

# Design and Characterization of Light Field and Holographic Near-eye Displays

Erdem Sahin<sup>[0000–0002–5371–6649]</sup>, Jani Mäkinen, Ugur Akpınar,  
Yuta Miyanishi<sup>[0000–0001–8305–2823]</sup>, and Atanas Gotchev<sup>[0000–0003–2320–1000]</sup>

Faculty of Information Technology and Communication Sciences, Tampere University, FI  
firstname.lastname@tuni.fi

**Abstract.** The light field and holographic displays constitute two important categories of advanced 3D displays that are aimed at delivering all physiological depth cues of the human visual system, such as stereo cues, motion parallax, and focus cues, with sufficient accuracy. As human observers are the end-users of such displays, the delivered spatial information (e.g., perceptual spatial resolution) and view-related image quality factors (e.g., focus cues) are usually determined based on human visual system characteristics, which then defines the display design. Retinal image formation models enable rigorous characterization and efficient design of light field and holographic displays. In this chapter the ray-based near-eye light field and wave-based near-eye holographic displays are reviewed, and the corresponding retinal image formation models are discussed. In particular, most of the discussion is devoted on characterization of the perceptual spatial resolution and focus cues.

**Keywords:** Light Field Display · Holographic Display · Display Characterization · Perceptual Resolution · Focus Cues

## 1 Introduction

The light field (LF) and holographic displays are ideally aimed at delivering the true LFs or wave fields, respectively, of three-dimensional (3D) scenes to human observers. Unlike conventional stereoscopic displays, they are capable of delivering all physiological cues of the human visual system (HVS), namely stereo cues, continuous motion parallax, and focus cues (accommodation and retinal blur), with sufficient accuracy [4,58]. In stereoscopic near-eye displays, the continuous motion parallax is readily available via head-tracking, though the parallax due to the eye movement is missing, which is not usually seen as a big problem. Enabling focus cues, on the other hand, has particularly been set as an important goal in the design of next generation advanced near-eye displays, to alleviate the so-called vergence-accommodation conflict [18]. From this perspective, the LF and holographic display technologies are expected to play a critical role in the future near-eye display technology.

The richness of required delivered information (correspondingly the resulting size of the visual data) has fundamentally limited the realization of fully immersive LF and holographic displays in practice, both computationally and optically. Full immersion is only possible with the delivery of sufficiently high (spatial) resolution image, in a

sufficiently wide field of view (FOV), and enabling accurate 3D perception, which together set very demanding constraints on the display design. Given the limited optical as well as computational resources, the problem is then to optimize these resources in the best possible way that would maximize the user experience and immersion. Usually one has to compromise between the spatial quality of the observed image, which can be characterized by perceivable resolution, and the accuracy of reproduced depth cues (in particular focus cues). Thus, to determine this compromise for optimal display design and further characterize a given design, computational frameworks enabling characterization of perceived spatial image quality as well as 3D perception are of critical importance. This Chapter aims at providing a general perspective of such frameworks (or models) in design and characterization of LF and holographic displays. Section 2 provides background information about HVS relevant for developing such models. Section 3 and Section 4, respectively, overviews the basics of LF and holographic display technologies, existing near-eye displays utilizing them, and computational models for their design and characterization.

## 2 Basics of human spatial vision

The flow of visual information from outside world till to the visual cortex can be described in three stages: retinal image encoding, neural representation of visual information, and perception [53]. This section briefly overviews several important functions of human vision in such stages to particularly elaborate on the *perceptual spatial resolution* and *focus cues*, which are further addressed in the following sections as two critical aspects in design and characterization of LF and holographic displays.

### 2.1 Perceptual spatial resolution

Retinal image encoding takes place in the eye. The optics of the eye is not that different from a camera: the cornea and crystalline lens of the eye are responsible of focusing the light (zoom lens), the iris works as an aperture, and the photoreceptors (pixels) on the retina (sensor) samples the incident light. The sampled optical information by the photoreceptors is turned to electrical signals and further carried to brain via retinal ganglion cells that are connected to photoreceptors via various interneurons. The retinal ganglion cells performs a subsequent sampling based on their receptive fields on the retina, and thus, in a sense, they are the output cells of the human eye [56]. Unlike the sensor in a conventional camera, the retina is curved, and the photoreceptors are non-uniformly distributed on the retina. Furthermore, there are different types of photoreceptors with different sensitivities to light. The densely arranged cones are responsible for high spatial color vision under photopic (high light) conditions, rather sparse set of rods are more sensitive to light and thus they are responsible for scotopic (low light) vision. Three different types of cones (S, M, L) have different sensitivities to wavelength, which enables color encoding and perception. The cones are densely arranged (as dense as 120 cones per degree for M and S cones [10]) only in the central retina (fovea), covering a central visual field of 5 degrees. In fovea the retinal ganglion cells are known to sample a single cone, which enables as high as 60 cycles per degree (CPD) foveal visual acuity.

Here CPD refers to the number of full cycles, in the observed pattern or stimulus (e.g., sinusoidal grating), in one degree of the visual angle. The visual acuity falls rapidly outside the fovea due to decrease in the density of cones. A formula for receptive field densities of retinal ganglion is given in [56], which can be utilized to explicitly describe eccentricity-dependent visual acuity.

Besides the sampling carried out by photoreceptors and subsequently by retinal ganglion cells, the lens, cornea and iris together also define the initial encoding of the visual information. The 2 mm - 8 mm range of pupil (aperture) size, is controlled by iris based on the amount of illumination. The total optical power of cornea and lens is around 60 diopters (D), when focused at infinity; and it can be increased up to 80 D to focus near objects. The distortions introduced by the optics is reduced for smaller aperture sizes. The diffraction, on the other hand, limits the resolution at small apertures. Below around 2.4 mm of aperture size, the optics of the eye does not introduce an extra blur compared to diffraction. The diffraction-limited resolution at 2.4 mm aperture size for the central wavelength of 550 nm is also around 60 CPD, which matches the density of cone photoreceptor sampling in fovea. Thus, the optics seems to do the necessary antialiasing, and the cones sample the incoming light in a most efficient way [53].

The neural representation starts just after the sampling of the signal by the photoreceptors on the retina. The visual information is transferred to visual cortex through various segregated visual pathways specialized for different visual tasks such as depth perception. The information carried in such pathways is more about the contrast of the signal captured on the retina, rather than mean illumination level, which explains our ability to response to a wide range of luminance levels [53]. For sinusoidal gratings, which are usually used in contrast measurements, the (Michelson) contrast is defined as

$$C = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \quad (1)$$

where  $I_{\max}$  and  $I_{\min}$  represent the highest and lowest luminance levels of the grating. The typical task used in measuring *contrast sensitivity function* of HVS is distinguishing gratings of a given spatial frequency from a uniform background. Below some level of contrast the observer is no longer able to detect the pattern. The contrast at this point is called the *threshold* contrast, and its reciprocal is defined as the *sensitivity* at that spatial frequency. The physiological methods enable measuring contrast sensitivity function (CSF) at single neuron level. The behavioral measurements, on the other, can provide the psychophysical CSF of the HVS. The CSF measured this way, usually with incoherent light sources, includes the response of whole system, i.e. the optics as well as neural response. In particular, the optical and neural modulation transfer functions (MTFs) determines the amount of contrast reduction in each stage depending on the spatial frequency of the input. By further utilizing the interferometric methods [35], which essentially avoids the blurring due to optics, the neural-only MTF and correspondingly the optics-only MTF of the HVS can be also estimated [53]. Methods such as wavefront sensing enables direct measurement of the optics-only MTF, modeling also possible aberrations [32]. Physiological data when combined with such behavioral measurements has lead rigorous models of eye optics [40]. Decades of such efforts have also established reliable methodologies to measure contrast thresholds [43]. A typical CSF of the HVS in photopic condition is shown in Figure 1. Based on Figure 1, one can

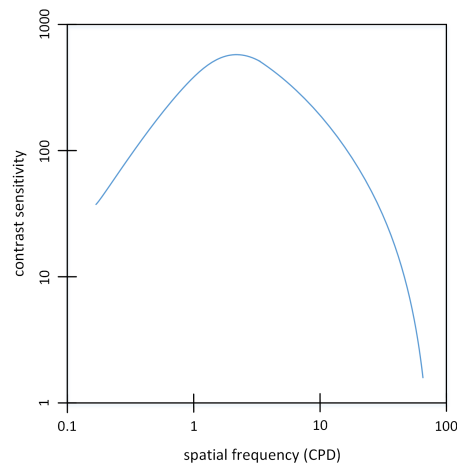


Fig. 1: A typical CSF in the central visual field (adapted from [30]).

observe that the bandwidth of the CSF, which is around 60 CPD, is mainly dictated by the initial sampling on the retina through the optics of the eye. On the other hand, based on the knowledge of optical MTF, the decrease of sensitivity in the lower bands can be attributed to neural factors.

The *perceptual resolution*, or visual acuity, of HVS can be thoroughly characterized based on the spatial CSF. That is, it can provide a more complete information compared to the visual acuity tests usually employed in ophthalmologic tests. In such tests the visual acuity is defined by determining the smallest size at which patient can detect the orientation of the target (E chart, Landolt C) or recognize a specific target among many others (Snellen letters), where the only variable is basically the size of the targets and furthermore the difference of brightness between the letters and the background is fixed. Such test provide good information about the limits of visual acuity, whereas the CSF characterizes the spatial vision over the entire spatial frequencies detectable by HVS.

It is important to note that, in practice, the CSF depends not only on the spatial frequency of the input but also many other factors such as the location of the stimulus in the visual field [48], mean illumination level [52], the temporal frequency of the target [26], the wavelength of the light [57], etc. Furthermore, due to the so-called (contrast) masking effect, a strong "masker stimulus" can also alter (reduce) the visibility of the test pattern under consideration, which typically has similar spatial frequency as the masker stimulus, thus reducing the contrast sensitivity at the test frequency [39]. Such a case can easily occur in natural scenes. Thus, in practice, all such factors need to be taken into account for complete characterization of CSF.

