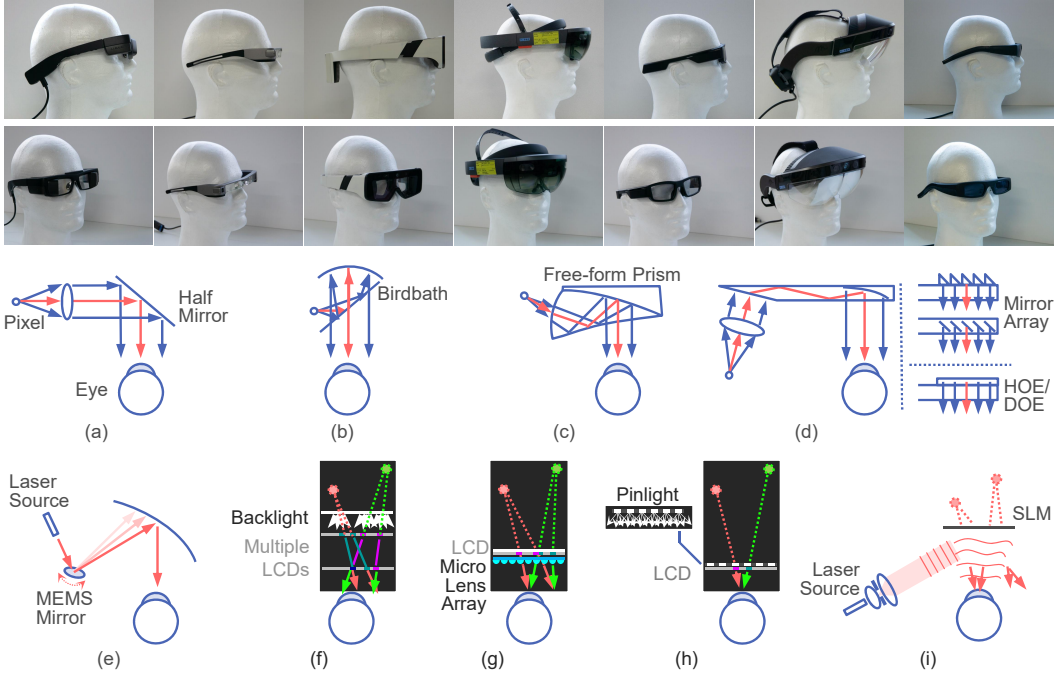


Fig. 3. Schematic overview of different optical designs used in existing OST-HMDs. Top row: commercial OST-HMDs. Bottom row: research prototypes. (a) Half mirror. (b) Birdbath combiner. (c) Free-form prism. (d) Waveguide [92] with a coated curved mirror, a wedge-shaped mirror array, cascaded mirrors or holographic/diffractive optical elements. (e) Retinal scanning. (f) Light-field OST-HMD with multiple liquid crystal display (LCD) layers (e.g. [114]). (g) Light-field OST-HMD with microlens arrays (e.g. [176]). (h) Pinlight display [116]. (i) 3D holographic display.

eyes, evoking a sensation of depth. A latent patent from the 1940s describes a version that displays artificially generated images [162]; many similar systems are described later. The first approach to demonstrate AR was the “Sword of Damocles” HMD (sometimes also referred to as the “ultimate display”) by Ivan Sutherland [160]. This HMD was the first to use the concept now referred to as optical see-through. The key idea was to use optical elements—in their case, half-silvered mirrors—installed in the optical path of the user’s eye to partially reflect a computer-generated image shown on a small cathode-ray tube (CRT) (Fig. 2). The user can still see the environment through the half-silvered mirror, but it is now superimposed (augmented) with the displayed image. Since then, many new prototypes have been demonstrated, but most are still based on this initial concept, even the optical design (Fig. 3).

Simple half-mirror designs 3.a) have a trade-off between the form factor and the size of the volume where an eye can see an AR image (the *eyebbox*). Birdbath optics is another common optical design 3.b) that combines a curved mirror lens to achieve a larger *field of view* (FoV), yet the optics have an essential issue in light efficiency since the optical path goes through half mirrors multiple times and the design tends to be larger.

Newer prototypes, however, replaced half-silvered mirrors with other optical elements that reflect a displayed image into a user’s eye while remaining transparent enough so the user can perceive the environment. One such approach is to use optical prisms or free-form optics to achieve results like those of half-silvered mirrors while allowing the light to enter from different angles, allowing



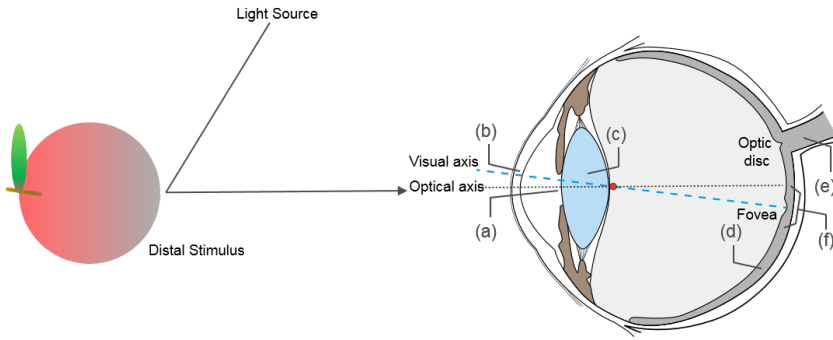


Fig. 4. Schematic drawing of a cross-section of the human eye as viewed from the top of the head. The drawing is inspired by [110]. (a) Pupil. (b) Lens. (c) Cornea. (d) Retina. (e) Optic nerve. (f) Macula.

improved form factors of the final HMD (Fig. 3.c). More recently, researchers have demonstrated how these free-form optics can even be 3D printed [138]. However, free-form optics usually require a certain thickness.

This can be reduced by using waveguides to reflect the displayed image into the user's eyes (Fig. 3.d). The key idea is to use internal reflection within the waveguide to guide the light towards the user's eyes. Different beam-combiner designs exist for waveguide optics. One forms a free-form mirror surface, such as Epson's Moverio Series, while others use a series of parallel mirrors or arrays of mirror layers, such as Lumus's products. More optically advanced approaches use a *holographic optical element* (HOE) as a waveguide and a combiner (Fig. 3.d). HOEs are holograms that are exposed so that light is transmitted or reflected depending on the incoming light's frequency. A general issue is that different light frequencies are reflected at different angles, which must be considered. A laser light source can also be directly projected into the user's eye. This is usually a specific laser projector or similar device (Fig. 3.e) and is commonly referred to as a retinal-projection display or a retinal-scanning display [36].

These standard designs usually have a fixed focal plane on which the virtual image is displayed, which causes vergence-accommodation conflict, which we address later with possible solutions that can be grouped as follows. Some designs use multiple stacked LCD layers to create different focal planes, as in a light-field display 3.f, which can also be created using a microlens array 3.g. A pinlight display modifies the design of the light source by placing an array of point light sources behind an LCD to create a pinhole projection that is always in focus 3.h. In contrast, holographic displays can create a true light-field display without using LCDs to mimic one 3.i.

This is only a rough overview of general approaches to implementing OST-HMDs. In the following, we briefly overview the physiological and perceptual factors relevant to designing OST-HMDs before identifying specific challenges and their current solutions.

### 3 PHYSIOLOGICAL AND PERCEPTUAL CONSIDERATIONS FOR OST-HMDS

When designing displays and, in particular, to display information indistinguishably from reality, we must first understand how we normally visually perceive our physical environment. Goldstein wrote one of the most widely used textbooks on this topic [42], but we only cover the physiological and perceptual foundations of our *human visual system* (HVS) and later focus on the properties and characteristics that define the requirements of OST-HMD design.

### 3.1 Light and the human eye

Visual light (with wavelengths from 390–750 nm) in the form of emitted photons is the basic requirement for visual perception. The HVS begins with the eyes [110] (Fig 4). Light is reflected from a distal stimulus (the stimulus providing information to the receiver) to the human eye. Once the light reaches the eye, it enters through the pupil (Fig 4a) and is focused onto the retina (Fig 4d) at the back of the eye by the cornea (Fig 4c) and lens (Fig 4b). The retina has receptors called rods and cones that contain light-sensitive chemicals known as visual pigments. Upon receiving light, an electrical signal is triggered by these pigments and sent to the brain by a network of neurons along the optical nerve (Fig 4e), where they are then processed. Rods are more sensitive to light (around 100 times more sensitive than cones) and thus contribute to seeing in dim environments. They contribute little to colour perception, however, and are most sensitive to wavelengths in the green and blue spectrum of visible light. They also react more slowly to changing light conditions, whereas cones are responsible for colour vision and react faster, allowing us to perceive fine details.

The distribution of these rod and cone receptors on the retina is non-uniform, which affects our visual perception. The macula is a small area on the retina that is responsible for perceiving a sharp image, with a tiny area called the fovea being the area with the sharpest vision. The fovea has a high density of cone receptors. This ratio of cones to rods changes quickly as the density of cones decreases towards the periphery of the macula. Outside the macula, we find mainly rods, although with decreasing density when moving towards the periphery of the retina.

### 3.2 Properties of the human visual system

The HVS has several notable properties that are of relevance to the production of visually coherent virtual content, such as perceivable wavelengths, dynamic range, FoV and resolution, latency and perceptible flicker, accommodation and perceptual criteria, such as the relevance of stereopsis or occlusions within the HVS and human perception.

**Colour:** The colour sensitivity of the human eye allows wavelengths between 390 nm and 750 nm to be perceived. Because of the distribution of cones, colour sensitivity is not uniform across the eye or across the spectrum of wavelengths. The fovea has a greater sensitivity to red and green chromatic stimuli and less across the periphery, while blue sensitivity is generally lower than that of red or green but more uniformly distributed between the fovea and the periphery [130].

**Dynamic range:** Research has found the static dynamic range of the human eye (assuming a constant aperture or pupil diameter) to cover 3.7 log units [94] in  $\text{cd/m}^2$ . However, the human eye has a focal length of 17 mm, and the pupil diameter can vary between 2 mm and 8 mm. This gives the human eye an equivalent aperture range of  $f/2.125$ – $f/6.5$ . Changing the pupil and consequently the aperture allows us to drastically increase the perceivable overall dynamic range. In fact, the dynamic range can be further increased by mechanical processes and photo-chemical, neural-adaptive processes that transcend the scope of this paper [42]. Research has shown that the cones in the eye are stimulated from 0.01 to  $10^8 \text{ cd/m}^2$  when the cones are active and that rods are stimulated from  $10^{-6}$  to  $10 \text{ cd/m}^2$  [94], leading to a practical overall dynamic range of approximately 46.5 stops. Note that while the human dynamic range can go beyond  $10,000 \text{ cd/m}^2$ , this can be damaging [126].

**Visual field:** Strasburger et al. [159] found that the human eye can receive light information of up to  $90^\circ$  eccentricity from the vertical meridian towards the periphery, known as the temporal visual field, and  $50^\circ$  to  $60^\circ$  from the meridian to the nasal, known as the nasal visual field. The vertical visual field extends  $50^\circ$  to  $60^\circ$  in both directions from the horizontal meridian. Overlap in the visual field occurs in the nasal visual field, allowing true binocular depth perception and giving an overall horizontal visual field of up to  $180^\circ$ . Other sources, such as Howard et al., have suggested that the FoV may be as wide as  $190^\circ$ , and some works have stated that when allowing the

movement of the eye within a stationary head, the visual field increases up to  $290^\circ$  [52]. It should be noted that the FoV of the human eye is known to decrease with increasing age, so numbers have been taken with caution, as age is often not reported.

**Acuity:** Within this visual field, the acuity of the human vision is often measured in *cycles per degree* (cpd), which is the number of cycles (a black circle with a white space around it) that can be distinguished within one degree of vision. Using a standard Snellen chart to assess visual acuity, a result of 20/20 (considered normal vision) is equivalent to 30 cpd [177]. However, the visual acuity of the HVS is largely dependent on two factors: eccentricity and environment luminance. Under low lighting where scotopic vision is used, cpd can fall to as low as 2, while under high luminance and photopic vision, it can reach 50–60 cpd [33, 166], and cpd falls off in an approximately linear fashion as eccentricity increases [120].

**Resolvable resolution:** The combination of the visual field and acuity gives indicates the required pixel resolution. To cover the visual field with static eyes in a singular display and have 20/20 vision (which is equivalent to 30 cpd or 1 arc minute [177]), a resolution of  $11400 \times 7200$  is required, which equates to 82 megapixels, although we assume a rectangular FoV, which we know is false. If we expand this to the full 60 cpd approximated as the maximum under ideal conditions, these values increase to  $22800 \times 14400$  and 328.3 megapixels. As the OST-HMD typically presents one image to eye, the requirements increase further, each eye requiring  $18000 \times 14400$  and 259.2 megapixels, giving 518.4 megapixels in total before considering motion; at 30 cpd, this total megapixel value falls to 129.6. Notably, these values are approximations that indicate what the equivalent pixel resolution to the human eye would be. To resolve this resolution, the human eye performs rapid movements to focus image areas in the fovea. These rapid movements are called saccades and help scan the environment. Saccades affect perceptions of scenes (e.g. saccadic masking), but a full exploration of this topic transcends the scope of this work.

**Latency:** VR studies have shown a limit to the latency which humans can recognise. Jerald and Whitton, for instance, suggested that immersive VR HMDs require a latency of 5 ms or less to be unnoticeable for the most sensitive subjects [68]. These findings confirmed earlier work from psycho-physical studies that found noticeable latency to be lower than 17 ms [69]. Other work on interface interaction by Jota et al. showed that noticeable latency was 20 s to 100 ms, with 85% of participants unable to perceive latency below 40 ms when finger tapping (e.g. tapping-based input, such as tapping on touchscreens), but according to Ng et al., user performance was affected by a latency as low as 2.38 ms with a mean of 6.04 ms [129].

**Critical flicker frequency (CFF):** CFF is the fastest rate at which a flicker is perceived and not viewed as a stable image. This value varies with the adaptation level of the retina, size of the stimulus, luminance and wavelength. It is not constant across the visual field, increasing up to  $55^\circ$  eccentricity before decreasing again toward the periphery, with a measured maximum value of approximately 90 Hz [120]. Under extreme cases and using specific patterns, however, people have perceived flickers up to 500 Hz [26].

**Depth perception:** The ability of humans to perceive the positions and sizes of objects is determined by several factors, including accommodation and occlusion. For the human eye to view objects at various distances in focus, the human eye has developed accommodation mechanisms. The shape of the lens can be adjusted by minute ciliary muscles in the eye, changing the curvature of the lens and subsequently moving the focal points. This allows distal stimuli from different distances to be focused on the retina. This is supported by vergence—the movement by both eyes to allow binocular vision—which is tightly coupled with eye accommodation. This works by triangulating the positions of objects in each eye with a fixed distance between the eyes, known as the interpupillary distance (IPD), which varies between people and thus needs to be adjustable. Occlusions comprise more important information in the HVS eye, determining whether objects lie

Table 1. Requirements for an OST-HMD to match the HVS. We also present the equivalent values as they have been reported for the Hololens 2. This represents a recent, commonly known commercial application of OST-HMD and gives an idea of the current state of publicly available devices. The values used were reported by Wired.com<sup>a</sup> and KGuttag.com,<sup>b</sup> and we used reported diagonal FoV and pixels per degree to calculate the combined FoV and resolution.

Requirement	To equal HVS	Hololens 2
Combined Static FoV	180–190° × 100–120° [52, 159]	43° × 29°
Per Eye Static FoV	140–150° × 100–120° [159]	
~Combined Resolution	11400 × 7200 @ 30 cpd 22800 × 14400 @ 60 cpd	calculated as ~2021 × 1363
~Per Eye Resolution	9000 × 7200 @ 30 cpd 18000 × 14400 @ 60 cpd	2k <sup>a</sup>
Static Dynamic Range	3.7 log units (cd/m <sup>2</sup> ) [94]	500 cd/m <sup>2</sup> <sup>b</sup>
Overall Dynamic Range	10 <sup>-6</sup> cd/m <sup>2</sup> - 10 <sup>8</sup> cd/m <sup>2</sup> [94]	
Latency	<2.38ms [129]	display latency 16.66ms (60 FPS)

<sup>a</sup><https://docs.microsoft.com/en-us/windows/mixed-reality/develop/platform-capabilities-and-apis/rendering>

<sup>b</sup><https://kguttag.com/2019/02/27/hololens-2-first-impressions-good-ergonomics-but-the-lbs-resolution-math-fails/>

in front of one another and how large objects may be compared to familiar objects in view. As we later show, the ability to create virtual occlusions (virtual objects occluding real-world objects) is a major challenge for OST-HMDs but is important for visual perception and scene understanding.

