공학박사 학위논문

Perceptual studies on holographic near-eye displays

홀로그래픽 근안 디스플레이에 대한 인지 연구

2023년 8월

서울대학교 대학원

전기·정보공학부

김 동 연

Perceptual studies on holographic near-eye displays

지도 교수 정 윤 찬

이 논문을 공학박사 학위논문으로 제출함

2023년 8월

서울대학교 대학원

전기·정보공학부

김 동 연

김동연의 공학박사 학위 논문을 인준함

2023년 8월

위 원 장: _____ (인)

부위원장: _____ (인)

위 원: _____ (인)

위 원: _____ (인)

위 원: _____ (인)

Abstract

Perceptual studies on holographic near-eye displays

Dongyeon Kim
Department of Electrical and Computer Engineering
College of Engineering
Seoul National University

Holographic near-eye displays have emerged as the most promising candidate for augmented reality and virtual reality applications, offering superior features to other advanced displays. Recent advancements in holographic image quality, boosted by sophisticated computer-generated holography (CGH) algorithms, have accelerated the potential to pass the visual Turing test. However, much of the research relies on camera-based evaluations, which might overlook certain perceptual impacts.

This dissertation signifies initial progress in the field of visual experience using holographic near-eye displays, including a perceptual evaluation of holographic contents. To improve the accommodation response induced by holographic near-eye displays, the author delicately simulates the contrast ratio of the 2D holographic stimuli considering the human visual system and human perception models. The author employs both an optical solution and a computational approach to overcome the limited provision of accommodation cues from conventional CGHs. These proposed solutions have been validated through comprehensive user studies, which demonstrated a significant improvement in accommodation response and negligible degradation in subjective image quality.

This dissertation extends the focus to include 3D holographic content with the aim of enhancing 3D perceptual realism through holographic near-eye displays. In practice, volumetric scenes, which serve as ground truth for CGH algorithms, are often approximated from ideal viewpoints since the eyebox of holographic near-eye displays is relatively small, and evaluations are typically conducted using stationary cameras. Given the constant eye movement in human visual behavior during navigation, the perceptual metrics of holographic scenes are simulated under various pupil states. User evaluations conducted with a high-quality perceptual testbed for holographic near-eye displays validate the superior performance of including parallax cues in CGH supervision in terms of 3D perceptual realism. The results maintained consistency across a range of natural viewing conditions, even when head movement was restricted.

The author believes that this thesis contributes significantly to the development of holographic near-eye displays, furthering the pursuit of perceptual realism. It also sets a milestone towards passing the visual Turing test with holographic near-eye displays.

keywords: holographic display, perceptual study, augmented reality, virtual reality

student number: 2017-29782

Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1 Overview of holographic displays

Holography is a technology that records an interference pattern, known as hologram, generated by two coherent beams and reproduces the wavefronts of light with reference beam illumination afterward. This complete reconstruction of the desired signal as shown in Fig. 1.1, encoded in a photoreactive material, has been extensively utilized in the domain of holographic storage. The ability to reconstruct the recorded light field intact showed the potential for three-dimensional (3D) visualization, yet, limited to a single static image.

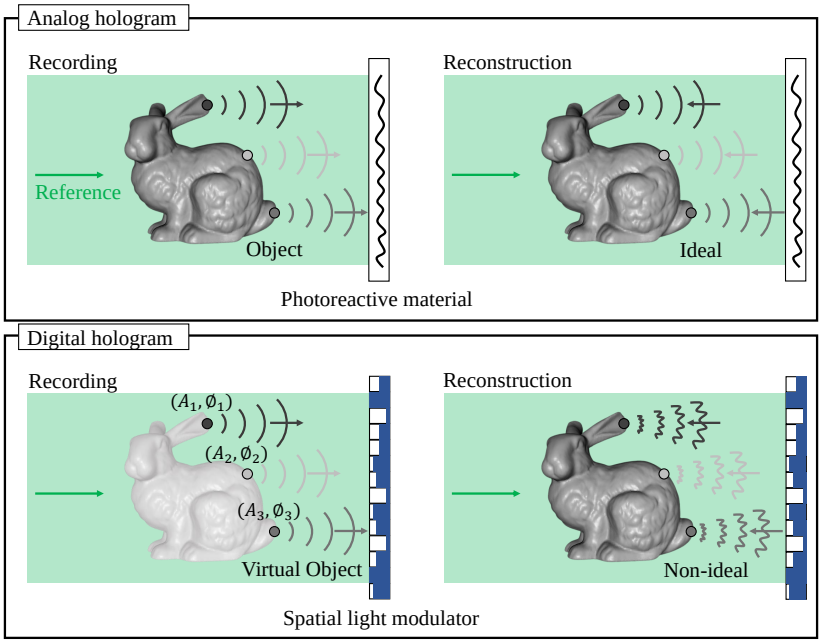


Figure 1.1 Recording and reconstruction procedure of analog, digital holograms.

The advent of the spatial light modulator (SLM) enabled the modulation of the amplitude or phase of the incident coherent light substituting the static photoreactive substance with a capability of dynamic update of holograms. Additionally, computer-generated holograms (CGHs) can be numerically calculated with amplitude and phase profiles, along with the positions of the virtual object, unlike the manual recording of hologram patterns. The numerical calculation of CGH is based on the theories established in wave optics. The capability to reconstruct the full light field along with the two aforementioned aspects has expanded the potential applications of holographic displays. These displays are now recognized as one of the next-generation platforms for augmented reality and virtual reality; however, they also present significant challenges that need to be addressed for widespread adoption.

One of the primary challenges faced by holographic displays is their form factor, as the display device is not self-emissive. However, the development of waveguide technology and polarization optics have successfully made holographic displays in a form factor of eyeglasses. The next challenge is the trade-off relationship between the eyebox (the range of eye position where scenes can be visualized) and the field of view. Unlike conventional displays, holographic displays rely on diffraction phenomena to generate the desired signal. To address the limitation, techniques such as spatial-angular multiplexing of holograms or the use of scattering masks have been employed. Lastly, holographic displays have long suffered from low image quality due to issues like speckle noise and imperfect representation of complex-valued fields with SLMs shown as Fig. 1.1. However, recent advancements in CGHs, incorporating machine learning-based approaches and calibration techniques, have significantly improved image quality but are mostly evaluated with camera-based experiments.

1.2 Motivation and purpose of this dissertation

The main objective of the display device is to successfully pass the visual Turing test, which was introduced a few years ago by Wetzstein and Lanman [1]. This test evaluates an individual's ability to distinguish between the virtual world seen through virtual reality headsets and the real world. Achieving success in the visual Turing test requires improvements in various aspects, not just obtaining high scores in image-based measurements.

While photo-realistic visualizations have been confirmed through camera-based experiments, they can potentially lead to unfavorable outcomes when experienced on holographic displays, which offer a higher degree of freedom compared to conventional displays. Since holographic displays provide more flexibility in visual representation, the effects and implications of photo-realistic visuals may differ when viewed through such displays. Consequently, the impact on user experience and potential drawbacks should be considered when implementing holographic displays and their associated visualizations.

The purpose of this dissertation is to investigate aspects that are frequently overlooked in camera-based evaluations but play a crucial role in the user experience by conducting perceptual evaluations using holographic near-eye displays. The findings from these perceptual evaluations will not only guide the community in the pursuit of photo-realistic visualization but also help to achieve the perceptual realism necessary to pass the visual Turing test in the near future. In this dissertation, the author was motivated to conduct perceptual evaluations focusing on different aspects of 2D and 3D holographic content.

Regarding 2D content, holographic near-eye displays need to provide high-quality 2D holographic scenes with a sufficiently narrow depth of field to provide adequate depth-dependent retinal blur to induce accommodation response.

However, CGHs that visualize high-quality 2D image typically represents a wide depth of field, resulting in minimal diffraction of light to realize all-in-focus images. On the other hand, CGHs that provide a narrow depth of field require a randomly distributed phase profile, which negatively impacts the image quality with severe speckle noise. Thus, the phase uniformity of the reconstructed field is directly related to the depth of field but inversely proportional to the 2D image quality. These relationships will be further explained in Section 3.1.3. To address this issue, the author manipulates the reconstructed image by optimizing CGH with an additional loss on contrast ratio to present a contrast curve similar to that of the incoherent case. The simulation of the contrast ratio takes into account the characteristics of human visual systems and human perception models, ensuring a more accurate representation.

In terms of 3D content, holographic near-eye displays are well-known for reconstructing scenes with varying depth information and quasi-natural parallax. However, CGHs that reconstructs light field tend to underperform in evaluations using traditional image-based metrics, and the limited size of the eyebox can result in parallax information being underestimated. It will be explained in Section 4.1.1 that humans are more sensitive to parallax and occlusion in depth perception, which can only be effectively reconstructed using light field-based CGHs. Additionally, light field-based CGHs have demonstrated competence with the robust visualizations performed under various pupil states, mimicking human ocular movement.

1.3 Scope and organization

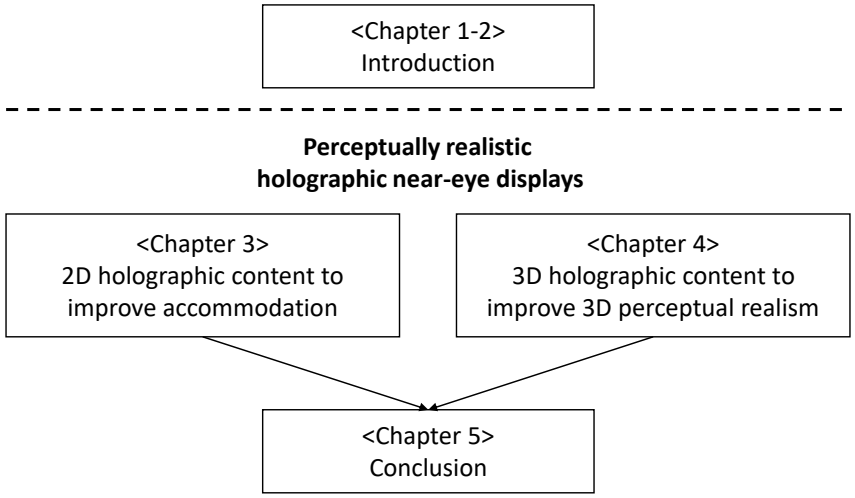


Figure 1.2 Block diagram for illustration of dissertation organization.

The organization of the dissertation is illustrated with a block diagram shown in Fig. 1.2. In Chapter 1, the principle of the holographic display, three major challenges of holographic displays, and the recent advancements to resolve the stated challenges are introduced. The author addresses some aspects that are overlooked in the camera-based experiments but can play an important role in improving perceptual realism through holographic near-eye displays. Then, the purpose and motivations that made the author investigate more on the 2D, and 3D holographic contents are introduced.

In Chapter 2, the background of the visual perception field is briefly introduced. Then, the author sequentially introduces traditional image-based metrics, perceptual quality metrics, and procedures to quantify the preference. Next, the depth perception mechanism is briefly summarized with the possible measurement strategies of ocular behavior that vary depending on perceived depth.

Lastly, the term '3D perceptual realism' is introduced with the ultimate goal of passing the visual Turing test with the visual device.

In Chapter 3, the prerequisite conditions of 2D CGHs to trigger monocular accommodation response properly are investigated and several approaches to improve the response are conceived. The author simulates the contrast gradients - change of the contrast over the change of the focal diopter of the eye - of holographic contents acquired with various CGH algorithms considering the optical configuration of the display system and the human visual perception models to validate their support of monocular accommodation cues. User studies that experimentally measure user data are designed and conducted. Specifically, the author quantitatively measures accommodation response and assesses the subjective image quality of the holographic contents with pairwise comparisons. The optimal form of 2D holographic contents inducing sufficient accommodation response is confirmed.

In Chapter 4, the scope of the dissertation is expanded to 3D content. The author investigates the perceptual realism of 3D scenes presented through holographic near-eye displays, while considering natural viewing conditions. The approach includes simulating the perceptual quality of the 3D holographic scenes with varying CGH supervision types and pupil conditions and accounting for the impact of eye movements on sampled signals. To find the best CGH supervision format for realistic 3D holographic scenes, user studies are conducted with the high-quality perceptual testbed of holographic near-eye displays. The experiment confirms CGH reconstructs parallax outperforms the other CGHs in terms of 3D perceptual realism even in the viewing condition with limited head movement. The results confirm the optimal target format of CGH supervision to reconstruct a 3D perceptually realistic holographic scene.

In Chapter 5, the author concludes the dissertation by summarizing the con-

tributions.

Chapter 2. Visual perception

2.1 Introduction

Virtual reality (VR) technology has made significant progress in overcoming hardware and software limitations to deliver immersive and realistic experiences. This progress has been driven by a deep understanding of the human visual system and the application of perceptual thresholds derived from extensive user studies. Important advancements in this domain include foveated graphics [2, 3], high-dynamic range image recovery [4], and other related advancements. The perceived image quality as perceived by our eyes can be estimated by leveraging recent advancements in visual difference predictors [3] established with various displaying configurations and observance models to quantify the preference. In this section, the author briefly introduces the field of visual perception from image perception to depth perception, to better understand the paper.

2.2 Image perception

The perception of images through the human visual system is limited in terms of spatial and temporal domains. Firstly, spatial sensitivity is primarily influenced by the eye's optical transfer function, which is affected by factors such as pupil size [5], lens aberration [6], and retinal eccentricity [7]. Additionally, the neural transfer function introduces further modulators generated by the retina-brain systems [8]. Consequently, the spatial contrast sensitivity function, encompassing the collective impact of these factors, characterizes the spatial resolution of

the human visual system [9–11]. The highest spatial frequency perceivable by the human eye is estimated to be approximately 48 cycles per degree (cpd) in the fovea region [12], with a conservative estimation of around 30 cpd.

Similarly, temporal sensitivity is assessed through visual stimuli with different temporal frequencies and varying luminance levels [13]. The temporal resolution, defined by the critical flicker fusion threshold, is approximately 50 Hz when a low-luminance, low spatial-frequency target is projected onto the near-fovea region [14]. When a visual stimulus updates faster than the temporal threshold, humans perceive a superimposed intensity profile.

In the subsequent subsection, the author presents image quality evaluation metrics, ranging from traditional image-based metrics to more recent perceptual quality metrics introduced publicly, along with a strategy to quantify user preferences.

2.2.1 Traditional evaluation metrics of image contents

Peak Signal-to-Noise Ratio (PSNR)

Peak Signal-to-Noise Ratio (PSNR) is a commonly employed objective metric that evaluates the quality of a reconstructed or compressed image by comparing it to a reference image. It quantifies the level of noise or distortion present in the image. PSNR is computed using the mean squared error (MSE) between the reconstructed image (I) and the reference image (I_{ref}). The formula for PSNR is expressed as:

$$PSNR = 10log_{10}\left(\frac{L^2}{MSE}\right), \tag{2.1}$$

where, L represents the maximum pixel value of the image (e.g., 255 for an 8-bit image). A higher PSNR value indicates a better-quality image.

Structural Similarity Index (SSIM)

Structural Similarity Index (SSIM) is a widely used metric that assesses the structural similarity between two images. In contrast to PSNR, SSIM considers not only the structural information but also the perceived variations in luminance, contrast, and structure between the original and distorted images.

SSIM evaluates three elements: luminance (brightness), contrast (intensity differences), and structure (correlation between local image patches). The resulting SSIM index is derived as the product of these three components. The SSIM value ranges from -1 to 1, with 1 indicating identical images. The calculation of the final SSIM index involves multiplying these three components, as represented by the following equation:

$$SSIM = (l(I, I_{ref})) \cdot (c(I, I_{ref})) \cdot (s(I, I_{ref})) \,. \tag{2.2}$$

SSIM is often considered more perceptually meaningful than PSNR because it takes into account human visual perception factors.

In addition to PSNR and SSIM, the field of image quality assessment encompasses several traditional metrics based on image properties. However, these metrics do not always align well with human perception, particularly when the type of distortion does not match human sensitivity or the display configurations are suboptimal. As a result, these metrics are frequently employed in combination with other perceptual metrics or subjective evaluations to obtain a more comprehensive evaluation of image quality.

2.2.2 Perceptual image quality metrics

Since the traditional image quality metrics do not correlate well with the visual perception arising from the high complexity of the visual perception mecha-

nism, there have been works to quantify the image quality metrics considering the human perception models.

HDR-VDP-2 [15] is a recently-introduced visual metric for predicting visibility and quality in all luminance conditions. This metric is based on a new visual model that was developed from contrast sensitivity measurements. The researchers calibrated and validated this metric against several datasets. The results showed that this new metric provides improved predictions compared to existing metrics, particularly for low luminance conditions.

As an updated version of HDR-VDP-2 [15], FovVideoVDP [3] is a new visual difference metric that is calibrated in physical units with different types of displays, models temporal aspects of vision, and accounts for foveated viewing. This metric is designed to generalize across a diverse range of contents and types of spatio-temporal artifacts. It was carefully calibrated using three independent video-quality datasets and a large image-quality dataset. These two metrics highly rely on the calibration and validations from the user study dataset and they require the procedure to scale the measured preference.

2.2.3 Metrics from pairwise comparison

The process of converting user data into a single measure involves a scaling procedure. The scaling technique mentioned in the recent study of Perez and Mantiuk [16] will be briefly explained and utilized throughout the dissertation. User data measurement is generally performed with the pairwise comparison with two-alternative forced choice (2-AFC) experiments [17] as the direct rating often leads to low sensitivity with high measurement errors [16].

The process of translating pairwise comparison data into a measurement scale involves several steps. First, the pairwise comparison data is collected by presenting two conditions at a time to the observers and asking them to choose

one according to specific criteria. The vote counts are then aggregated into a quality scale, where the distances between conditions can be interpreted as the probability of better-perceived quality.

To elaborate, when scaling the data, the main goal is to determine the distance between the quality scores q_i and q_j. This distance is associated with the probability of condition O_i being of higher quality than condition O_j. If we consider the observer model to be a Gaussian random variable with the same standard deviation and no correlation between the paired options, the difference between two Gaussian variables r_i and r_j is also a Gaussian random variable,

$$r_i - r_j \sim N\left(q_{ij}, \sigma_{ij}\right), \tag{2.3}$$

where, $q_{ij} = q_i - q_j$ and $\sigma_{ij}^2 = \sigma_i^2 + \sigma_j^2 = 2\sigma^2$. The probability of choosing O_i over O_j can be computed using the cumulative normal distribution Φ over the difference $r_i - r_j$:

$$P\left(r_i > r_j\right) = P\left(r_i - r_j > 0\right) = \Phi\left(\frac{q_i - q_j}{\sigma_{ij}}\right) \tag{2.4}$$

The relationship between the probabilities and the score differences is given as:

$$q_i - q_j = \sigma_{ij}\Phi^{-1}\left(P\left(r_i > r_j\right)\right). \tag{2.5}$$

The observation model functions on the premise that the noise parameter σ is fixed and recognized for every circumstance, leading to $\sigma_{ij} = \sqrt{2}\sigma$. A common technique involves configuring σ_{ij} such that it matches a 0.75 probability to a score difference of 1 Just-Noticeable-Difference (JND) unit [18]. When the standard deviation σ_{ij} is 1.48, the inverse cumulative distribution intersects the value of 1 JND.

The scaled metric with multiple axes of judgment

Usually, the outcomes of pairwise comparisons are measured in a unit of JND. A difference of 1 JND between two stimuli is defined as a difference that 75% of observers can discern. This unit is traditionally used to quantify the visual difference between distorted images along a single axis, such as when the images differ only in the degree of blur.

However, in many situations, especially when evaluating holographic content, the distortions vary between images. For instance, consider two distorted images: one is blurred and the other is affected by noise. The sensitivity to detect differences increases as changes occur along two axes instead of one. Despite the non-existence of the ground truth image, one can expect the ideal scene, and decisions can be made based on the distance from this ground truth image. Therefore, this dissertation alternatively uses the unit of Just-Objectionable-Difference (JOD) to quantify the visual difference of the provided content relative to the ideal case. However, in many instances, no ideal images are provided. As a result, the relative difference in JODs becomes the value of interest.

2.3 Depth perception

Depth perception mechanisms are crucial in assessing the perceptual realism of 3D scenes, as various cues such as binocular disparity, accommodation, vergence, and motion parallax contribute to depth perception [19]. The alignment of vergence and accommodation cues is important as it reduces visual fatigue [20] resulting from prolonged use of VR devices. Vergence and accommodation exist in oculomotor movements. Vergence is an eye rotation motion that fuses images seen through both eyes sharply, and inaccurate vergence results in double vision of the fixated object. Accommodation is the focal power adjustment of the eye

lens to obtain a sharp image with one eye. Incorrect accommodation of the eye lens causes blurs in the retinal image. These two oculomotor motions demonstrate a neural coupling [21, 22], and each response acts as a factor that triggers the other. However, vergence is primarily induced by retinal disparity, while accommodation is mainly triggered by retinal blur.

Retinal blur, as a visual cue to trigger accommodation, is directly affected by aberration and pupil size of the eye. Among eye aberrations, which vary among individuals, monochromatic aberrations, such as defocus and astigmatism can impede accommodation response. Whereas the intrinsic presence of chromatic aberration in the human eye can positively trigger accommodation [23, 24]. Furthermore, the eye pupil that manages the influx of light by modulating its size can change aberration, diffraction, and depth of focus, which further influences the accommodation response. Additionally, there are psychological factors such as texture gradient, object overlapping, shadowing, and motion-based factors, such as motion parallax, that affect monocular accommodation [25].

2.3.1 Measurements of depth perception

As stated above, there are two distinct ocular motor movements that can be observed and serve as measurements for depth perception.