The joint interdisciplinary effort in understanding the human spatial vision has led development of rigorous computational models. By providing the multi-channel (multiple spatial and temporal frequencies as well as different orientations of stimulus) response of HVS to given static or dynamic visual scenes, such models enable perceptual quality assessment of such visual data [9,46,6]. This can, then, guide design of

perceptually optimized displays as well as cameras, also image and video coding and compression algorithms, etc,

## 2.2 Stereo and focus cues

Among several other tasks accomplished in the visual cortex such as color perception, motion perception, *depth perception* is particularly important in the context of 3D displays in pursuing the reality of the recreated scene. In HVS not only ‘the signal from the sensor’, namely the retinal image, but also the signals related to eye movements and the accumulated information in the brain complexly contribute to depth perception. In improving the visual experience in VR, it is crucial to understand how human vision works to derive the perception from these various information sources. HVS relies on an extensive set of visual depth cues for depth perception. Below basics of the stereo and focus cues are discussed, which constitute an important set of physiological cues particularly relevant in the design and characterization of near-eye displays.

Stereo cues of vergence and, in particular, binocular disparity constitute the main source of information for depth perception for most of the people (but not all [11]). To be able to create sharp image of an object on the retina, the two eyes converge (or diverge) at it by rotating the two eyes in opposite directions. By this way, the object is fixated at the fovea, where the input can be sampled at highest spatial frequency. The fixated image points on the two retinæ does not have binocular disparity, i.e., exhibit zero disparity. An object nearer or further than the fixation point, on the other hand, is projected at different locations on the two retinæ (see Figure 2). The depth difference between the fixation point and the object corresponds to the magnitude of the binocular disparity of the object. Based on this disparity cue, the eyes can fastly converge (or diverge) at the closer (or further) object. Thus, both the oculomotor function vergence and the (stereo) image-related measure binocular disparity provide signals for depth perception. The just-noticeable change (JNC) in binocular disparity is as small as 10 arcsec. at fixated depth, which is inversely proportional to interpupillary distance (IPD), but it increases dramatically in front of or behind it due to decrease in spatial acuity at higher eccentricities [16].

The abovementioned stereo cues typically accurately delivered by all conventional stereoscopic near-eye displays. However, the monocular accommodation and retinal (defocus) blur cues are usually missed, which play also critical role in depth estimation in HVS. *Accommodation* refers to the oculomotor function that adjusts the refractive power of the crystalline lens in the eye, to obtain the sharp (focused) retinal image of the fixated object of interest (see Figure 3). The accommodation provides depth information based on the signals acquired from ciliary muscles controlling the crystalline lens.

Objects that are close enough to the accommodated object in depth, objects inside the depth of field (DoF), also produce sharp retinal images. The extend of the DoF is dependent on several factors such as pupil diameter, spatial frequency, wavelength, aberrations in the eye optics, etc. For a pupil size of 4mm diameter, typically it is around  $\pm 0.3$  D, whereas it can extend up to around  $\pm 0.8$  D for smaller pupil sizes [37]. The objects that are outside of this range can be imaged with *retinal blur*. The sensitivity of HVS to retinal defocus blur is limited by the extend of the DoF, i.e., the blurs perceived from different points within the DoF cannot be discriminated. The defocus blur itself is

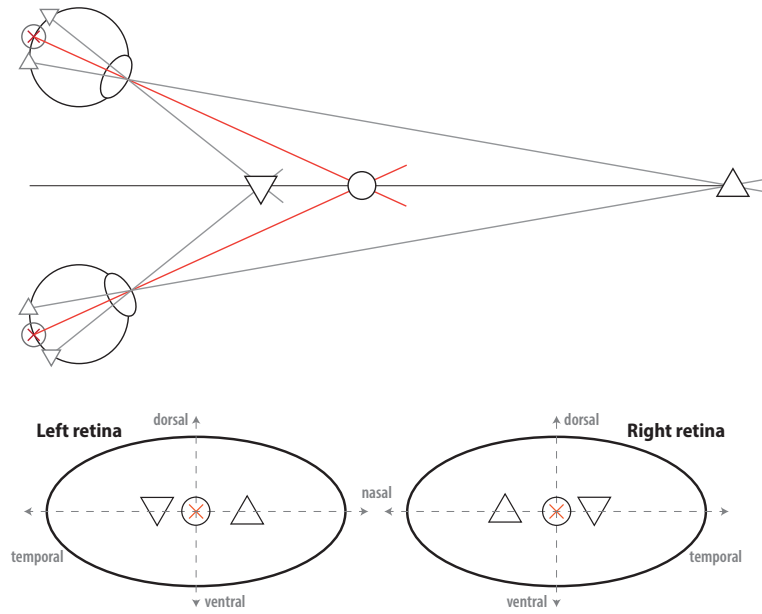


Fig. 2: Geometry of the vergence and the retinal images.

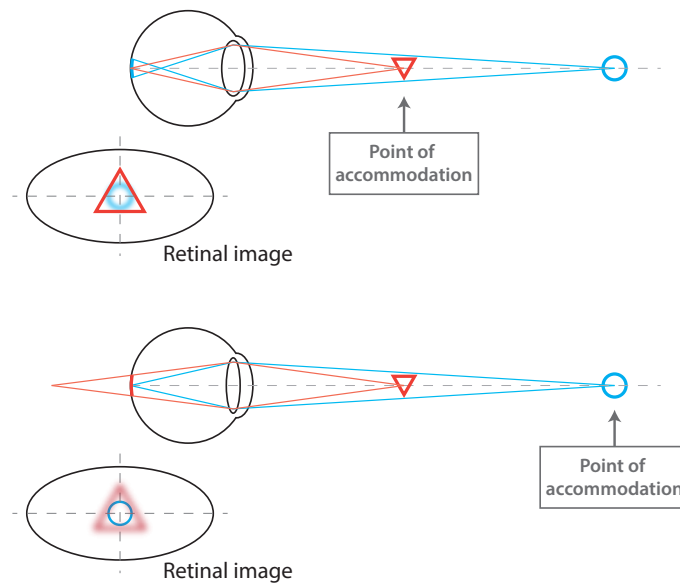


Fig. 3: Accommodation and the retinal image.

an even-error cue, i.e., it does not provide information regarding the direction of depth. However, in practice, the defocus blur is accompanied by the optical aberrations of the eye, which together result in optical blur with odd-error information. The role of chromatic aberration, which can extend up to two-dimensional (2D) in the visible spectrum, has especially dominant role in this, and thus it actually serves as an important depth cue [8]. The amount of optical blur encodes depth information, and it is actually the main source of information that drives accommodation. The magnitude and the gradient of the contrast in the retinal image are the key factors driving the accommodation. The eye is likely to be focused at a depth where both the magnitude and the gradient of the contrast is maximized. The strength of the stimulus also depends on the spatial frequency. The accommodation is most accurate at mid spatial frequencies around 3 – 5 CPD, like the contrast sensitivity itself (see Figure 1), [41]. Unlike the stereo cues, which provides depth information up to a few tens of meters, the accommodation cue is typically available within 2 meters of depth range, since the information coming from defocus blur beyond this range is limited.

The geometry of defocus blur (dictated by the pupil size) is analogous to the geometry of binocular disparity (dictated by the IPD). The JNC in defocus blur is, therefore, an order of magnitude or more bigger than the JNC in binocular disparity (at fixated depth), however it does not increase as rapidly as the JNC in binocular disparity. This way binocular disparity and defocus blur cues are complementary to each other: around the fixated depth, depth estimation mainly relies on the binocular disparity, whereas defocus blur provides more precise information away from it [16]. Furthermore, the oculomotor functions of vergence and accommodation are not only driven by binocular disparity and retinal blur, respectively. Instead, vergence and accommodation are actually coupled, and some part of the information is produced via cross links, i.e., disparity-driven accommodation and blur-driven vergence [45]. This explains the existence of accommodation cue, to some extent, even in conventional stereoscopic displays; and furthermore it is the main motivation of accommodation-invariant displays [27]. Beside such approaches, what is natural to HVS is that accommodation and vergence work in harmony, i.e., they both address the same depth. It is, however, worth to mention that this is usually the case only up to some age, till when the accommodative response is still strong, as by age 40 most of the strength is already lost [49]. In healthy eyes, breaking the harmony creates conflict between the accommodation and vergence and results in undesirable effects such as visual discomfort, fatigue, etc. [18]. Therefore, alleviating such conflict in the HVS has been always one of the main issues to be addressed in the near-eye display community. And this has been mostly aimed to be achieved via enabling the focus cues through more advanced 3D display technologies [19].

As two important categories of advanced 3D display technologies, the LF and holographic displays, and their characterization through perceptual spatial resolution and focus cues are discussed in the following sections. Generally speaking, the overall viewing process of such 3D displays can be formulated by modeling the light transport between the display light sources and the retina, which involves modulation by the display and eye optics. As discussed above, this early stage of retinal image encoding is actually followed by the subsequent neural functions that also play a critical

role in spatial vision and determining the actual perception. However, this is usually ignored or weakly considered and mainly the retinal image itself is taken as the main ingredient of the perceptual analysis and characterization of 3D displays. Furthermore, mostly the eye is usually simplified to constitute a thin lens and a planar retina with uniformly sampled pixels. Although such assumptions lead to suboptimal analysis and characterization, the employed basic framework can be simply extended by including, e.g., the necessary optical aberrations, eccentricity-dependent retinal resolution, neural transfer functions, etc., in the pipeline based on the discussion in this section. The following section provides a detailed discussion on LF display characterization based on the rather simple thin lens and planar retina model, whereas the following section also demonstrate a simulation model for holographic displays including foveation and curved retina surface.