**Other:** Other visual phenomena must be eliminated as well to create visual coherence. Scattering occurs when light is randomly deflected in random directions, causing vision degradation due to image blur [177]. Aberrations, of which there are many subcategories, occur when light incidents prevent image formation and otherwise limit human vision.

### 3.3 Summary and relevance for OST-HMD design

In summary, compared to state-of-the-art OST-HMDs, the key properties of the HVS as summarised in Table 1 far exceed the properties of most OST-HMDs (e.g. Microsoft's Hololens) and even most normal displays. Notably, the table does not report on other properties (e.g. accommodation and sensitivity of geometric distortions) that must be considered. All this provides many challenging requirements for an OST-HMD. Some of these requirements are not strictly necessary to meet, however, as demonstrated by foveated rendering discussed in Sec. 6.5 and dynamic range, as the provided values are extremes that we can perceive but to which we are not constantly exposed in our everyday lives. While these requirements remain challenging, research on individual aspects of these requirements has been done. In the following, we present these existing approaches, grouping them into characteristics and issues: *spatial* (optimising spatial registration and reducing geometric distortions and aberrations), *temporal* (addressing latency and flicker), and *contentual/visual* (addressing colour space, occlusion capability, accommodation issues, resolution and dynamic range).

## 4 MAINTAINING SPATIAL REALISM

This section covers the challenges of maintaining spatial realism between virtual images and the real world.

## Spatial

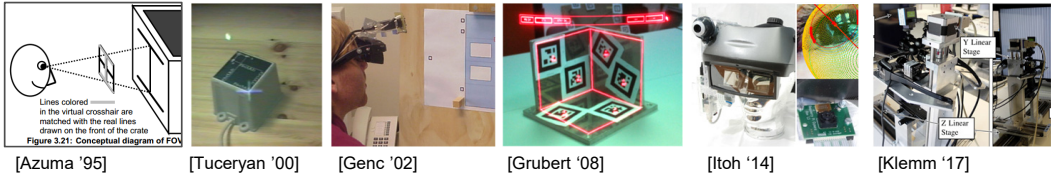


Fig. 5. Major works related to the spatial realism covered in Sec. 4.

### 4.1 Spatial consistency through spatial calibration

In any realistic AR application, virtual objects must appear as if they are registered in real space and maintain the correct perspective in the user's FoV against the real world (Fig. 5). This registration problem relates to two spatial relationships: the world to the OST-HMD and the OST-HMD to the eyes.

The former (the world to the HMD) is a well-established problem and is usually referred to as motion or position tracking. Popular solutions use vision-based tracking with one or several cameras facing the environment (*inside-out tracking*). Researchers once commonly used fiducial markers (e.g. [72, 165]) but increasingly rely on natural feature-based approaches, such as *simultaneous localisation and mapping* (SLAM) [83, 144]. However, the issue of tracking the spatial relationship between the world and the AR display (e.g. the HMD) is not specific to OST-HMDs but is common to nearly all AR techniques, and it is well covered (e.g. [13, 65]).

Consequently, we focus in this section on the spatial relationship of the OST-HMD to the eyes and, more specifically, on how to model a projection between the screen and the eyes. A basic approach is to only consider the IPD. While important for providing correct stereo vision, such as that used in VST-HMDs, it is insufficient when calibrating the eye-display relationship in OST-HMDs. In their recent survey of calibration methods, Grubert et al. introduced three categories of techniques for computing the spatial eye-display relationship [44] (Fig. 6): manual methods, semi-automated methods and automated methods. All these categories treat the eye-HMD system as an off-axis pinhole camera model in which a nodal point of the eyeball is the camera centre and the virtual image on the OST-HMD is the image plane [5] (Fig. 6 top left). Under this assumption, the calibration problem turns on estimating the 3D-to-2D projection matrix (Fig 7), and a calibration procedure becomes a data collection of pairs of corresponding 2D image points and 3D world points that match in the FoV of the user. The difficulty of the calibration lies in this data collection, as we cannot directly fetch the user's view from their eyes, unlike the calibration of eye-perspective cameras.

The manual methods rely on any sort of user input for the data collection. The most classic yet still-practical solution is the single-point active alignment method (SPAAM, Fig. 6 top right) [163], in which users align a 2D crosshair to a tracked 3D point (e.g. a fiducial marker) and notify the system. Repeating this procedure six times provides six pairs of 2D-3D correspondences, which lead to 12 linear equations. Solving the equations via direct linear transforms provides a projection matrix. For better estimation accuracy, in practice, we collect more pairs to reduce the estimation error.

In contrast to manual methods, semi-automatic methods separate the calibration models to reduce the number of parameters needed for the online calibration by users. Owen et al. were among the first to propose explicitly separating the eye-HMD calibration into the eye and the display parts, where the eye part is calibrated online by the user and the display part offline without



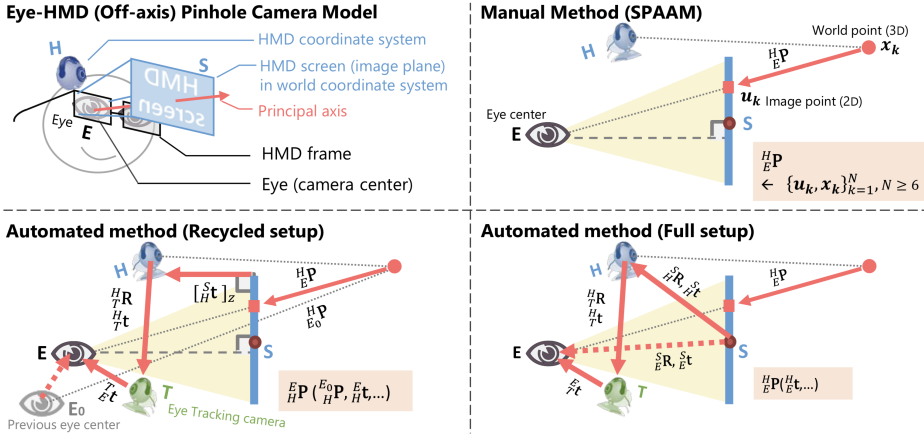


Fig. 6. General overview of calibration techniques for calibrating the spatial relationship between the eye and display [44].

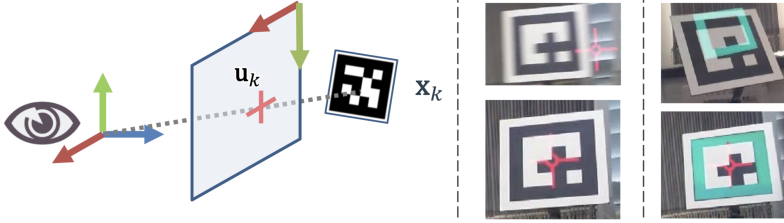


Fig. 7. Single-point active alignment method (SPAAM) [163] is probably the best-known approach to spatially calibrating OST-HMDs. (Left) Instead of a fully automatic calibration, SPAAM requires users to align 2D points displayed on the OST-HMD with known 3D points in the real world to compute the projection matrix describing the eye-display relationship. (Right) Visual overlay before and after calibration [44].

users [133]. Similarly, Itoh and Klinker [60] proposed interaction-free display calibration (INDICA), a fully automated calibration method for OST-HMDs with 3D eye localisation (Fig. 6 bottom row). Their method splits calibration phases into an offline step for the display and an online step for the current eye position. Their approach is fully automated, removing errors that normally arise from having the human in the loop. More importantly, having the user in the loop forces users to re-do the calibration every time they wear OST-HMDs, which is unrealistic for daily use.

**Distortion Correction.** One challenge often ignored in research on OST-HMDs that is important for achieving an experience indistinguishable from reality is the visual distortions caused by imperfections in the display and optics. This is like the optical system in cameras, which creates lines that are not straight or other effects (e.g. chromatic aberrations, out-of-focus image areas, etc.). However, the effect occurs in the image displayed to the user of the OST-HMD (Fig. 8).

Almost all distortions are highly view-dependent. However, unlike a camera with a fixed image plane, the user's eye can have various positions within the eye box given by the OST-HMD. One challenge is thus measuring and correcting the geometric distortions for all possible positions. Another oft-ignored fact is that the image plane where a displayed image appears is not actually a plane but a curved surface (Fig. 8) because of the optical elements used within the display [85, 133]. Researchers have proposed calibrating the optical distortion like distortions in cameras

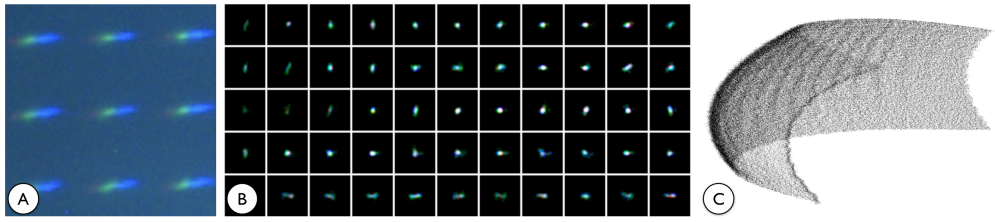


Fig. 8. Examples of points rendered to an OST-HMD (here, a Lumus DK-32) and the visual results as captured with a user-perspective camera. (A) A close inspection reveals a large chromatic aberration and image blur in the horizontal direction. (B) The distortions strongly depend on the position within the eyebox, as can be seen here for various patches from different 2D positions after noise filtering [63]. (C) The image plane of the OST-HMDs as measured and illustrated by Klemm et al. shows that it is a curved surface, not a plane [85]

are calibrated: by parametrically modelling the distortions and creating per-pixel calibrations, which can be automated using user-perspective cameras [85]. Initially, researchers proposed two-step approaches that first computed display parameters using a few point samples using a user-perspective camera and then calibrating the system for the individual position of the user's eye by considering the display parameters obtained in the first step [133]. Itoh et al. looked into measuring and correcting the view-dependent shift and blur in OST-HMDs [57, 61], highlighting the amount of chromatic aberrations and directional blur and again mentioning the strong view dependence (Fig. 8). They presented an approach to storing the aberrations as a *point spread function* (PSF) light field to model the view-dependent character, and they estimated the PSF for other unsampled positions using a few sample positions. Some studies moved away from approximating the display characteristics using a few point samples to calibrate every pixel individually [85]. Langlotz et al., for example, implicitly corrected the optical distortions of an OST-HMD, as their approach used per-pixel mapping from a world camera to the display [97].

Despite this work, it is fair to say that display distortions are an issue that is rarely addressed. They require careful calibration of the overall display characteristics. Maintaining small deviations within the production of the OST-HMD would allow a generic calibration, as in camera lenses, and we still require information about the user's eye position and gaze to correct for the individual user, which adds to the problem.

## 5 MAINTAINING TEMPORAL REALISM

As outlined in the previous section, we can achieve spatial realism for OST-HMDs by applying a spatial calibration and correcting for the distortion. However, as soon as we start moving the HMD, temporal realism comes into play, and most notably, the problems of image update rates and general latency (system and display latency) challenge the management of a temporally coherent augmentation of the environment (Fig. 9).

### 5.1 General system latency

Latency reduction for HMDs has already been investigated for immersive VR HMDs. Latency in VR can contribute to motion sickness and is consequently investigated often [21]. Cobb et al. identified that besides tracking inaccuracies, overall system latency is a source of simulator/motion sickness. Overall system latency thereby usually comprises 1) the tracking system delay (the time from a change of the physical pose and its measurement in the system), 2) the application delay (which can be attributed to general execution time), 3) image generation delay (the time it takes to generate a new image), and 4) the display delay (the time required to display the image in the HMD) [122].

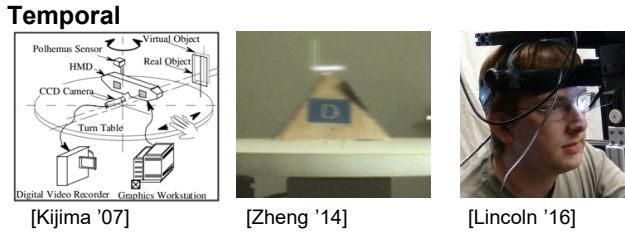


Fig. 9. Major works related to the temporal realism covered in Sec. 5.

Several works have investigated the maximum perceivable latency of an immersive VR system and have suggested that immersive VR HMDs require a latency of about 5 ms or less to be unnoticeable to most sensitive subjects [68]. These findings are consistent with other studies on noticeable latency, as outlined in Sec. 3.2.

In see-through AR systems, however, the latency requirement can be even more stringent dependent on the user's motion [113]. Because, unlike VR systems, where the latency mainly contributes to simulator/motion sickness, latency in AR also does to registration errors [109, 113], which are visible as misalignments between the virtual and the physical environment [6] and produce visual incoherence. In the work [6], Azuma categorised visible registration errors due to system latency when motion occurs as a dynamic error. Like immersive VR HMDs, latency can be caused by various components of an AR system (e.g. tracking, application processing time, rendering and display), so reducing the tracking delay can reduce overall latency. Itoh et al. [63] used this by converting a commercial immersive VR system with low-latency tracking (Oculus Rift) into an OST-HMD. Similarly, rendering techniques such as foveated rendering can decrease latency by prioritising image areas currently observed by the user (e.g. by only rendering them with a high resolution). We discuss this further when discussing the physical display resolution of OST-HMDs (see Sec. 6.5).

Buker et al. investigated motion sickness using a monoscopic OST-HMD and confirmed earlier studies in VR by showing that reducing latency also reduces motion sickness [12]. In their system, they reduced the latency by correcting the computed visual output using the measured system latency. Like Buker et al., the well-known game developer John Carmack presented several programmatic approaches to reducing latency [16] within a VR system which can easily be applied to AR and OST-HMDs, such as time warping, which predicts motion during rendering and applies it in late rendering stages. He also mentioned the benefit of directly accessing the display matrix, something we address below when discussing display latency.

## 5.2 Flicker and display latency

Within general system latency, the display itself and how individual pixels can be written to it can be a significant source of latency. As we discussed earlier, 90 Hz is often recommended as a display update frequency, but depending on the displayed patterns and eye movements, higher frequencies can affect flicker fusion (sequential renderings are perceived as two distinct images), as mentioned above [26]. Consequently, VR displays tend to aim for display update frequencies beyond 90 Hz, with examples such as the Valve Index being capable of 144 Hz in experimental mode. A further reduction depends mainly on the display panel, as some technologies require pixels to switch to an off or neutral state, while others can switch colours directly [16].

Research has also addressed the issue of latency, particularly display latency, in OST-HMDs. Zheng et al., for instance, focused entirely on reducing display latency [182]. Their idea was to use a digital micromirror device (DMD) imaging chip for image creation like those in Digital Light

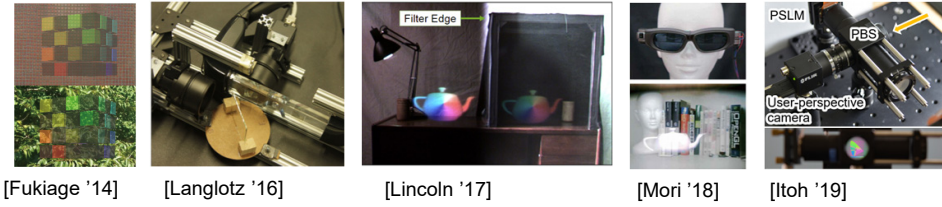
**Color & HDR**

Fig. 10. Major works related to the colour reproduction covered in Sec. 6.1.

Processing (DLP) Projectors to allow the direct control of individual pixels (only on or off) with an update rate of  $50 \mu\text{s}$ . Instead of displaying a full image, the projector projects a new binary pattern as fast as the DMD can change its state. The displayed binary pattern decreases the difference between the image that the user perceives and the desired image given the latest tracking data, so it reduces the error between the desired image and the “is” (representing the current state computed using a sliding window over the last 64 frames, which represent the integration interval). The prototype is only greyscale, uses precomputed data and is specifically a low-latency tracker. Unfortunately, the sliding-window approach can also introduce noticeable flicker. Lincoln et al. later introduced a head-tracked setup using a “zero latency” mechanical tracker and proposed a new coding scheme that avoided the flicker [109]. Instead of computing the binary pattern based on the difference between the “is” and the desired image, the authors proposed a statistical model that chose the current binary output based on a probability based on the desired grey value to be displayed. Using this approach when displaying 50% grey, the projectors distribute the “0” and “1” evenly over the 64 frames, which reduces the perceived flicker even though the displayed image is still only greyscale. The authors also extended the idea to support high dynamic-range (HDR) images while maintaining low latency [108].