Vergence response

The vergence response of the eye refers to the ability of both eyes to turn inward to maintain single binocular vision as an object moves closer. This is a crucial aspect of depth perception and is particularly important for near-vision tasks.

To measure the vergence response of the eye, a few methods are typically used. Cover test [26] is a simple and commonly used method in clinical settings. The examiner alternately covers each eye while the person fixates on a near

object. The examiner observes the uncovered eye for movement. An eye that moves inward when uncovered indicates a vergence response.

More advanced methods use eye-tracking systems to measure the vergence response. These systems use cameras and infrared light to track the position and movement of the eyes in real-time as the person fixates on a target moving in depth. The infrared light is directed toward the eyes, causing visible reflections in the pupil and cornea. These reflections are captured by the cameras and processed by software to determine the gaze direction and eye movement.

If the left eye rotates inward with the angle of θ_L and the right eye rotates inward with the angle of θ_R and the interpupillary distance between both eyes is set as ipd, the dioptric depth of the fixated object (D_{object}) can be estimated as

$$D_{object} = \frac{\tan\theta_L + \tan\theta_R}{ipd}. \tag{2.6}$$

The depth estimation of the object being fixated relies on binocular disparity, so it can only be achieved with binocular displays. In addition, the angle of vergence doesn't necessarily confirm if the user has achieved a clear focus on the gazed object. Therefore, measuring the accommodation response, which confirms an actual change in focus with a single eye, accurately indicates the depth of the object in the air.

Accommodation response

Common metrics used to quantify accommodation response are the accommodation response time and accommodation amplitude.

The accommodation response time measures the time the eye takes to change its focus from one distance to another. This can be further broken down into reaction time, which is the time from the start of the stimulus to the start of the

accommodation response, and response time, which is the time from the start of the accommodation response to when the eye reaches its final focus. This indicates how acutely the user's eye focuses on the visual stimuli, so it does not guarantee the depth perceived by the eye.

Another metric is the accommodation amplitude, which measures the change in the eye's optical power during the accommodation response. To quantify the perceived depth with accommodation response, some specialized equipment, such as an autorefractor or an aberrometer is utilized. These devices can measure the eye's refractive error in real-time as the eye adjusts its focus from one distance to another.

2.4 Three-dimensional perceptual realism

The assessment of three-dimensional (3D) realism through visual devices is subjective as it involves various cues for image and depth perception. In addition, there are no objective measurements to predict how much the user perceives 3D realism with the visual stimuli. Moreover, there are several psychological qualities, such as presence and immersion. The evaluation of such qualities also requires subjective ratings [27]. However, one of the ultimate goals for the researchers in this field is to achieve a level of *3D perceptual realism* with a display device and pass the visual Turing test [1] using real-world volumetric scenes. Recent studies [28, 29] have conducted a visual Turing test using a dual-plane stereo display to advance next-generation displays.

Unlike other incoherent displays, conducting meaningful user studies with holographic displays has been challenging due to low image quality caused by speckles and imperfect representation of the complex-valued field, which results in low contrast. However, recent advancements in CGH and SLMs have

significantly improved image quality through time-multiplexing and calibration techniques [30–35], which have made it possible to conduct more accurate and robust user studies with holographic displays.

Chapter 3. Depth of field and speckle noise in 2D holographic displays and their effects on accommodation response

3.1 Introduction

In Chapter 3, the author focuses on a crucial but often overlooked aspect in the field of holographic near-eye displays: supporting accommodation cues. The research explores the necessary conditions for 2D CGHs to effectively trigger a monocular accommodation response. The ultimate goal is to achieve perceptual realism with holographic near-eye displays. The author examines how the optical setup of the display system and human visual perception models contribute to supporting monocular accommodation cues. To do this, the author employs holographic content generated through various CGH algorithms to simulate contrast gradients, which correspond to changes in the eye's focal diopter.

To empirically analyze the effectiveness of the proposed strategies, the author conducts user studies and collects empirical user data. Specifically, the author quantifies the accommodation response and evaluates the subjective image quality of the holographic content using pairwise comparisons. These evaluations are carried out using a benchtop prototype of the holographic near-eye display. The results demonstrate that current CGHs used in holographic near-eye displays only occasionally or weakly trigger a monocular accommodation response. However, the suggested strategies significantly improve the accommodative gain without causing any notable deterioration in subjective image quality.

3.1.1 Holographic wave propagation

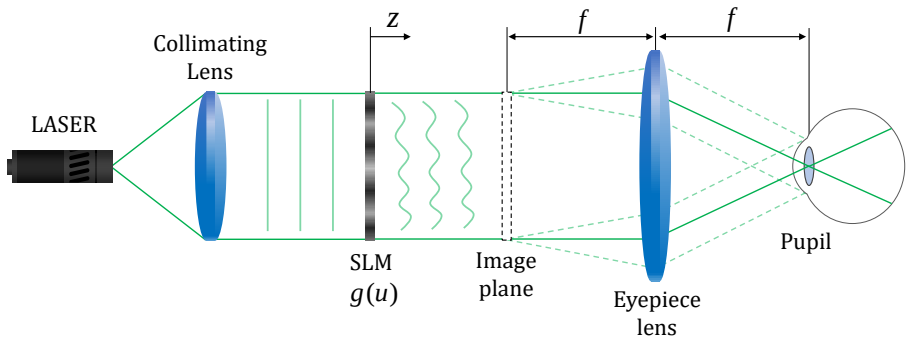


Figure 3.1 Concise schematic of conventional holographic near-eye displays.

In the holographic near-eye display configuration illustrated in Figure 3.1, a coherent laser beam is collimated using a lens and directed towards the SLM. The SLM modulates either the amplitude or the phase component to represent the complex-valued field $u = ae^{i\phi} \in \mathbb{C}^{M\times N\times 3}$. The field distribution in the plane after the SLM can be numerically calculated using wave diffraction theory. The angular spectrum method [36] is a commonly employed wave propagation model, and the propagated field can be described using this model as

$$\mathcal{P}_z\left(g(u)\right) = \iint \mathcal{F}\left(g\left(u\right)\right)\mathcal{H}(\nu_x,\nu_y;z,\lambda)e^{i2\pi(\nu_x x+\nu_y y)}d\nu_x d\nu_y, \tag{3.1}$$

$$\mathcal{H}(\nu_x,\nu_y;z,\lambda) = \begin{cases} e^{i\frac{2\pi z}{\lambda}\sqrt{1-(\lambda\nu_x)^2-(\lambda\nu_y)^2}}, & if\ \sqrt{\nu_x^2+\nu_y^2}<\frac{1}{\lambda}, \\ 0, & otherwise \end{cases}$$

where, x, y are coordinates of SLM domain, $\mathcal{P}_z(\cdot)$ is a propagation operator with a distance of z, $\mathcal{F}(\cdot)$ is the two-dimensional Fourier transform operator, $g(\cdot)$ is the SLM decoding operator, λ is the wavelength of the laser beam, ν_x, ν_y

are the spatial frequencies and $\mathcal{H}(\cdot)$ is the transfer function.

3.1.2 CGH encoding

The complete complex representation of a coherent field using a single SLM ($g(u)$) does not suffer from information loss, which would necessitate additional CGH optimization. However, due to the phase-only or amplitude-only modulation capability of SLMs, additional procedures are required to reconstruct the desired complex amplitude of light. Two methods are commonly used for complex-valued field reconstruction with SLMs.

Direct encoding involves encoding the complex field to a phase-only or amplitude-only SLM pattern with a single operation. For phase-only CGHs, the simplest solution is the amplitude discard method, which directly extracts the phase component from the complex-valued field, represented as $g(u) = e^{i\phi}$. However, this approach leads to noticeable artifacts in the reconstructed image as it ignores the amplitude component. Another straightforward solution is double phase-amplitude encoding [37–40], where a pair of adjacent pixels on the SLM represents a single complex value.

In addition to direct phase retrieval strategies, several iterative methods employ phase patterns with optimization to achieve better image quality [41, 42]. Recent works on obtaining high-quality 2D phase-only CGHs have utilized a differentiable framework [34, 35, 43], as shown in Figure 3.2.

On the other hand, SLMs that incorporate a stack of a polarizer and an analyzer can modulate the amplitude of the incident light. In the case of amplitude-type SLMs, the target field is combined with its complex conjugate as $g(u) = \text{real}(u) = \frac{ae^{i\phi}+ae^{-i\phi}}{2}$. However, amplitude modulation SLMs aim to improve their frame rates through binary modulation and are typically implemented using micromirrors [30, 44] or ferroelectric liquid crystals [45]. The operator is

expressed as $g(u) = \text{sign}\,(\text{real}(u))$. The incomplete representation of complex-valued fields solely with the binarized amplitude results in a low signal-to-noise ratio (SNR) in the reconstructed holographic image. The rapid operation speed is a significant advantage of binary SLMs compared to conventional 8-bit SLMs. Recent research has employed the stochastic gradient descent method for optimizing binary amplitude CGHs to mitigate the degradation of the SNR [33, 46].

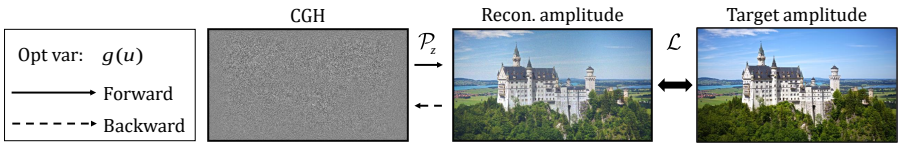


Figure 3.2 Illustration of differentiable 2D CGH optimization pipeline.

The acquisition of the complex-valued hologram is usually performed by solving the minimization problem formulated in the amplitude basis as follows:

$$\underset{u}{minimize}\,\mathcal{L}\left(s \cdot \left|\mathcal{P}_z\left(g(u)\right)\right|, a_{target}\right). \tag{3.2}$$

In this context, the scale factor s is employed to balance the overall value difference between $\left|\mathcal{P}_z\left(g(u)\right)\right|$ and a_{target}. To acquire a phase pattern, the optimization variable is converted into real values, represented as $\phi \in \mathbb{R}^{M\times N\times 3}$. The loss function $(\mathcal{L})$ typically takes the form of l_1 or l_2 error, and the solutions to these problems can be obtained using first-order gradients, as recently introduced by Peng et al. [34], employing the SGD solver.

3.1.3 Phase uniformity and depth of field

The complex-valued field is a combination of both amplitude and phase profiles. Fig.3.3 displays complex-valued fields with an identical amplitude profile of the cameraman but varying phase profiles. In the case of the uniform phase distribution, the phase is constant and set to zero. However, for the non-uniform

phase distribution, the phase is randomly distributed within a range from $-\pi$ to π. To propagate the coherent field, holographic wave propagation is employed using the angular spectrum method, as shown in Eq.3.1. The amplitude at a distance of 5 mm from the original plane is illustrated in the images of the third column in Fig. 3.3.

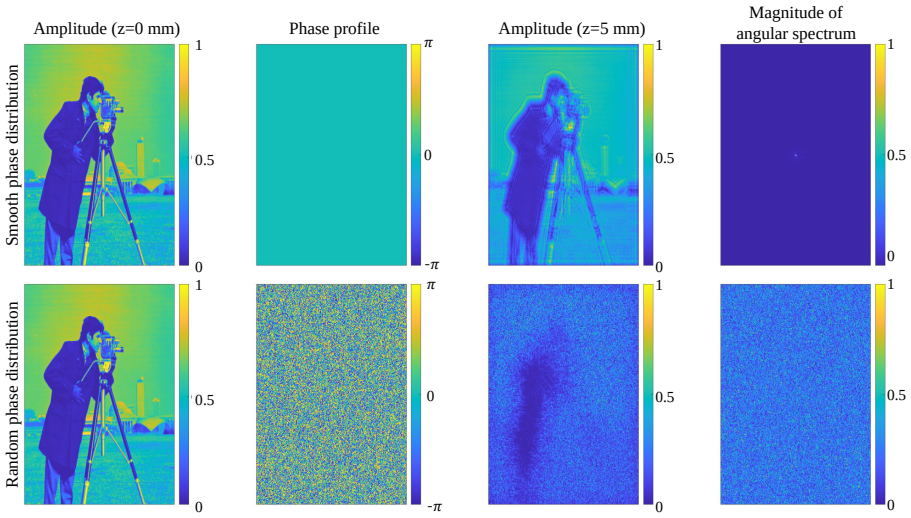


Figure 3.3 Relationship between the phase uniformity and the depth of field.

For a complex field with a smooth phase profile, the image of the cameraman rarely gets defocused with the fringe patterns generated from the amplitude boundary. However, the image gets defocused much in the case of the field with a random phase distributed profile. The difference in the defocus blur mainly stems from the magnitude of the angular spectrum of the original field shown in the fourth column of Fig. 3.3. Therefore, the phase distribution of the hologram affects the depth of field.

3.1.4 Accommodation measurement with holographic displays

The investigation of whether holographic displays effectively induce monocular accommodation responses and its validation through user experiments have not been extensively conducted. Prior research by Takaki and Yokouchi [47], and Ohara et al. [48] involved measuring static accommodation responses on their table-top holographic displays. In these studies, the vergence cue was not eliminated as the images were presented to both eyes, although the measured diopter when viewing a holographic stimulus was consistent with that measured when viewing a real-world scene. Additionally, Nozaki et al. [49] failed to demonstrate that the holographic stimulus elicited a similar accommodation response to that of the real target for a monocular eye, despite the reconstructed images being free of speckle noise with light-emitting diode (LED) illumination. As a result, none of the existing works have yet confirmed whether a holographic display can effectively trigger an appropriate monocular accommodation response or identified potential factors that may limit or enhance the accommodation response.

3.2 Motivation

3.2.1 Limited depth of field in high-quality 2D CGH

The accommodation response is mainly influenced by retinal blur, which changes based on the focal states, as depicted in Figure 3.4. Consequently, the focus-dependent retinal blur is determined by the depth of field provided by the display system. In the case of holographic near-eye displays, the depth of field varies depending on the CGH. Figure 3.5(a) illustrates a holographic image of CGH acquired using the naive stochastic gradient descent (SGD) algorithm. The SGD hologram produces a high-contrast holographic image similar to the ground truth in the focused state, while the defocused image exhibits a blur effect with a limited size compared to the ground truth case of an incoherent display. This limitation occurs because the reconstructed field exhibits high energy concentration within a restricted angular spectrum, as demonstrated in Fig. 3.5(b). The white dashed line represents the pupil of the human eye placed at the eyebox, while the dashed lines in different colors indicate the wavelength-dependent size of the Fourier plane.

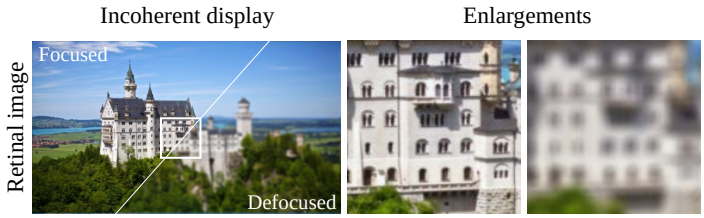


Figure 3.4 Reconstructed images of an incoherent display with two different focal states: focused and defocused with 0.6 diopter (D, reciprocal of the metric distance from eye) under a pupil diameter of 3 mm.

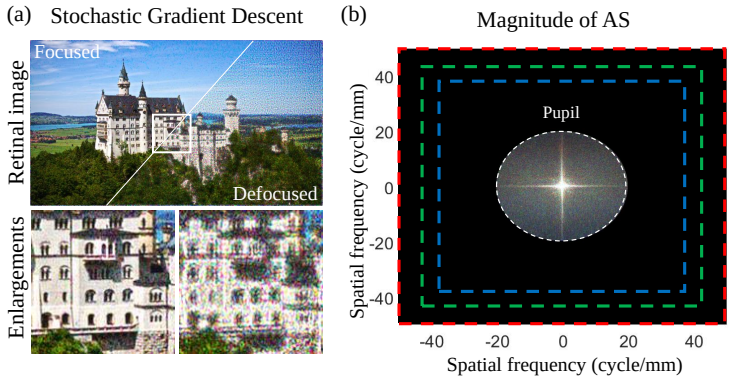


Figure 3.5 Reconstructed holographic images of 2D CGH acquired with SGD algorithm and the magnitude of the field's angular spectrum. The dashed line in red, green, blue color denotes the eyebox of the near-eye display system.

3.2.2 Low image quality in classical 2D CGH

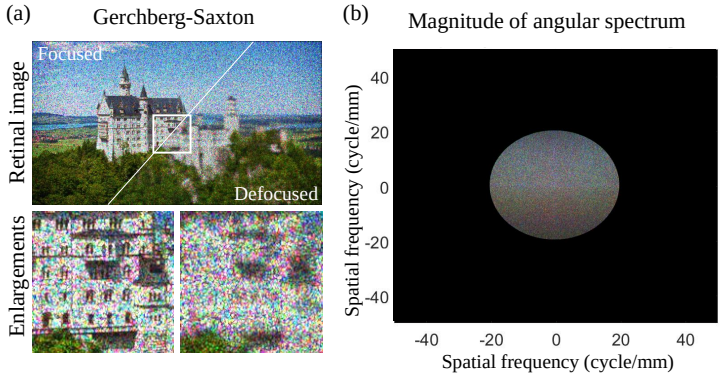


Figure 3.6 Reconstructed holographic images of 2D CGH acquired with classical GS algorithm and the magnitude of the field's angular spectrum.

On the contrary, the image produced by the Gerchberg-Saxton (GS) hologram [41] exhibits low image quality even in the focused state, as illustrated in Fig. 3.6.

When the image is defocused with a difference of 0.6 D, it appears sufficiently blurry, but severe speckle noise is present in both the focused and defocused images. The reconstructed results shown in Fig. 3.5 and Fig. 3.6 suggest that the SGD CGH realizes a field with a relatively smoother phase distribution, whereas the GS CGH reconstructs a field with a random phase distribution. It is reasonable to assume that the GS CGH can induce the accommodation response as in the incoherent case, which can be inferred from the angular spectrum profiles showing uniformity over the eye pupil. However, to ensure that the stimuli induce a sufficient accommodation response, the contrast gradients - the change in contrast concerning the change in the focal diopter of the eye - of holographic contents must be examined.

3.3 Improving accommodation response

3.3.1 Contrast ratio of holographic stimuli

Suppose the complex field distribution $u \in \mathbb{C}^{M\times N\times 3}$ is reconstructed at the focal length (f) of an eyepiece lens, and an eye with a dioptric error of ΔD is positioned at the opposite focal plane of the eyepiece. The amplitude transfer function, which describes the optical relationship in a coherent imaging system, is obtained as follows:

$$ATF_{\Delta D}(\nu_x, \nu_y) = \mathcal{A}(f\nu_x, f\nu_y)e^{i\frac{2\pi}{\lambda}\mathcal{W}_{\Delta D}(f\nu_x, f\nu_y)}, \tag{3.3}$$

where, $\mathcal{A}$ is the apodization function, and $\mathcal{W}_{\Delta D}$ is the aberration function of the given system with additional eye dioptric error of ΔD.

If the eye pupil is diffraction-limited and circular with a radius of r_{ep}, it acts as a finite passband in the frequency domain, and the corresponding apodiza-

tion functions are denoted as $\mathcal{A} = circ\left(f\sqrt{\nu_x^2 + \nu_y^2}/rep\right)$. Similarly, the aberration function, resulting from the path-length error of the incident beam on the pupil, is expressed in a quadratic form of spatial frequencies as $\mathcal{W}_{\Delta D} = f^2\pi\Delta D\left(\nu_x^2 + \nu_y^2\right)/\lambda$. In a coherent optical system, the intensity profile at the plane optically conjugated to the retinal plane is the absolute square of the reconstructed field, given by:

$$I_{c,\Delta D}(u) = \left|\mathcal{F}^{-1}\left(\mathcal{F}(u)ATF_{\Delta D}\left(\nu_x, \nu_y\right)\right)\right|^2, \tag{3.4}$$

where, $\mathcal{F}^{-1}(\cdot)$ is a two-dimensional inverse Fourier transform. The processed image is resized to a lower resolution to simulate the projected image in the retinal plane with a unit cell of 2 $\mu m \times 2\ \mu m$.

Humans do not directly perceive the visual stimulus; thus, the author employs perceptually plausible assumptions to simulate the perceived image. The simulation involves multiplying a frequency-dependent weight, corresponding to the contrast sensitivity function (CSF), with the frequency components of the retinal image reconstructed under the diffraction-limited condition. The CSF approximates a high-level visual function. The contrast ratio, which represents the ratio of the perceived image with a dioptric error to the focused image, is expressed for each spatial frequency band $\mathcal{S} : [\nu min, \nu max]$ as

$$CR_{\mathcal{S}}(I_{\Delta D}) = \frac{\iint\limits_{\mathcal{S}} \mathcal{F}\left(I_{\Delta D}\right) CSF(\nu_x, \nu_y) d\nu_x d\nu_y}{\iint\limits_{\mathcal{S}} \mathcal{F}\left(I_0\right) CSF(\nu_x, \nu_y) d\nu_x d\nu_y}. \tag{3.5}$$

The CSF model investigated by Barten et al. [11] is applied by the author, as this model includes additional parameterization with respect to the background luminance level. For the current study, the luminance conditions of a conventional VR device, HTC Vive Pro, are considered, with a reported luminance level of 133.3 cd/m^2 [50], and a background brightness level of 0.1 cd/m^2. The

CSF is determined accordingly. Additionally, the author assumes that the CSF remains constant across the retinal eccentricity, given that the display prototype provides an image with a limited field of view. In this paper, the term “contrast gradient” refers to the change in the contrast ratio concerning the change in the focal diopter of the eye.