### 3 Characterization of LF displays

#### 3.1 Basics of the LF displays

Under the geometric (ray) optics formalism, the light distribution due to a 3D scene can be fully described via the seven-dimensional plenoptic function  $L(x, y, z, \theta, \phi, \lambda, t)$ , which assumes the light as a collection of rays and describes their intensities through parametrization over the propagation point  $(x, y, z)$ , propagation direction  $(\theta, \phi)$ , spectral content (color)  $\lambda$ , and time  $t$  [1]. The four-dimensional LF is a reduced form of the plenoptic function, describing the intensities of rays at a given time instant and with a given color together with extra assumptions to satisfy unique intensity mappings [31,14]. That is, it is mostly parametrized through two-plane  $L(x, y, s, t)$  or space-angle  $L(x, y, \theta, \phi)$  representations, as shown in Figure 4.

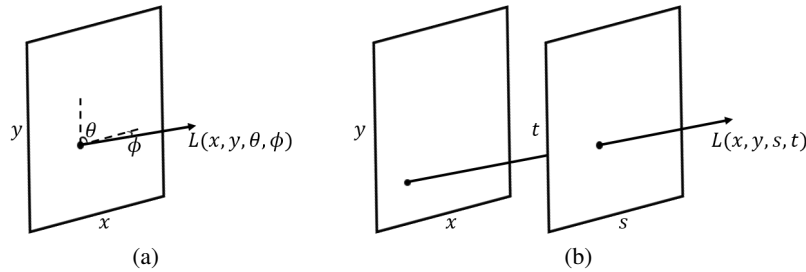


Fig. 4: Two different parametrization of the 4D LF: the space-angle (a), and the two-plane (b).

The performance of the LF display systems can be characterized through their reconstruction capabilities of the continuous 4D LF, i.e., the deliverable spatio-angular resolution. Considering the space-angle parametrization in Figure 4 (a), if the  $(x, y)$  plane is chosen at (or around) the depth of the object of interest, the spatial resolution



corresponds to the minimum resolvable distance between two object points, while the angular resolution refers to the minimum separation angle between the two light rays emitted from the same object point. Equivalently in the two-plane parametrization of Figure 4 (b), one can denote  $(x, y)$  and  $(s, t)$  as the spatial and the angular resolutions, respectively. For instance, if  $(x, y)$  plane is set at the scene depth and  $(s, t)$  is chosen to be the viewing plane, representing the camera or eye pupil location. In both cases, the spatial resolution determines to which extent the spatial information of the scene is reproduced, i.e., it dictates the perceivable spatial resolution, while the angular resolution defines the quality of view-related image aspects, such as occlusions, (direction-dependent) reflectance, (motion) parallax, and focus cues as a particularly important aspect for this Chapter.

Ideally, the LF displays aim to provide the immersive visual experience by stimulating all physiological depth cues of the HVS. In the context of near-eye LF displays, this practically means accurate delivery of accommodation and defocus blur cues to each eye, which are both dictated by the angular resolution of the LF display as explained above. That is, the angular resolution of the LF display should be sufficiently high. On the other hand, the perceivable spatial resolution is another important aspect that can be maximized by increasing the spatial resolution of the LF display. However, the device limitations, such as the total number of available display pixels, display pixel pitch, pitch of lens etc.) as well as the diffractive nature of light do not permit arbitrarily increasing both the spatial and angular resolutions. All traditional LF display techniques, such as integral imaging and super multiview, suffer from such trade-off [4,7]. In the most general form, consistent with the two-plane parametrization, traditional LF displays two layers of optics: an image source plane (e.g., liquid crystal display) and an optical modulator plane (e.g., lens array or parallax barrier), which is placed on top of the image source plane and responsible of directing the light rays emitted from the image source pixels to desired directions. The conventional multiplexing-based technique (distributing the pixels between spatial and angular dimensions) used in such traditional near-eye LF display techniques implements the Nyquist-based sampling and reconstruction of LF [28,20]. This has been recently advanced with the more sophisticated approaches that apply modern compressive sensing methods on multilayer multiplicative displays to benefit from the redundant structure of the LF [21,33,29]. Despite some challenges still to overcome, such as computational complexity, diffraction blur and reduced brightness, such methods have a big potential to significantly improve the delivered spatio-angular resolution.

The efforts on perceptually optimizing the near-eye stereoscopic displays through foveated rendering and optics is particularly important for LF displays, since basically the amount of data is far more in this case. The perceptual optimization of near-eye LF displays not only involves perceptual spatial resolution, which is addressed via foveated rendering, but also optimization of eccentricity-dependent angular LF sampling. This is fundamentally linked to how the focus response of the HVS varies with eccentricity for such displays [50]. The optical as well as computational methods exploiting these perceptual aspects in near-eye LF displays have reported significantly reduced data rate and computational complexity [33,29,24,50].

Rigorous modeling of LF perception is critical to further advance the state of the art in near-eye LF displays and address the abovementioned issues in the area. The following section discusses such simulation models, and further presents a particular case on analysis and characterization of perceptual resolution and focus response in integral imaging type of near-eye display.

### 3.2 Retinal image formation models for LF displays

The existing perceptual simulation models used for characterization and design of the LF displays can be categorized into two main categories. The first category of approaches aim to linking the LF parametrizations of the display and eye (defined over lens and retina planes) through ray propagations and utilize the properties of the eye optics and retina to obtain the so-called perceivable or perceptual LF [47,50]. The perceivable LF defines a perceptually sampled reduced form of the incident LF that is (perceptually) equivalent to any other higher spatial and angular resolution LF incident at the pupil. The second category of approaches model the forward retinal image encoding to find the retinal image itself [22,23,44,2]. The most rigorous approach is to simulate the retinal image formation utilizing the wave optics, which accounts for possible diffraction effects due to both display and eye optics.

In the remaining part of this section, an example wave optics based simulation method [2] is presented in more details to elaborate more on the characterization of perceptual resolution and focus cues in integral imaging (InIm) type of near-eye LF displays.

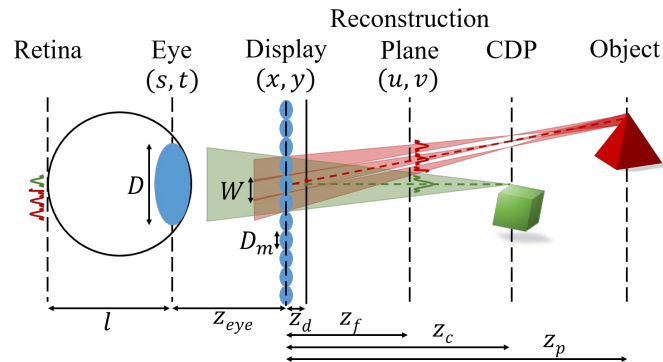


Fig. 5: InIm display operation principle.

The overall InIm display setup is illustrated in Figure 5. The virtual 3D scene is reconstructed through the elemental images behind each microlens, which are focused at the central depth plane (CDP). Assuming aberration-free thin lens model for the eye lens and planar surface for the retina, a conjugate retina plane can be defined at the scene depth where the eye is accommodated. Such plane is denoted as the reconstruction

plane as in Figure 5. Defining the resolution of the reconstruction as the diffraction-limited resolution dictated by the eye pupil, i.e.,  $\Delta_u = 1.22\lambda(z_{eye} + z_f)/D$  where  $D$  is the diameter of the eye pupil, the retinal image can be equivalently analyzed on the reconstruction plane.

An important characterization criterion of LF displays is the angular resolution or the view density, which is defined as the number of distinct view images delivered within the eye pupil. The effect of view density on the accommodative response has been studied in the literature. It is required by the so-called super multiview condition [25] that there should be at least two views delivered within the eye pupil to evoke the accommodation. As will be demonstrated below this is actually satisfied by creating natural defocus blur cue. In the context of InIm, one can define the view density through the number of beams propagating from neighbouring microlenses corresponding the same scene point, as the directional information of an object point is reconstructed by different microlenses. As illustrated in Figure 5, the width  $W$  of each beam at the eye plane is defined by the microlens pitch as  $W = D_m(z_{eye} + z_c)/z_c$ , from which the number of views within the eye pupil can be found as  $N = D/W$ . The retinal image formation at the reconstruction plane can be simulated through integration of the views visible to the eye at a given position. As shown in Figure 5, for a given scene point, each view is contributed by a different elemental image beneath the corresponding microlens, and the contribution is characterized by the corresponding point spread function (PSF). At a given eye position and focused depth, each such contribution is masked by its corresponding elemental field of view (FOV), as illustrated in Figure 5 for the three red beams. The overall image formations can then be formulated as

$$I_p(u, v, \lambda) = \sum_{m=1}^M \sum_{n=1}^N w_{m,n}(u, v) \sum_{k=1}^K \sum_{l=1}^L I(m, n, k, l, \lambda) h_{m,n,k,l}(u, v, \lambda) \quad (2)$$

where  $I(m, n, k, l, \lambda)$  is the intensity of the pixel  $[k, l]$  of the elemental image behind the lens  $[m, n]$ ,  $h_{m,n,k,l}(u, v, \lambda)$  is the PSF of the same pixel at the reconstruction plane  $(u, v)$  (i.e., it accounts also the finite size of the pixel),  $w_{m,n}(u, v)$  is the FoV of the lens  $[m, n]$ , and  $I_p(u, v, \lambda)$  is the final perceived image at wavelength  $\lambda$ . For a given microlens, the elemental FoV  $w_{m,n}(u, v)$  can be calculated via the convolution of the projected eye pupil through the microlens center, with the projected microlens through the center of the eye pupil, both calculated at the reconstruction plane [2]. The PSF  $h_{m,n,k,l}(u, v, \lambda)$  of the display pixels can be derived using the Fresnel diffraction theory as [2]

$$h_{m,n,0,0}(u, v, \lambda) = \text{rect}(M\Delta_s, m\Delta_t) * p_{m,n,0,0}(u, v, \lambda) \quad (3)$$

where

$$p_{m,n,0,0}(u, v, \lambda) \sim \frac{1}{z_f^2} \left| \mathcal{F} \left\{ A_{m,n}(x, y) \exp \left[ \frac{j\pi}{\lambda} \left( \frac{1}{z_d} - \frac{1}{z_f} - \frac{1}{f_m} \right) (x^2 + y^2) \right] \right\} \right|^2 \quad (4)$$

is the actual response for a point on the display plane,  $\Delta_s \times \Delta_t$  denotes the display pixel size, and  $M = z_f/z_d$  defines the magnification.