We conclude that latency is an issue when creating a coherent visual representation. However, research has only started to explore this problem, particularly considering the overall requirement of staying at or below 5 ms, which can only be demonstrated in research prototypes. Unfortunately, miniaturising these prototypes and the practicality of “mechanical zero latency trackers” is questionable. While we can expect improvements in tracking and rendering speed through generally faster CPUs and GPUs and techniques such as foveated rendering, limitations remain, such as display technologies with sufficient update rates.

## 6 MAINTAINING VISUAL REALISM

In our discussion of visual realism, we leave out issues common to the whole field of AR, including coherent lighting (e.g. light estimation [43]) and realistic rendering of the AR scene (e.g. [84]). Instead, we focus on the topics of colour reproduction, dynamic range, creating occlusions, support depth accommodation and wide FoV, which require solutions unique to OST-HMDs (Fig. 10 and Table 2).

### 6.1 Colour reproduction

Colour is critical to achieving visual realism. For example, if we assume two red apples on a table, one of which is a real physical apple and the other a virtual one seen through an OST-HMD, it should be impossible to distinguish between them by means of visual comparison; we would expect the virtual red apple to have the same colour and appearance as the real apple (visual coherence). For video see-through displays, the blending of the real and virtual happens in the image space

Table 2. Different colour modification methods and approaches explored with OST-HMDs.

Reference	Goal	Method	Software/Hardware	User Study	Portable Prototype
Itoh et al. [58]	Color reproduction	parametrised colour response model	software	no	no
Sridharan et al. [158]	Colour blending	colour transmission database	software	no	no
Hincapié-Ramos et al. [25]	Color blending	color transmission database	software	no	no
Fukiage et al. [35]	Color blending	content visibility model	software	no	simulation
Weiland et al. [167]	Color blending	user adaptation modelling	software	no	no
Langlotz et al. [97]	Color blending	radiometric compensation	software	yes	yes
Wetzstein et al. [168]	View Modification	light filtering via SLM	hardware	yes	yes
Itoh et al. [62]	View Modification	light filtering via SLM	hardware	no	no
Kaminokado et al. [71]	View Modification	light filtering via SLM	hardware	no	no
Mori et al. [124]	Brightness adjustment	adjustable liquid crystal shutter	hardware	yes	yes
Xu and Hua [174]	HDR imaging	two SLMs	hardware	no	no
Zhao et al. [181]	HDR imaging	two SLMs	hardware	no	no
Lincoln et al. [108]	HDR imaging	high speed LED with custom high speed controllers	hardware	no	no
Itoh et al. [64]	HDR imaging	HDR projector with dihedral corner reflection array	hardware	no	no

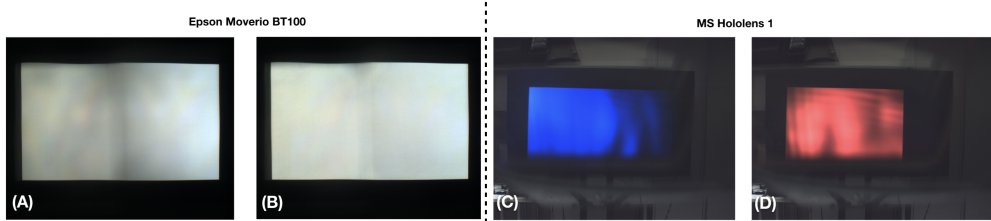


Fig. 11. Examples of colour reproduction on different OST-HMDs when captured with a camera placed at the eye position. (A) Epson Moverio BT100 when displaying uniform white; changes in intensity caused by the display optics are clearly visible. (B) The same display when calibrated for uniformity [97]. (C) and (D) show examples of colour artefacts when displaying single colours on a Microsoft Hololens 1, demonstrating that these effects are also colour-dependent.

within the video image. We can achieve visual coherence by mimicking the characteristics of the camera used to capture the scene in the rendering step [84]. In spatial AR, this blending is more challenging and requires a careful radiometric calibration of projectors [8–11] to consider the background or colour response of the projectors. When using OST-HMDs, several factors come into play that affect the visual appearance of the virtual image, particularly colour. One factor is the colour response of the display used. When sending a specific colour to the display, we assume it reproduces exactly this colour. We know from other displays, however, that this is usually not the case; examples of this in Fig. 11 show the non-uniform intensity and colour response when showing solid colours or intensities in front of a black background. This is caused by the specific colour range we can display (e.g. full RGB) but also by individual discrepancies within normally uncalibrated displays, which are not only caused by the actual display (the light-emitting element or the colour filter) but, in OST-HMDs, also often by the optics that are part of the display.

Similarly, OST-HMDs blend together the real and virtual using an optical combiner (Fig. 3), so the visible colour is always a result of this blending affecting correct colour reproduction and consequently the perceived visual realism. Finally, integrated displays are usually unable to produce intensities with a dynamic range, as in physical work. In the following, we give a more detailed overview of these specific issues and present proposed solutions.





Fig. 12. Incorrect colour reproduction in OST-HMDs. (Left) A physical scene captured by a user-perspective camera representing the intended target colours. (Middle) Rendering the captured image with an OST-HMD in front of a fully black background produces a different result. (Right) The result of the correction method proposed by Itoh et al. produces a result that is more consistent with the captured physical appearance [58]

*Display colour reproduction.* Common to most displays, not just OST-HMDs, is that colour reproduction is not accurate: when sending a certain colour to the display, the displayed colour will not match the intended colour exactly. Itoh et al. demonstrated that even when ignoring the background (e.g. by having a uniform black background), the colour of a virtual object does not match the real environment [58] (Fig. 12). They identified two main reasons for this effect: a) light absorption by the imperfect optical medium of the OST-HMD and b) a lossy digital-analogue colour conversion.

Sridharan et al. proposed colour profiles based on look-up tables (LUTs) to correct colour rendering in OST-HMDs. In addition to the work by Itoh et al., they proposed a method for creating profiles for multiple background colours and showed that this does improve the perceived colours in an OST-HMD [158]. However, their approach usually assumes a coherent background and cannot compensate for complex backgrounds—nor do they say how the system knows about the background.

Similarly, Itoh et al. proposed applying a semi-parametric approach to compute a more coherent colour display profile like a display colour calibration on standard screens to improve the perceived colour and support visual coherence. Again, all these works assumed a black or a uniformly coloured background, as their focus was the colour characteristics of the used display, not the colour blending with the background caused by the see-through character of the display.

*Colour blending.* In OST-HMDs, the real and virtual blend together using an optical system—usually, a beam combiner. Regardless of the optical design used, the effect of this blending is that the displayed image (e.g. virtual object) blend with the physical world based on the characteristics of the used optical combiner. The result is that depending on the physical environment, we perceive visual artefacts (Fig. 13 (A)–(D)). Another side effect of this is that we cannot make the environment darker using standard OST-HMDs. Many commercial solutions work around this problem by adding shades to their HMDs to reduce the brightness of the background (e.g. the Microsoft Hololens). Some research and actual products even proposed shades that can change their global opacity and thus better accommodate different environmental conditions [124]. Depending on the characteristics of the optical combiner and the opacity of the used shades, the background will always add a certain intensity to each pixel (e.g. in the case of a combiner that is 50% transparent and 50% reflective without shades used, the background would always contribute 50% to the resulting colour and intensity).

This general problem of colour blending in OST-HMDs was first described by Gabbard et al. [37], who gave an overview of the problem and empirically analysed the effect with different backgrounds. While their results were not evaluated with actual users, their findings suggested problems in the usability of OST-HMDs due to the shift in colour intensities and hues, which

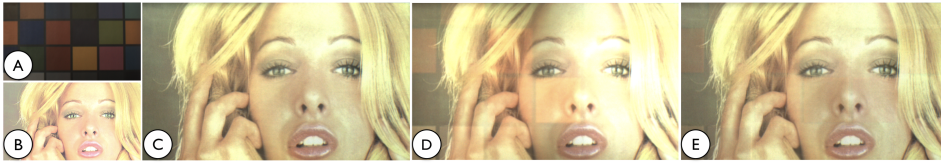


Fig. 13. Examples of colour blending in OST-HMDs resulting in the decreased quality of the displayed image. (A) A coloured background with the OST-HMD switched off, (B) the test image to be displayed, (C) the test image displayed in an OST-HMD with a black background, (D) the test image displayed in front of the coloured background showing the effect of colour blending, which depends on the optical combiner, (E) radiometrical compensation reducing the effect of the background and colour response of the display [97].

produce legibility issues and constrain the design of user-interface elements, as only a few colours can easily be discriminated under various backgrounds.

Several research groups have since looked into addressing the problem of colour blending. A theoretical solution was presented in the Smart Colour system [25] that extended earlier work by Sridharan et al. [158] and Fukiage et al. [35]. They both presented approaches simulating and controlling colour blending per pixel in real time. They succeeded in this goal but only demonstrated in a simulated environment: no camera was used to collect environmental information nor was an OST-HMD used to display results. Furthermore, the fundamental problem of how to capture the pixel-precise colour information of an environment and align it with the display, which is needed for actual applications, remains unanswered. For this reason, it is impossible to determine the extent to which their method corrects blending errors on real hardware.

Weiland et al. [167] presented the first approach that tried to neutralise the background by integrating a camera capturing the environment. The captured camera feed is then used to compute a compensation image to reduce the effect of colour blending, but it only works for a static, calibrated system, as they use a camera calibrated for one calibrated position.

Langlotz et al. also proposed neutralising the environmental light when blended with the display image by measuring the environment and modifying the displayed image accordingly [97]. In contrast to earlier work by Weiland et al. [167], they also discussed a practical approach for integrating the camera into the OST-HMD and calibrating the overall system. They demonstrated that they could map each pixel from the camera-image space to the display space, even for arbitrary environments, allowing pixel-precise compensation. The work by Langlotz et al. also introduced more complex compensation algorithms that consider the display and camera characteristics and minimise clipping artefacts, creating an overall more visually coherent result (Fig. 13 (E)).

All these approaches assume a constant semi-transparent display. However, if one can change the transparency, one can also reduce the effect of background light. To successfully employ this solution, one must be able to switch the transparency on a pixel level. If a pixel is used to display a virtual object, it should be opaque to block out the environmental light, but if the pixel is not used for displaying virtual objects, it should be transparent to allow seeing the environment. There are several approaches that allow blocking the environment on a per-pixel base (e.g. [14, 81]), but they have not been introduced in the context of reducing colour-blending artefacts. We introduce them in the next chapter when discussing occlusion capability—for now, it is important to remember that they can improve colour reproduction and allow the blocking of environmental light.

One can not only control the transparency of each pixel but also filter the incoming environmental light on a per-pixel level. Wetzstein et al. [168], Itoh et al. [62], and Kaminokado et al. [71] used this in their prototypes. While Wetzstein et al. used LCD panels and optical lenses to modulate scene light (e.g. to change the saliency of the environment), Itoh et al. and Kaminokado et al. used

Table 3. Properties of different display designs that present virtual content at different focal distances. Moderate FoV equals 40–80° and moderate optical complexity requires 3–4 optical elements.

Reference	Mechanism	Classification	See-Through	Field of View	Resolution	Eye-tracking Needed	Optical Complexity	Form Factor	Computational Demand	Miniature Prototype
Lanman & Luebke [101]	microlenses	light-field	no	33.3°	low	yes	simple	thin	moderate	yes
Yamaguchi & Takaki [176]	microlenses	light-field	yes	4.3°	low	yes	simple	thin	moderate	no
Otao et al. [132]	microlenses	light-field	no	narrow	low	yes	simple	bulky	moderate	yes
Maimone et al. [116]	pinlight display	light-field	yes	110°	low	recommended	simple	thin	moderate	yes
Akşit et al. [2]	pinhole display	light-field	no	83°	low	recommended	simple	thin	moderate	yes
Song et al. [155]	pinhole display	light-field	yes	narrow	low	recommended	simple	thin	moderate	yes
Jang et al. [66]	laser w/ steering mirror	holographic	yes	68°	moderate	yes	complex	thin	high	yes
Maimone et al. [115]	SLM	holographic	yes	80° hori.	high	yes	complex	thin	moderate	yes
Shi et al. [154]	SLM	holographic	no	N/A	high	yes	complex	moderate	high	no
Moon et al. [123]	SLM	holographic	yes	narrow	N/A	yes	complex	moderate	high	yes
Gao et al. [41]	amplitude SLM	holographic	yes	1.7°	high	no	complex	bulky	low	no
Gao et al. [40]	SLM	holographic	yes	5.4°	high	no	complex	moderate	low	yes
Huang et al. [56]	stacked LCD panels	multi-plane	no	125°	high	no	simple	bulky	high	yes
Lee et al. [103]	savart plate	multi-plane	yes	37°	high	no	moderate	moderate	low	yes
Yoo et al. [180]	polarised lenses	multi-plane	yes	7.5° hori.	high	no	moderate	moderate	low	no
Liu et al. [112]	polymer stabilised scattering shutters	multi-plane	yes	narrow	N/A	no	simple	moderate	low	no
Maimone & Fuchs [114]	stacked LCD panels	multi-plane	yes	65°	low	no	simple	moderate	high	yes
Rolland et al. [146]	stacked planar displays	multi-plane	yes	N/A	high	no	simple	moderate	low	no
Liu et al. [111]	focus tunable lens	varifocal	yes	28°	moderate	no	complex	moderate	low	no
Rathinavel et al. [139]	focus tunable lens	varifocal	yes	15.3°	moderate	no	complex	bulky	low	yes
Xia et al. [172]	focus tunable lens	varifocal	yes	37°	moderate	no	complex	bulky	low	no
Dunn et al. [30, 31]	focus tunable membrane	varifocal	yes	103°	moderate	yes	simple	bulky	moderate	yes
Wilson and Hua [171]	alvarez lens	varifocal	yes	103°	moderate	no	complex	moderate	moderate	no
Matsuda et al. [119]	SLM	focal surface	no	18°	high	yes	complex	bulky	high	yes

phase-only spatial light modulators (SLMs) to implement a display. Ideally, one would combine an additive display (current OST-HMDs) with a subtractive approach that allowed almost full control of the incoming environmental light and of the displayed image, thus also solving the issue of colour blending. However, to the best of our knowledge, this has not been achieved.

*High dynamic range.* The human eye is capable of perceiving a large range of intensities (dynamic range). While research often addresses extending the dynamic range of normal displays (e.g. [153]), producing near-eye displays as used in AR and VR is challenging because precisely aligning multiple modulation planes (a popular concept to produce HDR displays) inaccuracies in the alignment and calibration are magnified [174, 181]. In particular, when looking at OST-HMDs, including commercial and research prototypes, not much work has focused on increasing dynamic range. One of the few studies that have specifically targeted the issue of dynamic range is the work by Lincoln et al. [108], which supported displaying high dynamic range images by utilising a DMD, such as that used in DLP projectors. Being able to control the timing of each mirror exactly extends the dynamic range of the image that can be displayed. In their work, they achieved 16-bit intensity control per colour. However, their setup is still relatively large and would require extensive miniaturisation.

This is also true for another prototype aiming for a high dynamic range in near-eye displays [64]: Itoh et al. used a high-contrast projector for retinal projection via a dihedral corner reflector array as an optical element and a neutral-density (ND) filter to reduce the amount of light to safe limits. While retinal displays are occasionally used to implement OST-HMDs, the developed prototype is not see-through and thus has not been demonstrated for AR.

Overall, increasing the dynamic range of near-eye displays and, more specifically, OST-HMDs is a topic not often addressed in research despite its relevance when considering human perception and the overall aim of a convincing integration of the real and virtual.