In a similar manner, if the incoherent image is denoted as $i \in \mathbb{R}^{M\times N\times 3}$ and is placed in an optically equivalent condition, the resulting reconstructed image is expressed as

$$I_{i,\Delta D}(i) = \mathcal{F}^{-1}\left(\mathcal{F}(i)\left(ATF_{\Delta D}\left(\nu_x,\nu_y\right) \star ATF_{\Delta D}\left(\nu_x,\nu_y\right)\right)\right), \tag{3.6}$$

where, the symbol $\star$ denotes the autocorrelation integral. To estimate the contrast ratio of the incoherent stimulus, Eq.3.5 is used, and the intensity term is replaced with Eq.3.6. An incoherent visual stimulus is treated as a ground truth case, similar to real-world scenes.

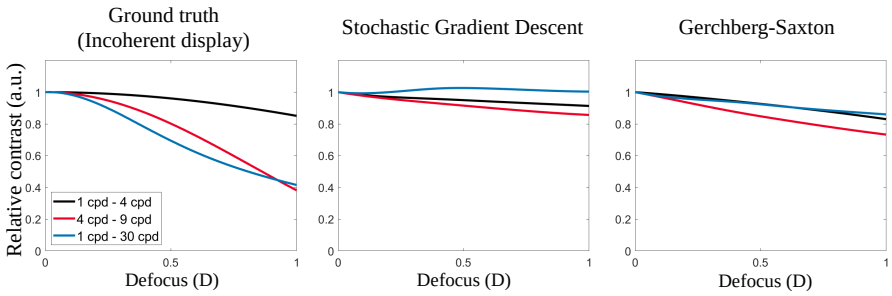


Figure 3.7 Comparison of contrast ratio of holographic images realized with SGD CGH, GS CGH with ground truth case.

Using Eqs. 3.4 to 3.6, the contrast ratio curve for 2D holographic stimuli can be estimated, as shown in Fig.3.7. The contrast curves for each scheme are provided, depicting different spatial frequency regions (1-4 cpd in black, 4-9 cpd in red, and 1-30 cpd in blue). However, none of the cases exhibit a

contrast curve similar to the contrast curve simulated with the ground truth case of an incoherent display. Particularly, the contrast ratio of regions corresponding to the middle spatial frequency range (4-9 cpd) or broadband stimulus (1-30 cpd) varies significantly from the ground truth case. It is worth noting a slight increase in the contrast ratio for the SGD CGH in the frequency range from 1 cpd to 30 cpd. This is due to the image being over-fitted to the focused plane, leading to noticeable noise in the holographic images reconstructed in the out-of-focus region.

The reconstructed images with SGD holograms demonstrate slight contrast gradients because of the narrow size of the effective bandwidth. However, in the case of GS holograms, the deficient contrast change over the defocused diopter can be explained by the existence of the speckle noise shown in Fig. 3.8. Here, speckle contrast (C_s) is estimated as a ratio of the standard deviation of the reconstructed intensity over the mean intensity. The existence of speckle contrast limits the contrast gradient, although GS CGH showed a relatively uniform angular spectrum as demonstrated in Fig. 3.6(b).

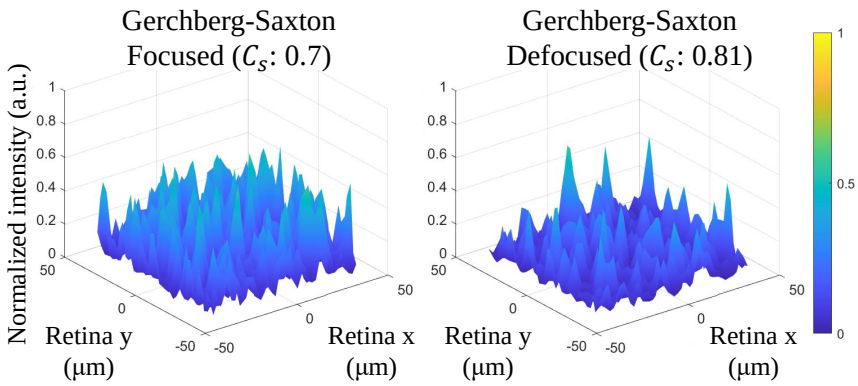


Figure 3.8 Speckle noise present in both focused image (left) and defocused image (right).

3.3.2 Principle

In this subsection, the author presents two approaches for improving the accommodation response in 2D holographic near-eye displays. The goal is to provide holographic stimuli with sufficient contrast gradients similar to incoherent stimuli, as illustrated in Fig. 3.9. To achieve this, the CGHs are generated to match the effective depth of focus provided by the incoherent stimuli simulated in the retinal plane of the user's eye.

The first approach involves reducing speckle noise in the realized holographic image. To address this, the author adopts an optical solution known as temporal multiplexing as a practical approach. The second approach focuses on providing CGHs that are optimized using a regularization strategy. This optimization allows for the realization of holographic images with manipulated contrast ratios to smoothly guide accommodation, even in the presence of speckle in the holographic images.

3.3.3 Speckle reduction through temporal multiplexing

The intensity profile realized with temporal multiplexing (TM) of the holograms can be described as an average of the reconstructed intensity profiles defined in linear color space as

$$I_{TM} = \frac{1}{J} \sum_{j=1}^{J} I_c \left(s \cdot \mathcal{P}_d \left(g(z_j) \right) \right), \tag{3.7}$$

where, J represents the number of holograms obtained using different orthogonal random distributions of phase. Within the TM frames, the scale factor remains constant. It is noteworthy that the speckle contrast, which measures the standard deviation of intensity over the mean intensity, decreases proportionally to the square root of the number of frames used in TM [30]. The TM

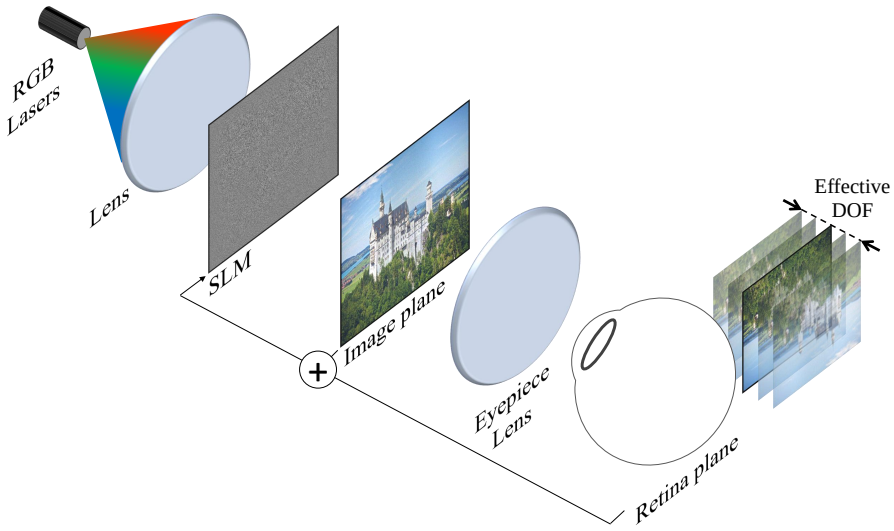


Figure 3.9 Concept of holographic near-eye displays with narrowed effective DOF to improve accommodation response.

technique effectively reduces speckle noise without sacrificing spatial resolution. However, in practice, it can only be implemented using binary SLMs with fast operation speeds. Therefore, the author adopts the recent work on binary hologram optimization with the binary stochastic gradient descent (B-SGD) algorithm [46] as a method to implement holographic TM. This allows for the visualization of a holographic image with minimal contrast degradation resulting from the additional binarization process.

Figure 3.10 illustrates the reconstructed holographic images along with their corresponding contrast curves. The presence of the TM frame in the holograms is shown to reduce speckle noise in both in-focus and out-of-focus images, as depicted in the two images on the left side of Fig. 3.10. The figure provides the reconstructed holographic images with TM frames of 1 and 24, showcasing their focused and defocused states. The contrast curves for these reconstructed

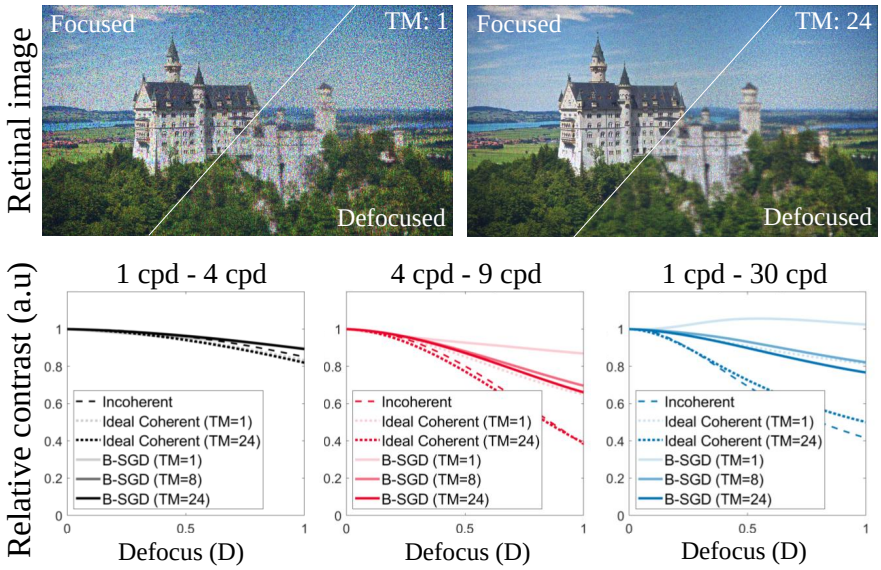


Figure 3.10 Reconstructed holographic images of B-SGD CGHs and corresponding contrast curves with different TM conditions.

images are represented in different colors, each denoting a specific spatial frequency range (first column: 1-4 cpd (black), second column: 4-9 cpd (red), third column: 1-30 cpd (blue)). The contrast curves for *Incoherent*, *Ideal Coherent*, and B-SGD holograms correspond to the dashed line, the dotted line, and the solid lines without a marker, respectively. The transparencies of the lines are adjusted based on the TM conditions. The holographic image reconstructions are conducted under conditions with a pupil diameter of 3 mm and an eyebox of 5.27 mm $\times$ 2.64 mm.

The figure also presents the contrast curves of the reconstructed images for two different schemes (*Ideal Coherent*, B-SGD) under various TM conditions, with the ground truth case of an incoherent display (*Incoherent*) shown in the second row of Fig. 3.10. The term *Ideal Coherent* is used when a complex-

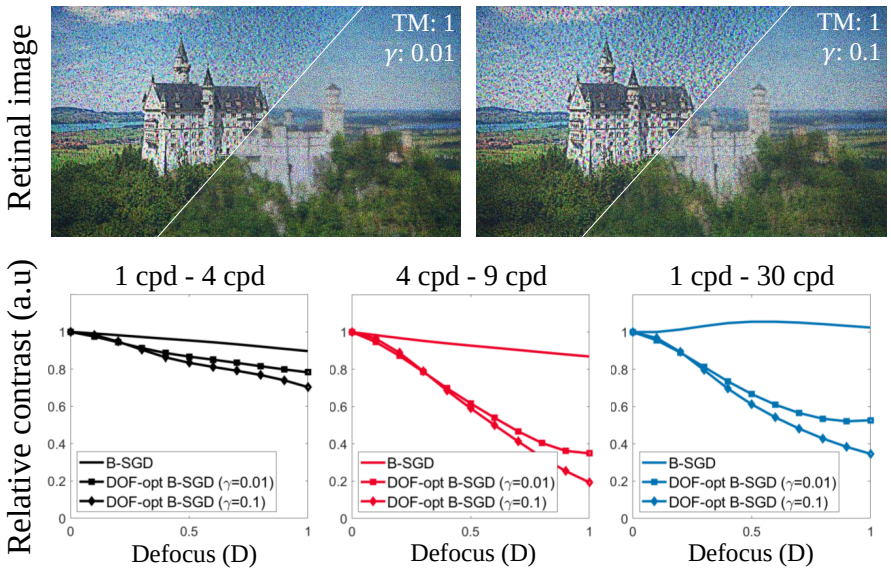


Figure 3.11 Reconstructed single-frame holographic images and corresponding contrast curves with different contrast ratio regularizers.

valued field with a random phase distribution is reconstructed without additional procedures affecting the original field, such as encoding and propagation. The contrast curves of *Ideal Coherent* are similar to those of *Incoherent* when speckle noise is reduced using TM. Additionally, the contrast curves of the images obtained with B-SGD holograms are presented under different TM conditions. Although these contrast curves differ from the ideal cases, speckle reduction through TM effectively decreases the contrast in the defocused image, particularly in the middle spatial frequency region (4-9 cpd) and the broadband range (1-30 cpd).

3.3.4 CGH optimization with contrast ratio regularization

Mitigating speckle noise using TM proves to be a highly effective approach for achieving smooth contrast degradation across focal states. Nevertheless, this method can become burdensome due to the increase in hologram acquisition time, which scales with the number of frames needed for TM. To address this challenge, the author presents a CGH optimization strategy in this subsection. The primary objective of this strategy is to design the contrast ratio of the holographic stimulus in a manner that minimizes quality degradation in the focused image. The optimization process involves a loss function comprising two key components: the mean squared error between the reconstructed and target amplitude and a regularization term concerning the contrast ratio as

$$\begin{aligned}\mathcal{L} &= \mathcal{L}_a + \mathcal{L}_{CR} = \left\| s \cdot \left| \mathcal{P}_z \left(g(u) \right) \right| - a_{target} \right\|_2^2 \\ &+ \frac{\gamma}{N} \sum_{n=1}^{N} \left\| CR_{\mathcal{S}} \left(\Gamma \left(I_{c,\Delta D_n} \left(s \cdot \mathcal{P}_z \left(g(u) \right) \right) \right) \right) - CR_{\mathcal{S},\Delta D_n,target} \right\|_1\end{aligned} \tag{3.8}$$

where, $|\cdot|^1$ and $|\cdot|^2$ correspond to the l_1 and l_2 norm operators, respectively. The regularization coefficient γ is defined by the user to strike a balance between the two loss terms. The optimization procedure takes into account a total of N focal states. The operator $\Gamma(\cdot)$ represents the sRGB gamma correction operator, which is approximately equal to $(\cdot)^{1/2.2}$.

The amplitude loss ($\mathcal{L}_a$) is calculated as the square of the l_2 norm of the difference between the reconstructed and target amplitude. In the regularization term, the contrast ratio loss ($\mathcal{L}_{CR}$) is defined as the average of the l_1 norm of the difference between the contrast ratio of the reconstructed holographic image and that of the incoherent stimulus estimated at each focal state. Here, $a_{target} = \sqrt{i_{lin}} = \left(\Gamma^{-1}(i) \right)^{1/2} \approx i$ represents the target amplitude for CGH optimization, which approximately corresponds to the target image in the sRGB color space.

The target contrast ratio, denoted as $CR_{\mathcal{S},target,\Delta D_n} = CR_{\mathcal{S}}\left(I_{i,\Delta D_n}\left(i\right)\right)$, is obtained using the incoherent image and estimated within a spatial frequency band ranging from 1 cpd to 30 cpd.

The optimization process with an additional regularization term leads to quality degradation in the focused image, as demonstrated in the reconstructed images shown in the first row of Fig. 3.11. The figure displays the reconstructed holographic images with regularization coefficients of 0.01 and 0.1, both in focused and defocused states. The contrast curves of the reconstructed holographic images are represented in different colors, each indicating a specific spatial frequency range (first column: 1-4 cpd (black), second column: 4-9 cpd (red), third column: 1-30 cpd (blue)). The hologram cases of B-SGD, DOF-opt B-SGD (γ=0.01), and DOF-opt B-SGD (γ=0.1) correspond to the solid lines without a marker, with a square marker, and a diamond marker, respectively. These reconstructed images are created under the same conditions as in Fig. 3.10.

As the regularization coefficient increases, the issue becomes more pronounced, leading to images with more artifacts. The corresponding contrast curves of the holographic images are provided alongside the case of the B-SGD hologram in the second row of Fig. 3.11. The optimized hologram with a larger regularization coefficient (γ=0.1) exhibits a contrast curve in the broadband frequency range (1-30 cpd) similar to that of the ground truth case. However, the contrast curve of the DOF-opt B-SGD (γ=0.01) stimulus in the middle spatial frequency region (4-9 cpd) is closer to the incoherent case compared to DOF-opt B-SGD (γ=0.1). It is worth noting that accommodation and blur perception are primarily influenced by spatial frequencies ranging from 4-9 cpd [51, 52].

3.4 Implementation

In this section, the author discusses the execution of the work, which can be divided into two primary components: hardware and software. A significant amount of work has been dedicated to creating a system that is suitable for conducting offline user experiments. Prior to these experiments, the recorded 2D holographic scenes are demonstrated for further analysis and assessment.

3.4.1 Hardware

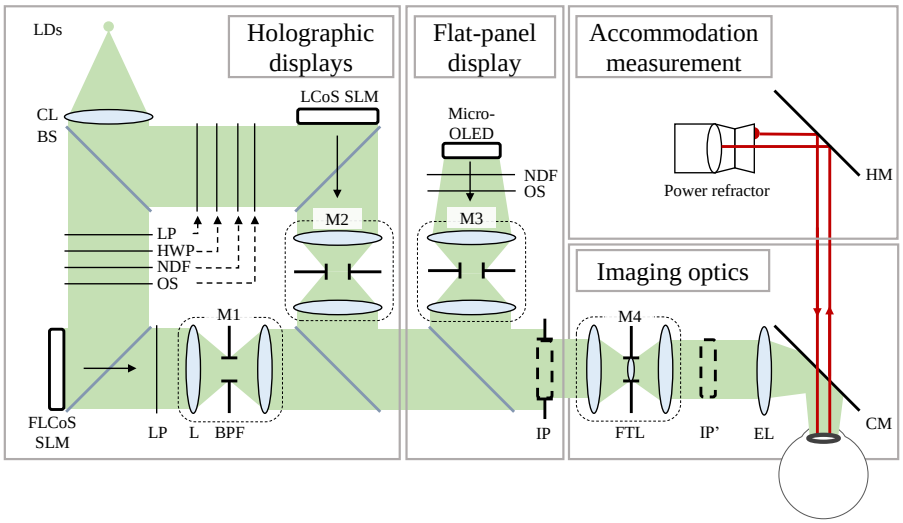
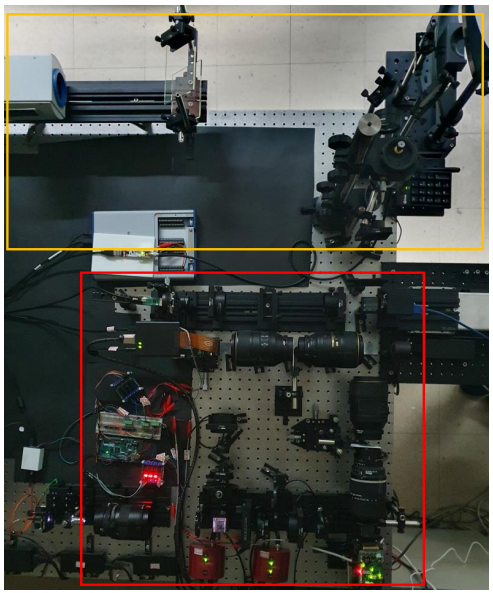
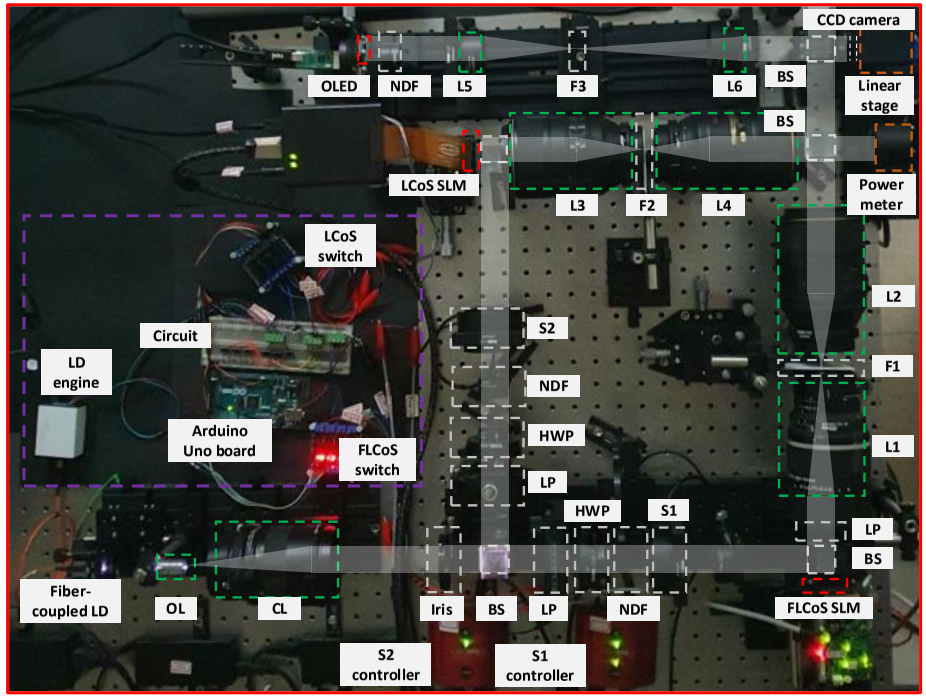
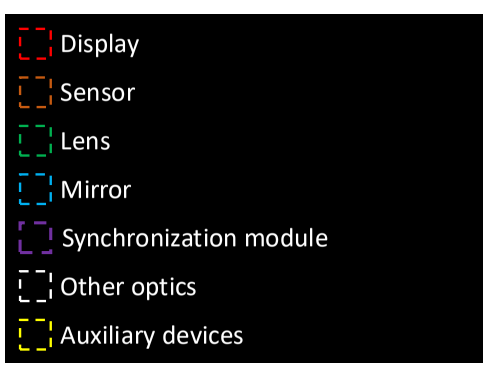


Figure 3.12 Schematic of the apparatus built for the experimental assessments. (LD: laser diode, CL: collimating lens, BS: beam splitter, LP: linear polarizer, HWP: half wave plate, NDF: neutral density filter, OS: optical shutter, FLCoS SLM: ferroelectric liquid crystal on silicon spatial light modulator, LCoS SLM: liquid crystal on silicon spatial light modulator, M: Magnifying optics, BPF: bandpass filter, IP: imaging plane, FTL: focus tunable lens, EL: eyepiece lens, HM: hot mirror, CM: cold mirror.)