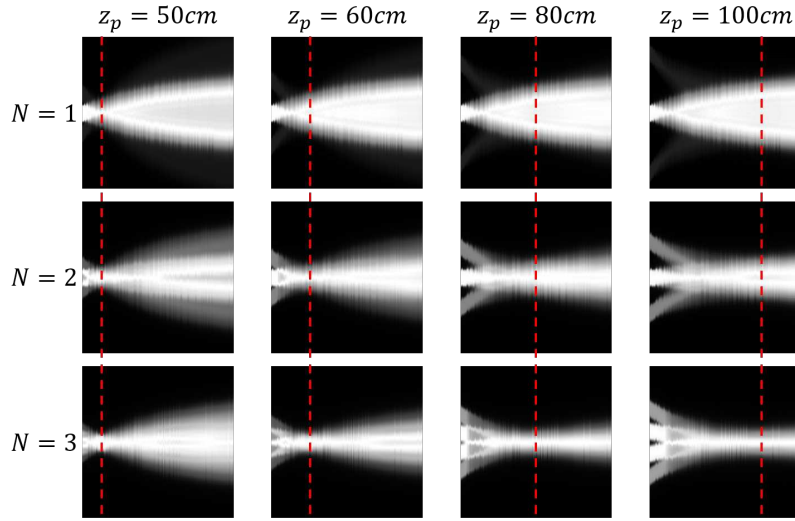


Fig. 6: PSFs at different focused distances for different display setups. A test point point is located at  $z_p$ , where the reconstruction plane  $z_f$  is changed from 40 cm to 120 cm. The CDP is located at  $z_c = 40$  cm. Dashed red lines indicate the locations of the test points.

**Characterization of perceivable spatial resolution and focus cues** The retinal image formation model given by Equation 2 can be used to characterize a near-eye InIm display, e.g., depending on the delivered number of views within the eye pupil. As an example, assume the following simulation parameters according to definitions in Figure 5:  $z_{eye} = 2$  cm,  $D = 5$  mm;  $z_c = 40$  cm,  $D_m = 4$  mm, 2.4 mm, 1.5 mm corresponding to  $N = 1, 2, 3$ , respectively.

The quantitative analysis of perceptual resolution and accommodative response can be performed through evaluating PSFs and the MTFs at different reconstruction distances. Please note that here PSF refers to the total response of the eye and display to an object point. The columns of Figure 6 illustrate the one-dimensional cross-sections of PSFs, obtained through sweeping the scene by changing the focused (reconstruction) plane of the eye from  $z_f = 40$  cm to  $z_f = 120$  cm, for four different test point depths of  $z_p = 50$  cm,  $z_p = 60$  cm,  $z_p = 80$  cm and 100 cm. Such PSF stack is a useful tool for qualitative analysis of focus response. As pointed out in 2, accommodative response is expected at depth where the blur is minimized. As seen in Figure 6, correct accommodative response is observed only when the number of views within the pupil is sufficiently large and the test point is sufficiently close to the CDP of InIm display.

A more rigorous quantitative analysis can be further performed through evaluating the MTFs. Figure 7 illustrates the corresponding results at 5 CPD spatial frequency, which is a good representative case as the CSF of HVS has a peak around such mid spatial frequencies (see Figure 1). Ideally, when comparing different focused depths, the magnitude of the MTF should be maximized at the intended object depth. Together with

this, it is desirable for a 3D display that the gradient of the MTF also maximized at this depth. These are the main factors to evaluate when analyzing the accuracy of delivered accommodative response as well as defocus blur cue. In line with the observations from PSF stack, the results in Figure 7 also demonstrates that the display delivering 3 views into the pupil can achieve maximum frequency response near the actual object depth for each case. The display with 2 views can preserve the accommodative response up to 60 cm, after which the maximum frequency shifts towards the CDP. Finally, as expected, the one-view display cannot deliver correct accommodative response in any case.

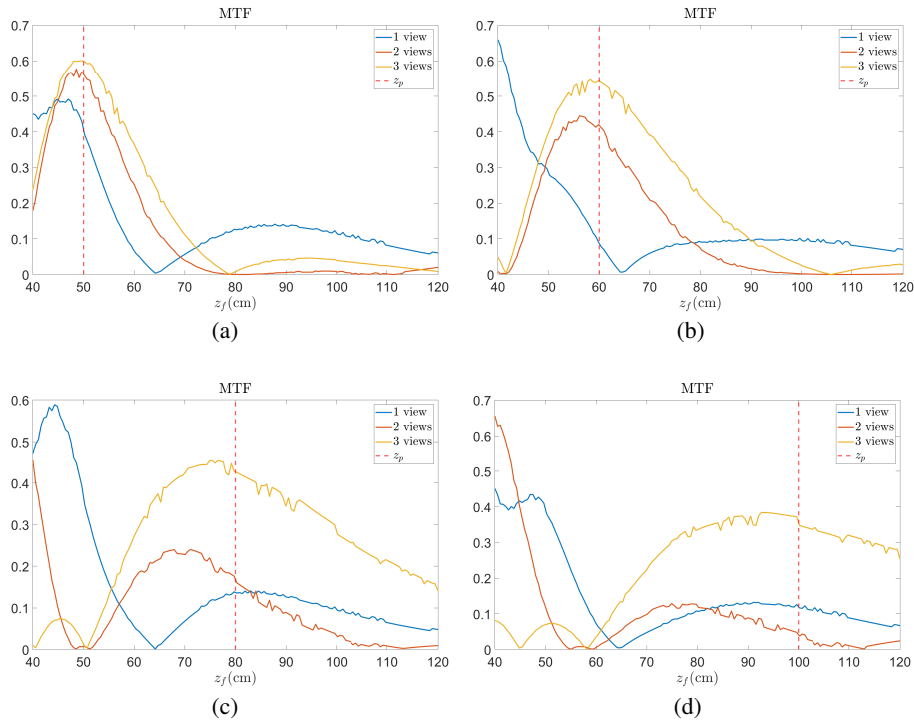


Fig. 7: Comparison of MTF magnitudes at 5 CPD for focused stack images in  $z_f = 40$  cm and  $z_f = 120$  cm. Each subfigure corresponds to a different test point distance indicated by the red dashed line, and shows the results for different number of views within the pupil.

As mentioned before, spatio-angular trade-off is an inherent trade-off in the traditional LF displays such as InIm. This trade-off can be observed in Figure 8 when analyzed together with the above PSF and MTF stack figures. For  $N = 3$ , the higher magnitude and bandwidth of the MTF at  $z_f = 60$  cm compared to  $z_f = 40$  cm explains the correct accommodative response, whereas in the case of  $N = 1$  the MTF magnitude and bandwidth is higher at the CDP  $z_f = 40$  cm, i.e., accommodation cue is not

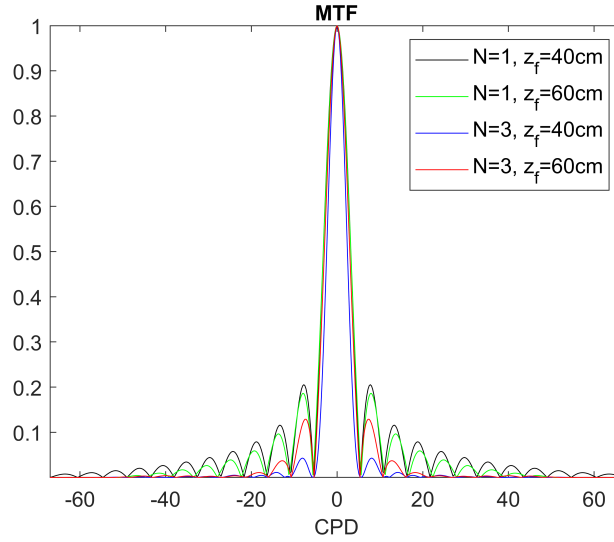


Fig. 8: MTFs corresponding to different number of views  $N = 1$  and  $N = 3$ , and point depths  $z_f = 40$  cm,  $z_f = 60$  cm.

evoked. On the other hand, by comparing the MTFs for  $N=3$  and  $z_f = 60$  cm with  $N=1$  and  $z_f = 40$  cm, it can be observed that the cost of having accurate accommodative response in the former case is the loss in the perceivable spatial resolution.

The qualitative analysis of the accommodative response can be also performed on a 3D scene by superposing the PSFs of scene points given by Equation 2. Figure 9 shows the retinal image, for a scene in depth range from 40 cm to 90 cm, at different focused depths and number of views within the eye pupil. Consistent results can be deduced with the above discussions. In particular, the three-view display clearly delivers the desired accommodative response and defocus blur, whereas the one-view display fails in that.

#### 4 Characterization of holographic displays

Holographic displays provide a desirable alternative solution for 3D displays. They are often considered to be the ultimate display technology due to their ability to provide all of the important 3D related human visual cues. Significant research efforts have undergone to develop both hardware and computational methods for such displays. Particularly due to the constraints imposed by current state of technology, notable focus has been on near-eye display configurations. Below the fundamentals of holographic displays are briefly described, the existing research on holographic near-eye displays is overviewed, and methods for analyzing holographic displays via retinal image formation models are discussed.