## 6.2 Accommodation capability

Accommodation is a physiological depth cue that occurs when human eyes focus on objects at different depths by changing the curvature of the lens. The cue plays an important role in humans' natural 3D perception [24]. A problem with conventional OST-HMD design is that the display screen

### Accommodation

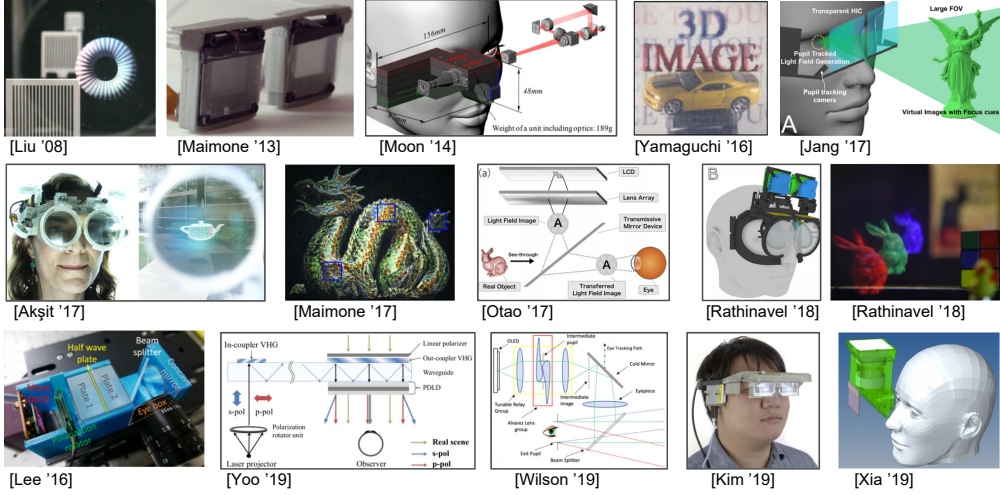


Fig. 14. Some major works related to the accommodation capability are covered in Sec. 6.2.

appears at a fixed focus distance. While we pointed out earlier that this display screen is actually not planar, as often stated in the literature, because individual pixels appear at different distances, these distances are still fixed. This fixed focus distance is commonly somewhere between 3–7 m and is determined by the optics and hardware because the optics merely expand images from the microdisplay panels used to display the image. As a result, users of OST-HMDs have various issues in depth sensation. First, there is an accommodation mismatch between the physical and virtual objects, and virtual objects' vergence and accommodation are not coherent. This controversial sensation, known as the vergence–accommodation mismatch or conflict [29, 51, 89, 96], affects users' 3D perceptions and is a major source of VR sickness.

A typical example of this issue in OST-HMDs is the following: Imagine we want to render a virtual object (e.g. an apple) in our open hand. While the hand appears roughly within arm's reach, for the apple to be in focus, our eyes must focus on a virtual image plane several metres away. Consequently, as we cannot have our hand and the apple in focus, the illusion of holding the apple in our hand is destroyed and the scene feels visually incoherent.

The solution to the conflict is to render virtual objects at correct depths so that the light rays from the object appear as if they are coming from the proper distances. This requirement essentially requires additional optics or different display technologies in OST-HMDs [54]. We categorise existing solutions into three main categories: adjustable-focus OST-HMDs, multifocal, varifocal OST-HMDs and OST-HMDs utilising light fields or holographic displays. Common to almost all these approaches is that they not only require different hardware but also different rendering algorithms. For example, one needs specific rendering steps for multifocal displays to compute the image area display on each focal plane [128]. Similarly, one needs to compute phase images for holographic near-eye displays [175]. All this needs to happen in real time, which is still challenging. Xiao et al. [173] presented a general overview of creating imagery for accommodation-supporting HMDs independent of hardware. Table 3 and Fig. 14 give overview of major related works.

*Adjustable-focus OST-HMDs.* A relatively basic approach is to allow the focus plane to be adjusted. This could work manually but also automatically. While this does not fully mitigate the effect, it would allow adjusting the focal plane based on context (e.g. adjusted for arm's reach or the distance

of a table surface) or based on a measured vergence depth. Willson and Hua proposed combining a free-form Alvarez lens pair that can smoothly tune its focal depth from 0–3 diopters [171].

*Multifocal or Varifocal OST-HMDs.* In contrast to all these approaches, some OST-HMDs have several focus planes (multifocal [148]) or can mimic continuous focus planes in a time-multiplex manner (varifocal). This allows content to be displayed on different focus planes and thus reduces the vergence–accommodation conflict. However, one downside of both approaches is that the number of focal planes is limited, so they cannot provide all focus cues but only approximate them by showing objects at the focal plane closest to their real position. Furthermore, multifocal approaches tend to introduce perceivable flicker caused by switching.

Rendering on different focus planes has been attempted outside HMDs. For example, Akeley et al. proposed a desktop setup of something that could be seen as a giant HMD prototype that produced three focus planes using several stacked half-silvered mirrors. Each reflected a certain part of a screen onto the user’s view [3].

As a multifocal OST-HMD, Liu et al. demonstrated a benchtop multi-plane OST-HMD by projecting an image onto a stack of liquid crystal layers that could switch scattering properties in a time sequence [112]. Lee et al. combined a fast polarisation rotator and Savart plate so that an image from a microdisplay could be formed at two focal distances [103]. Yoo et al. instead used a geometric phase holographic lens as a dual-focal waveguide [180]. With the Magic Leap One, the first commercial OST-HMD with a multifocal display is available but limited to two focal planes that are switched based on the data in the integrated eye tracker.<sup>1</sup>

The focal surface display is another approach to generalising the multifocal approach [119]. Although designed for VR displays, their method utilises a phase-only SLM, which acts as a dynamic free-form lens. The system can display multiple focal plane images as continuous focal surfaces.

Like multifocal systems, the first approaches using varifocal lenses to approximate several focus planes were demonstrated using desktop prototypes. For example, Suyama et al. showed a prototype using a liquid crystal varifocal lens that was synchronised with a 2D display at 60 Hz and gave the illusion of different focus planes (ranging from -1.2 to +1.5 diopters) [161]. Much research has focused on the investigation of different varifocal lenses. For example, Liu et al. used a liquid lens as a varifocal element (ranging from -5 to +20 diopters) [111]. They demonstrated a monoscopic see-through display prototype, but the overall performance was limited by the time it took to change the optical power. In their example, they could only demonstrate a dual focus–plane display (two depth levels) with a 7 fps update rate, although they showed the potential to approach 56 fps with different varifocal lenses. Still, this is only for a dual-focus display; more focus planes reduce the performance. Another drawback of the liquid-lens approach is that it is hard to extend the FoV of the display due to the limited lens aperture.

Xia et al. combined a shutter glass with a focus-tunable lens to realise an OST-HMD switchable to VR mode [172]. The system also utilised the user’s corrective prescription to compensate for the focus error by the tunable lens. Rathinavel et al. further investigated the varifocal design with a focus-tunable lens and proposed a volumetric OST-HMD that could render 280 perceptually simultaneous depth planes [139].

Konrad et al. proposed the unique idea of cancelling the accommodation effect with a focus-tunable system so that depth perception was only triggered by vergence information [86]. Their system tracked the user’s accommodation and drove the focus-tunable lens to create a constant point spread function for the image input, significantly reducing the vergence–accommodation mismatch. As an alternative approach to employing refractive optical components, Dunn et al.

<sup>1</sup>/www.magicleap.com



introduced a deformable-membrane mirror as a beam combiner [30, 31], which realised both a large FoV and accommodation capability.

*Light-field and Holographic OST-HMDs.* Contrary to these aforementioned techniques, display techniques such as light-field displays [66, 101, 116, 118, 132] and holographic displays [17, 40, 41, 49, 123, 154] theoretically offer all depth cues simultaneously, so the user can focus on any depth at any time.

Both light-field displays and holographic displays replicate the optical 3D geometry of a target surface, yet their image-forming principles are completely different [102]. Light-field displays typically represent light scattering from the surface of a virtual object as 4D light rays. To create 4D light rays, two major approaches exist: a microlens array [155, 176] and stacked displays using transmissive LCDs [56, 114].

Holographic displays use SLMs to control the amplitude or phase of an input collimated laser beam so that the wavefront of the propagated light forms a desired image at the viewpoint [178, 179]. Maimone et al. proposed a holographic OST-HMD using a phase-only SLM with laser light sources [115]. They also demonstrated cancelling user astigmatism by taking its optical behaviour into account in the computation of an input image. To achieve a high-quality holographic image, optimising the source modulation image is critical but computationally expensive [19, 135]. As an attempt at real-time holographic rendering, Peng et al. proposed neural holography, a machine-learning method for computational holography which creates a 1080p colour image in real time [175].

Jang et al. proposed a light-field projection display [66] consisting of multiple laser light sources with a steering mirror. By tracking the pupil position with an eye-tracking camera, the system rendered 3D images at the pupil position. Cem et al. extended the FoV of a holographic display [17]. Kuo et al. proposed a unique approach to expand the *étendue* of holographic OST-HMDs by inserting a binary scattering mask [95].

*In-focus OST-HMDs.* This type of display has virtually infinite depth of focus. Typical approaches are retinal-scanning displays and virtual retinal displays [36] based on the Maxiwellian view, on which light rays from a point light source are relayed and focused at the pupil. Each ray thus directly hits the retina without being affected by the crystalline lens, so the user perceives an image which is always in focus regardless of the eye accommodation. A retinal-scanning display typically uses a laser projector on the Maxiwellian view to directly scan the retina with the input image [36]. Do et al. built a retinal-scanning display by combining a pinhole projector with retro-reflective film [28].

Some light-field displays offer an extended depth of field in some configurations leading to all-in-focus view at all depth planes for users. An example is the pinlight display by Maimone et al. that utilises near-eye pinlight projectors [116]. The key idea is inspired by pinhole cameras, which capture photos where everything is in focus thanks to the small aperture. Instead of one pinhole, however, the authors created a prototype consisting of many pinholes etched into acrylic glass, each pinhole representing a small aperture–defocused light source that could be used to project an image (displayed on an LCD panel) directly on the retinal plane appearing always in focus. This has been demonstrated for a non-see-through near-eye display as well [2], but the acrylic glass used by Maimone et al. was transparent while serving as a waveguide for the light leaving via the pinholes. While their approach allowed showing images that were always in focus, it needed to blur the images in the rendering step if out-of-focus effects were required for the displayed image. This has been demonstrated by works that tried to create out-of-focus effects in software or sharpen the image [23, 131, 151] to simulate different focus planes.

Finally, we highlight a specific problem when accommodating: refractive errors (e.g. different forms of short- or long-sightedness) affecting the use of OST-HMDs [18]. The authors pointed out

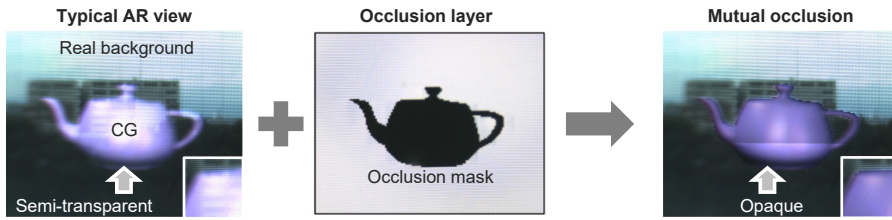


Fig. 15. Examples of the occlusion issue on OST-HMDs and mutual occlusion as a solution. The left figure is a typical AR rendering where a virtual teapot appears semi-transparent. By combining an occlusion layer (the middle figure) in the optical path between the display and the eye, we can achieve mutual occlusion where the virtual object appears opaque, blocking background light like a real object does (the right figure).

Table 4. Reported properties of different occlusion-capable HMDs adapted from [48]. Designs of moderate optical complexity use 3–4 optical elements (including the SLM), and a moderate FoV is 40–80°.

Reference	Occlusion Generation	Adaptive Depth	Varifocal Image	Mask Appearance	Computational Demand	Optical Complexity	Form Factor	Field of View	Portable Prototype
Kiyokawa et al. [81]	LCD	no	no	sharp	low	complex	bulky	25°	yes
Kiyokawa et al. [80]	LCD	no	no	sharp	low	complex	bulky	30°	yes
Wilson and Hua [170]	LCoS	no	no	sharp	low	complex	thin	30.58°	no
Gao et al. [38, 39]	LCoS	no	no	sharp	low	complex	thin	40°	no
Cakmakci et al. [15]	LCD	no	no	sharp	low	moderate	thin	40°	no
Santos et al. [15]	LCD	no	no	dull	low	N/A	thin	20°	yes
Maimone & Fuchs [114]	LCD	no	yes	dull	high	moderate	thin	65°	yes
Maimone et al. [116]	LCD	yes (in-focus)	yes	dull	moderate	simple	moderate	80°	yes
Itoh et al. [59]	LCD	partially	no	sharp	low	simple	thin	70–80°	no
Yamaguchi & Takaki [176]	LCD with Microlenses	yes	yes	sharp	high	moderate	thin	moderate	no
Uchida et al. [176]	DMD	no	no	sharp	low	complex	moderate	narrow	no
Kwangsoo et al. [176]	DMD	no	no	sharp	low	complex	moderate	narrow	no
Krajancich et al. [88]	DMD	yes	no	sharp	high	complex	moderate	8.7°	no
Hamasaki & Itoh [48]	LCD on linear stage	yes	no	sharp	low	moderate	moderate	narrow	no
Rathinavel et al. [140]	LCD with refocusable lens	yes	yes	sharp	moderate	complex	moderate	15.3°	no

that presbyopia (loss of near vision with age) is particularly critical for the adoption of OST-HMDs given the general age distribution of many societies and the occurrence pattern of presbyopia. Padmanabana et al. also investigated how gaze-contingent and adaptive-focus displays could overcome issues in VR displays through intensive user studies [134].

### 6.3 Occlusion capability

Occlusion is a perceptual cue that allows users to understand the 3D relationship between objects and the general depth of a scene (Fig. 15). Table 4 and Fig. 16 lists major related works. Unlike other depth cues that modern graphics engines can simulate in software (e.g. vergence, motion parallax, perspective, gradient and colour), realising occlusion on OST-HMDs requires extra hardware due to their optical see-through nature. To realise realistic occlusion in OST-HMDs, we must block environmental light that would normally blend with the displayed content. As discussed in the section on colour blending, however, blocking the background light is not available with currently available OST-HMDs (Fig. 16).

A key optical element in existing occlusion-capable OST-HMDs is the SLM, a device that can selectively change the property of incoming light. For example, an LCD is an SLM that can change the transparency of the light rays passing through each pixel.

## Occlusion

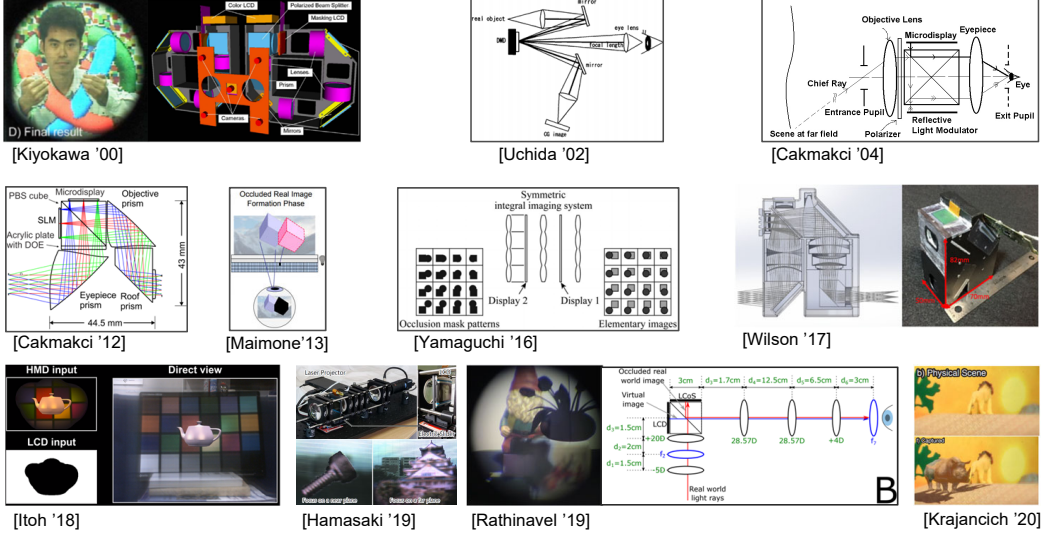


Fig. 16. Example works related to the occlusion issues covered in Sec. 6.3.

Kiyokawa et al. proposed an OST-HMD system called ELMO-1 [81], seminal work for occlusion-capable OST-HMDs. Their OST-HMD inserted an LCD between two coupling lenses so that background light first focused at the LCD plane before reaching the eyes. The system then modulated the LCD to block the light. Because of the optical pathways, ELMO-1 shifted the viewpoint of the users relative to their natural viewpoint. The authors later addressed this issue with a revised prototype (ELMO-4) [80] that allowed maintaining the original viewpoint. Santos et al. proposed a binocular-type OST-HMD with occlusion but did not disclose the optical design [152].