<Overview of benchtop prototype>





<Classifications>

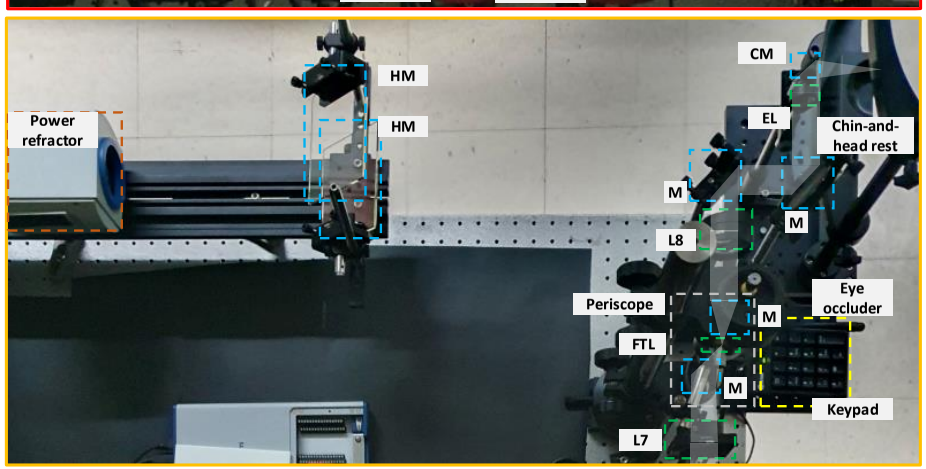


Figure 3.13 Overview of the benchtop prototype utilized in the experimental evaluations.

Figure 3.12 illustrates the schematic of the experimental apparatus used in the assessment. The entire system can be divided into four main parts: Fresnel-type holographic displays employing two different SLMs, a flat-panel display, accommodation measurement devices, and other imaging optics.

In the holographic display section, the hardware setup utilizes a fiber-coupled laser diode (LD) of Wikioptics as the laser source, emitting a full-color beam with central wavelengths of 638, 520, and 450 nm. Full-color holographic visualization is achieved by synchronizing the color sequence with the SLMs. The beam is collimated with a lens and split into two paths using a beam splitter (BS). Two different spatial light modulators are employed for the assessment. One beam path passes through a linear polarizer (LP) and a half-wave plate (HWP) to match the polarization angle required by the ferroelectric liquid crystal on silicon spatial light modulator (FLCoS SLM). An additional neutral density filter (NDF) and optical shutter (OS) are placed in the optical path to attenuate and block the beam, respectively. The FLCoS SLM, manufactured by Fourth Dimension Displays, with a pixel pitch of 8.2 $\mu m \times 8.2$ μm and a resolution of 1920×1200, provides 24 different full-color binary patterns at 50 Hz and modulates the binary amplitude of the field with an additional LP placed after the SLM. The FLCoS SLM plane is optically shifted using relay optics built with two camera lenses with a magnification ratio (M1) of 0.78. A bandpass filter (BPF) is positioned at the Fourier plane to block high-order signals. The physical size of the BPF is determined based on the blue signal bandwidth and is vertically halved due to the complex encoding of the amplitude hologram. Similarly, the other optical path of the laser beam passes through LP, HWP, NDF, and OS for identical purposes.

The Holoeye LETO LCoS SLM, with a pixel pitch of 6.4 $\mu m \times 6.4$ μm, a resolution of 1920×1080, and a full-color operation frame rate of 60 Hz, is used

to modulate phase components with a bit depth of eight. The modulated beam passes through a 4-f system (M2(=1)) and a BPF. Both SLMs are provided with the hologram to reconstruct the desired intensity in the target image plane (IP). Additionally, both SLMs are vertically rotated at an angle corresponding to half of the diffraction angle of the blue signal to prevent undiffracted DC noise from entering the eye. The computer-generated holograms (CGHs) undergo an additional shift in the frequency domain for off-axis reconstruction. The complete system used for evaluation is depicted in Fig.3.13. For a more detailed description, please refer to the Supplementary Material of the publication [53].

3.4.2 Software

The implementation of CGH acquisition was carried out using PyTorch. The author devised a forward propagation model, and the automatic differentiation capability of PyTorch facilitated CGH optimization by monitoring the gradient flow. A learning rate of 0.1 was utilized to update the gradient of the loss function during the optimization process. The implementation was conducted on an Nvidia Geforce RTX 3080 Ti graphic card with 12 GB RAM.

For the iterative CGH algorithms (GS, SGD, B-SGD) constructed with a plane-to-plane model, a total of 500 iterations took approximately 20 seconds to acquire a single hologram frame. However, when optimizing the contrast ratio with the simulation involving ten defocused images and a focused image (ranging from -1.0 to 1.0 D with a unit step of 0.2 D), it took around 10 minutes for 500 iterations. Although further improvements in computational speed can be achieved through parameter tuning and code optimization, the relatively slow computation does not compromise the significance of the study.

3.4.3 Camera-based assessments

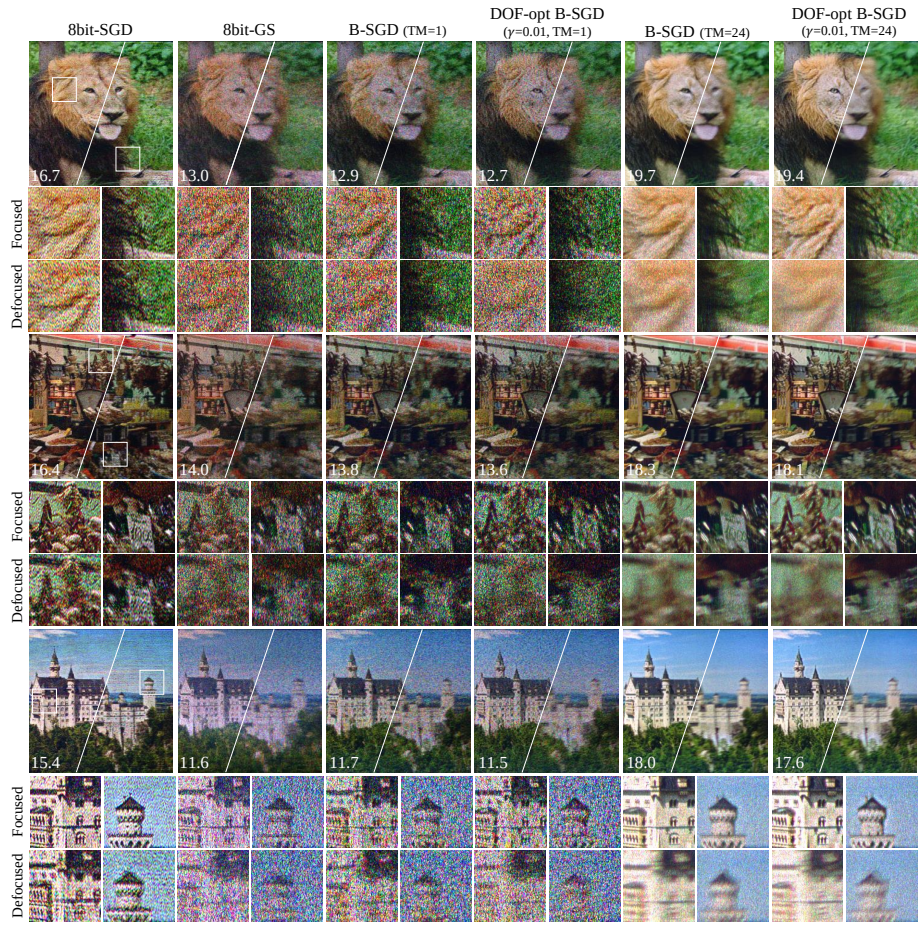


Figure 3.14 Experimental results of holographic images acquired with various algorithms.

Before conducting user evaluations on holographic content, the author conducted assessments of the display prototypes. To capture the individual images of various CGH algorithms, a charge-coupled device (FLIR, GS3-U3-91S6C) with a resolution of 3376 $\times$ 2704 and a pitch of 3.69 μm was placed at the image plane without an additional attached lens, as depicted in Fig.3.14. The holo-

graphic images obtained with different algorithms and various tone-mapping (TM) conditions were photographed at the image plane of the display prototypes. The left column shows captured images of 8bit-SGD, 8bit-GS, B-SGD (TM=1), DOF-opt B-SGD (γ=0.01, TM=1), B-SGD (TM=24), and DOF-opt B-SGD (γ=0.01, TM=24), both in focus and out of focus with a 0.6 D difference. The peak signal-to-noise ratio (PSNR) estimated from the captured result of the focused image is provided in the bottom left corner of each image. The image sources are derived from the DIV2K dataset [54].

The image is located at a distance of 130 mm from the relayed LCoS SLM plane and 200 mm from the relayed plane of FLCoS SLM. To account for the dioptric difference of 0.6 D (1.0-1.6 D) in the near-eye display prototype condition, it is converted to a metric distance of 2.8 mm for capturing defocused results. Among the images created using single-frame CGHs, the results of 8bit-SGD exhibited the highest PSNR, although they were not adequately blurred in the defocused state. The artifacts caused by pupil apodization were eliminated by capturing holographic images at the image plane. Additional captured results can be found in Fig. 3.15-3.17.

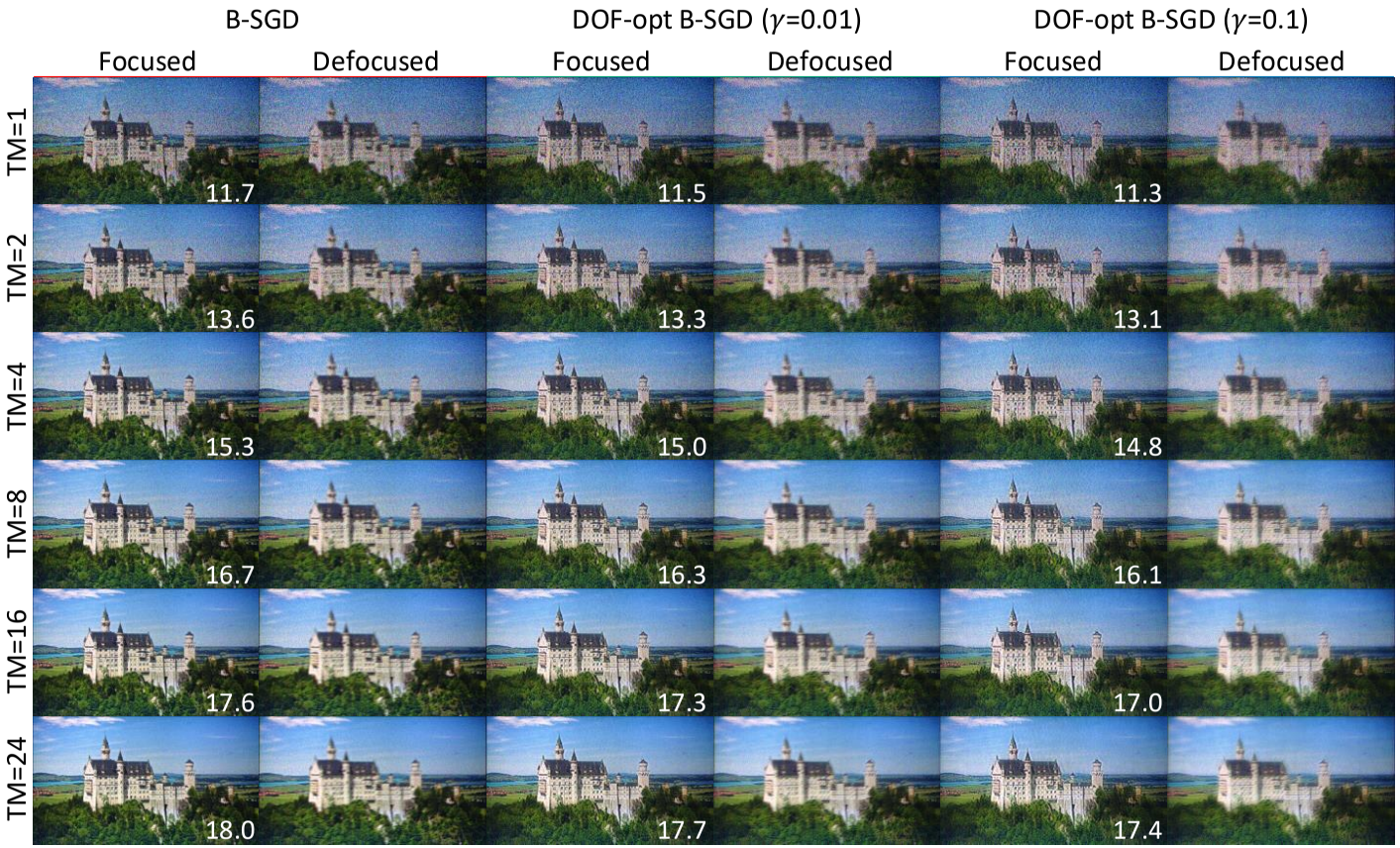


Figure 3.15 Experimental results of *castle* scene of various binary CGH algorithms and TM conditions are provided with PSNR.

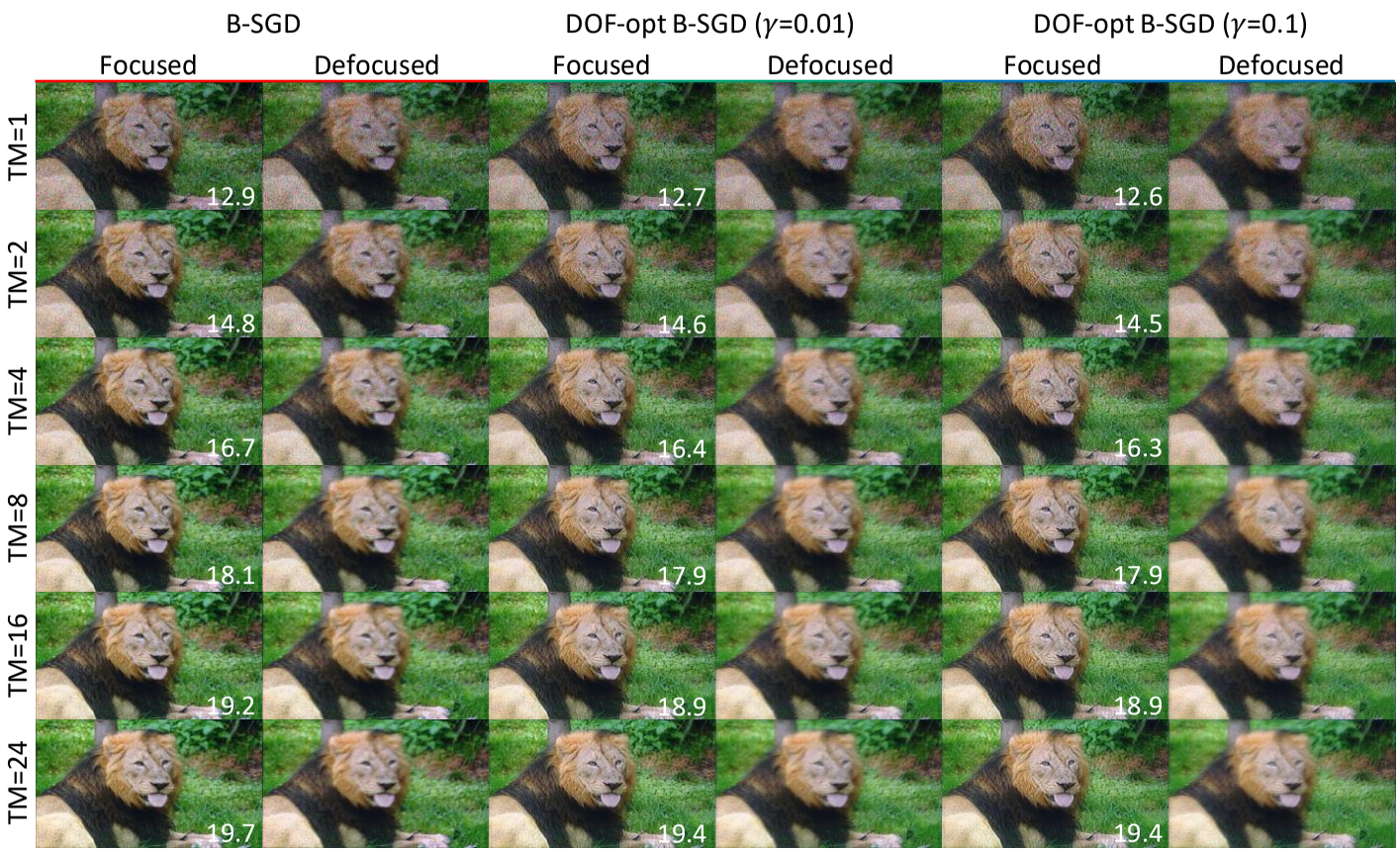


Figure 3.16 Experimental results of *lion* scene of various binary CGH algorithms and TM conditions are provided with PSNR.

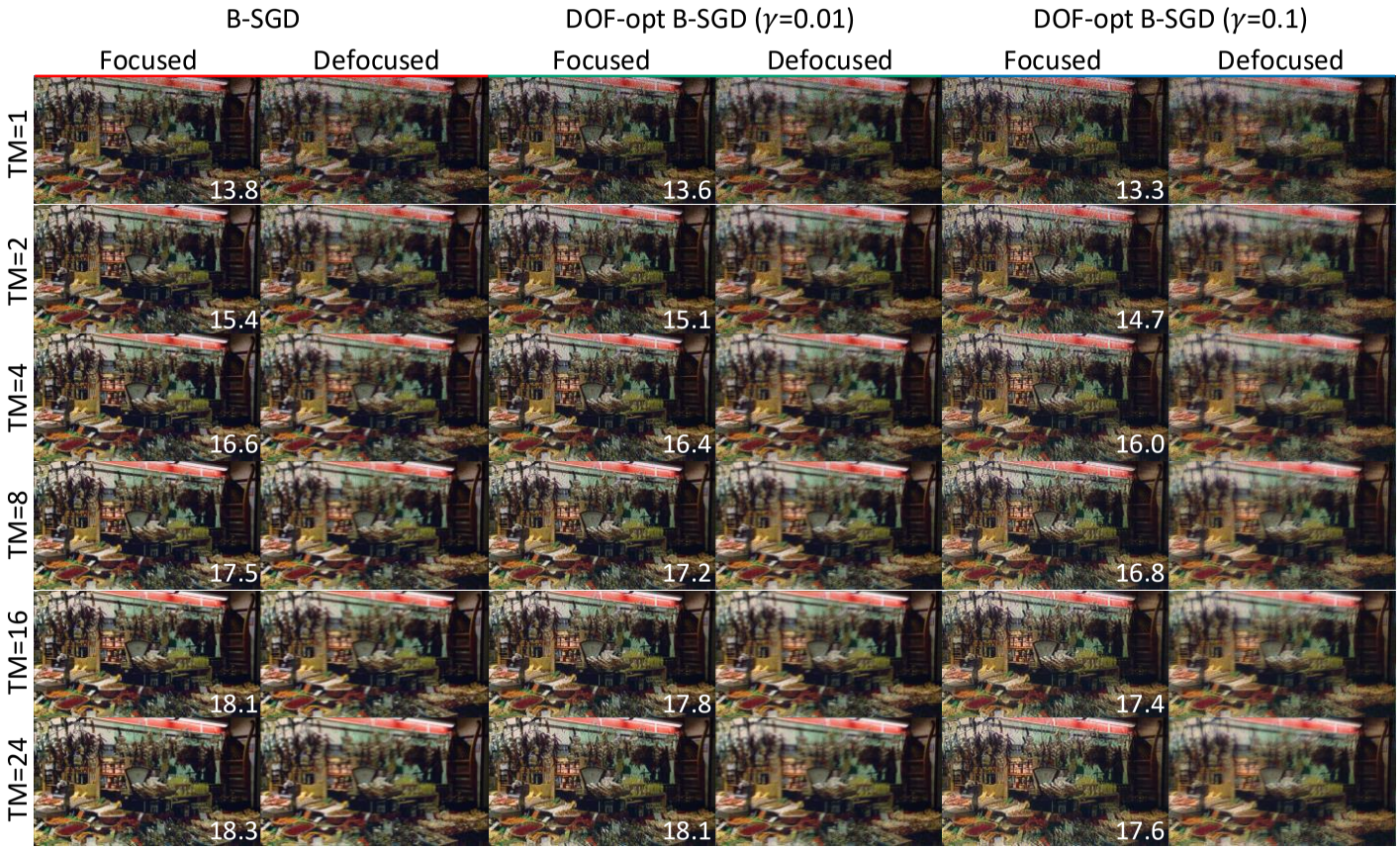


Figure 3.17 Experimental results of *market* scene of various binary CGH algorithms and TM conditions are provided with PSNR.