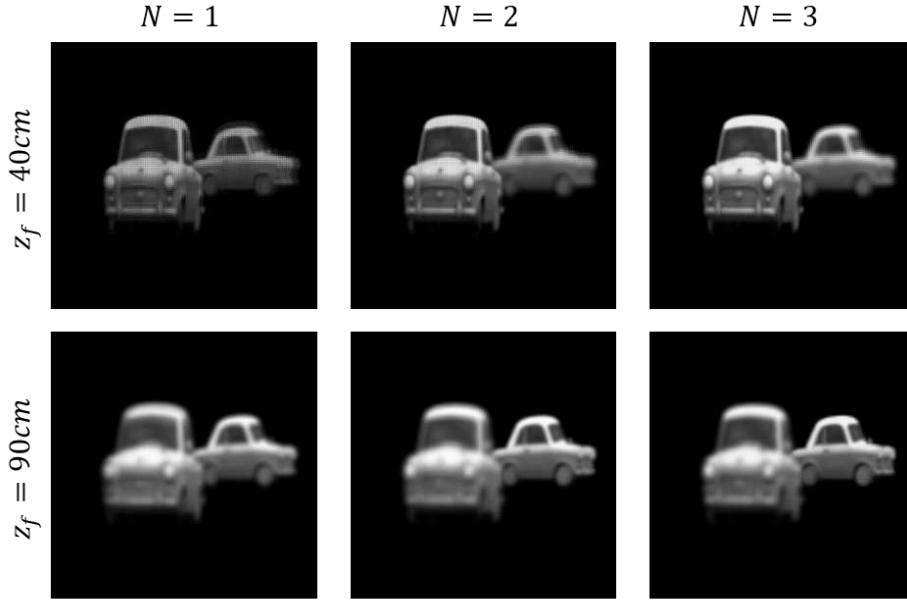


Fig. 9: Simulation results for a 3D scene consisting of two objects. The closer object is located around the CDP,  $z_c = 40$  cm, and the further one is around 90 cm. Slava Z. ©2014 www.sketchfab.com, used under the Creative Commons Attribution license.

#### 4.1 Basics of holographic displays

Holography is a technique that enables recording the wavefront from a 3D object, the object field  $O(x, y)$ , and later reconstruct it without the original object. The wave field contains both the amplitude and the phase information, i.e. the complex amplitude. The interference between the object and a reference wave  $R(x, y)$  (used for reconstructing the field) is recorded on the hologram as

$$I_H(x, y) = |R(x, y) + O(x, y)|^2 = RR^* + OO^* + OR^* + O^*R \quad (5)$$

where the asterisk denotes a complex conjugate operator. The +1 and -1 diffraction orders, i.e. the last two terms in Equation 5, contain the relevant information of the scene, whereas the zero-order term can be discarded in a computer-generated hologram (CGH). Encoding only these diffraction orders results in the bipolar intensity distribution  $\tilde{I}_H(x, y)$  [34]:

$$\tilde{I}_H(x, y) = 2\text{Re}\{O(x, y)R^*(x, y)\} \quad (6)$$

The object field is reconstructed by multiplying  $I_H(x, y)$  with  $R(x, y)$ .

**General characteristics of holographic displays** Holographic displays employ the principles of holography using digital devices, such as spatial light modulators (SLM)

and digital micromirror devices (DMD). By modulating either the amplitude or the phase of the light emitted by the display light source, the desired wave field can be generated and thus displayed to the viewer. Generally such displays are constructed either as amplitude-only or phase-only construction, however both amplitude and phase modulation can be combined to achieve a full complex representation. Mathematically, a holographic display can be described fully by a single plane of complex values and the reference light. In computational simulations of such displays, however, the reference light is often omitted to avoid issues with noise from the conjugate object wave during the reconstruction step.

The visual characteristics of holographic displays are one of the main reasons for their desirability. The displays are able to provide continuous parallax, correct focus cues and high spatial resolution. Unlike LF displays, most of the holographic display methods do not have a trade-off between spatial and angular resolution and they can reproduce the desired 3D visual cues even for deep scenes. Moreover, holographic displays do not suffer from the vergence-accommodation conflict, which makes them more comfortable to use.

The use of coherent imaging techniques makes holographic displays susceptible to speckle noise, an important issue to discuss in the context of such displays and their characterization. The noise is the result of high contrast and frequency speckle patterns of random nature. In CGHs, the speckle patterns originate from utilizing random phase distributions to avoid concentration of light on the hologram and to simulate diffused diffraction of the object wave. As the noise heavily degrades the perceived quality, suppression of speckle noise is crucial for maintaining satisfactory quality in holographic displays. The speckle suppression methods can be broadly categorized into two groups. The first group of solutions rely on altering the display optics and techniques. Commonly these approaches aim at reducing either the spatial or temporal coherence of the reconstruction light, e.g. by diverging the illumination light with a diffuser [59] or by utilizing LEDs as the light source [60]. The second group of speckle suppression methods are algorithmic in nature and modify the computational synthesis of holograms. Such methods range from time-multiplexing hologram frames with statistically independent speckle patterns [3] or sparse sets of light rays [51] to cyclic sequential shifting of the hologram [12].

**Computer-generated holography** Computational synthesis of holograms is a difficult task, yet crucial for holographic displays, requiring compromises between reconstruction image quality and computational burden. Particularly challenging is the accurate reconstruction of view-dependent properties, such as occlusions, shading and reflections. Achieving realistic reconstruction quality, i.e. accurate view-dependent properties and high spatial resolution, requires huge amount of data to process, thus facilitating the need for computationally efficient CGH synthesis methods. The methods can be broadly divided into two categories: wavefront- and ray-based. The former methods utilize the 3D positional information of the scene, whereas the latter methods rely solely on captured images of the scene. In the former approach, mostly the scene is described as a collection of independently acting, self-emitting point sources of light [55]. The CGH is computed via superposition of each points contribution on the hologram



plane, usually assuming Fresnel diffraction model. Such methods usually deliver the abovementioned quality aspects of holography, including focus cues, however they are usually very demanding in terms of computational complexity. The alternative methods from the ray-based approaches, such as holographic stereogram (HS), can address the computational complexity by requiring only a set multiperspective images (i.e., LFs) [38]. On the other hand, due to the nature of the underlying LF sampling, HSs also suffer from the spatio-angular resolution trade-off; objects far away from the hologram plane cannot be sharply reconstructed [17]. There is vast literature on CGH [42]. Especially the efforts on reducing the computational complexity is crucial for realization of holographic display technology.

**Design characteristics of near-eye holographic displays** When designing holographic near-eye displays, characteristics common with LF displays need to be addressed. These include critical properties for accurate and realistic 3D vision, such as perceived spatial resolution, focus response provided by the display, FOV, etc. Currently the main limitations regarding the practical implementation of near-eye holographic displays are due to the frame rate and resolution of SLMs as well as computational efficiency of the CGH synthesis methods. The main physical design parameters of a holographic display are the pixel pitch  $\Delta$  and resolution  $N$ . The pixel pitch determines the maximum diffraction angle  $\theta_d$  according to the grating equation as

$$\theta_d = 2 \arcsin \left( \frac{\lambda}{2\Delta} \right) \quad (7)$$

Smaller pixel pitch increases the angle and is therefore preferred. However, the existing SLM technology does not still meet the desired pitch levels, which is also an important limitation in practical realization of the holographic displays. This has led to the research of new materials and elements, such as photorefractive polymers [5] and reconfigurable metasurfaces [54]. The number of pixels is critical for the space-bandwidth product and is generally desirable to be maximized.

The eyebox, or the viewing range of the display, is for the majority dictated by the diffraction orders. The extent of a single period of the wave field at a distance of  $z$  from the hologram is defined by the pixel pitch as  $\lambda z/\Delta$ , within which the eye should fit. Thus, the small size pixel pitch is also desired to ensure sufficiently large eyebox. In general, the display FOV is determined by the physical size of the display and the maximum diffraction angle. However, the FOV of a holographic display can be extended at the cost of the eyebox by utilizing spherical illumination instead of the planar one [15].

## 4.2 Retinal image formation models for holographic near-eye displays

In the most general case, regardless of the utilized CGH method, all holograms can be described using the complex object field or the intensity of interference between the object the reference waves as discussed in 4.1. Besides the HVS model itself, the key component of the retinal image formation models for holographic displays is thus the employed numerical wave propagation algorithms used to propagate the wave field

due to hologram till to the retina. Although the more rigorous algorithm for numerical wave propagation is the Rayleigh-Sommerfeld diffraction, and equivalently the angular spectrum method (ASM) in the spectral domain, due to computational issues mostly its paraxial approximation Fresnel diffraction is utilized, which can be implemented using fast Fourier transform techniques [13].

In the context holographic imaging, in forward (retinal image formation) models it is also common to model the eye as a simple camera with a thin lens and a planar sensor. The accommodation is simulated by varying the focal length of the lens chirp function, which models diffraction-limited imaging in the paraxial Fresnel regime. The wave field on the hologram plane can also be propagated back to the scene to reconstruct an image at the desired plane. Such back propagation methods ignore the effects of the human eye optics and mainly target to evaluate the effects of the holographic encoding.

**Characterization of perceivable spatial resolution and focus cues** In this part an example retinal image formation model is presented for near-eye holographic displays, with an eye model consisting of curved retina that has eccentricity-dependent spatial resolution. Such (foveated) eye models play a key role, when analyzing the perceptual spatial resolution as well as focus cues as a function of content eccentricity, which are vital when developing, e.g., foveated rendering methods. The model is illustrated in Figure 10, where the additional lens in front of the hologram, the eyepiece, is used as a magnifier as usually employed near-eye displays.

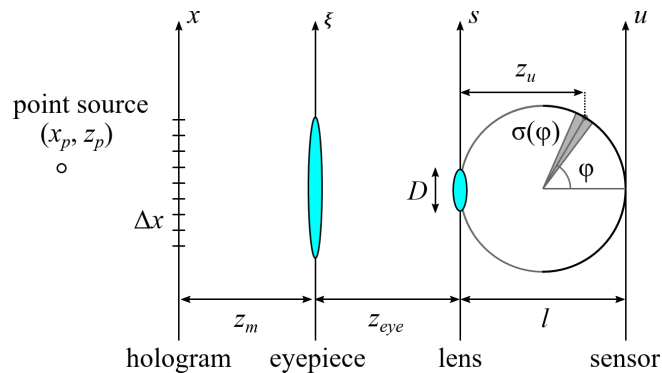


Fig. 10: The retinal image formation model for a holographic near-eye setup. The eye model includes non-uniform sampling on a curved retina surface.