While these approaches create a sharp occlusion mask, they tend to be bulky. Another approach simply places the occlusion layer in a near-eye fashion. Maimone et al. proposed a light-field display based on pinlight projection to realise a wide-FoV display [116]. Their system had an LCD in front of the display to render a mask image for occlusion. The approach partially worked with a blurry outline in the mask due to the depth mismatch between the LCD and virtual objects overlaid in the scene. To resolve this blurry occlusion while maintaining the simple optics with a near-eye LCD layer, Itoh et al. proposed a method to compensate for the blurry occlusion by fusing the scene image through an OST-HMD [59]. Their approach computed the amount of blur caused by the near-eye LCD given the eye aperture size before creating a compensation image from the scene camera optically aligned with the user's view.

Unlike these approaches with a single occlusion mask, Yamaguchi and Takaki proposed an integral light-field display using three microlens-array layers [176]. They inserted two LCD panels between each gap of the three layers. One panel with a transparent organic LED (OLED) served as a raw display screen for the integral display, and the other was an occlusion layer. Thanks to the integral imaging approach, the occlusion was perceived as 3D.

As another light-field approach, Maimone and Fuchs proposed a multilayer see-through display [114] that comprised a shutter, transparent illumination and a minimum of two LCD panels. The system temporally drove the LCD panels to simulate a light field of the desired image while activating the shutter and illumination so that a user perceived the image. By switching this rendering mode and the see-through mode, that is, deactivating the shutter and the illumination, the

user could achieve a see-through view. Furthermore, in see-through mode, the system turned off the LCD pixels corresponding to the rendered light-field rays, so the realised occlusion effect was perceived as 3D occlusion.

Unlike these systems using transmissive SLMs, others have developed occlusion-capable OST-HMDs using reflective SLMs. Uchida et al. proposed using a DMD, a reflective mirror device, in a see-through display system [164]. Their DMD system could switch if a light path from the viewpoint went to the background or a microdisplay, enabling per-pixel selection of transparent or opaque state thus allowing mutual occlusion between the real and virtual content. Kim et al. demonstrated a similar DMD-based system [74]. Krajancich et al. further enhanced the DMD-based approach by using the DMD itself as a display screen by combining LEDs [88]. Their approach realised a simpler optical design yet came with an optimised rendering algorithm based on binary image factorisation.

Instead of using DMDs, Cakmakci et al. used a liquid-crystal-on-silicon (LCoS) chip, that is, a reflective LCD [14]. Their system unified the optical path from the real scene and a microdisplay via an LCoS with an X-cube prism. Gao et al. designed two-layer folded optics with free-form prisms to reduce the size and improve the optical performance of the display [38, 39]. Wilson and Hua later demonstrated a similar folded occlusion-capable OST-HMD with affordable lenses instead of free-form lenses [170].

*Varifocal Occlusion.* Similar to the accommodation issue in OST-HMDs (Sec. 6.2), the occlusion methods mentioned above suffer from the vergence–accommodation conflict, as occlusions appear on a plane with a fixed distance based on the optics and the display used to render the occlusions. Hamasaki and Itoh proposed a hardware approach to make an occlusion layer varifocal [48]. Their system comprised relay optics consisting of three lenses with the LCD layer placed between the first two lenses. The LCD was mounted on a linear stage so that the system could optically shift the depth of the occlusion mask in the viewing direction. Rathinavel et al. proposed another varifocal-occlusion approach with two focus-tunable lenses placed in the see-through path of the display. By optimally controlling the focus of the lenses, the system could optically tune the focus depth of the LCD layer, which was placed between the lenses [140].

## 6.4 Wide Field of View

One of the most discussed downsides of current OST-HMDs is their limited FoV. In most cases, only a small head or eye movement would make the digital overlay disappear, making the difference between the real and virtual obvious. Human vision typically has a horizontal FoV of  $180^\circ$ – $200^\circ$  when combining both eyes, a range that needs to be covered. However, in practice, most OST-HMDs do not even come close, lagging far behind most prescription glasses [15, 116]. In the following, we revisit some optical designs for OST-HMDs presented above to examine their FoVs before discussing dedicated optical designs. We also give an overview of different optical designs for HMDs classified by their FoVs [15].

Using conventional optical combiners, such as half-silvered mirrors and prisms [160], would theoretically allow a wide FoV; however, this comes at the cost of size and weight. The optical elements would be relatively large because they would need to cover the entire FoV, and the required angled placement of the optical elements would add to the bulkiness of the overall design (Fig. 2). Furthermore, depending on the design, additional lenses would be required to magnify the displayed image or relay it, for example (depending on the position of the light source).

Instead of planar mirrors, several groups have investigated curved mirrors to reflect the light emitted by a display into the human eye. This design can lead to a larger FoV and smaller form factors than designs using planar mirrors [76, 127]. The commercial Meta 2 AR Glasses use a curved

optical element, reflecting the image of an integrated display to achieve a relatively large FoV ( $90^\circ$ ). Prototypical implementations have demonstrated displays with  $180^\circ$  horizontal FoVs [76, 127], but they cannot reassemble the form factors of normal glasses or currently available commercial solutions.

The literature contains approaches that allow wide-FoV displays using membranes instead of solid optical elements [31]. While the membrane has a similar purpose as a conventional half-transparent mirror, it can also rapidly change its curvature and consequently the focal plane, allowing varifocal displays. However, current prototypical designs are bulky when building displays with wide FoVs, and while they have advantages in terms of accommodation capability, they are comparable to the optical design using half-silvered mirrors when targeting wide FoVs.

Only recently have advances in manufacturing allowed the mass production of free-form optics, which can reduce the number and overall size of optical elements by producing complex integrated optics with specific shapes and curvatures that magnify and relay images to the human eye. However, free-form optics usually have a thickness substantially greater than prescription glasses, and their volume grows proportional to the FoV, limiting practical uses. One work improved the design by tiling free-form optics [20]. Their system achieved a large resolution and a relatively wide FoV ( $56^\circ \times 45^\circ$ ), yet without flat lens systems.

An alternative to using free-form optics is embedding metasurfaces into an eyepiece [104]. Each metasurface is only a few nm in size and is encoded with an encoding a spatial phase, allowing the overall eyepiece to function as a spherical lens of a desired focal length. Although Lee et al. have demonstrated a prototype with an FoV of  $76^\circ$ , the prototype was too bulky for a practical application.

Waveguide-based systems use specific diffraction patterns or holographic elements to guide light to the eye [125]. The advantages of these approaches are that the waveguide can be flat and allow form factors closer to traditional prescription glasses. Unfortunately, this technology allows only a limited FoV when keeping this form factor. Using current technology, a wide-FoV OST-HMD using waveguides will result in ghost images and colour bleeding [125].

One approach from the commercial sector avoids OST-HMD design altogether [157] by using glasses with a small microdisplay that displays an unfocused image using a narrow band of RGB frequencies. The users must also wear a specific contact lens that acts like a pinhole but is made of filters. Most of the contact lens is a filter that blocks out the frequencies used by the microdisplay but is transmissive to the frequencies of the environmental light.

However, the centre of the contact lens (the pinhole) is made of a different filter material that is transmissive to the narrow band of frequencies of the microdisplay while blocking all other frequencies. A user using the OST-HMD without the specific contact lens will only see the environment but not the image from the image displayed in the microdisplay integrated into the glasses. However, when also using the contact lens, the previously unfocused image from the microdisplay is visible through the pinhole and is consequently in focus. Using this approach enables building wide-FoV displays but always requires wearing contact lenses. Maimone et al. extended this idea of using unfocused light but removed the requirement of wearing contact lenses [116]. In their approach, pinlight displays, the authors proposed using a see-through point light array and a transmissive SLM (here, a transparent LCD panel) placed directly in front of the eye.

The point light array is placed in the optical path of the eye but closer than the typical accommodation range. As such, each point light is defocused (a pinlight). The light emitted by the point light array is modulated by a single transmissive SLM placed between the pinlights and the eye. As the point lights emit light not only in a single direction but over a wide angle so that each is visible to the eye, in this configuration, the SLM acts as a virtual aperture that controls the emitted light.



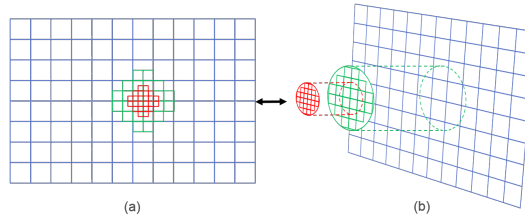


Fig. 17. Example of foveation of a pixel grid inspired by Guenter et al. [46]. (a) Grid foveated around a point with the highest detail/acyuity in red, reduced in green and low acuity in blue. (b) central “fovea” area with the highest level of detail in red, reducing as eccentricity increases in green and low detail in the blue periphery.

The advantage of this approach is that it allows building wide-FoV displays that do not require normal optical combiners to guide an image to the eye (e.g. free-form optics or waveguides). Instead, the authors demonstrated that the point light array can be built from a flat acrylic sheet with small divots. While the acrylic plate acts as a waveguide, it is only used to illuminate the SLM where the actual image is created. Consequently, optical inaccuracies that occur when using a waveguide to guide full images, as more traditional approaches do, are less of a problem. As the SLM is also flat, the authors demonstrated that their pinlight displays could be used for a wide-FoV display while maintaining a small form factor.

Benko et al. presented a completely different approach [7]. Their work took up an idea previously presented for normal displays where the screen size was extended with projectors. Instead of having normal displays, however, a wide FoV was achieved using projector-based spatial AR to generate the peripheral vision and the OST-HMD for focus areas. They needed a spatial AR system and ideally precise tracking for the pixel blending of focus and peripheral vision [137].

Maimone et al. employed a polarization-based optical folding technique and demonstrated a thin VR-HMD [117]. Their prototype achieved less than 9 mm thickness with  $90^\circ$  horizontal FoV. Lee et al. proposed a similar approach yet combined polarisation selective diffuser, enabling an OST-HMD prototype with  $80^\circ$  diagonal FoV [107].

## 6.5 Resolution

The resolution of the human eye provides a significant challenge for any display, especially near-eye displays, such as OST-HMDs. To create an image that has perceptually the same resolution as the physical world, we must accommodate 60 cpd across an FoV greater than  $190^\circ \times 120^\circ$ , requiring a resolution of  $22800 \times 14400$  and 328.3 megapixels to be processed in real time and displayed to the user, as detailed in Sec. 3.1.

Compared to most research prototypes, which display at most  $1920 \times 1080/1200$  pixels [55, 62], and the commercial HoloLens 2, which claims a 2k resolution, these only require up to 3.7 megapixels (assuming  $2560 \times 1440$ ), which is 0.01% of the postulated value. The actual cpd produced by a display of a given resolution varies, as it is re-projected onto various FoVs.

While display resolutions are slowly improving, even for near-eye displays, the problem of processing the required information is still pressing. To overcome this, several techniques to reduce the resolution can be utilised [156]. As previously mentioned, resolution is not uniform across the visual field at any one time. Therefore, to match the resolution of the human eye, a high resolution is only required for the fovea, while the periphery can be presented at a much lower resolution. This is the basic idea of foveated rendering to improve the computational time and bandwidth while still creating high resolution in areas where it is needed.

The general idea of foveated rendering is not new and has applications in general computer graphics (Fig. 17), where it has been covered in detail [46], and in HMDs [136]. Early work by Howlett et al. and Rolland et al. first explored this idea for near-eye displays [53, 150], but they were limited to VST-HMDs and HMDs in general, while more recent research with OST-HMDs has seen work by J. Lee et al., S. Lee et al., Kim et al. and Askit et al. [1, 73, 105–107].

Howlett et al. and Rolland et al. both used two displays to achieve this effect [53, 150] by projecting a smaller, high-resolution inset onto the low-resolution background image. Where Rolland et al. detailed an enclosed HMD using a microlens array, Howlett et al. detailed a VST-HMD system using two scene cameras. Rolland et al. designed a microlens array system able to project a small, high-resolution inset onto a lower-resolution background image, with the location of the inset determined by eye tracking. The projection of the inset was achieved via a beam splitter placed between the user's eye and the lower-resolution display, redirecting light from the microlens array towards the viewer [150]. Howlett et al. used two real-world, wide-angle cameras to capture the lower-resolution scene, while some light passing through the camera lens was redirected to a higher-resolution camera via a beam splitter. The video signals from both cameras were then displayed on two different displays, with the high-resolution image being projected through the Large Expanse Extra Perspective (LEEP) display using the wide-angle image [53]. The system proposed by Rolland et al. was not an OST-HMD, but replacing the primary display with an OST-HMD and integrating the microlens array at right angles, as described, could theoretically create a foveated OST-HMD.

Lee et al. proposed an algorithm to enable foveation for multilayer displays [106]. The algorithm emulated foveated rendering without the need for eye tracking by optimising layers such that high contrast was maintained when the eye was rotated to view an area of the display—however, this came at the cost of peripheral contrast.

Lee et al. and Kim et al. both used holographic optical elements in their work [73, 105]. Like Howlett et al. and Rolland et al. integrating two displays into their systems [53, 150], Lee et al. [105] used a retinal-scanning display based on the Maxwellian view to provide a lower-resolution, wide-FoV image at a lower computational cost than the smaller holographic display for a high-resolution fovea display. The choice of a Maxwellian view and a holographic display enabled foveated rendering while avoiding the accommodation–convergence conflict. Kim et al. [73] used the holographic element to produce a virtual retinal display in the periphery, however, and used a more traditional beam splitter and OLED to deliver a high-resolution inset for the fovea. The display position is adjusted based on real-time gaze data by a linear stage.

Lee et al. reported a peripheral FoV of  $22.6^\circ$  and a foveal FoV of  $1.02^\circ$  without reporting the peripheral or foveal cpd [106]. Kim et al. reported a peripheral FoV estimate of  $101.4^\circ$  and a foveal FoV of  $33^\circ$  to  $16^\circ$  depending on the prototype with a peripheral cpd of 3 and a foveal cpd of 29–59 depending on the prototype, respectively [73].

Akşit et al. worked differently to produce optical elements that created foveation via hardware with mechanically movable lenses [1]. They detailed a manufacturing process for producing custom near-eye displays, including printing optics, and they described the use of a foveated lens which distributed pixels non-uniformly to the viewer and could be adjusted to move the area of foveation.

In keeping with the modern growth of foveated display research, Spjut et al. [156] have looked to provide a classification for foveated displays, offering a definition of the term *foveated display* and providing a simplified model of the human visual acuity with which foveated displays can be evaluated. The given classification looks at resolution (A–D) and gaze (1–4) to classify displays. They would classify Kim et al.'s prototype [73], for example, as a 2B.

As we can see, while the resolution problem has been considered for near-eye displays, only recently has research started to deal with the need for high resolutions and the problems they introduce. The proposed solution is foveated displays. Of modern work on foveated displays, the

only one to bring this concept beyond a bench prototype has been Kim et al., whose work included a wearable prototype. However, it was untested by users [73]. Their work also demonstrated that one of their prototypes could achieve 98% of the resolution of the human eye at the fovea. These works show the potential of foveated rendering and the interest in the field, although it is clearly novel work.

## 7 DISCUSSION AND CONCLUSION

When looking into reviews and reports of OST-HMDs, we often see comments that the FoV is the main technical factor preventing visually coherent presentations and mass adoption in professional contexts. This work highlights the challenges of spatial, temporal and visual realism and highlights that besides limited FoV, several other factors contribute to a convincing augmentation of the physical environment. We also show that research addresses these factors differently.

While a large body of work exists to solve accommodation issues, only a few have addressed latency, colour representation or resolution issues. Furthermore, many of the presented techniques have only been explored in large benchtop prototypes; the required miniaturisation remains challenging. Similarly, most of the techniques focus only on one aspect, and integration with other requirements is not addressed and is sometimes impossible.

To that end, we highlight recent works that have combined and advanced several aspects, such as the work by Kim et al. [73] addressing the issue of high resolution while also supporting accommodation. They even showed possible miniaturisation into a wearable prototype. More work like this is needed, but the technical skills and equipment required are significant; not many research labs combine 3D printing and prototyping skills, optical expertise (lens design and optical coating) and equipment (laser projectors, linear stages, optical components and displays), knowledge in electronics, programming skills and a background in visual computing (computer graphics, computer vision and human-computer interaction).

*Research opportunities.* Several research directions arise from the current work. Spatial calibration is essential to achieve some results demonstrated in the research, but a usable calibration supporting a continuous (re)calibration is still challenging. A human-centred research approach is necessary to really understand how a spatial calibration can be achieved by non-technical users and what kind of continuous calibration techniques are acceptable. Eye tracking seems key to achieving this goal.