3.5 Accommodation experiments

In this section, the author conducts user experiments to evaluate various CGH algorithms. The main focus of these experiments is to measure the accommodation responses and validate the insufficient support of accommodation cues in existing CGH algorithms. Additionally, the experiments aim to demonstrate the effectiveness of the proposed approaches in enhancing accommodation cues for holographic displays.

3.5.1 Methods

Subjects

The experiments involved thirty naïve subjects, ranging from 20 to 30 years old, with a mean age of 24.2 years. Among the participants, 18 were female, and the remaining were male. To ensure consistency in the results, we recruited individuals with normal or corrected-to-normal visual acuity and aged under 40, as both visual acuity and age can impact the depth of focus and accommodation range, thereby influencing the overall accommodative gain. Before the tests, we measured the spherical equivalents (SEs) of both eyes of each participant using an autorefractor (Huvitz, HRK-8000A). The participants were encouraged to take part in the experiment using the eye with the smaller SE. All participants had normal or corrected-to-normal acuity, with an average SE of -0.52 D, and none of them reported any color deficiency or color-blind vision. The studies adhered to the principles outlined in the Declaration of Helsinki, and all subjects provided voluntary written and informed consent. The research was approved by the Institutional Review Board at the host institution.

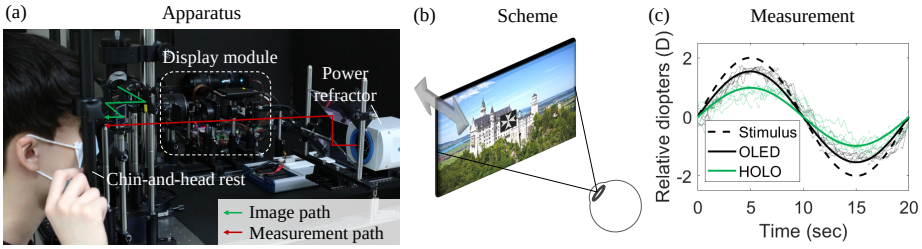


Figure 3.18 Accommodation response measurement experiments with a holographic near-eye display.

Apparatus

Figure 3.18 provides an overview of the entire accommodation experiments. Figure 3.18(a) illustrates the apparatus used in the experimental setup, where the participants view a depth-varying 2D stimulus (a castle scene with a gray Maltese cross at the center) with one eye, as described in Fig. 3.18(b). The holographic near-eye display prototype presents the image with a resolution of 1600×900, offering a corresponding field of view of $7.8^\circ \times 4.4^\circ$ with a 2-inch eye lens (EL) and a focal length of 75 mm. The eyebox size of the near-eye display system is specified as 5.27 mm $\times$ 2.64 mm.

It is worth noting that using eye lenses with short focal lengths increases the field of view but reduces the size of the eyebox, making the holographic near-eye display system optically unsuitable for evaluating accommodative response. The Nyquist frequency of the holographic image is estimated to be 102 cpd, which exceeds the maximum spatial resolution that the human eye can perceive [12]. For the experiments, full-color images are utilized, as color serves as one of the accommodation cues [24, 55]. During the tests, the room is kept sufficiently dark, except for the stimuli provided by the display prototype.

The optical powers of the individual holographic scenes are measured using

an optical power detector (Newport,918D-SL-OD3R) with a circular aperture of 11.3 mm connected to the power meter (Newport, 2936-R). The light fluxes of the two holographic displays are balanced by placing neutral density filters (NDFs) with different transmittance ratios. The average optical powers of the reconstructed scenes (lion, market, castle, and castle with a Maltese cross) in the red (638 nm), green (520 nm), and blue (450 nm) channels are measured as 1.6 nW, 1.8 nW, and 2.5 nW, respectively. Then, the luminance level is calculated considering the luminosity functions and estimated to be 0.2 cd/m^2, which is below the permissible level of laser exposure [56]. The experiments are conducted only with this low luminance level to address potential eye safety concerns associated with holographic displays. The luminance levels of the three different displays are balanced by placing NDFs with different transmittance ratios in each beam path.

Procedure

A participant viewed the 2D stimulus presented at a distance of 0.33 meters (3 D) using either their left or right eye, ensuring proper alignment with the eyebox of the display system by adjusting the chin-and-head rest. The participant was sequentially presented with three different images: one reconstructed with FLCoS SLM, another with LCoS SLM, and a third with an OLED panel, and their observation for each was verified. The experimental stimulus was a castle scene with a gray Maltese cross at the center. The target moved sinusoidally in depth, ranging from 0.2 meters (5 D) to 1 meter (1 D) and back, over two periods with each period lasting 20 seconds, preceded by a 5-second buffer for each trial. The total range of motion was 4 D. Subsequently, the dynamic accommodation response was measured three times for each condition.

There were one reference mode and five different hologram modes: OLED,

where a micro-OLED displays the stimulus; 8bit-SGD and 8bit-GS, where LCoS SLM displays a single frame of 8bit hologram acquired with SGD and GS algorithms, respectively; B-SGD, where FLCoS SLM shows a binary hologram acquired with the SGD algorithm; DOF-opt B-SGD (γ=0.01 and γ=0.1), where FLCoS SLM displays a binary hologram optimized with the proposed algorithm. Additionally, the modes providing binary holograms were tested under various TM conditions: TM=1, 2, 4, 8, 16, 24. Thus, a total of 21 conditions were evaluated, with 63 trials per subject. The conditions were presented randomly, and a 30-second break was given between trials. Due to the significant fatigue induced by the long-duration dynamic accommodation task, the tests were conducted over two separate days, as the experiment lasted approximately 1.5 hours.

Data processing

The author collected the measured data in the form of a comma-separated value (csv) file for further analysis. Data corrupted by eye blinking was excluded, and the response data was divided into two periods, each lasting 20 seconds (1000 points), measured after a 5-second buffer for every trial. Six trials of each viewing condition are depicted as thin black lines in Fig. 3.18(c). The figure also shows the accommodative responses of an individual user, along with representative instances measured when a stimulus is presented by an incoherent display (OLED, black) or a holographic display (HOLO, green) with dioptric modulation, as indicated by the dashed black line. Each set of 1000 points was fitted to a sinusoidal curve. The measured data was fitted with a sinusoid at the target frequency, and the amplitude, phase delay, and DC offset were treated as free parameters using the Levenberg-Marquardt damped least-squares method. Periods with a valid measurement ratio below 70% and a residual norm of curve-fitting

greater than 0.3 were excluded. Additionally, any fitted curve with an amplitude greater than three standard deviations from the median amplitude was considered an outlier and excluded from the analysis. After these exclusions, the amplitudes for each condition were averaged.

Traditionally, the accommodative gain is calculated as the ratio of the amplitude of the fitted sinusoidal curve to the stimulus sinusoid [57]. However, in this study, the accommodative gain was normalized with the mean amplitude of the fitted sinusoids obtained when viewing the OLED. This was done because the highest observed accommodative gains measured under optimal conditions were saturated at a level of 0.8-0.9 [24, 58], and the accommodative gains varied significantly between subjects. Nine participants' results were excluded for various reasons: 5 due to a lack of valid measurements caused by small or large pupils (valid measurement ratio $< 70\%$), 2 due to low dioptric amplitude when viewing OLED ($<$ 1.0 D), 1 due to misaligned stimuli depth, and 1 due to different OLED gains on two separate days. Such exclusions are common in accommodation measurement experiments with untrained subjects [59].

3.5.2 Results

The author calculated the normalized accommodative gains for 21 subjects (14 female, mean age: 24.5, mean SE: -0.59 D) for each viewing condition and then averaged the results across all subjects. The accommodation outcomes are depicted in Fig.3.19. Fig.3.19(a) shows the normalized accommodative gains with varying TM frames, while Fig. 3.19(b) compares the accommodative gains measured when viewing holographic images generated by different single-frame CGHs. Asterisks in the figure indicate significant differences in the paired conditions, as evaluated using the Wilcoxon rank-sum test (*: p <0.05, **: p <0.01, ***: p <0.001).

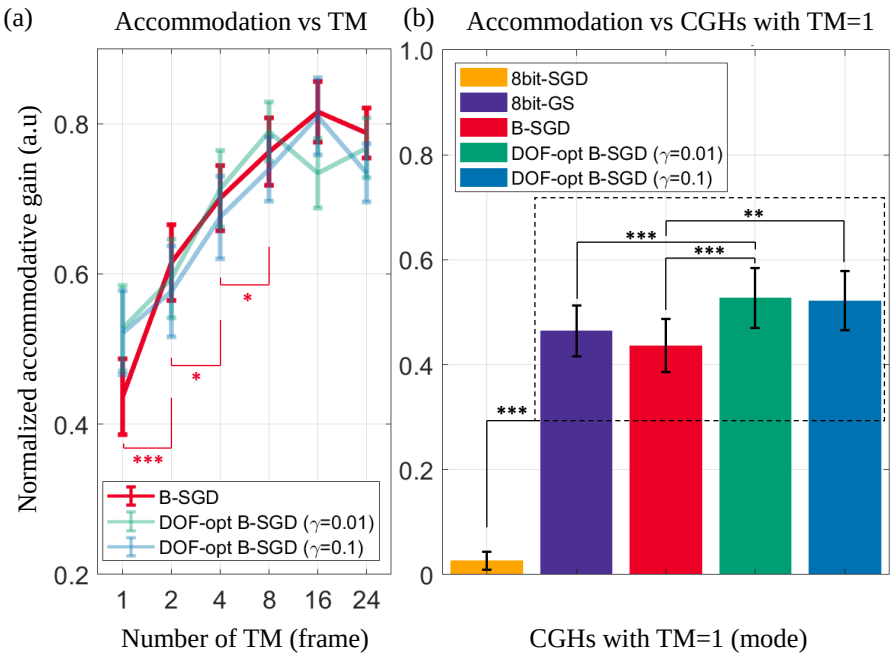


Figure 3.19 Results of accommodation experiments.

The results demonstrated a significant effect of speckle reduction on the improvement of accommodation response. The mean gains evaluated on B-SGD CGHs with different TM conditions (TM=1, 2, 4, 8, 16, 24) were 0.44, 0.61, 0.70, 0.76, 0.81, and 0.79, respectively. For DOF-opt B-SGD (γ=0.01), the mean gains were measured as 0.53, 0.59, 0.71, 0.79, 0.73, and 0.77. Similarly, for DOF-opt B-SGD (γ=0.1), the mean gains were 0.52, 0.58, 0.68, 0.74, 0.81, and 0.73. Nonparametric paired tests were conducted to ensure the statistical reliability of the experiments, as the measured data did not exhibit normality based on the Shapiro-Wilk test. One-tailed Wilcoxon tests indicated statistical significance in the measured gains of B-SGD pairs under consecutive TM conditions, such as TM=1 versus TM=2 (p <0.001), TM=2 versus TM=4 (p <0.05), and TM=4 versus TM=8 (p <0.05). The paired conditions of TM=8 versus

TM=16 and TM=16 versus TM=24 did not show significant differences as the accommodative gain began to saturate. The mean normalized accommodative gains and speckle contrast ratio ($C_s/C_o = 1/\sqrt{N}$;(in case of fully-developed speckle), C_o: speckle contrast at TM=1) exhibited a strong negative correlation with a Pearson coefficient of -0.99 (p <0.001). The absolute value of the speckle contrast at TM=1 varied depending on factors such as the propagation distance from SLM, the numerical aperture of the display system, pupil size, and coherence characteristics of light sources. The tests were conducted using the Scipy package in Python.

Similarly, the author conducted one-tailed Wilcoxon tests with the mean normalized gains measured when viewing holographic images of single-frame CGHs (8bit-SGD, 8bit-GS, B-SGD, DOF-opt B-SGD (γ=0.01), DOF-opt B-SGD (γ=0.1)). They were estimated as 0.03, 0.46, 0.44, 0.53, and 0.52, respectively. There were strong statistical differences between the measured gains of 8bit-SGD CGHs and the other CGHs (p <0.001). The CGHs acquired with the proposed algorithm (DOF-opt B-SGD (γ=0.01 and γ=0.1)) showed significant improvements in accommodative gains over the case of primitive B-SGD under the condition of TM=1 (DOF-opt B-SGD (γ=0.01) versus B-SGD: p <0.001, DOF-opt B-SGD (γ=0.1) versus B-SGD: p <0.01). DOF-opt B-SGD (γ=0.01) showed a strong significant improvement in normalized accommodation gain over 8bit-GS (p <0.001), whereas the other modes of binary CGHs failed to show statistical significance (B-SGD versus 8bit-GS: p=0.82, DOF-opt B-SGD (γ=0.1) versus 8bit-GS: p=0.056). These results show the efficacy of the proposed CGH optimization algorithm.

While conducting statistical tests on the pairs of binary holograms displayed in different TM conditions (TM=2, 4, 8, 16, 24), none of the tests resulted in statistical significance. This could be attributed to the fact that the proposed

CGH acquisition focuses on optimizing a single hologram frame rather than a set of hologram frames used in TM. To address this, the CGH acquisition algorithm could be expanded to incorporate multi-frame optimization [33].

To summarize, speckle reduction through TM led to significant improvements in accommodative gains, and the proposed CGH optimization strategy demonstrated significance in the TM=1 condition. However, the overall CGH acquisition time increases in proportion to the number of hologram frames used in TM. Additionally, the reconstructed holographic image obtained with the proposed algorithm sacrifices the quality of the focused image, as evident in Fig. 3.14. To address potential concerns, the author conducted user experiments to evaluate the subjective image quality of each CGH scheme.

3.6 Subjective image quality assessments

In this section, the author conducts subjective quality evaluations of holographic contents using pairwise comparisons to address two main questions: 1) which CGH algorithm provides a high-quality image in the real holographic viewing experience? and 2) does the holographic image created by the proposed algorithm exhibit noticeable artifacts?

3.6.1 Methods

Subjects

The participants who took part in the accommodation experiments also participated in these experiments. As the previous experiments were conducted on two separate days, the remaining time after the accommodation experiments was utilized for these evaluations. However, three out of the thirty participants

who completed the accommodation experiment were unable to complete these experiments, mainly due to time constraints.

Apparatus

The identical apparatus was utilized as in the accommodation experiments without accommodation measurement.

Procedure

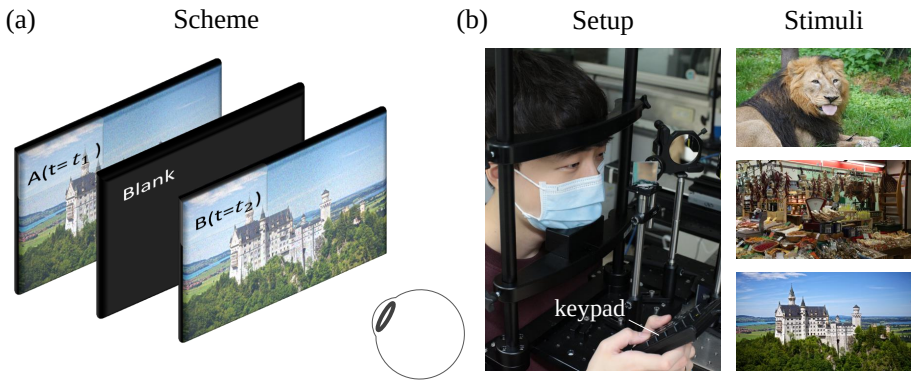


Figure 3.20 Subjective image quality evaluation on holographic contents through pairwise comparisons.

Figure 3.20 provides an overview of the subjective image quality evaluation experiments. In Figure 3.20(a), the experimental setup for subjective image quality evaluation is shown, which employs the two-interval forced choice (2-IFC) method [17]. Each subject views two holographic images sequentially, presented 0.33 m (3 D) away, for 1.5 s per image, with a blank screen displayed for 0.5 s between them. Subsequently, the subject is asked to select the 'high-contrast and less-noisy' image using a keypad, with no 'tie' option available, as depicted in Figure 3.20(b). The stimuli used for pairwise comparisons include

lion, market, and castle images from the DIV2K dataset [54]. To avoid underestimating the differences, we do not include a 'no preference' option [16]. Prior to the experiments, each subject views the stimuli presented by the OLED panel as reference images. Moreover, the sequence of the two options in the 2-IFC experiments is randomly shuffled to prevent any decision bias towards either the former or latter option.

Due to a total of 20 conditions (8bit-SGD, 8bit-GS, B-SGD (TM=1, 2, 4, 8, 16, 24), DOF-opt B-SGD (γ=0.01, 0.1 / TM=1, 2, 4, 8, 16, 24)), it is not feasible to perform complete pairwise comparisons. As a result, the full-scale experiment is split into two parts. The first experiment involves conventional CGHs (8bit-SGD, 8bit-GS, and B-SGD (TM=1, 2, 4, 8, 16, 24)), resulting in 8 options and 28 pairs. Three trials are conducted for each pair, making a total of 84 trials per stimulus for the first experiment. A three-minute break is given between each session. For the second experiment, the author partially compares holographic images realized by binary holograms with identical TM conditions. This experiment is performed in six different TM conditions, with three pairs of binary holograms and three trials per pair, resulting in 54 trials per stimulus. In total, the pairwise comparisons of various CGH algorithms with three different stimuli consist of 414 trials, taking approximately 1.5 hours to complete the entire task.

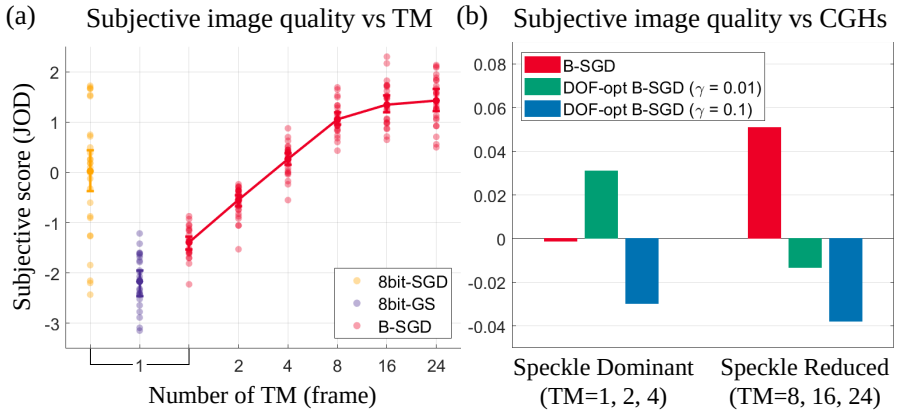


Figure 3.21 Results of subjective image quality evaluation.

3.6.2 Results

The results of the subjective image quality evaluation of holographic contents (8bit-SGD, 8bit-GS, B-SGD with TM=1, 2, 4, 8, 16, 24) through pairwise comparisons were scaled using the Thurston model V [16], and the matrices with accumulated vote counts were converted to a JOD unit [3, 15], as depicted in Fig. 3.21(a). The error bars in the figure represent 95% confidence intervals estimated using bootstrapping. The comparison matrices of the three stimuli were combined into a single matrix. One subject's responses were excluded from the analysis due to being classified as an outlier based on the outlier analysis [16]. This subject's responses showed an inter-quartile-normalized score of 6.59, which exceeds the customary threshold of 1.5. Consequently, data from 27 subjects (2: incomplete data, 1: outlier) were used, resulting in a total of 6,804 accumulated vote counts. The JOD values were regularized to have a mean JOD value of zero. The holographic contents exhibited JOD values of 0.02, -2.17, -1.39, -0.55, 0.26, 1.05, 1.34, and 1.43. The JOD difference increased by approximately 0.8 as the number of TM frames doubled (corresponding to a speckle

noise reduction ratio of the square root of two) until it reached eight, indicating a preference of about 40% over the paired option. However, the JOD difference between B-SGD (TM=8) and B-SGD (TM=16) noticeably decreased to 0.3, representing a preference probability of 16%.

Figure 3.22 Effects of artifacts such as dust in the captured holographic images with different conditions.

In contrast to the PSNR evaluations based on captured images shown in Fig. 3.14-3.17, the 8bit-SGD holographic images exhibited a relatively low JOD average with significantly large deviations. This unexpected degradation in subjective image quality of the 8bit-SGD case may be attributed to the lack of tolerance to artifacts such as scratches or dirt that may undesirably exist in the optical system. While Camera-in-the-loop calibration [34,35] could potentially address this issue, it is limited to artifacts present in the display system. Some subjects reported observing moving particles with ringing patterns in color on 8bit-SGD images. This type of defect in the image may be caused by dust or debris present in the tear film of the user's eye, as the proportional coherent beam is localized in a small section of the pupil for 8bit-SGD images. The tolerance of each CGH algorithm to artifacts present in the pupil plane and the image plane was tested and shown in Fig. 3.22. The results indicate that 8bit-SGD images are susceptible to artifacts. Additionally, the noticeable JOD difference between 8bit-GS and B-SGD TM=1 may result from the undesirable color distortions in 8bit-GS images, as shown in Fig.3.14.