The non-uniform sampling on the retina is based on the model proposed in [56] describing the density of the retinal ganglion cell receptive field as a function of retinal

eccentricity  $r = \sqrt{\phi^2 + \theta^2}$  and meridian  $m$

$$\rho(r, m) = 2\rho_{cone} \left(1 + \frac{r}{41.03}\right)^{-1} \times \left[ a_m \left(1 + \frac{r}{r_{2,m}}\right)^{-2} + (1 - a_m) \exp\left(\frac{r}{r_{e,m}}\right) \right] \quad (8)$$

The constants  $\rho_{cone}$ ,  $a_m$ ,  $r_{2,m}$ ,  $r_{e,m}$  fit the model along the four different meridians (temporal, superior, nasal, inferior). The eccentricity-dependent sampling of the retina is then defined from the density as

$$\sigma(\phi, \theta) = \frac{1}{r} \sqrt{\frac{2}{\sqrt{3}} \left( \frac{\phi^2}{\rho(r, 1)} + \frac{\theta^2}{\rho(r, 2)} \right)} \quad (9)$$

Both the lens of the eye and the eyepiece are assumed as aberration-free thin lenses with phase transmittance functions of

$$T(s, t) = \exp \left[ \frac{-j\pi}{\lambda f} (s^2 + t^2) \right] \quad (10)$$

where  $\lambda$  is the wavelength of the monochromatic light and  $f$  is the focal length of the corresponding lens. The simulated eye can be set to focus (accommodate) at a certain distance  $z_f$  by changing  $f$ , based on the paraxial lens imaging as

$$\frac{1}{f} = \frac{1}{z_f} + \frac{1}{l} \quad (11)$$

For a given hologram, the computation of the retinal image involves a set of plane-to-plane wave propagation (between the hologram and eyepiece as well as between the eyepiece and eye lens), modulation of the propagating waves by the eyepiece and the lens, and also plane-to-curved surface (retina) propagation. The end-to-end light transport between the hologram and retina can be rigorously calculated using ASM [13], since, unlike some other methods, it can compute the 3D field due to a given planar field, which naturally provides the relation between planar wave field just after the eye lens and the wave field sampled by the curved retina. That is, taking into account Figure 10, the retinal image is found for a given (complex-valued) hologram  $O(x, y)$  as

$$I(u, v, z_{uv}) = \left| A_{u,v,z_{uv}} \{ T_l(s, t) A_{s,t,z_{eye}} \{ T_e(\xi, \eta) A_{\xi,\eta,z_m} \{ O(x, y) \} \} \} \right|^2 \quad (12)$$

where  $A_{x,y,z_{xy}} \{ \}$  is the ASM propagation operator computing the scalar diffraction on the curved surface  $z_{xy}$  due to an input wave field defined on a plane;  $z_{uv}$  represents the depth of the curved surface of retina with respect to eye lens plane ( $s, t$ );  $T_l(s, t)$  and  $T_e(\xi, \eta)$  denote the lens transmittance functions of the eye lens and eyepiece, respectively.

Similar to the analysis demonstrated in Section 3 for LF displays, the PSFs and MTFs of the retinal image formation model for object points at intended depths can be used as reliable tools to analyze and characterize accommodative response, defocus blur as well as perceptual spatial resolution in holographic near-eye displays. Let

us consider comparative analysis of Fresnel hologram and a HS, as two widely used wave-based and ray-based CGH methods, respectively. The test point source is set at two different locations such that they are imaged by the eyepiece at  $z_p = 50$  cm and  $z_p = 80$  cm (the hologram plane itself is imaged at 40 cm) to compare a shallow and a deep scene. Both a Fresnel hologram and a HS of each point is generated with a sampling step  $\Delta x = 2 \mu\text{m}$ , number of samples  $N = 8192$  and wavelength  $\lambda = 534$  nm. Additionally, the segment size of the HS is chosen as 32 pixels ( $64 \mu\text{m}$ ). Though this sampling step is far larger than what the human eye is capable of resolving, it ensures multiple plane wave segments (the equivalent of rays in LF displays) within the extent of the pupil. Similar to LF displays, increasing this value is expected to improve the accommodative response of HS, at the cost of decrease in perceived spatial resolution [36]. The eye relief and the pupil size are fixed to  $z_{eye} = 2$  cm and  $D = 5$  mm, respectively. The simulated eye is set to focus at the exact distance of the point source, as well as around it at various shifted depths. For simplicity, a 2D cross-section of the 3D space is considered in the analysis, i.e., including only  $x$ - and  $z$ -axis. Thus, the eye model and the wave propagations are implemented for 1D signals. The stack of MTF magnitudes are evaluated at five different spatial frequencies (1, 2.5, 5, 10 and 15 CPD) and within the range of accommodation shift  $[-1 \text{ D} : 1 \text{ D}]$  around the test point. The maximum value of MTF in a given stack is denoted as the estimated accommodation distance  $\hat{z}_f$ .

Initial conclusions can be drawn by observing the behaviour of the PSFs at different accommodation shifts. As shown in Figure 11, in all cases the sharpest PSF is obtained near the OD shift (please note that the slight shifts are smaller than the DoF of the eye). In Figure 11 (a) and Figure 11 (c) the conjugate hologram surface is at 0.5 D shift, whereas in Figure 11 (b) and Figure 11 (d) it is at 1.25D. It is also observed that in each case, within the depth range from the conjugate hologram plane to the actual point depth, the envelope of the defocus blur follows a natural trend. Although in terms of focus cues the HS seem to be compatible with the Fresnel hologram, it is clearly seen from the sharpest PSFs that the perceptual spatial resolution in the case of HS is expected to be significantly lower than Fresnel case. As mentioned above, this is due to spatio-angular resolution trade-off in HSs.

The results of the comparative simulations in Figure 12 provide further insights for the difference between the Fresnel hologram and HS. When the test point source is relatively close to the hologram (see Figure 12 (a,c)), the focus estimate  $\hat{z}_f$  is near-correct for both methods. For the deeper scene case where  $z_p = 80$  cm, the near-correct accommodation distance is still maintained with the Fresnel hologram (see Figure 12 (b)) and the behaviour of the MTFs is almost invariant to changes in the scene depth. Similarly, the HS can produce near correct accommodation cues in terms of the accommodation estimate  $\hat{z}_f$  (see Figure 12 (d)) due to the large hogel size. However, the contrast gradient, or the sharpness of the MTF magnitude peak, is particularly noteworthy: the low gradient of the HS results indicates a weaker response trigger for accommodation in comparison to the Fresnel results.

Alternatively, the effects of foveation can be analyzed using the same framework to extend the characterization for larger eccentricities. Now the horizontal position of the point source is altered and the results are compared between varying amounts of

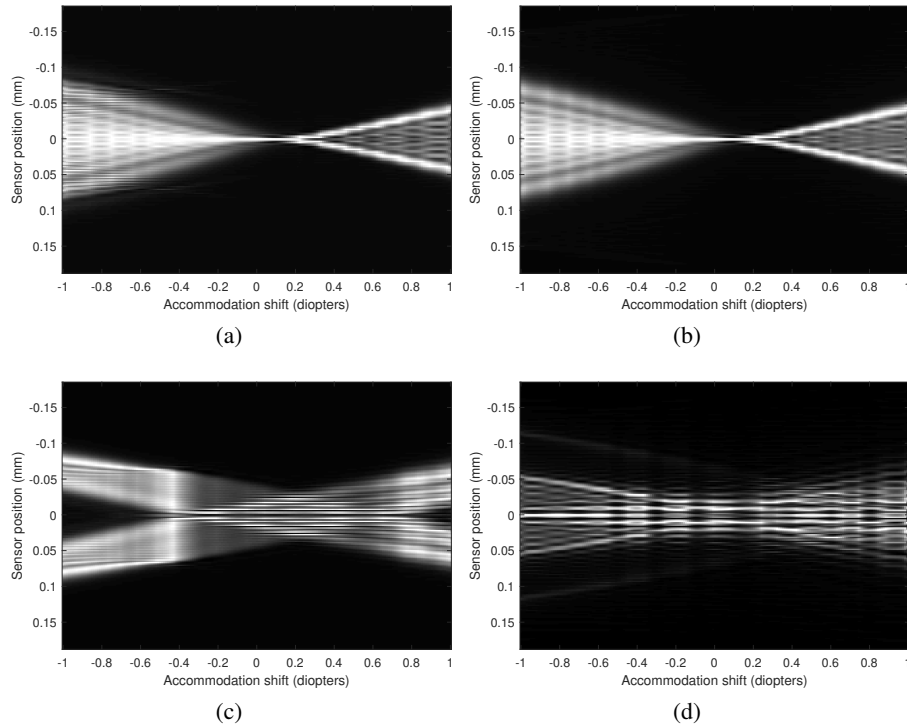


Fig. 11: A set of PSFs for different accommodation shift values, each displayed in a single column of the image. Fresnel (a,b) and HS (c,d) hologram synthesis methods are utilized. The point is placed at 50 cm (a,c) and 80 cm (b,d).

eccentricity. Since the eye model considers both the non-uniform nature of the sampling on the retina and its curvature, it is expected that a point further away from the direction of the gaze is perceived differently which likely to alter the focus response, in addition to more obvious change in perceptual spatial resolution. The test point is placed at visual angles of 0, 10 and 20 degrees with respect to central gaze direction. Figure 13 illustrate the effects of eccentricity on the contrast magnitude and gradient. In order to minimize the effect of possible outliers, a unimodal Gaussian is fitted to the obtained data points and its maximum value is estimated as the accommodated distance  $\hat{z}_f$ . The estimated accommodation distance, though changes slightly across the eccentricity range, remains within the typical values of HVS depth-of-field. Furthermore, one can observe that the contrast gradient reduces at larger eccentricities. This suggests that the strength of the defocus blur cue also decreases in the peripheral vision creating weaker accommodative response. The decrease in the contrast magnitudes at higher eccentricities is also consistent with the well-known eccentricity-dependent perceptual spatial resolution behaviour.