Furthermore, it remains unclear if pixel precision can be achieved in unprepared environments, even with the most accurate tracking techniques and careful spatial calibration. However, if one had an image of the environment exactly as seen by the user, one could perform image-based corrections [27, 82] that could achieve pixel precision. Recent papers on vision augmentation using OST-HMDs followed this idea by approximating the user's view with calibrated cameras in the optical path of the user [50, 100] to establish a per-pixel mapping from camera pixels to display pixels. This would theoretically allow pixel-precise calibration and spatial calibration.

Temporal realism and particularly latency are rarely addressed in AR research, even though we know about the need for low latency from VR. Perhaps people think that reducing latency is a low-reward, purely engineering task. We argue against that notion. Approaches like foveated rendering obviously might reduce rendering load and latency, particularly for higher resolutions. However, to overcome display latency as well, novel approaches are needed that re-consider how we send and display images while considering, of course, the final form factor of the OST-HMD, something current research struggles with (e.g. [109]).

The inability to fully control the light within an OST-HMD seems a core issue that prevents visual realism. Some existing approaches do not control environmental light and only adapt the displayed image based on the measured incoming light (e.g. approaches focused on colour blending

and display colour reproduction). Other approaches do not control the displayed light but only the incoming environmental light (e.g. approaches to handling occlusions). Approaches that do both and offer full control over the light within an OST-HMD are missing. Recent work by Itoh et al. [62] explored a light-attenuation display that filtered incoming environmental light on a per-pixel level. If this could be integrated with a normal display adding light on a per-pixel level, it would come close to that vision of controlling (filtering or adding) light on a per-pixel level, potentially increasing visual realism. This is not considering that stray light and unwanted reflections can still be problematic but require different solutions.

Finally, an enormous remaining research challenge is to consider approaches that satisfy all requirements raised in this paper. Now that we know that we can address individual issues, future research should accept the challenge that practical solutions would address all of them, which is often hard if not impossible.

*Turning research into commercial products.* In this paper, we focused on results reported in published research. However, in particular, when aiming for commercial products, other factors come into play that are often ignored by research, including human factors such as weight, ergonomics and general design. Many of these require the miniaturisation of research prototypes. While electronics often can easily be miniaturised, this is often not true for optical elements. In fact, miniaturising optical elements often changes optical parameters and significantly undermines achievable quality (e.g. possible FoV, eye box size, light efficiency and the general quality of optical components). Thus, aiming for a form factor closely resembling sunglasses usually prohibits a large FoV. These challenges are obvious in research that tried to produce miniaturised prototypes [107, 117]. They are often not used to create the visual results presented in the final papers and lag behind in functionality. It is important to understand that this cannot be easily solved, even if money is no object.

Another practical issue is the spatial calibration of OST-HMDs [44]. Calibrating an OST-HMD specifically for individual users, as often applied in research, is usually not possible in commercial solutions apart from manually adjusting the IPD. The integration of eye-tracking cameras in some recent commercial OST-HMDs offers some potential for continuous calibration but seems to be of low priority. Cameras are nowadays often calibrated within the factory. This is theoretically also possible for the displays used in OST-HMDs, but many of the effects in displays are highly view-dependent (e.g. distortions). Knowing the exact eye position would allow at least compensating for some of these geometric effects, as outlined in the research. Still, it is important to realise that the careful, time-intensive calibration to individuals that we often see in research is not easily applicable to commercial products. Likewise, commercial products must satisfy higher standards of eye safety and ergonomics, particularly those that could also cause additional strain to the eye (e.g. eye relief, divergence and binocular overlap) while also considering individual differences in human anatomy and perception.

*Limitations.* Despite our best intentions, this work has limitations. Foremost, it is not a systematic review, as the general breadth of the topic spanning many research communities (e.g. optics, computer graphics, virtual and augmented reality, computer vision, etc.) and the different terminologies used synonymously make a systematic approach difficult. As an example, OST-HMDs are sometimes referred to as HMDs, near-eye displays, see-through HMDs, AR or MR eyeglasses, and digital eyewear, just to name a few. However, the authors are well established within the field and conducted extensive searches to identify the works highlighted in this paper. Furthermore, this paper does not attempt to give priorities, such as a wider FoV being more important than a higher resolution. We believe this really depends on the specific aims and use cases. As is common for most of these kinds of papers, we could not test or verify all the prototypes in this paper, and all reported

data are from the original papers. Finally, we are aware of several commercial prototypes and patent applications but intentionally focused on the academic literature. When including patents, we often speculated or touched on non-disclosure agreements. However, it is still worth mentioning Karl Gutttag's blog,<sup>2</sup> which is a good source of information on displays.

## ACKNOWLEDGMENTS

Yuta Itoh are partially supported by JST PRESTO Grant Number JPMJPR17J2 and JSPS KAKENHI Grant Number JP17H04692, JP17K19985, and JP20H04222, Japan. Tobias Langlotz and Jonathan Sutton are partially supported by the Marsden Fund Council from Government funding and a Catalyst Seed Grant, both administered by the Royal Society of NZ and by Callaghan Innovation, host of the Science for Technological Innovation National Science Challenge, Seed Project 52421, New Zealand.

## REFERENCES

- [1] K. Akşit, P. Chakravarthula, K. Rathinavel, Y. Jeong, R. Albert, H. Fuchs, and D. Luebke. 2019. Manufacturing Application-Driven Foveated Near-Eye Displays. *IEEE TVCG* 25, 5 (May 2019), 1928–1939.
- [2] Kaan Akşit, Jan Kautz, and David Luebke. 2015. Slim near-eye display using pinhole aperture arrays. *Appl. Opt.* 54, 11 (Apr 2015), 3422–3427. <http://ao.osa.org/abstract.cfm?URI=ao-54-11-3422>
- [3] Kurt Akeley, Simon J Watt, Ahna Reza Girshick, and Martin S Banks. 2004. A stereo display prototype with multiple focal distances. In *ACM TOG*, Vol. 23. ACM, 804–813.
- [4] Ron Azuma. 1995. *"Predictive Tracking for Augmented Reality"*. Ph.D. Dissertation. Computer Science, University of North Carolina, Chapel Hill, NC.
- [5] Ronald Azuma and Gary Bishop. 1994. Improving static and dynamic registration in an optical see-through HMD. In *Proc. of ACM SIGGRAPH 1994*. 197–204. <https://doi.org/10.1145/192161.192199>
- [6] Ronald T. Azuma. 1997. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments* 6, 4 (Aug. 1997), 355–385.
- [7] Hrvoje Benko, Eyal Ofek, Feng Zheng, and Andrew D Wilson. 2015. Fovear: Combining an optically see-through near-eye display with projector-based spatial augmented reality. In *Proc. of the 28th ACM UIST*. ACM, 129–135.
- [8] Oliver Bimber, Franz Coriand, Alexander Kleppe, Erich Bruns, Stefanie Zollmann, and Tobias Langlotz. 2005. Superimposing pictorial artwork with projected imagery. In *ACM SIGGRAPH 2005 Courses*. 6–es.
- [9] Oliver Bimber, Andreas Emmerling, and Thomas Klemmer. 2005. Embedded entertainment with smart projectors. *Computer* 38, 1 (2005), 48–55.
- [10] Oliver Bimber, Daisuke Iwai, Gordon Wetzstein, and Anselm Grundhöfer. 2008. The Visual Computing of Projector-Camera Systems. In *Computer Graphics Forum*, Vol. 27. Wiley Online Library, 2219–2245.
- [11] Oliver Bimber and Ramesh Raskar. 2005. *Spatial augmented reality: merging real and virtual worlds*. CRC Press.
- [12] Timothy J. Buker, Dennis A. Vincenzi, and John E. Deaton. 2012. The effect of apparent latency on simulator sickness while using a see-through helmet-mounted display: reducing apparent latency with predictive compensation. *Human factors* 54, 2 (2012), 235–249. <http://dx.doi.org/>
- [13] Cesar Cadena, Luca Carlone, Henry Carrillo, Yasir Latif, Davide Scaramuzza, José Neira, Ian Reid, and John J Leonard. 2016. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on Robotics* 32, 6 (2016), 1309–1332.
- [14] Ozan Cakmakci, Yonggang Ha, and Jannick P. Rolland. 2004. A Compact Optical See-Through Head-Worn Display with Occlusion Support. In *IEEE ISMAR 2004*. IEEE Computer Society, 16–25. <https://doi.org/10.1109/ISMAR.2004.2>
- [15] Ozan Cakmakci and Jannick Rolland. 2006. Head-worn displays: a review. *Journal of Display Technology* 2, 3 (2006), 199–216.
- [16] John Carmack. 2013 (accessed October, 2019). *Latency mitigation strategies*. <https://danluu.com/latency-mitigation/>
- [17] Ali Cem, M Kivanc Hedili, Erdem Ulusoy, and Hakan Urey. 2020. Foveated near-eye display using computational holography. *Scientific reports* 10, 1 (2020), 1–9.
- [18] P. Chakravarthula, D. Dunn, K. Akşit, and H. Fuchs. 2018. FocusAR: Auto-focus Augmented Reality Eyeglasses for both Real World and Virtual Imagery. *IEEE TVCG* 24, 11 (Nov 2018), 2906–2916.
- [19] Praneeth Chakravarthula, Yifan Peng, Joel Kollin, Henry Fuchs, and Felix Heide. 2019. Wirtinger holography for near-eye displays. *ACM Transactions on Graphics (TOG)* 38, 6 (2019), 1–13.

<sup>2</sup>/[www.kgutttag.com](http://www.kgutttag.com)



- [20] Dewen Cheng, Yongtian Wang, Hong Hua, and Jose Sasian. 2011. Design of a wide-angle, lightweight head-mounted display using free-form optics tiling. *Opt. Lett.* 36, 11 (Jun 2011), 2098–2100. <https://doi.org/10.1364/OL.36.002098>
- [21] Sue VG Cobb, Sarah Nichols, Amanda Ramsey, and John R Wilson. 1999. Virtual reality-induced symptoms and effects (VRISE). *Presence: Teleoperators and Virtual Environments* 8, 2 (1999), 169–186.
- [22] Jonny Collins, Holger Regenbrecht, and Tobias Langlotz. 2017. Visual Coherence in Mixed Reality: A Systematic Enquiry. *Presence* 26, 1 (2017), 27–36.
- [23] Trey Cook, Nate Phillips, Kristen Massey, Alexander Plopski, Christian Sandor, and J Edward Swan. 2018. User preference for sharpview-enhanced virtual text during non-fixated viewing. In *IEEE VR 2018*. IEEE, 1–400.
- [24] James E Cutting. 1997. How the eye measures reality and virtual reality. *Behavior Research Methods* 29, 1 (1997), 27–36.
- [25] Juan David Hincapié-Ramos, Levko Ivanchuk, Srikanth Kirshnamachari Sridharan, and Pourang Irani. 2014. Smart-Color: Real-time color correction and contrast for optical see-through head-mounted displays. In *IEEE ISMAR*. 187–194.
- [26] James Davis, Yi-Hsuan Hsieh, and Hung-Chi Lee. 2015. Humans perceive flicker artifacts at 500 Hz. *Scientific reports* 5 (02 2015), 7861. <https://doi.org/10.1038/srep07861>
- [27] Stephen DiVerdi and Tobias Hollerer. 2006. Image-space correction of AR registration errors using graphics hardware. In *IEEE VR 2006*. IEEE, 241–244.
- [28] Hyeonkyu Do, Young Min Kim, and Sung-Wook Min. 2019. Focus-free head-mounted display based on Maxwellian view using retroreflector film. *Applied optics* 58, 11 (2019), 2882–2889.
- [29] David Dunn. 2019. Required Accuracy of Gaze Tracking for Varifocal Displays. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 1838–1842.
- [30] David Dunn, Praneeth Chakravarthula, Qian Dong, Kaan Akşit, and Henry Fuchs. 2018. 10-1: Towards Varifocal Augmented Reality Displays using Deformable Beamsplitter Membranes. *SID Symposium Digest of Technical Papers* 49, 1 (2018), 92–95.
- [31] David Dunn, Cary Tippets, Kent Torell, Petr Kellnhofer, Kaan Akşit, Piotr Didyk, Karol Myszkowski, David Luebke, and Henry Fuchs. 2017. Wide field of view varifocal near-eye display using see-through deformable membrane mirrors. *IEEE transactions on visualization and computer graphics* 23, 4 (2017), 1322–1331.
- [32] R. W. Evans, A. P. Ramsbottom, and D. W. Sheel. 1989. Head-up displays in motor cars. In *1989 Second International Conference on Holographic Systems, Components and Applications*. 56–62.
- [33] James A. Ferwerda, Sumanta N. Pattanaik, Peter Shirley, and Donald P. Greenberg. 1996. A Model of Visual Adaptation for Realistic Image Synthesis. In *Proc. of the 23rd SIGGRAPH (SIGGRAPH '96)*. ACM, New York, NY, USA, 249–258.
- [34] Jan Fischer and Dirk Bartz. 2005. Stylized Augmented Reality for Improved Immersion. In *IEEE VR 2005 (VR '05)*. IEEE Computer Society, Washington, DC, USA, 195–202, 325. <https://doi.org/10.1109/VR.2005.71>
- [35] T. Fukiage, T. Oishi, and K. Ikeuchi. 2014. Visibility-based blending for real-time applications. In *IEEE ISMAR 2014*. 63–72. <https://doi.org/10.1109/ISMAR.2014.6948410>
- [36] Thomas A Furness III and Joel S Kollin. 1995. Virtual retinal display. US Patent 5,467,104.
- [37] J. L. Gabbard, J. Zedlitz, J. E. Swan, and W. W. Winchester. 2000. More than meets the eye: An engineering study to empirically examine the blending of real and virtual color spaces. *IEEE VR 2000 (2000)*, 79–86.
- [38] Chunyu Gao, Yuxiang Lin, and Hong Hua. 2012. Occlusion capable optical see-through head-mounted display using freeform optics. In *11th IEEE ISMAR*. IEEE, 281–282.
- [39] Chunyu Gao, Yuxiang Lin, and Hong Hua. 2013. Optical see-through head-mounted display with occlusion capability. In *Proc. SPIE*, Vol. 8735. 87350F.
- [40] Qiankun Gao, Juan Liu, Xinhui Duan, Tao Zhao, Xin Li, and Peilin Liu. 2017. Compact see-through 3D head-mounted display based on wavefront modulation with holographic grating filter. *Optics express* 25, 7 (2017), 8412–8424.
- [41] Qiankun Gao, Juan Liu, Jian Han, and Xin Li. 2016. Monocular 3D see-through head-mounted display via complex amplitude modulation. *Optics express* 24, 15 (2016), 17372–17383.
- [42] E.B. Goldstein and J. Brockmole. 2016. *Sensation and Perception*. Cengage Learning.
- [43] Lukas Gruber, Tobias Langlotz, Pradeep Sen, Tobias Hoherer, and Dieter Schmalstieg. 2014. Efficient and robust radiance transfer for probeless photorealistic augmented reality. In *IEEE VR 2014*. IEEE, 15–20.
- [44] J. Grubert, Y. Itoh, K. Moser, and J. E. Swan. 2018. A Survey of Calibration Methods for Optical See-Through Head-Mounted Displays. *IEEE TVCG* 24, 9 (Sept 2018), 2649–2662. <https://doi.org/10.1109/TVCG.2017.2754257>
- [45] Anselm Grundhöfer and Daisuke Iwai. 2018. Recent advances in projection mapping algorithms, hardware and applications. In *Computer Graphics Forum*, Vol. 37. Wiley Online Library, 653–675.
- [46] Brian Guenter, Mark Finch, Steven Drucker, Desney Tan, and John Snyder. 2012. Foveated 3D Graphics. *ACM TOG* 31, 6, Article 164 (Nov. 2012), 10 pages. <https://doi.org/10.1145/2366145.2366183>
- [47] Rolf R Hainich and Oliver Bimber. 2016. *Displays: fundamentals & applications*. CRC press.