The author compared the binary holograms acquired with different algorithms (B-SGD, DOF-opt B-SGD (γ=0.01, and γ=0.1)) as depicted in Fig. 3.21(b). The results are categorized into two conditions based on the level of speckle reduction for visibility: *Speckle Dominant* (TM=1, 2, 4) and *Speckle Reduced* (TM=8, 16, 24). The total number of trials was insufficient to derive individual results for each image and TM condition. Therefore, the cases of binary holograms reconstructed with TM=1, 2, 4 conditions, and TM=8, 16, 24 conditions were combined and named *Speckle Dominant* and *Speckle Reduced*, respectively. The vote counts of 28 subjects (1: incomplete data, 1: outlier) corresponding to each category were combined to derive JOD values. In the case of *Speckle Dominant*, the participants could not distinguish speckle noise from

additional regularization, resulting in JOD values of 0, 0.031, and -0.030 for B-SGD, DOF-opt B-SGD (γ=0.01), and DOF-opt B-SGD (γ=0.1), respectively. Conversely, in the case of *Speckle Reduced*, the subjects were able to distinguish the artificial noise, leading to measured JOD values of 0.051, -0.013, and -0.038 for B-SGD, DOF-opt B-SGD (γ=0.01), and DOF-opt B-SGD (γ=0.1), respectively. Moreover, a large weight on the l_1 regularization term in the optimization process can lead to noticeable image quality degradation, as DOF-opt B-SGD (γ=0.1) was ranked last among the three candidates. Thus, the proposed CGH algorithm delivers holographic images that can hardly be distinguished from the images provided by the primitive method when speckle is dominantly present.

3.7 Discussion

Model mismatch by human eye

In this study, CGH optimization was conducted assuming a diffraction-limited eye with a fixed pupil diameter. However, in reality, the subjects' pupils continuously dilated and contracted over time, resulting in varying sizes ranging from 4-8 mm even in a uniform luminance setting. Furthermore, eye aberrations were present, even in normal eyes, and were particularly exaggerated under conditions of a large pupil. This discrepancy between the simulated and actual experimental conditions may have limited the validation of this study. Chakravarthula et al. [60] introduced speckle reduction in holographic displays by providing an optimized CGH based on a target optical model of an individual's eye. While this approach is theoretically sound, it is currently challenging to apply practically due to the reoccurrence of speckle even with a slight mismatch in the model, which is almost unavoidable in real-world scenarios.

CGH rendering speed

The process of acquiring holograms is currently time-consuming due to the reconstruction of images for various focal states to estimate the contrast ratio of the reconstructed hologram. In this study, the main emphasis was on user evaluations of holographic contents, rather than achieving real-time CGH rendering, which is crucial for advanced holographic displays. However, it is worth noting that the speed of CGH acquisition can be expected to improve in the near future, as recent developments have started to utilize deep learning technology to generate holograms at real-time frame rates [34, 40, 61, 62].

Holographic contents

The evaluations in this study were conducted using 2D holographic contents, even though there have been recent advancements in achieving high-quality holographic representation of 3D scenes [40, 46, 63]. The main reason for this limitation in scope was the lack of well-defined quantitative criteria for evaluating autostereoscopic 3D displays, particularly holographic displays. Evaluations of 3D contents are often based on subjective questionnaires due to the current ambiguity in assessment methods [20].

Speckle reduction with other approaches

The author utilized temporal multiplexing as a method to reduce speckle without compromising spatial and angular resolution. However, achieving robust speckle reduction in holographic displays may require additional physical considerations. One alternative solution is to use partially coherent light sources, especially in conditions with limited frame rates. However, determining the acceptable level of resolution loss caused by partial coherence of light sources re-

mains an unanswered question, and conducting user experiments to investigate this aspect would be a valuable and intriguing area of research for the relevant scientific community.

3.8 Conclusion

Accommodation cues play a vital role in creating a realistic viewing experience, especially in the context of emerging technologies like virtual reality and augmented reality. This study focuses on understanding and improving accommodation response in holographic near-eye displays, which are becoming increasingly important in these applications. The author introduces strategies to enhance accommodation cues and validates their effectiveness through experimental evaluations, including user studies. Special attention is given to the speckle phenomenon, a characteristic intrinsic to holographic displays with coherent sources. The study provides valuable insights for user evaluations with holographic displays and contributes to the development of diverse evaluation metrics for holographic content.

The chapter evaluates 2D holographic content generated using various CGH algorithms, considering both accommodative gain and subjective image quality. Through experiments with a prototype holographic near-eye display, it is demonstrated that ensuring an appropriate bandwidth size of holographic content is essential to provide monocular accommodation cues effectively. In a holographic viewing environment with a wide eyebox, speckle reduction significantly improves accommodative gain. Additionally, the proposed CGH optimization algorithm, focusing on contrast ratio, proves to be highly effective in enhancing accommodative gains. Among the tested single-frame CGHs, it ranks first and demonstrates minimal quality degradation compared to binary holograms.

Chapter 4. Holographic parallax and its effects on 3D perceptual realism

4.1 Introduction

In the previous chapter, the author investigated the prerequisite aspects of 2D CGHs that could improve accommodation responses with holographic near-eye displays. Firstly, it's crucial that the total energy is uniformly distributed across the eyebox. A 2D holographic scene achieving this characteristic necessitates speckle reduction, as the perceived image's mid-high spatial frequency region is adversely affected by speckle noise, which obstructs the accommodation response. Lastly, the 2D CGH optimized with an added emphasis on contrast ratio revealed its significance in enhancing accommodative gains among the single-frame CGHs.

Chapter 4 extends these findings from the work carried out in the previous chapter and enhances the perceptual realism of 3D scenes displayed through holographic near-eye displays in natural viewing conditions. The methodology entails simulating the perceptual quality of the 3D holographic scenes using various CGH supervision formats and pupil conditions, and factoring in the impact of eye movements on the signals sampled. The author conducts user studies to identify the most effective CGH supervision format for creating lifelike 3D holographic scenes. The results highlight that the inclusion of parallax cues significantly elevates the 3D user experience, even when head movement is minimal. This research marks an initial stride in the domain of 3D visual experience with holographic near-eye displays, and offers guidelines for crafting perceptu-

ally realistic 3D holographic scenes.

4.1.1 Parallax and occlusion as depth perception cues

As Cutting and Vishton's work demonstrates [19] (see Fig. 4.1), humans are most sensitive to images featuring disparity in perceiving depth. This is because binocular disparity is the primary cue for depth perception among various physiological cues, especially for objects within arm's reach. However, when it comes to objects roughly a meter away or more, motion parallax provides minimal depth contrast. Extending the consideration to pictorial cues, humans show a strong response to occlusion as a key factor in determining depth.

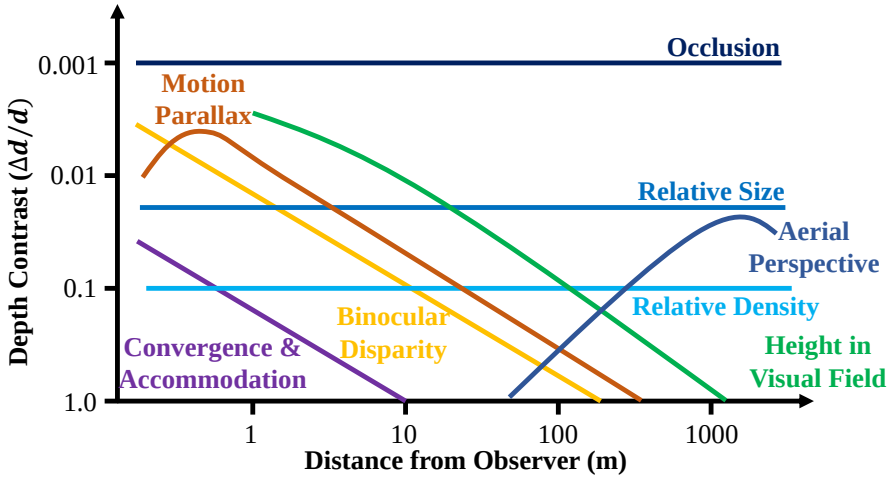


Figure 4.1 Depth perception cues and its sensitivity.

Figure 4.2 illustrates the concepts of motion parallax and dynamic occlusion. Motion parallax is a visual phenomenon experienced when the eyes move from side to side; objects that are further away seem to move slower than those that are closer. Conversely, dynamic occlusion comes into effect with eye rotation, as obscured parts of an object can become visible when viewed from

different angles.

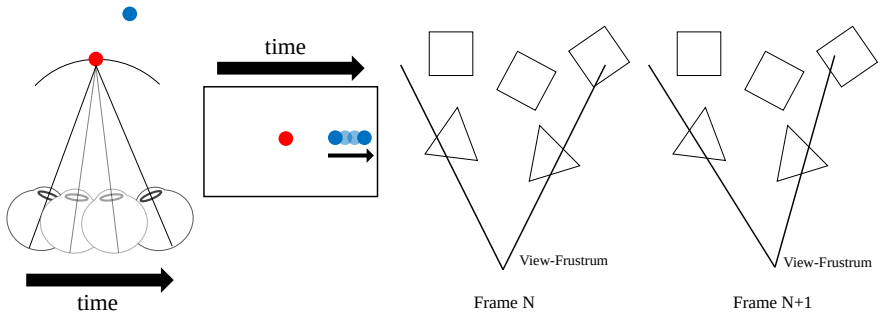


Figure 4.2 Motion parallax and dynamic occlusion.

4.1.2 Gaze-contingent VR rendering

The human eye is constantly in motion, experiencing involuntary pupil contractions and relaxations [64]. To address this, incoherent displays have developed gaze-contingent strategies. For instance, Mercier et al. [65] proposed a multi-plane display with a rapid multi-plane image decomposition algorithm, specifically designed to reflect tracked eye movement. Furthermore, gaze-contingent adaptive-focus displays have been developed to modulate focus-tuning optics in sync with the depth of the objects being gazed at, ensuring sharp focus. Recently, Guan et al. [66] highlighted the perceptual significance of dynamic distortion correction due to pupil movement, even in conventional VR displays with a wide field of view.

In contrast to incoherent displays, holographic displays have unique capabilities in manipulating the plenoptic function of light [31, 67]. However, the limited size of the eyebox and the computational burden of holographic displays often result in approximations of the 3D scene based on the center view,

which can neglect the impact on other views [40].

4.2 Holographic near-eye displays with 3D CGH

4.2.1 Various 3D target formats

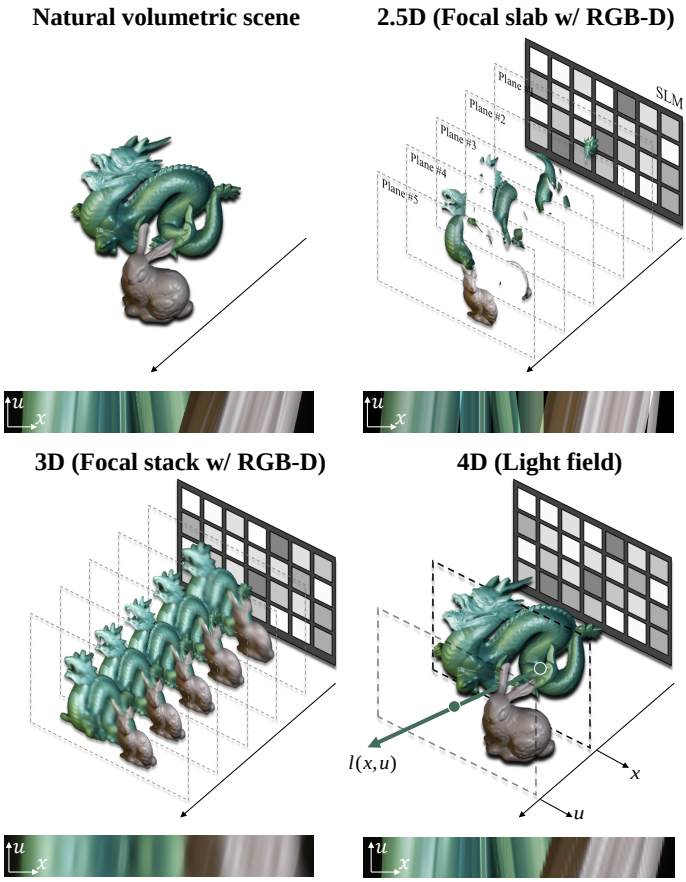


Figure 4.3 Various 3D target formats and the reconstructed epipolar plane image (EPI) corresponding to the target type.

The visual experience of a display device depends on the content it presents, especially for 3D content. To represent the intensity of light passing through

a point in space, variables such as the 3D position of the point through which light passes, the direction of light travel, the wavelength of light, and time are required. Thus, the real-world scenes are a superposition of numerous plenoptic functions, which mandates the approximation of the scenes into a low-dimension target format to alleviate the overall memory capacity and computation load.

Fig. 4.3 demonstrates three representative data formats to realize volumetric scenes. In the reconstruction of the natural volumetric scene, a plane-to-plane model employs the focal slabs (2.5D) or focal stacks (3D) placed at discrete axial distances to reconstruct volumetric scenes. However, a plane-to-perspective model utilizes the spatial information of the sampled views, known as the light field (4D), that contain view-dependent information. $l(x, u)$ demonstrated in Fig. 4.3 denotes the two-dimensional spatial-angular light field defined with a space (x) and a direction (u).

The epipolar plane image (EPI) in Fig. 4.3 is a horizontal cross-section image of ray-space ($f(x, u)$). Here, each data format has its limitations: focal slabs lack spatial information in angles other than the normal angle ($u = 0$) [68]. Focal stacks tend to overfit to the central view, and while the light field (4D) provides angle-dependent spatial information, it does so with sparse sampling. The intrinsic nature of the light field allows for the reproduction of angle-dependent visual effects such as occlusion and shading.

Typically, 3D target formats are standardized in metric units. However, the spatial perception formed by the human eye operates in diopters. Consequently, it's necessary to establish a correlation between these two physical spaces beforehand. This correlation can be determined through the optical setup of the near-eye display system.

4.2.2 System scheme for 3D holographic near-eye displays

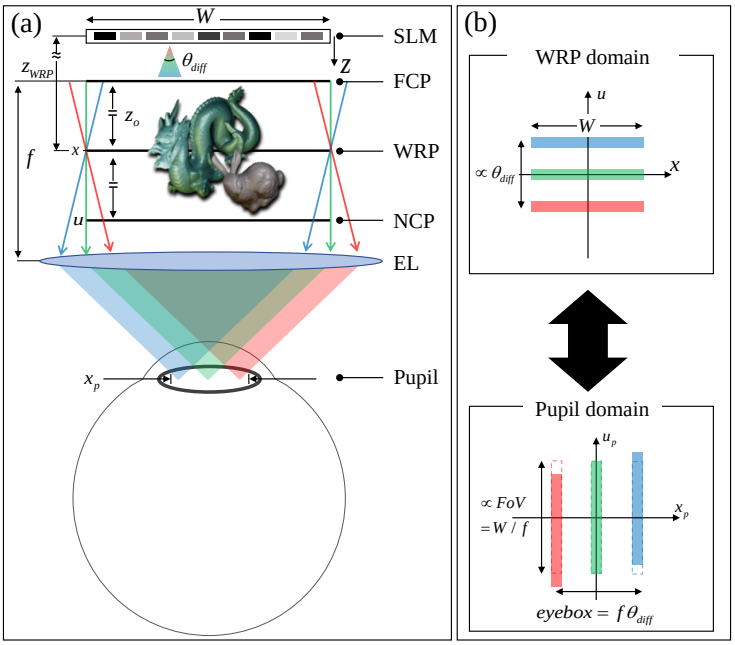


Figure 4.4 Illustration demonstrating (a) schematic of holographic near-eye display and the (b) relationship between wavefront recording plane (WRP) domain and pupil domain.

The setup of a holographic near-eye display is briefly outlined in Fig. 4.4(a). This setup utilizes a spatial light modulator (SLM) with a pitch of p and illuminates it with a coherent light source of wavelength λ, thus generating a wave field within the diffraction angle of $\theta_{diff} = 2\sin^{-1}(\lambda/2p)$. This SLM field propagates and reconstructs a wave field at a specific distance, characterized by a width similar to that of the SLM and an angle within the diffraction angle.

The goal is to reproduce the intensity profile of $I(x, y, z), z \in [z_{FCP}, z_{NCP}]$, where x, y denote the horizontal and vertical positions, respectively, and z signifies the axial distance from the SLM, situated within the axial distance be-

tween the far clipping plane (FCP, $z = z_{FCP}$) and the near clipping plane (NCP, $z = z_{NCP}$). Also, the wavefront recording plane (WRP, $z = z_{WRP}$) corresponds to the reference plane of orthographic light fields, placed midway between the FCP ($z_{FCP} = z_{WRP} - z_o$) and NCP ($z_{NCP} = z_{WRP} + z_o$). The FCP of the rendered volume is positioned at the eyepiece lens's (EL) focal length (f), with the FCP virtually placed at optical infinity and the NCP located at the dioptric distance of $D_{NCP} = 1/(f - 2z_o) - 1/f$. The overall scheme is in accordance with Fig 3.1 in Chapter 3, the IP in Fig. 3.1 is WRP in Fig. 4.4.

As depicted in Fig. 4.4(b), the relationship between the WRP domain and the pupil domain is such that a beam with a limited diverging angle forms an eyebox, which serves as the system's exit pupil. The three beams propagating in different directions with a small angular bandwidth in the WRP plane are remapped in the pupil domain, illustrating the inversion of spatial and angular dimensions as the beams pass through the lens. If the WRP domain is filled with a light field possessing the spatio-angular size of (W, θ_{diff}), the pupil domain will accommodate a spatio-angular size of $(f\theta_{diff}, W/f)$. The spatial dimension corresponds to the near-eye display's eyebox, and the field of view (FoV) is proportional to the angular dimension. Note that the product of the eyebox and FoV is proportional to the display resolution and the beam's wavelength.

However, if the WRP is not placed at the focal length of EL, the projected light field tilts, causing a FoV discrepancy depending on the pupil's position within the eyebox. Nevertheless, centering the WRP plane is beneficial, as the resolution degradation of the LF-based hologram is proportional to the distance between the WRP and an object's depth.

4.2.3 CGH supervision with various 3D target formats

In this subsection, the author describes the image formation model and CGH techniques the author uses in the setup, including 2.5D, 3D, 4D supervisions. For more comprehensive algorithms, the author refers to the works [31, 69]. All software is implemented in PyTorch [70].

Image formation model

In a holographic near-eye display, a coherent light source is incident on an SLM with a source field u_{src}. The amplitude or phase of the source field is delayed by a spatially-varying input u_{in}. The manipulated field further propagates, creating a target intensity volume at the desired volume at a distance z off the SLM. We use the angular spectrum method as the free space wave propagation model f with the single sideband encoding [36, 71]. The resulting complex-valued field u_z is formulated as follows:

$$\begin{aligned} u_z\left(x,y\right) &= f\left(u_{\text{SLM}}\left(x,y\right),z\right), \\ u_{\text{SLM}}\left(x,y\right) &= u_{\text{in}}\left(x,y\right) u_{\text{src}}\left(x,y\right). \end{aligned} \tag{4.1}$$

$$\begin{aligned} f\left(u,z\right) &= \iint \mathcal{F}\left(u\right) \cdot \mathcal{H}\left(f_x,f_y,z\right) e^{i2\pi\left(f_x x+f_y y\right)} df_x df_y, \\ \mathcal{H}\left(f_x,f_y,z\right) &= \begin{cases} e^{i\left(\frac{2\pi}{\lambda} z\sqrt{1-\left(\lambda f_x\right)^2-\left(\lambda f_y\right)^2}\right)} & \text{if } f_y \geqslant 0 \\ 0 & \text{if } f_y < 0, \end{cases}, \end{aligned} \tag{4.2}$$

where f_x, f_y denotes the spatial frequency, λ is the wavelength of the light, and $\mathcal{F}$ is the 2D Fourier transform.

Optimization for binary amplitude SLMs

An SLM modulates the complex-valued field with an input amplitude or phase pattern, and the input is usually quantized into a set of levels $\mathcal{Q}$, (e.g. $\{0, 1\}$). Here, the author uses a 1-bit SLM in amplitude mode, which only supports output of $q_{\text{in}} \in \{0, 1\}^{M \times N}$. This SLM can operate at 3600 Hz so the user perceives the time-averaged intensity. In other words, the CGH algorithms aim to obtain the optimal amplitude pattern q_{in} for desired target intensity distributions. Since optimizing binary values is a combinatorial optimization problem that is NP-hard, the author relaxes the binary value q_{in} as an output of quantization function q that takes float value a_{in} as input which we optimize for a specific loss function according to the target data:

$$u_{\text{in}}(x, y) = q_{\text{in}}(x, y) = q\left(a_{\text{in}}(x, y)\right), \tag{4.3}$$

The quantization process is non-differentiable, which does not allow us to use gradient-descent-based methods. To overcome this, the author uses the Gumbel-softmax trick [72] for approximating the gradient of the quantization function. Specifically, amplitude values are updated using the following equation:

$$a_{\text{in}}^{(k)} \leftarrow {a_{\text{in}}}^{(k-1)} - \alpha \left(\frac{\partial \mathcal{L}}{\partial q} \cdot \frac{\partial \widehat{q}}{\partial a_{\text{in}}}\right)^{T} \mathcal{L}\left(s \cdot \left| f\left(a_{\text{in}}^{(k-1)}\right)\right|, a_{\text{target}}\right), \tag{4.4}$$

where α is the step size, $\mathcal{L}$ is the loss function, q is the quantization function, $\widehat{q}$ is the relaxed quantization function obtained using the Gumbel-softmax layer, and s is a scaling factor.

While Choi et al. [31] previously demonstrated the effectiveness of this surrogate gradient method using the Gumbel-softmax for phase SLMs, this work represents the first application of this technique to binary amplitude SLMs. Figure 4.5 demonstrates that this Gumbel-softmax-based optimization outperforms the previous state-of-the-art binary CGH [32].