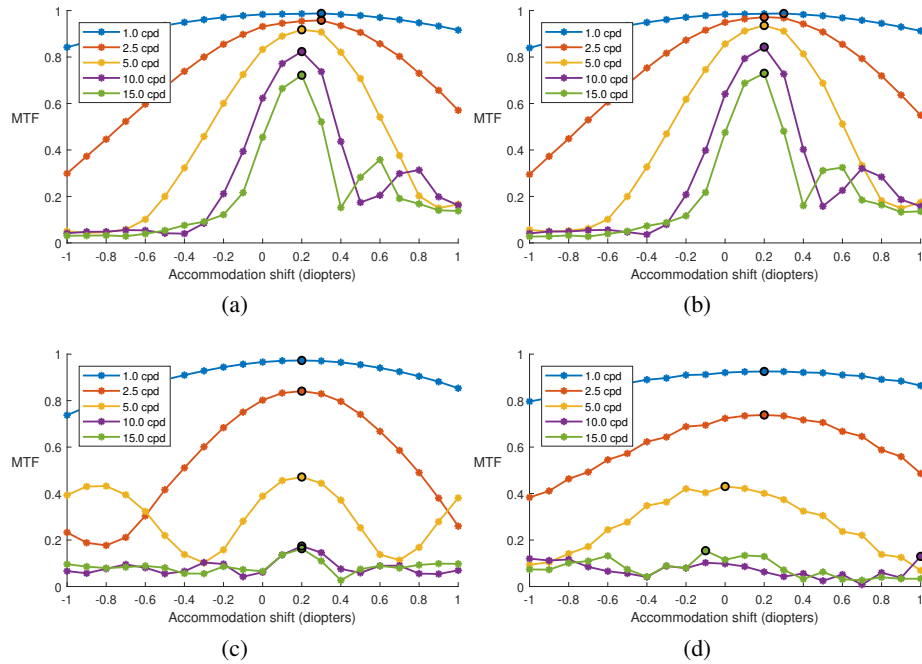


Fig. 12: Magnitudes of the MTFs as a function of accommodation shift for five different spatial frequencies. Fresnel (a,b) and HS (c,d) hologram synthesis methods are utilized. The point is placed at 50 cm (a,c) and 80 cm (b,d). The largest magnitude, i.e. the estimate  $\hat{z}_f$ , is marked with a black circle.

Finally, it is also important to evaluate the perceptual spatial resolution and the focus cues on 3D scenes to infer about general visual quality. The notable difference of 3D scene analysis in comparison to the single point source case is the presence of speckle noise. As seen in Figure 14, the noise severely degrades the visual quality of the retinal images for the corresponding Fresnel holograms and prohibits any meaningful analysis from them. Specifically, the effects of accommodation cannot be observed properly. Such extreme levels of noise necessitate the use of speckle reduction methods. Including even a relatively simple method, such as random averaging [3], significantly improves the visual clarity by reducing the speckle noise. Furthermore, such analysis also suggests the role of speckle suppression when implementing holographic displays in practice.

## 5 Conclusion

Their capabilities of delivering all physiological depth cues of HVS make LF and holographic displays strong candidates for the next-generation near-eye displays in creating the desired realistic visualization with full-immersion. In this chapter the retinal image

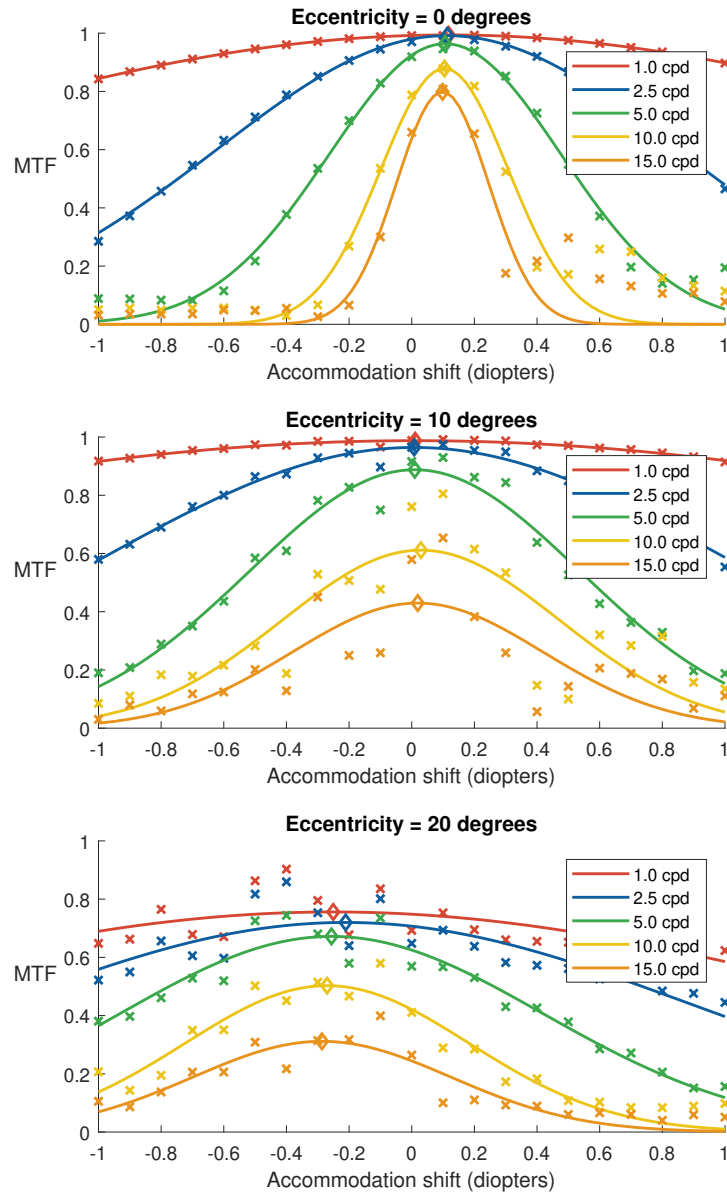


Fig. 13: Magnitudes of the MTFs from Fresnel hologram reconstructions and their Gaussian fit as a function of accommodation shift for five different spatial frequencies. The point is placed at 50 cm depth and horizontally at the center of the gaze (top), shifted 10 degrees (center) or 20 degrees (bottom) from the gaze. The diamond marker denotes the peak of the Gaussian, i.e. the estimate  $\hat{z}_f$ .

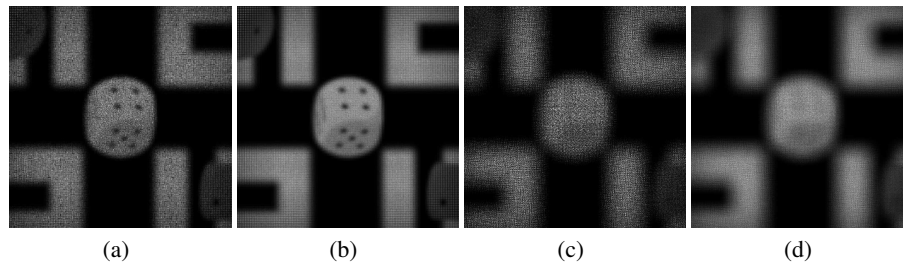


Fig. 14: Retinal images reconstructed from a Fresnel hologram of a 3D scene, without speckle reduction (a,c) and with random averaging (b,d). The simulated eye is set to focus on the die in the foreground in (a)–(b) and behind the scene in (c)–(d).

formation models are discussed for such displays. In particular, among several other aspects of spatial vision, characterization of perceptual spatial resolution as well accommodation and defocus blur cues are addressed. Thorough analysis of perceived images with such models are necessary to rigorously characterize and optimize the capabilities of the display under consideration. The presented framework, for instance, reveals some of the well-known aspects such as the trade-off between the perceptual resolution and the accuracy of focus cues in LF display, or rarely addressed properties of HSs that they are capable of delivering focus cues.

It is common in the 3D displays literature that the retinal image formation models have many assumptions, i.e., they mostly adopt simple (reduced) eye models, which consists of a thin lens and uniformly sampled planar retina. The analysis framework presented in this chapter with such models can be extended to include more rigorous eye models, e.g., including aberrations in the eye optics, which is likely to lead better optimized displays through more realistic characterization. Furthermore, the characterization and optimization of LF and holographic displays (in general all 3D displays) should further include the neural factors in the HVS to fully exploit its characteristics, which is also likely to reveal new aspects to be taken into account in the design of novel display optics as well as computational (rendering) algorithms.

## References

1. Adelson, E.H., Bergen, J.R.: The plenoptic function and the elements of early vision (1991)
2. Akpınar, U., Sahin, E., Gotchev, A.: Viewing simulation of integral imaging display based on wave optics. In: 2018-3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON). pp. 1–4. IEEE (2018)
3. Amako, J., Miura, H., Sonehara, T.: Speckle-noise reduction on kinoform reconstruction using a phase-only spatial light modulator. *Applied Optics* **34**(17), 3165–3171 (Jun 1995)
4. Banks, M.S., Hoffman, D.M., Kim, J., Wetzstein, G.: 3d displays. *Annual Review of Vision Science* **2**(1), 397–435 (2016), pMID: 28532351
5. Blanche, P.A., Bablumian, A., Voorakaranam, R., Christenson, C., Lin, W., Gu, T., Flores, D., Wang, P., Hsieh, W.Y., Kathaperumal, M., Rachwal, B., Siddiqui, O., Thomas, J., Norwood,