- [48] Takumi Hamasaki and Yuta Itoh. 2019. Varifocal Occlusion for Optical See-Through Head-Mounted Displays using a Slide Occlusion Mask. *IEEE TVCG* 25, 5 (2019), 1961–1969.
- [49] Zehao He, Xiaomeng Sui, Guofan Jin, and Liangcai Cao. 2019. Progress in virtual reality and augmented reality based on holographic display. *Applied optics* 58, 5 (2019), A74–A81.
- [50] Yuichi Hiroi, Takumi Kaminokado, Atsushi Mori, and Yuta Itoh. 2020. DehazeGlasses: Optical Dehazing with an Occlusion Capable See-Through Display. In *Proceedings of the Augmented Humans International Conference*. 1–11.
- [51] David M Hoffman, Ahna R Girshick, Kurt Akeley, and Martin S Banks. 2008. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision* 8, 3 (2008), 33–33.
- [52] Ian P. Howard and Brian J. Rogers. 1996. *Binocular Vision and Stereopsis*. Oxford University Press.
- [53] Eric M Howlett. 1992. High-resolution inserts in wide-angle head-mounted stereoscopic displays. In *Stereoscopic Displays and Applications III*, Vol. 1669. International Society for Optics and Photonics, 193–203.
- [54] Hong Hua. 2017. Enabling focus cues in head-mounted displays. *Proc. IEEE* 105, 5 (2017), 805–824.
- [55] Hong Hua and Bahram Javidi. 2014. A 3D integral imaging optical see-through head-mounted display. *Opt. Express* 22, 11 (Jun 2014), 13484–13491. <https://doi.org/10.1364/OE.22.013484>
- [56] Fu-Chung Huang, Kevin Chen, and Gordon Wetzstein. 2015. The light field stereoscope: immersive computer graphics via factored near-eye light field displays with focus cues. *ACM TOG* 34, 4 (2015), 60.
- [57] Y. Itoh, T. Amano, D. Iwai, and G. Klinker. 2016. Gaussian Light Field: Estimation of Viewpoint-Dependent Blur for Optical See-Through Head-Mounted Displays. *IEEE TVCG* 22, 11 (Nov 2016), 2368–2376.
- [58] Yuta Itoh, Maksym Dzitsiuk, Toshiyuki Amano, and Gudrun Klinker. 2015. Semi-Parametric Color Reproduction Method for Optical See-Through Head-Mounted Displays. *IEEE TVCG* 21, 11 (Nov. 2015), 1269–1278. *Proc. of ISMAR*.
- [59] Yuta Itoh, Takumi Hamasaki, and Maki Sugimoto. 2017. Occlusion leak compensation for optical see-through displays using a single-layer transmissive spatial light modulator. *IEEE TVCG* 23, 11 (2017), 2463–2473.
- [60] Yuta Itoh and Gudrun Klinker. 2014. Performance and Sensitivity Analysis of INDICA: Interaction-Free Display Calibration for Optical See-Through Head-Mounted Displays. In *IEEE ISMAR 2014*. 171–176.
- [61] Yuta Itoh and Gudrun Klinker. 2015. Light-Field Correction for Spatial Calibration of Optical See-Through Head-Mounted Displays. *IEEE TVCG (Proc. VR 2015)* 21, 4 (April 2015), 471–480.
- [62] Yuta Itoh, Tobias Langlotz, Daisuke Iwai, Kiyoshi Kiyokawa, and Toshiyuki Amano. 2019. Light Attenuation Display: Subtractive See-Through Near-Eye Display via Spatial Color Filtering. *IEEE TVCG* 25, 5 (May 2019), 1951–1960.
- [63] Yuta Itoh, Jonathan Orlosky, Manuel Huber, Kiyoshi Kiyokawa, and Gudrun Klinker. 2016. OST Rift: Temporally consistent augmented reality with a consumer optical see-through head-mounted display. In *IEEE VR 2016*. 189–190.
- [64] Yuta Itoh, Kenta Yamamoto, and Yoichi Ochiai. 2018. Retinal HDR: HDR Image Projection Method onto Retina. In *SIGGRAPH Asia 2018 Posters (SA'18)*. ACM, New York, NY, USA, Article Article 82, 2 pages.
- [65] Redhwan Jamiruddin, Ali Osman Sari, Jahanzaib Shabbir, and Tarique Anwer. 2018. RGB-Depth SLAM Review. *CoRR* abs/1805.07696 (2018). arXiv:1805.07696
- [66] Changwon Jang, Kiseung Bang, Seokil Moon, Jonghyun Kim, Seungjae Lee, and Byoungcho Lee. 2017. Retinal 3D: Augmented Reality Near-eye Display via Pupil-tracked Light Field Projection on Retina. *ACM TOG* 36, 6, Article 190 (Nov. 2017), 13 pages. <https://doi.org/10.1145/3130800.3130889>
- [67] Seokhee Jeon, Seungmoon Choi, and Matthias Harders. 2015. Haptic Augmented Reality: Taxonomy, Research Status, and Challenges. *Fundamentals of Wearable Computers and Augmented Reality* (2015), 227–255.
- [68] Jason Jerald and Mary Whitton. 2009. Relating Scene-Motion Thresholds to Latency Thresholds for Head-Mounted Displays. In *IEEE VR 2009 (VR '09)*. IEEE Computer Society, Washington, DC, USA, 211–218.
- [69] Ricardo Jota, Albert Ng, Paul Dietz, and Daniel Wigdor. 2013. How Fast is Fast Enough?: A Study of the Effects of Latency in Direct-touch Pointing Tasks. In *CHI 2013 (CHI '13)*. ACM, New York, NY, USA, 2291–2300.
- [70] M. Carmen Juan, Mariano Alcaniz, Carlos Monserrat, Cristina Botella, Rosa M. Banos, and Belen Guerrero. 2005. Using Augmented Reality to Treat Phobias. *IEEE Comput. Graph. Appl.* 25, 6 (Nov. 2005), 31–37.
- [71] Takumi Kaminokado, Yuichi Hiroi, and Yuta Itoh. 2020. StainedView: Variable-Intensity Light-Attenuation Display with Cascaded Spatial Color Filtering for Improved Color Fidelity. *IEEE Transactions on Visualization and Computer Graphics* 26, 12 (2020), 3576–3586.
- [72] Hirokazu Kato and Mark Billinghurst. 1999. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *2nd IEEE and ACM International Workshop on Augmented Reality (IWAR 1999)*. IEEE, 85–94.
- [73] Jonghyun Kim, Youngmo Jeong, Michael Stengel, Kaan Akçit, Rachel Albert, Ben Boudaoud, Trey Greer, Joohwan Kim, Ward Lopes, Zander Majercik, Peter Shirley, Josef Spjut, Morgan McGuire, and David Luebke. 2019. Foveated AR: Dynamically-foveated Augmented Reality Display. *ACM TOG* 38, 4, Article 99 (July 2019), 15 pages.
- [74] Kwangsoo Kim, Daerak Heo, and Joonku Hahn. 2019. Occlusion-capable Head-mounted Display. In *Proc. of the 7th PHOTOPTICS 2019, Prague, Czech Republic, February 25-27, 2019*. SciTePress, 299–302.
- [75] Kiyoshi Kiyokawa. 2007. *An Introduction to Head Mounted Displays for Augmented Reality*. IGI Global, 43–63. <https://doi.org/10.4018/9781599040660.ch003>

- [76] Kiyoshi Kiyokawa. 2007. A Wide Field-of-view Head Mounted Projective Display Using Hyperbolic Half-silvered Mirrors. In *Proc. of the 2007 6th IEEE and ACM ISMAR*. IEEE Computer Society, Washington, DC, USA, 1–4.
- [77] Kiyoshi Kiyokawa. 2011. Display Technologies for Augmented Reality. *Proceedings of the International Display Workshops 3* (01 2011), 1971–1974. <https://doi.org/10.1201/b18703-7>
- [78] K. Kiyokawa. 2012. Trends and Vision of Head Mounted Display in Augmented Reality. In *2012 International Symposium on Ubiquitous Virtual Reality*. 14–17. <https://doi.org/10.1109/ISUVR.2012.11>
- [79] Kiyoshi Kiyokawa. 2016. *Occlusion Displays*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1–9.
- [80] Kiyoshi Kiyokawa, Mark Billinghurst, Bruce Campbell, and Eric Woods. 2003. An occlusion-capable optical see-through head mount display for supporting co-located collaboration. In *ACM/IEEE ISMAR*. 133.
- [81] Kiyoshi Kiyokawa, Yoshinori Kurata, and Hiroyuki Ohno. 2000. An optical see-through display for mutual occlusion of real and virtual environments. In *Proceedings IEEE and ACM ISAR*. IEEE, 60–67.
- [82] G. Klein and T. Drummond. 2004. Sensor fusion and occlusion refinement for tablet-based AR. In *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*. 38–47.
- [83] G. Klein and D. Murray. 2007. Parallel Tracking and Mapping for Small AR Workspaces. In *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*. 225–234. <https://doi.org/10.1109/ISMAR.2007.4538852>
- [84] Georg Klein and David Murray. 2008. Compositing for Small Cameras. In *Proc. Seventh IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'08)*. Cambridge.
- [85] M. Klemm, F. Seebacher, and H. Hoppe. 2016. Non-parametric Camera-Based Calibration of Optical See-Through Glasses for AR Applications. In *2016 International Conference on Cyberworlds (CW)*. 33–40.
- [86] Robert Konrad, Nitish Padmanaban, Keenan Molner, Emily A. Cooper, and Gordon Wetzstein. 2017. Accommodation-invariant Computational Near-eye Displays. *ACM TOG* 36, 4, Article 88 (July 2017), 12 pages.
- [87] George Alex Koulieris, Kaan Aksit, Michael Stengel, Rafał K Mantiuk, Katerina Mania, and Christian Richardt. 2019. Near-Eye Display and Tracking Technologies for Virtual and Augmented Reality. In *Computer Graphics Forum*, Vol. 38. Wiley Online Library, 493–519.
- [88] B. Krajancich, N. Padmanaban, and G. Wetzstein. 2020. Factored Occlusion: Single Spatial Light Modulator Occlusion-capable Optical See-through Augmented Reality Display. *IEEE TVCG* 26, 5 (2020), 1871–1879.
- [89] Gregory Kramida. 2016. Resolving the vergence-accommodation conflict in head-mounted displays. *IEEE TVCG* 22, 7 (2016), 1912–1931.
- [90] B.C. Kress. 2020. *Optical Architectures for Augmented-, Virtual-, and Mixed-Reality Headsets*. SPIE.
- [91] Bernard Kress and Thad Starner. 2013. A review of head-mounted displays (HMD) technologies and applications for consumer electronics. In *Proc. SPIE*, Vol. 8720. 87200A.
- [92] Bernard C. Kress and Ishan Chatterjee. 07 Oct. 2020. Waveguide combiners for mixed reality headsets: a nanophotonics design perspective. *Nanophotonics* 0 (07 Oct. 2020), 20200410. <https://doi.org/10.1515/nanoph-2020-0410>
- [93] Ernst Kruijff, J Edward Swan, and Steven Feiner. 2010. Perceptual issues in augmented reality revisited. In *Mixed and Augmented Reality (ISMAR)*, 2010 9th IEEE International Symposium on. IEEE, 3–12.
- [94] Timo Kunkel and Erik Reinhard. 2010. A Reassessment of the Simultaneous Dynamic Range of the Human Visual System. In *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization (APGV '10)*. ACM, New York, NY, USA, 17–24. <https://doi.org/10.1145/1836248.1836251>
- [95] Grace Kuo, Laura Waller, Ren Ng, and Andrew Maimone. 2020. High resolution étendue expansion for holographic displays. *ACM Transactions on Graphics (TOG)* 39, 4 (2020), 66–1.
- [96] Marc Lambooi, Marten Fortuin, Ingrid Heynderickx, and Wijnand IJsselstein. 2009. Visual discomfort and visual fatigue of stereoscopic displays: A review. *Journal of Imaging Science and Technology* 53, 3 (2009), 30201–1.
- [97] T. Langlotz, M. Cook, and H. Regenbrecht. 2016. Real-Time Radiometric Compensation for Optical See-Through Head-Mounted Displays. *IEEE TVCG* 22, 11 (Nov 2016), 2385–2394. <https://doi.org/10.1109/TVCG.2016.2593781>
- [98] Tobias Langlotz, Thanh Nguyen, Dieter Schmalstieg, and Raphael Grasset. 2014. Next-Generation Augmented Reality Browsers: Rich, Seamless, and Adaptive. *Proc. IEEE* 102, 2 (feb 2014), 155–169.
- [99] Tobias Langlotz, Holger Regenbrecht, Stefanie Zollmann, and Dieter Schmalstieg. 2013. Audio stickies: visually-guided spatial audio annotations on a mobile augmented reality. In *Proceedings of the 25th OzCHI 2013*. ACM Press, New York, New York, USA, 545–554. <https://doi.org/10.1145/2541016.2541022>
- [100] Tobias Langlotz, Jonathan Sutton, Stefanie Zollmann, Yuta Itoh, and Holger Regenbrecht. 2018. ChromaGlasses: Computational Glasses for Compensating Colour Blindness. In *Proc.CHI'18*. ACM, New York, NY, USA, 390:1–390:12.
- [101] Douglas Lanman and David Luebke. 2013. Near-eye light field displays. *ACM TOG* 32, 6 (2013), 220.
- [102] Byoungcho Lee, Jong-Young Hong, Changwon Jang, Jinsoo Jeong, and Chang-Kun Lee. 2017. Holographic and light-field imaging for augmented reality. In *Proc. SPIE*, Vol. 10125. 101251A–1.
- [103] Chang-Kun Lee, Seokil Moon, Seungjae Lee, Dongheon Yoo, Jong-Young Hong, and Byoungcho Lee. 2016. Compact three-dimensional head-mounted display system with Savart plate. *Opt. Express* 24, 17 (2016), 19531–19544.

- [104] Gun-Yeal Lee, Jong-Young Hong, SoonHyoung Hwang, Seokil Moon, Hyeokjung Kang, Sohee Jeon, Hwi Kim, Jun-Ho Jeong, and Byoung-ho Lee. 2018. Metasurface eyepiece for augmented reality. *Nature comm/s* 9, 1 (2018), 1–10.
- [105] Jin Su Lee, Yoo Kwang Kim, Mu Young Lee, and Yong Hyub Won. 2019. Enhanced see-through near-eye display using time-division multiplexing of a Maxwellian-view and holographic display. *Opt. Express* 27, 2 (Jan 2019), 689–701.
- [106] S. Lee, J. Cho, B. Lee, Y. Jo, C. Jang, D. Kim, and B. Lee. 2018. Foveated Retinal Optimization for See-Through Near-Eye Multi-Layer Displays. *IEEE Access* 6 (2018), 2170–2180. <https://doi.org/10.1109/ACCESS.2017.2782219>
- [107] Seungjae Lee, Mengfei Wang, Gang Li, Lu Lu, Yusufu Sulai, Changwon Jang, and Barry Silverstein. 2020. Foveated near-eye display for mixed reality using liquid crystal photonics. *Scientific Reports* 10, 1 (2020), 1–11.
- [108] Peter Lincoln, Alex Blate, Montek Singh, Andrei State, Mary C. Whitton, Turner Whitted, and Henry Fuchs. 2017. Scene-adaptive High Dynamic Range Display for Low Latency Augmented Reality. In *Proceedings of the 21st ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (I3D '17)*. ACM, New York, NY, USA, Article 15, 7 pages.
- [109] Peter Lincoln, Alex Blate, Montek Singh, Turner Whitted, Andrei State, Anselmo Lastra, and Henry Fuchs. 2016. From Motion to Photons in 80 Microseconds: Towards Minimal Latency for Virtual and Augmented Reality. *IEEE TVCG* 22, 4 (apr 2016), 1367–1376. <https://doi.org/10.1109/TVCG.2016.2518038>
- [110] Peter H Lindsay and Donald A Norman. 2013. *Human information processing: An introduction to psychology*. Academic press.
- [111] Sheng Liu, Dewen Cheng, and Hong Hua. 2008. An optical see-through head mounted display with addressable focal planes. In *IEEE ISMAR 2008*. 33–42. <https://doi.org/10.1109/ISMAR.2008.4637321>
- [112] Shuxin Liu, Yan Li, Pengcheng Zhou, Xiao Li, Na Rong, Shuaijia Huang, Wenqing Lu, and Yikai Su. 2016. A multi-plane optical see-through head mounted display design for augmented reality applications. *Journal of the Society for Information Display* 24, 4 (2016), 246–251.
- [113] Mark A Livingston and Zhuming Ai. 2008. The effect of registration error on tracking distant augmented objects. In *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*. IEEE, 77–86.
- [114] Andrew Maimone and Henry Fuchs. 2013. Computational augmented reality eyeglasses. In *IEEE ISMAR 2013*. IEEE, 29–38.
- [115] Andrew Maimone, Andreas Georgiou, and Joel S. Kollin. 2017. Holographic Near-eye Displays for Virtual and Augmented Reality. *ACM TOG* 36, 4, Article 85 (July 2017), 16 pages. <https://doi.org/10.1145/3072959.3073624>
- [116] Andrew Maimone, Douglas Lanman, Kishore Rathinavel, Kurtis Keller, David Luebke, and Henry Fuchs. 2014. Pinlight displays: wide field of view augmented reality eyeglasses using defocused point light sources. In *ACM SIGGRAPH 2014 Emerging Technologies*. ACM, 20.
- [117] Andrew Maimone and Junren Wang. 2020. Holographic optics for thin and lightweight virtual reality. *ACM Transactions on Graphics (TOG)* 39, 4 (2020), 67–1.
- [118] Belen Masia, Gordon Wetzstein, Piotr Didyk, and Diego Gutierrez. 2013. A survey on computational displays: Pushing the boundaries of optics, computation, and perception. *Computers & Graphics* 37, 8 (2013), 1012–1038.
- [119] Nathan Matsuda, Alexander Fix, and Douglas Lanman. 2017. Focal surface displays. *ACM TOG* 36, 4 (2017), 1–14.
- [120] Allison McKendrick and Chris Johnson. 2011. *Temporal Properties of Vision*. Elsevier Health Sciences.
- [121] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. 1995. Augmented reality: A class of displays on the reality-virtuality continuum. In *Photonics for Industrial Applications*. International Society for Optics and Photonics, 282–292.
- [122] Mark R. Mine. 1993. *Characterization of End-to-End Delays in Head-Mounted Display Systems*. Technical Report. Chapel Hill, NC, USA.
- [123] Eunkyong Moon, Myeongjae Kim, Jinyoung Roh, Hwi Kim, and Joonku Hahn. 2014. Holographic head-mounted display with RGB light emitting diode light source. *Optics express* 22, 6 (2014), 6526–6534.
- [124] S. Mori, S. Ikeda, A. Plopski, and C. Sandor. 2018. BrightView: Increasing Perceived Brightness of Optical See-Through Head-Mounted Displays Through Unnoticeable Incident Light Reduction. In *IEEE VR 2018*. 251–258.
- [125] Hiroshi Mukawa, Katsuyuki Akutsu, Ikuo Matsumura, Satoshi Nakano, Takuji Yoshida, Mieko Kuwahara, Kazuma Aiki, and Masataka Ogawa. 2008. 8.4: Distinguished Paper: A Full Color Eyewear Display Using Holographic Planar Waveguides. *SID Symposium Digest of Technical Papers* 39, 1 (2008), 89–92. <https://doi.org/10.1889/1.3069819>
- [126] Robert L. Myers. 2002. *Display Interfaces: Fundamentals and Standards*. John Wiley & Sons.
- [127] Hajime Nagahara, Yasushi Yagi, and Masahiko Yachida. 2006. Super Wide Field of View Head Mounted Display Using Catadioptrical Optics. *Presence: Teleoper. Virtual Environ.* 15, 5 (Oct. 2006), 588–598.
- [128] Rahul Narain, Rachel A. Albert, Abdullah Bulbul, Gregory J. Ward, Martin S. Banks, and James F. O'Brien. 2015. Optimal Presentation of Imagery with Focus Cues on Multi-Plane Displays. *ACM Trans. Graph.* 34, 4, Article 59 (July 2015), 12 pages. <https://doi.org/10.1145/2766909>
- [129] Albert Ng, Julian Lepinski, Daniel Wigdor, Steven Sanders, and Paul Dietz. 2012. Designing for Low-latency Direct-touch Input. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology (UIST '12)*. ACM, New York, NY, USA, 453–464. <https://doi.org/10.1145/2380116.2380174>