Figure 4.5(a) presents a convergence graph for binary amplitude SLMs using unit gradient and Gumbel-Softmax gradient methods and Fig. 4.5(b) 1D energy distributions across the exit pupil by combining and summing up the intensities at the Fourier plane, with a different number of light field views supervised ($3\times3, 3\times5, 3\times7,$ and 3×9, respectively). One of the views from light field reconstruction results is additionally demonstrated as Fig. 4.5 for a qualitative comparison. This approach offers a promising new direction for optimizing binary amplitude SLMs.

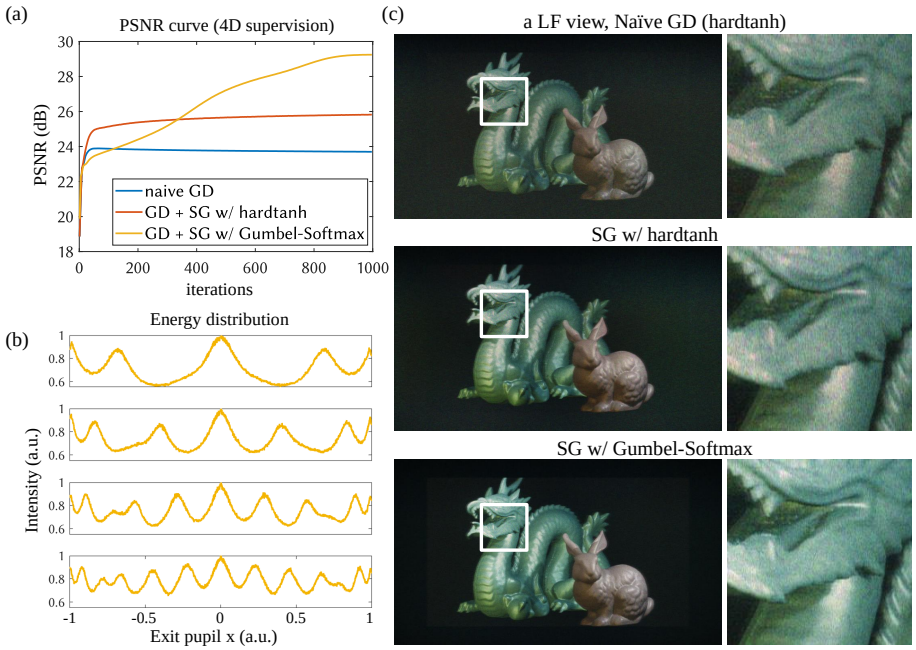


Figure 4.5 A direct comparison of the surrogate gradient approaches for 4D supervision. (a) convergence graph of different binary CGH optimization approaches. (b) 1D energy distribution across the exit pupil of the near-eye displays with different numbers of light field views supervised ($3 \times 3, 3 \times 5, 3 \times 7,$ and 3×9, respectively). (c) One of the views is reconstructed for qualitative comparison among the optimization approaches.

2.5D supervision

By leveraging the image formation model and utilizing a gradient descent-based update rule, the author can optimize the binary amplitude SLM pattern to accommodate various loss functions as described in [31]. First, the author produces the 2.5D supervision results in the work employing the multiplane loss function in Eq. 4.6. To implement this approach, the closest distance matching

technique is employed to create a set of binary masks $M^{(k)}$, corresponding to various distances $z^{(k)}$ from the SLM, using the depth map D obtained from an RGB-D input.

$$M^{(k)}(x,y) = \begin{cases} 1, & \text{if } |z^{(k)} - D(x,y)| < |z^{(l)} - D(x,y)|, \forall l \neq k, \\ 0, & \text{otherwise.} \end{cases} \tag{4.5}$$

Subsequently, the author uses these binary masks for the multiplane loss, which constrains the wavefront to reconstruct the desired RGB amplitude, denoted as a_{target}, at the relevant distances from the SLM, where $\circ$ represents the element-wise product.

$$\mathcal{L}_{\text{2.5D}} = \frac{1}{K}\sum_{k=1}^{K} \mathcal{L}\Bigg(M^{(k)} \circ s\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left| \left(f(q\left(a_{\text{in}}^{(t)}\right)), z^{(k)}\right)\right|^2}, \\ M^{(k)} \circ a_{\text{target}} \Bigg). \tag{4.6}$$

3D supervision

The 2.5D loss function only restricts the positioning of objects and does not necessarily result in a natural defocus blur for the unconstrained part. To address this, one can assume the amount of defocus occurring at each plane based on the pupil size and penalize all focal slices throughout the volume, ultimately pushing the wavefront toward the desired focal stack using the following loss function:

$$\mathcal{L}_{\text{3D}} = \mathcal{L}\left(s\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left| f\left(q\left(a_{\text{in}}^{(t)}\right), z^{\{j\}}\right)\right|^2}, \text{fs}_{\text{target}} \right). \tag{4.7}$$

The target focal stack can be generated using various techniques, such as RGB-D data, off-the-shelf 3D computer graphics software, or light field data. In this work, the author differentiates between 3D supervision techniques based on how

the focal stack is produced. Specifically, the focal stack generated from RGB-D data is labeled as 3D w/ RGB-D supervision. In contrast, when the focal stack target is generated from light field data, which offers more realistic occlusion handling, the author refers to it as 3D w/ LF supervision.

4D supervision

It is also possible to obtain an observable light field from the wavefront utilizing the short-time Fourier transform [73,74]. The short-time Fourier transform computes the Fourier transform over a small patch surrounding each pixel, providing information about how each pixel appears from different directions. By exploiting this analytical forward relationship between the observable light field and the wavefront, one can directly penalize the wavefront to create the observable light field, incorporating the short-time Fourier transform into the loss function as presented by [31]:

$$\mathcal{L}_{\text{4D}} = \mathcal{L}\left(s\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left|\text{STFT}\left(f\left(q\left(a_{\text{in}}^{(t)}\right),z\right)\right)\right|^2}, \text{lf}_{\text{target}}\right). \tag{4.8}$$

4.3 Optimal CGH supervision targets for 3D perceptual realism

Most conventional CGH algorithms depend on a plane-to-plane model. This model employs wave propagation techniques like the angular spectrum method [36] to reconstruct the complex-valued hologram, taking into account the complex field at a specific distance. Adhering to the Huygens-Fresnel principle, this approach models each plane as a superimposition of a set of complex-valued point sources. However, since real scenes are inherently incoherent, this principle cannot be applied directly.

The light field, comprising a set of intensity profiles with different propagation directions, intrinsically carries view-dependent information. Yet, the direct conversion of an accurate light field into a single hologram presents an ill-posed problem since the light field does not contain phase information. As an alternative, the author employs a differentiable hologram-to-light-field transform using the Short-time Fourier transform for 4D CGH supervision [73, 74].

While 4D-supervised CGH achieves view-dependent visuals, it remains unclear whether a holographic near-eye display requires the reconstruction of light fields rather than approximated 3D information. This question arises because the eyebox size of a holographic near-eye display is intrinsically limited, minimizing the visuals arising from the support of view-dependent effects. In this section, the author seeks to validate the CGH supervision data format that best reconstructs 3D holographic content through simulations and experiments performed with the perceptual testbed of a holographic near-eye display.

4.3.1 Simulation

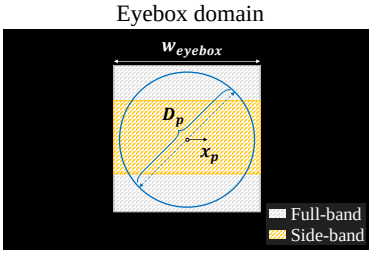


Figure 4.6 Illustration of the eyebox domain of holographic near-eye display.

In practice, the eye rotates to gaze at the objects located across the field of view leading to pupil displacement. Moreover, the size of the pupil varies depending on the intensity of light entering it, and there is significant variation in pupil size among individuals. Fig. 4.6 illustrates the eyebox domian of holographic near-eye display with the width of w_{eyebox} and the circular pupil with a diameter of D_p and displacement of $(x_p, 0)$. The side-band eyebox (yellow) is vertically halved in size relative to the full-band eyebox (blue) for complex modulation with a single amplitude SLM. As shown in the figure, the human eye pupil located in the eyebox domain acts as a low-pass filter and it can partially sample the signal. This effect is dominant since the eyebox size of holographic near-eye displays is limited and the size is additionally narrowed with band-limitation of signal for complex modulation.

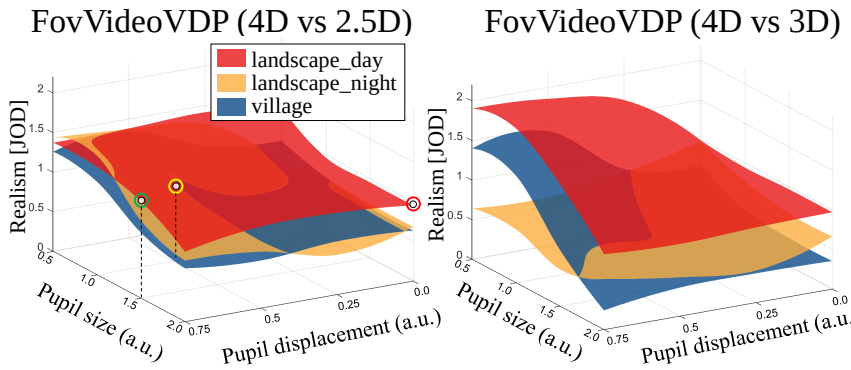


Figure 4.7 Comparison of 4D CGH supervision with 2.5D (left), 3D (right) CGH supervision in various pupil states.

The holographic images were reconstructed to examine the impact of different pupil states (size and displacement). Based on the different pupil states, the level of perceptual realism achieved through 4D CGH supervision is evaluated and compared with cases of 2.5D supervision and 3D supervision, as illustrated in Fig. 4.7. It is evaluated in various normalized pupil displacement ($x_{p,norm} = x_p/w_{eyebox}$) and normalized pupil size ($D_{p,norm} = D_p/w_{eyebox}$) conditions. In detail, FovVideoVDP [3] (v1.2.0) is estimated in a non-foveated mode with the condition of 86.2 [pix/deg], Lpeak=100, Lblack=0.1 [cd/m^2] and the results are scaled in a unit of JOD. Sixteen distinct pupil states are sampled, and the visual difference predictor (VDP) differences in units of JOD are interpolated. The simulation assumes the field of view identical to the experimental setup, with luminance levels assumed to be similar to conventional head-mounted displays. The metric is obtained by comparing the simulated holographic images and the ground truth images processed with the light field having dense views, both of which are generated under the same pupil conditions.

It can be observed that 4D-supervised CGH visualizes images with superior

perceptual quality across the eyebox compared to the images reconstructed with 2.5D, 3D-supervised CGHs as it offers positive JOD values and a VDP of more than 1 JOD can be noticed in some portions of the sampled pupil states. The difference can be observed with the holographic images of landscape_day scene shown in Fig. 4.8. If the pupil gets decentered in a certain amount, the occluded part demonstrates invalid information in the 2.5D and 3D-supervised cases. And this phenomenon gets emphasized when the pupil is shifted in a large amount when the image starts to get vignetted. On the other hand, the 4D-supervised holographic images show intact sampled view images.

The superior perceptual quality of 4D-supervised CGHs compared to other cases can be thought of in two ways. First, the ground truth images are generated with a dense light field of 25x25 views. The additional view-dependent information in 4D-supervised CGHs has led to improvements in the simulated FovVideoVdp. Furthermore, the 2.5D and 3D-supervised cases are realized with plane-to-plane models, which results in inferior quality in the reconstructed results when viewed from perspectives other than the central view.

Additional reconstruction results are provided with evaluation metrics in Fig. 4.9. Fig. 4.9(a) provides near-depth reconstructed image of landscape_day scene, and the white box sections of the images reconstructed with various pupil states (pupil displacement, and pupil radius) are shown as Fig. 4.9(b). The blank section with a dashed boundary in the figure indicates that visualization is not capable as the state is fully vignetted. PSNR, SSIM, and FovvideoVDP quality metric in a unit of JOD are consecutively provided on the bottom of every inset. In Fig. 4.9(c), the identical parts of images reconstructed with three different pupil states (colored in Fig. 4.9(b)) are provided with four sampled focal depths.

It is noteworthy that there is a difference in estimated VDP depending on the scene since the depth distribution of individual scenes is different. For instance,

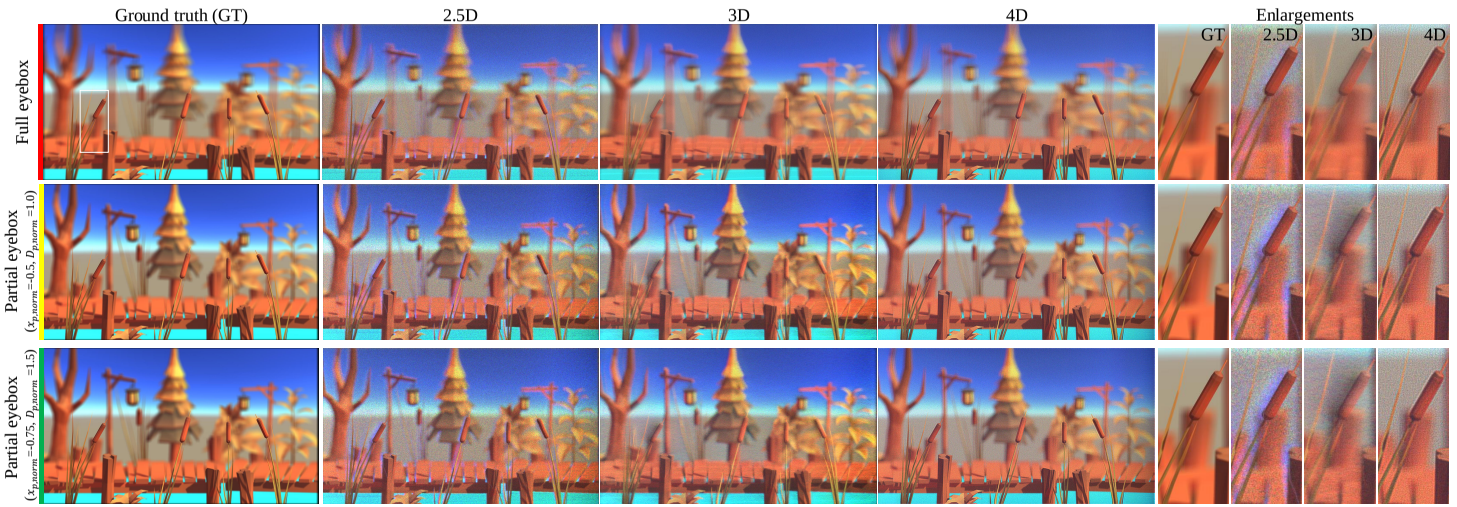


Figure 4.8 Near-depth focused holographic images with different CGH supervision targets reconstructed under three different pupil settings.

landscape_day scene has a majority of objects located near or far relative to the reference plane. However, most of the objects lie near the middle depth of the village scene. Moreover, landscape_night scene seems less susceptible to occlusion and parallax problems with the dark background.

In summary, the VDPs estimated with various pupil states show the competence of 4D CGH supervision over the 2.5D and 3D CGH supervision. However, as the perceptual metric is established for the evaluation of 2D contents, it does not guarantee the perceptual quality of 3D contents. Note that the provision of depth cues (accommodation, parallax, occlusion etc.) could affect the 3D visual experience even with a monocular eye. Due to this major difference, the evaluation of 3D holographic contents should be performed with user studies.

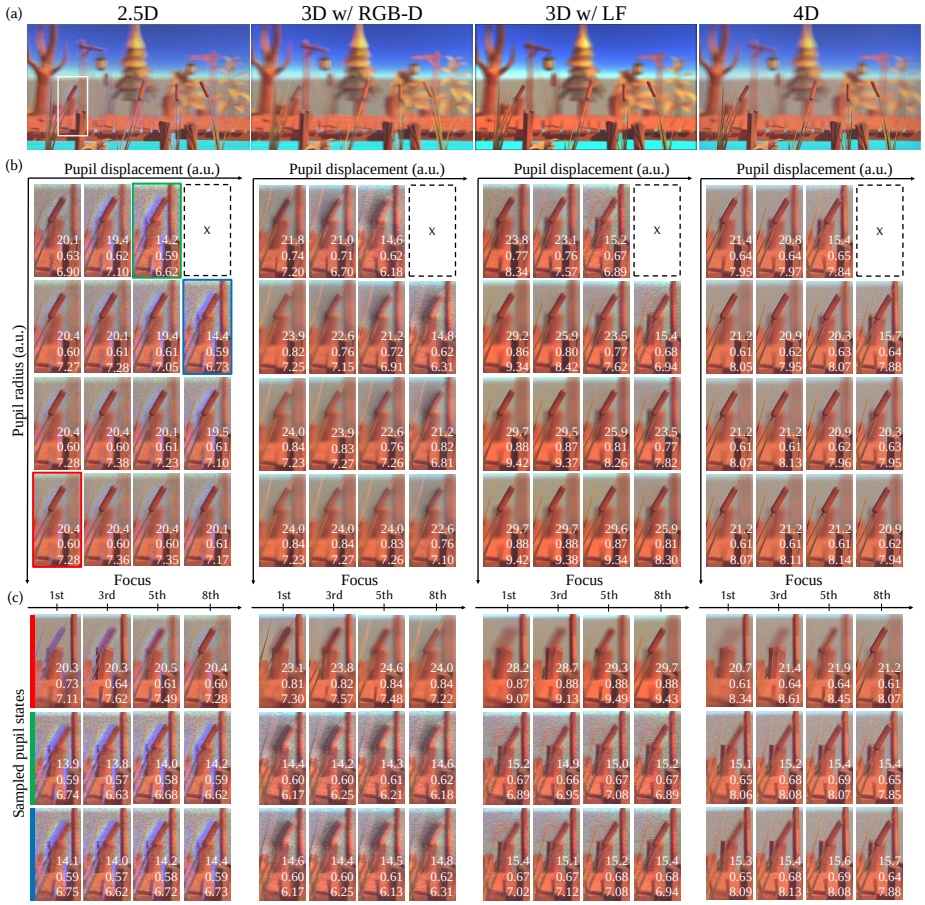


Figure 4.9 Extensive assessments on the reconstructed holographic images with various 3D CGH supervision targets.

4.3.2 Implementation

Software

The CGH acquisition with various target formats is implemented with Pytorch based on the previous works of the differentiable CGH optimization frameworks. Software implementation is described in Sec. 4.2.3.

Light field rendering

The author utilized a total of five different scenes in the work, and these scenes were rendered using Unity-C#. Fig. 4.10 presents the rendered light field maps, RGB-D images, and the corresponding epipolar plane images (EPIs) of the scenes (landscape_day, landscape_night, village, village_mirror, dragon_bunny) used in this work are provided. For the light field maps, 25×25 orthographic views with a resolution of 1600×900 identical to that of the SLM, are rendered for each color channel as the diffraction angle differs by the color primary. However, in the figure, only a subset of only 9×9 sampled views has been provided due to space limitations.

The EPI of the horizontal section depicted with a green dashed line is provided. The EPIs are drawn based on orthographic light fields and the upright slope implies the object is placed at the WRP. (Low Poly Series: Landscape, Fantastic-Village Pack: purchased unity asset, and Dragon, Bunny: credit to Stanford Computer Graphics Laboratory)

The EPIs provide insights into the angular distribution of the scenes. As depicted in the EPIs shown in Fig. 4.10, the individual slices exhibit distinct spatial information along the angular dimension. Notably, certain objects are only visible within specific angular ranges, while they disappear in other angular regions. This disparity in information across angles signifies the presence of “valid par-

allax", and it emphasizes that parallax containing meaningful information can be obtained when working with data formats that have four or more dimensions.

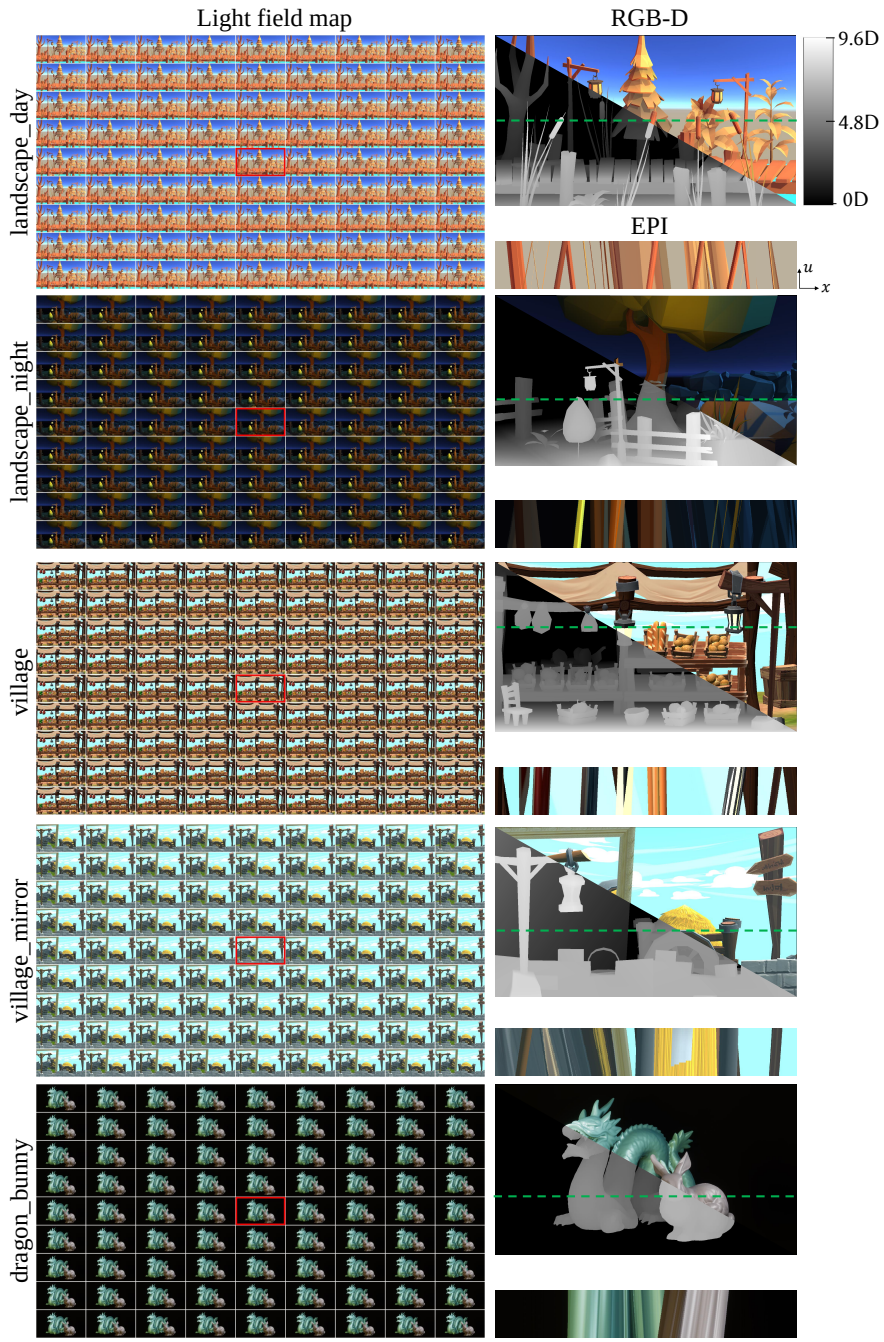


Figure 4.10 Light field map, RGB-D image, and epipolar plane image (EPI) of the scenes.