- R.A., Yamamoto, M., Peyghambarian, N.: Holographic three-dimensional telepresence using large-area photorefractive polymer. *Nature* **468**(7320), 80–3 (Nov 2010)
6. Boev, A., Poikela, M., Gotchev, A.P., Aksay, A.: Modelling of the stereoscopic hvs (2009)
  7. Bregovic, R., Sahin, E., Vagharshakyan, S., Gotchev, A.: *Signal Processing Methods for Light Field Displays*, pp. 3–50. Springer International Publishing, Cham (2019)
  8. Cholewiak, S.A., Love, G.D., Banks, M.S.: Creating correct blur and its effect on accommodation. *Journal of Vision* **18**(9), 1–1 (09 2018)
  9. Cottaris, N.P., Jiang, H., Ding, X., Wandell, B.A., Brainard, D.H.: A computational observer model of spatial contrast sensitivity: Effects of wavefront-based optics, cone mosaic structure, and inference engine. *bioRxiv* (2018)
  10. Curcio, C.A., Allen, K.A.R., Sloan, K.R., Lerea, C.L., Hurley, J., Klock, I.B., Milam, A.H.: Distribution and morphology of human cone photoreceptors stained with anti-blue opsin. *The Journal of comparative neurology* **312** 4, 610–24 (1991)
  11. Dorman, R., van Ee, R.: 50 years of stereoblindness: Reconciliation of a continuum of disparity detectors with blindness for disparity in near or far depth. *i-Perception* **8**(6), 204166951773854 (dec 2017)
  12. Golan, L., Shoham, S.: Speckle elimination using shift-averaging in high-rate holographic projection. *Optics Express* **17**(3), 1330–1339 (Feb 2009)
  13. Goodman, J.W.: *Introduction to Fourier Optics*. McGraw-Hill, 2nd edn. (1996)
  14. Gortler, S.J., Grzeszczuk, R., Szeliski, R., Cohen, M.F.: *The lumigraph* (1996)
  15. Haupt, C., Kolodziejczyk, A., Tiziani, H.J.: Resolution and intensity distribution of output images reconstructed by sampled computer-generated holograms. *Appl. Opt.* **34**(17), 3077–3086 (Jun 1995)
  16. Held, R.T., Cooper, E.A., Banks, M.S.: Blur and disparity are complementary cues to depth. *Current biology : CB* **22**(5), 426–431 (March 2012)
  17. Hilaire, P.S.: Modulation transfer function and optimum sampling of holographic stereograms. *Applied Optics* **33**(5), 768–774 (Feb 1994)
  18. Hoffman, D.M., Girshick, A.R., Akeley, K., Banks, M.S.: Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision* **8**(3), 33–33 (03 2008)
  19. Hua, H.: Enabling focus cues in head-mounted displays. *Proceedings of the IEEE* **105**(5), 805–824 (May 2017)
  20. Hua, H., Javidi, B.: A 3d integral imaging optical see-through head-mounted display. *Opt. Express* **22**(11), 13484–13491 (2014)
  21. Huang, F.C., Chen, K., Wetzstein, G.: The light field stereoscope: Immersive computer graphics via factored near-eye light field displays with focus cues. *ACM Trans. Graph.* **34**(4), 60:1–60:12 (2015)
  22. Huang, H., Hua, H.: Systematic characterization and optimization of 3d light field displays. *Opt. Express* **25**(16), 18508–18525 (2017)
  23. Huang, H., Hua, H.: Effects of ray position sampling on the visual responses of 3d light field displays. *Opt. Express* **27**(7), 9343–9360 (apr 2019)
  24. Jang, C., Bang, K., Moon, S., Kim, J., Lee, S., Lee, B.: Retinal 3d: augmented reality near-eye display via pupil-tracked light field projection on retina. *ACM Transactions on Graphics (TOG)* **36**(6), 190 (2017)
  25. Kajiki, Y., Yoshikawa, H., Honda, T.: Hologramlike video images by 45-view stereoscopic display. In: Fisher, S.S., Merritt, J.O., Bolas, M.T. (eds.) *Stereoscopic Displays and Virtual Reality Systems IV*. vol. 3012, pp. 154–166. International Society for Optics and Photonics, SPIE (1997)
  26. Kelly, D.H.: Motion and vision. ii. stabilized spatio-temporal threshold surface. *J. Opt. Soc. Am.* **69**(10), 1340–1349 (Oct 1979)

27. Konrad, R., Padmanaban, N., Molner, K., Cooper, E.A., Wetzstein, G.: Accommodation-invariant computational near-eye displays. *ACM Trans. Graph.* **36**(4), 88:1–88:12 (Jul 2017)
28. Lanman, D., Luebke, D.: Near-eye light field displays. *ACM Trans. Graph.* **32**(6), 220:1—220:10 (2013)
29. Lee, S., Cho, J., Lee, B., Jo, Y., Jang, C., Kim, D., Lee, B.: Foveated retinal optimization for see-through near-eye multi-layer displays. *IEEE Access* **6**, 2170–2180 (2018)
30. Legge, G.E.: A power law for contrast discrimination. *Vision Research* **21**(4), 457 – 467 (1981)
31. Levoy, M., Hanrahan, P.: Light field rendering. In: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques. pp. 31–42. ACM (1996)
32. Liang, J., Williams, D.R.: Aberrations and retinal image quality of the normal human eye. *J. Opt. Soc. Am. A* **14**(11), 2873–2883 (Nov 1997)
33. Liu, M., Lu, C., Li, H., Liu, X.: Near eye light field display based on human visual features. *Optics Express* **25**(9), 9886–9900 (2017)
34. Lucente, M.E.: Diffraction-specific Fringe Computation for Electro-holography. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, USA (1994)
35. Macleod, D.I.A., Williams, D.R., Makous, W.: A visual nonlinearity fed by single cones. *Vision Research* **32**, 347–363 (1992)
36. Mäkinen, J., Sahin, E., Gotchev, A.: Analysis of accommodation cues in holographic stereograms. In: 2018 - 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON). pp. 1–4 (June 2018)
37. Marcos, S., Moreno, E., Navarro, R.: The depth-of-field of the human eye from objective and subjective measurements. *Vision Research* **39**(12), 2039 – 2049 (1999)
38. McCrickerd, J.T., George, N.: Holographic stereogram from sequential component photographs. *Applied Physics Letters* **12**(1), 10–12 (1968)
39. Nadenau, M.J., Reichel, J., Kunt, M.: Performance comparison of masking models based on a new psychovisual test method with natural scenery stimuli. *Signal Processing: Image Communication* **17**(10), 807 – 823 (2002)
40. Navarro, R.: The optical design of the human eye: a critical review. *Journal of Optometry* **2**, 3–18 (12 2009)
41. Owens, D.A.: A comparison of accommodative responsiveness and contrast sensitivity for sinusoidal gratings. *Vision Research* **20**(2), 159 – 167 (1980)
42. Park, J.H.: Recent progress in computer-generated holography for three-dimensional scenes. *Journal of Information Display* **18**(1), 1–12 (2017)
43. Pelli, D.G., Bex, P.: Measuring contrast sensitivity. *Vision Research* **90**, 10–14 (2013)
44. Qin, Z., Chou, P.Y., Wu, J., Chen, Y., Huang, C., Balram, N., Huang, Y.P.: Image formation modeling and analysis of near-eye light field displays. *Journal of the Society for Information Display* (2019)
45. Schor, C.M.: A dynamic model of cross-coupling between accommodation and convergence: simulations of step and frequency responses. *Optometry and vision science : official publication of the American Academy of Optometry* **69** **4**, 258–69 (1992)
46. Seshadrinathan, K., Pappas, T.N., Safranek, R.J., Chen, J., Wang, Z., Sheikh, H.R., Bovik, A.C.: Chapter 21 - image quality assessment. In: Bovik, A. (ed.) *The Essential Guide to Image Processing*, pp. 553 – 595. Academic Press, Boston (2009)
47. Stern, A., Yitzhaky, Y., Javidi, B.: Perceivable light fields: Matching the requirements between the human visual system and autostereoscopic 3-d displays. *Proceedings of the IEEE* **102**(10), 1571–1587 (2014)
48. Strasburger, H., Rentschler, I., Jüttner, M.: Peripheral vision and pattern recognition: A review. *Journal of Vision* **11**(5), 13–13 (12 2011)

49. Sun, F.C., Stark, L., Nguyen, A., Wong, J., Lakshminarayanan, V., Mueller, E.: Changes in accommodation with age: static and dynamic. *American journal of optometry and physiological optics* **65** 6, 492–8 (1988)
50. Sun, Q., Huang, F.C., Kim, J., Wei, L.Y., Luebke, D., Kaufman, A.: Perceptually-guided foveation for light field displays. *ACM Trans. Graph.* **36**(6), 192:1–192:13 (Nov 2017)
51. Utsugi, T., Yamaguchi, M.: Speckle-suppression in hologram calculation using ray-sampling plane. *Optics Express* **22**(14), 17193–17206 (Jul 2014)
52. Van Nes, F.L., Bouman, M.A.: Spatial modulation transfer in the human eye. *Journal of the Optical Society of America* **57**(3), 401–406 (mar 1967)
53. Wandell, B.: *Foundations of Vision*. Sinauer Associates (1995)
54. Wang, Q., Rogers, E.T.F., Gholipour, B., Wang, C.m., Yuan, G., Teng, J., Zheludev, N.I.: Optically reconfigurable metasurfaces and photonic devices based on phase change materials. *Nature Photonics* **10**(1), 60–65 (Jan 2016)
55. Waters, J.P.: Holographic image synthesis utilizing theoretical methods. *Applied Physics Letters* **9**(11), 405–407 (1966)
56. Watson, A.B.: A formula for human retinal ganglion cell receptive field density as a function of visual field location. *Journal of Vision* **14**(7), 1–17 (June 2014)
57. Williams, D., Sekiguchi, N., Brainard, D.: Color, contrast sensitivity, and the cone mosaic. *Proceedings of the National Academy of Sciences of the United States of America* **90**(21), 9770–9777 (1993)
58. Yamaguchi, M.: Light-field and holographic three-dimensional displays. *J. Opt. Soc. Am. A* **33**(12), 2348–2364 (Dec 2016)
59. Yamaguchi, M., Endoh, H., Honda, T., Ohyama, N.: High-quality recording of a full-parallax holographic stereogram with a digital diffuser. *Optics Letters* **19**(2), 135–137 (Jan 1994)
60. Yaraş, F., Kang, H., Onural, L.: Real-time phase-only color holographic video display system using led illumination. *Applied Optics* **48**(34), H48–H53 (Dec 2009)