- [130] Cornelis Noorlander, Jan J. Koenderink, Ron J. Den Olden, and B. Wigbold Edens. 1983. Sensitivity to spatiotemporal colour contrast in the peripheral visual field. *Vision Research* 23, 1 (1983), 1 – 11.
- [131] Kohei Oshima, Kenneth R Moser, Damien Constantine Rompapas, J Edward Swan, Sei Ikeda, Goshiro Yamamoto, Takafumi Taketomi, Christian Sandor, and Hirokazu Kato. 2016. SharpView: Improved clarity of defocused content on optical see-through head-mounted displays. In *2016 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 173–181.
- [132] Kazuki Otao, Yuta Itoh, Hiroyuki Osone, Kazuki Takazawa, Shunnosuke Kataoka, and Yoichi Ochiai. 2017. Light field blender: designing optics and rendering methods for see-through and aerial near-eye display. In *SIGGRAPH Asia 2017 Technical Briefs*. 1–4.
- [133] Charles B. Owen, Ji Zhou, Arthur Tang, and Fan Xiao. 2004. Display-Relative Calibration for Optical See-Through Head-Mounted Displays. In *Proc. 3rd IEEE/ACM ISMAR (ISMAR 2004)*. IEEE Computer Society, USA, 70–78.
- [134] Nitish Padmanaban, Robert Konrad, Tal Stramer, Emily A Cooper, and Gordon Wetzstein. 2017. Optimizing virtual reality for all users through gaze-contingent and adaptive focus displays. *Proceedings of the National Academy of Sciences* 114, 9 (2017), 2183–2188.
- [135] Nitish Padmanaban, Yifan Peng, and Gordon Wetzstein. 2019. Holographic near-eye displays based on overlap-add stereograms. *ACM Transactions on Graphics (TOG)* 38, 6 (2019), 1–13.
- [136] Anjul Patney, Marco Salvi, Joohwan Kim, Anton Kaplanyan, Chris Wyman, Nir Benty, David Luebke, and Aaron Lefohn. 2016. Towards foveated rendering for gaze-tracked virtual reality. *ACM TOG* 35, 6 (2016), 179.
- [137] L. Qian, A. Plopski, N. Navab, and P. Kazanzides. 2018. Restoring the Awareness in the Occluded Visual Field for Optical See-Through Head-Mounted Displays. *IEEE TVCG* 24, 11 (Nov 2018), 2936–2946.
- [138] Kishore Rathinavel, Praneeth Chakravarthula, Kaan Akşit, Josef Spjut, Ben Boudaoud, Turner Whitted, David Luebke, and Henry Fuchs. 2018. Steerable application-adaptive near eye displays. In *ACM SIGGRAPH 2018 Emerging Technologies*. ACM.
- [139] K. Rathinavel, H. Wang, A. Blate, and H. Fuchs. 2018. An Extended Depth-at-Field Volumetric Near-Eye Augmented Reality Display. *IEEE TVCG* 24, 11 (Nov 2018), 2857–2866. <https://doi.org/10.1109/TVCG.2018.2868570>
- [140] Kishore Rathinavel, Gordon Wetzstein, and Henry Fuchs. 2019. Varifocal Occlusion-Capable Optical See-through Augmented Reality Display based on Focus-tunable Optics. *IEEE TVCG* 25, 11 (2019), 3125–3134.
- [141] Holger Regenbrecht, Graham McGregor, Claudia Ott, Simon Hoermann, Thomas Schubert, Leigh Hale, Julia Hoermann, Brian Dixon, and Elizabeth Franz. 2011. Out of reach?-A novel AR interface approach for motor rehabilitation. In *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on*. IEEE, 219–228.
- [142] Holger Regenbrecht, Jung-Woo (Noel) Park, Claudia Ott, Steven Mills, Matthew Cook, and Tobias Langlotz. 2019. Preaching Voxels: An Alternative Approach to Mixed Reality. *Frontiers in ICT* 6 (2019), 7.
- [143] Holger Regenbrecht, Arne Reepen, Katrin Meng, Stephan Beck, and Tobias Langlotz. 2017. Mixed Voxel Reality: Presence and Embodiment in Low Fidelity, Visually Coherent, Mixed Reality Environments. In *IEEE ISMAR 2017*. IEEE, 219–228.
- [144] Gerhard Reitmayr, Tobias Langlotz, Daniel Wagner, Alessandro Mulloni, G. Schall, D. Schmalstieg, and Qi Pan. 2010. Simultaneous Localization and Mapping for Augmented Reality. In *Ubiquitous Virtual Reality (ISUVR), 2010 International Symposium on*. 5–8. <https://doi.org/10.1109/ISUVR.2010.12>
- [145] JP Rolland and Hong Hua. 2005. Head-mounted display systems. *Encyclopedia of optical engineering* (2005), 1–13.
- [146] Jannick P Rolland and Henry Fuchs. 2000. Optical versus video see-through head-mounted displays in medical visualization. *Presence: Teleoperators and Virtual Environments* 9, 3 (2000), 287–309.
- [147] Janick P Rolland, Richard L Holloway, and Henry Fuchs. 1994. A Comparison of Optical and Video See-Through Head-Mounted Displays. *SPIE Telemanipulator and Telepresence Technologies* 2351 (1994), 293–307.
- [148] Jannick P. Rolland, Myron W. Krueger, and Alexei Goon. 2000. Multifocal planes head-mounted displays. *Appl. Opt.* 39, 19 (Jul 2000), 3209–3215. <https://doi.org/10.1364/AO.39.003209>
- [149] Jannick P. Rolland, Kevin P. Thompson, Hakan Urey, and Mason Thomas. 2012. *See-Through Head Worn Display (HWD) Architectures*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2145–2170.
- [150] Jannick P. Rolland, Akitoshi Yoshida, Larry D. Davis, and John H. Reif. 1998. High-resolution inset head-mounted display. *Appl. Opt.* 37, 19 (Jul 1998), 4183–4193. <https://doi.org/10.1364/AO.37.004183>
- [151] Damien Constantine Rompapas, Aitor Rovira, Alexander Plopski, Christian Sandor, Takafumi Taketomi, Goshiro Yamamoto, Hirokazu Kato, and Sei Ikeda. 2017. EyeAR: Refocusable augmented reality content through eye measurements. *Multimodal Technologies and Interaction* 1, 4 (2017), 22.
- [152] Pedro Santos, Thomas Gierlinger, Oliver Machui, and André Stork. 2008. The daylight blocking optical stereo see-through HMD. In *Proceedings of the 2008 workshop on immersive projection technologies*. ACM, 4.
- [153] Helge Seetzen, Wolfgang Heidrich, Wolfgang Stuerzlinger, Greg Ward, Lorne Whitehead, Matthew Trentacoste, Abhijeet Ghosh, and Andrejs Vorozcovs. 2004. High Dynamic Range Display Systems. *ACM ToG* 3 (2004), 760–768.
- [154] Liang Shi, Fu-Chung Huang, Ward Lopes, Wojciech Matusik, and David Luebke. 2017. Near-eye Light Field Holographic Rendering with Spherical Waves for Wide Field of View Interactive 3D Computer Graphics. *ACM TOG* 36, 6, Article



- 236 (Nov. 2017), 17 pages. <https://doi.org/10.1145/3130800.3130832>
- [155] Weitao Song, Yongtian Wang, Dewen Cheng, and Yue Liu. 2014. Light field head-mounted display with correct focus cue using micro structure array. *Chinese Optics Letters* 12, 6 (2014), 060010.
  - [156] J. Spjut, B. Boudaoud, J. Kim, T. Greer, R. Albert, M. Stengel, K. Akşit, and D. Luebke. 2020. Toward Standardized Classification of Foveated Displays. *IEEE TVCG* 26, 5 (2020), 2126–2134.
  - [157] R.B. Sprague. 2010. Method and apparatus to process display and non-display information. WO Patent App. PCT/US2009/055,758.
  - [158] Srikanth Kirshnamachari Sridharan, Juan David Hincapié-Ramos, David R Flatla, and Pourang Irani. 2013. Color correction for optical see-through displays using display color profiles. In *ACM VRST 2013*. 231–240.
  - [159] Hans Strasburger and Ernst Pöppel. 2002. *Visual field*.
  - [160] I. E. Sutherland. 1968. A head-mounted three dimensional display. *Fall Joint Computer Conference* (1968), 757–764.
  - [161] Shiro Suyama, Munekazu Date, and Hideaki Takada. 2000. Three-dimensional display system with dual-frequency liquid-crystal varifocal lens. *Japanese Journal of Applied Physics* 39, 2R (2000), 480.
  - [162] Mccollum Thelma. 1945. Stereoscopic television apparatus. <http://www.freepatentsonline.com/2388170.html>
  - [163] Mihran Tuceryan, Yakup Genc, and Nassir Navab. 2002. Single-Point Active Alignment Method (SPAAM) for Optical See-through HMD Calibration for Augmented Reality. *Presence* 11, 3 (June 2002), 259–276.
  - [164] Takayuki Uchida, K Sato, and S Inokuchi. 2002. An optical see-through MR display with digital micro-mirror device. *Transactions of the Virtual Reality Society of Japan* 7, 2 (2002).
  - [165] Daniel Wagner, Tobias Langlotz, and Dieter Schmalstieg. 2008. Robust and Unobtrusive Marker Tracking on Mobile Phones. In *Proc. of the 7th IEEE/ACM ISMAR (ISMAR '08)*. IEEE Computer Society, Washington, DC, USA, 121–124.
  - [166] Brian A. Wandell. 1995. *Foundations of Vision*. Sinauer Associates.
  - [167] C. Weiland, A. Braun, and W. Heiden. 2009. Colorimetric and Photometric Compensation for Optical See-Through Displays. (2009), 603–612.
  - [168] Gordon Wetzstein, Wolfgang Heidrich, and David Luebke. 2010. Optical image processing using light modulation displays. 29, 6 (2010), 1934–1944.
  - [169] Charles Wheatstone. 1838. XVIII. Contributions to the physiology of vision. – Part the first. on some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical transactions of the Royal Society of London* 128 (1838), 371–394.
  - [170] Austin Wilson and Hong Hua. 2017. Design and prototype of an augmented reality display with per-pixel mutual occlusion capability. *Opt. Express* 25, 24 (2017), 30539–30549.
  - [171] Austin Wilson and Hong Hua. 2019. Design and demonstration of a vari-focal optical see-through head-mounted display using freeform Alvarez lenses. *Optics express* 27, 11 (2019), 15627–15637.
  - [172] Xinxing Xia, Yunqing Guan, Andrei State, Praneeth Chakravarthula, Kishore Rathinavel, Tat-Jen Cham, and Henry Fuchs. 2019. Towards a Switchable AR/VR Near-eye Display with Accommodation-Vergence and Eyeglass Prescription Support. *IEEE TVCG* 25, 11 (2019), 3114–3124.
  - [173] Lei Xiao, Anton Kaplanyan, Alexander Fix, Matthew Chapman, and Douglas Lanman. 2018. DeepFocus: Learned Image Synthesis for Computational Displays. *ACM Trans. Graph.* 37, 6, Article 200 (Dec. 2018), 13 pages.
  - [174] Miaomiao Xu and Hong Hua. 2017. High dynamic range head mounted display based on dual-layer spatial modulation. *Opt. Express* 25, 19 (Sep 2017), 23320–23333. <https://doi.org/10.1364/OE.25.023320>
  - [175] N. Padmanaban G. Wetzstein Y. Peng, S. Choi. 2020. Neural Holography with Camera-in-the-loop Training. *ACM Trans. Graph. (SIGGRAPH Asia)* (2020).
  - [176] Yuta Yamaguchi and Yasuhiro Takaki. 2016. See-through integral imaging display with background occlusion capability. *Applied optics* 55, 3 (2016), A144–A149.
  - [177] Myron Yanoff and Jay Duker. 2018. *Ophthalmology* (5th edition. ed.). Elsevier Saunders, Edinburgh.
  - [178] Fahri Yaraş, Hoonjong Kang, and Levent Onural. 2010. State of the art in holographic displays: a survey. *Journal of display technology* 6, 10 (2010), 443–454.
  - [179] Fahri Yaraş, Hoonjong Kang, and Levent Onural. 2011. Circular holographic video display system. *Optics Express* 19, 10 (2011), 9147–9156.
  - [180] Chanhjung Yoo, Kiseung Bang, Changwon Jang, Dongyeon Kim, Chang-Kun Lee, Geeyoung Sung, Hong-Seok Lee, and Byoungcho Lee. 2019. Dual-focal waveguide see-through near-eye display with polarization-dependent lenses. *Optics letters* 44, 8 (2019), 1920–1923.
  - [181] Yang Zhao, Nathan Matsuda, Xuan Wang, Marina Zannoli, and Douglas Lanman. 2020. High dynamic range near-eye displays. In *Optical Architectures for Displays and Sensing in Augmented, Virtual, and Mixed Reality*, Vol. 11310. International Society for Optics and Photonics, SPIE, 268 – 279. <https://doi.org/10.1117/12.2546687>
  - [182] Feng Zheng, Turner Whitted, Anselmo Lastra, Peter Lincoln, Andrei State, Andrew Maimone, and Henry Fuchs. 2014. Minimizing latency for augmented reality displays: Frames considered harmful. In *IEEE ISMAR 2014*. 195–200.

Received May 2020; revised Nov. 2020, Jan 2021; accepted Feb. 2021