Hardware

Figure 4.11 demonstrates an overview of the holographic near-eye display prototype. A fiber-coupled laser diode of Wikioptics emanates a full-color beam with a central wavelength of 638 nm, 520 nm, and 450 nm. The beam is collimated with a lens (AC-508-200-A, Thorlabs) and the beam is linearly polarized with a series of linear polarizer (LPVISE200-A, Thorlabs) and an achromatic half-wave plate (AHWP10M-600, Thorlabs). The author additionally placed a half wave plate to maintain the color balance as there are difference in polarization states by color. The beam is redirected with a 1-inch beam splitter and modulated with a reflective-type spatial light modulator.

The author uses a binary ferroelectric liquid crystal on silicon spatial light modulator (FLCoS SLM) to modulate the incident coherent beam. This SLM (QXGA-R10, a product of Forth Dimension Display) operates 1920 $\times$ 1200 pixels with a pitch of 8.2 μm at a speed of 3600 Hz to serve 24 full-color binary frames within 1/50 seconds. Placing an analyzer in the beam path allows the operation of SLM in amplitude mode. The field at the SLM plane is relayed with a 4-f system built with two identical camera lenses (AF Nikkor 50mm f/1.4D, Nikkon) facing opposite each other. A filter is placed in the Fourier domain to filter out the high-order signals arising from diffraction and the conjugate noise from complex representation with an amplitude SLM. The filter is fabricated with a rectangle aperture in an aspect ratio of 2:1 with a size determined by the SLM's diffraction angle in blue and the focal length of the 2-f lens. The relayed field is virtually floated by a 2-inch eyepiece lens (AL5040M, Thorlabs) having a focal length of 40 mm to guarantee a wide field of view. Thus, the eyebox size of the near-eye display is 2.2 mm $\times$ 1.1 mm.

The author made an additional beam path by placing a beam splitter after the

4-f system to capture the experimental results and monitor the user experiment. For this arm, a lens (AC508-100-A-ML, Thorlabs) having a focal length of 100 mm is used as an eyepiece lens, and the scenes are captured with a c-mount lens with a 25 mm focal length and charge-coupled device (CCD) camera (BFS-U3-51S5C-C, FLIR) having a resolution of 2448×2048 and a pitch of $3.45 \mu m$. The CCD camera is placed on the two single-axis motorized stages (M-112.1DG1, PI) to capture the image in distinct viewpoints with high accuracy. Note that the eyebox size of this arm is 2.5 times larger than the actual user experiment settings. Thus, the CCD is translated with the converted geometry. Additional spatial filters are placed at the relayed WRPs to eliminate the noise present in the peripheral region. Additional components shown in Fig. 4.11 but not described in this subsection will be explained in the subsequent subsection with the description of user study implementation.

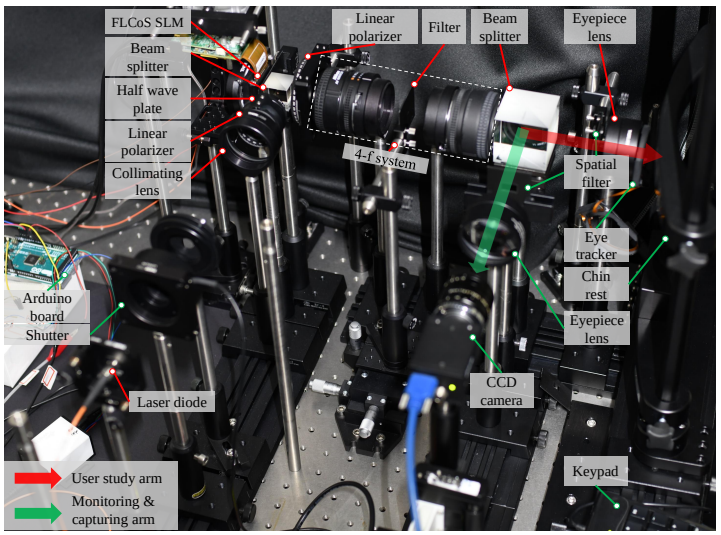


Figure 4.11 The photograph depicts the testbed of a holographic near-eye display prototype used for user validation.

User study setup

The overall setup for the user study is implemented with an eye tracker, and a keypad for data acquisition. In addition, a chin-and-head rest, a shutter, an Arduino board, and a luminance meter are equipped to improve the study's accuracy and guarantee the subjects' eye safety.

The setup utilized the Add-on eye tracker, developed by Pupil Labs, to measure the displacement of the subject's eye while they viewed the stimulus. Since the pupil displacement of a single eye was recorded, the author couldn't utilize the built-in calibration functions designed for tracking both eyes simultaneously. Instead, the measured data, which represented the center of the detected pupil, is calibrated using a scale factor obtained through a pre-calibration procedure. This pre-calibration involved an eye figure with a black pupil that was moved laterally at the eye relief of the near-eye display. The collected data was obtained using the Pupil Labs Network API and saved in comma-separated value (csv) file format. Each trial involved a 2-second recording session, capturing the data at a speed of approximately 120 frames per second. We only included data points with a confidence value higher than 0.85 for further analysis. The response for each pair of options from the participants was received using a keypad and saved in csv file format, with values indicating the options that were compared.

To ensure the absolute position of the subject's head, a chin-and-head rest was employed to restrict head movement. The chin-and-head rest was positioned on a stage that allowed lateral movement. Subjects were instructed to adjust the initial position of their pupil by translating the stage. To address differences in intensity levels between holographic images created with various CGH supervision targets, as well as to ensure safety, the author incorporated an

Arduino board (Arduino Mega 2560) for two purposes. First, it helped balance the intensity levels by adjusting the pulse width of the light source. The reconstructed images displayed different intensity levels due to variations in the scale factor for CGH supervision. By modulating the width of the rectangular pulse generated by the Arduino board, the intensity levels across the images were standardized. Next, the Arduino board controlled a shutter placed in front of the light source, preventing uncontrolled light emission during the initialization process. The luminance of the scenes was measured to ensure eye safety. However, direct measurement of the luminance using the luminance meter (LS-150, Konica Minolta) was not feasible. Instead, the luminance of the eyebox plane was measured, which covered a 2-degree angle at the measurement distance, and obtained a measurement of approximately 1 cd/m^2. Considering that the luminance is inversely proportional to the subtended angle, it can be estimated that the luminance of each scene realized in the user experiments is below 0.1 cd/m^2. This value is significantly lower than the permissible level of laser exposure stated in the cited reference [56].

4.3.3 User validation

Stimuli and Conditions

Apparatus

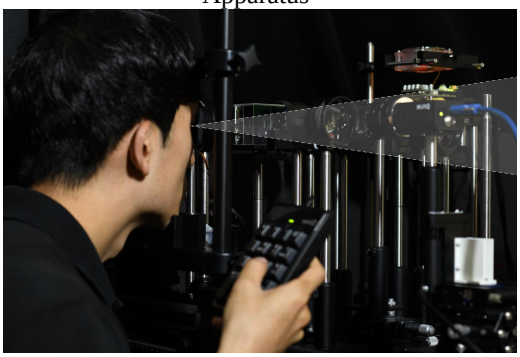


Figure 4.12 The apparatus for the user study.

For the user validation depicted as Fig. 4.12, three volumetric scenes (landscape_day, landscape_night, and village), out of five presented in Fig. 4.10, are used as stimuli. The resolution of the rendered images was set to 1600 $\times$ 900, and the orthographic images were rendered with equal angle spacing within the diffraction angle. The reference plane for the orthographic light field was positioned 120 mm from the SLM, which lies at the midpoint between the near-clipping plane (+5.5 mm) and the far-clipping plane (-5.5 mm). The depth range extends from 0 D to 9.57 D, spanning from optical infinity to 11 cm from the eye. This range sufficiently covers the average accommodation range of young adults [75] and the scheme is explained in Fig. 4.4. The luminance of each scene is estimated below 0.1 cd/m^2, and the room was kept dark during the experiment.

In the case of 2.5D and 3D-supervised CGHs, nine planes equally spaced in a unit of diopter are sampled. For 3D-supervised CGHs, we prepared two scenarios; *3D w/ RGB-D* and *3D w/ LF*. For *3D w/ RGB-D*, the focal stacks

are generated with a single RGB-D map blending occlusion boundaries. For *3D w/ LF*, LF with 25×25 orthographic views was utilized to generate nine focal stacks to handle occlusion naturally. Lastly, LF with 9×9 orthographic views is utilized for 4D CGH supervision. The captured scenes of each stimulus are provided in Fig. 4.13.

Fig. 4.13 demonstrates the holographic scenes supervised with 2D, 2.5D, 3D w/ RGB-D, 3D w/ LF, 4D targets that are captured with different pupil positions. The eyebox, where the valid signal is present, is demonstrated with an orange dashed plane. The scenes are photographed with different focal states (landscape_day: 7th, landscape_night: 7th, village: 7th, village_mirror: 3rd) out of 9 distinct focal states equally sampled in diopter. The colors of each row for the enlargements indicate the pupil positions (red: center, yellow: decentered, green: vignetted).

Variations in artifacts stemming from occlusion, discontinuous depths, erroneous depth of reflected objects, and image quality degradation can be observed across different types of CGH supervision. Additionally, an examination of the brightness uniformity under each condition reveals that energy distribution across the eyebox is limited. However, this factor does not significantly impact the evaluation of 3D perception.

For the capturing system, the near-eye display is equipped with an eyepiece lens that has a focal length of 100 mm. The resulting images are captured with a c-mount lens that features a 25 mm focal length and an F-number of 4. Considering the increased size of the eyebox, the image acquisition settings can be easily converted to simulate an eye with a pupil diameter of 2.5 mm. The centers of each condition are horizontally shifted by 0 mm, 1.875 mm, and 3.75 mm, respectively. These displacements can be converted to 0 mm, 0.75 mm, and 1.5 mm. Images are captured in horizontal positions with less vignetting effect com-

pared to actual user validation conditions (0 mm, 1.25 mm, and 2.5 mm). This is done to ensure high-fidelity image capture.

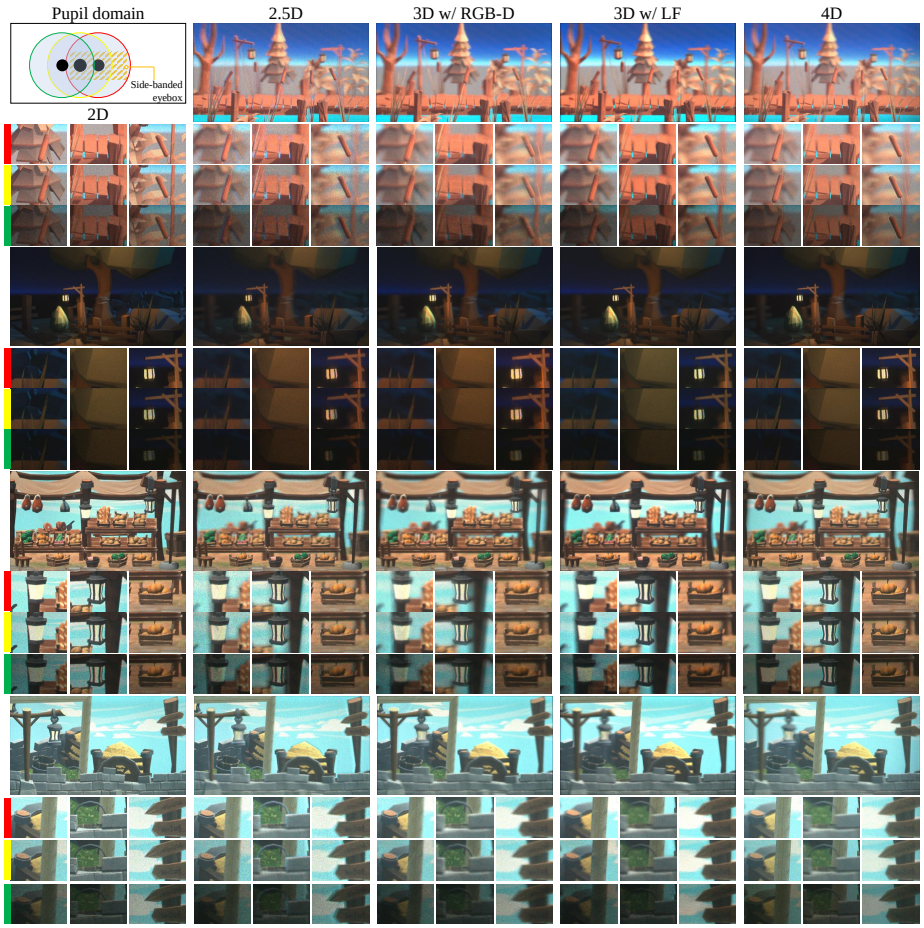


Figure 4.13 Experimental results with different pupil positions.

The user validation is performed with four different viewing conditions; *Center* refers to the case when the subjects view 3D contents while placing the pupil at the 'sweet spot' of the eyebox. *Decentered* refers to the condition when the eye is horizontally decentered about 1.25 mm from the center but the subject can see the whole image without vignetting. *Vignetted* refers to the case

when the center of the pupil is horizontally decentered about 2.5 mm and almost one-third of the image is vignetted. *w/ head movement* refers to the viewing condition when the subjects perform the task without head movement restriction.

Before the experiment, six complete pairs with four different CGH supervision cases (*2.5D*, *3D w/ RGB-D*, *3D w/ LF*, *4D*) were prepared, the order was randomly shuffled to eliminate the potential decision bias and each pair was repeatedly provided three times. The complete pairwise comparison was held with three different scenes (landscape_day, landscape_night, and village) in four different viewing conditions (*Center*, *Decentered*, *Vignetted*, *w/ head movement*). The whole number of trials was 216 (6 pairs $\times$ 3 repetitions $\times$ 3 scenes $\times$ 4 viewing conditions).

Subjects

The author recruited 28 naïve participants under the age of 40 (ranging from 23 to 36 with an average of 27.6, 12 female) since the age could potentially decrease the accommodation range. All participants were normal or corrected-to-normal vision and had normal color vision. The participants were rewarded for their participation. The studies adhered to the Declaration of Helsinki. All the subjects gave voluntary written and informed consent. The experiment was conducted after the approval from the Institutional Review Board of the host institution.

Procedure

Before each session, a precise head alignment was performed. The subjects were requested to restrict their head movement in viewing conditions other than *w/ head movement*. The head position of the subjects was controlled by adjusting the components of the chin-and-head rest. In this adjustment procedure, the sub-

jects were asked to keep their gaze over the center object of the sample scene to employ the eye-tracked data monitored in real-time. For the viewing condition of *w/ head movement*, the chin-and-head rest was removed from the setup, and the subjects were asked to freely move the head in a range where the scenes are observable. In addition, the subjects were asked to get a sharp focus on the nearest object of each scene to check the potential presbyopic eye.

In every viewing condition, two-interval forced choice (2-IFC) [17] was performed asking the subjects to choose the 'more realistic 3D' option after showing a pair of stimuli in a sequence. The subjects were requested to gaze at different objects simultaneously trying to get a sharp focus on the gazed object to evaluate 3D quality and to rule out the subjects who maintain the focus at a single plane. A pair of stimuli was displayed for 8 seconds and a second of a gray noisy image was provided between. The responses were made with a keypad, and the next pair is provided after the valid input. After each session, the subjects were encouraged to take a break for at least a minute, and the whole experiment took around an hour.

Results

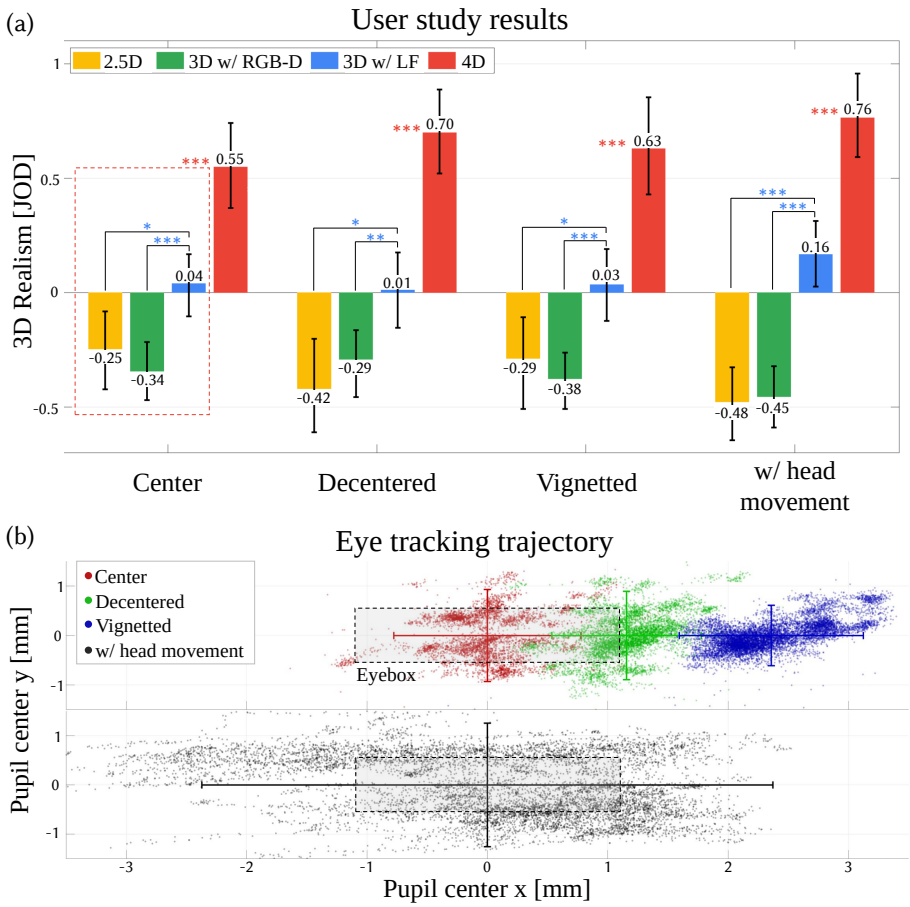


Figure 4.14 3D perceptual realism user evaluation results.

The CGHs supervised with different targets were evaluated in four distinct viewing conditions and compared in terms of perceived 3D realism, as depicted in Fig. 4.14(a). The vote counts from a total of 24 subjects are scaled in a unit of JOD depending on the viewing conditions and indicated for each viewing condition. The mean JOD is set as zero for each viewing condition. The errorbars indicate 95% confidence intervals estimated by bootstrapping. The asterisks (blue:

3D w/ LF vs the paired cases, red: 4D vs the other cases) indicate the statistical significance of the difference (*: p <0.05, **: p <0.01, ***: p <0.001). The responses of four subjects were excluded after distance-based outlier analysis introduced by the work of Perez and Mantiuk [16].

Upon conducting the two-tailed z-test with the scaled JOD scores for each CGH supervision target in each viewing condition, the results indicate that 4D-supervised CGHs exhibit significant improvements in perceived 3D quality across all viewing conditions compared to other forms of CGH supervision. Additionally, *3D w/ LF* is significantly preferred over *2.5D* and *3D w/ RGB-D*, and this preference is even more pronounced in the viewing condition involving head movement. Notably, no significant differences were observed between *2.5D* and *3D w/ RGB-D* in any of the viewing conditions.

In detail, in the viewing condition of *Center*, statistically significant differences in JOD scores were found in most of the paired conditions (p <0.001: 4D vs the other cases, 3D w/ LF vs 3D w/ RGB-D, 3D w/ LF vs 2.5D, p <0.05: 3D w/ LF vs 2.5D) except the 2.5D vs 3D w/ RGB-D (p= 0.37). In the viewing condition of *Decentered*, the significant results were found in the paired conditions (p <0.001: 4D vs the other cases, p <0.01: 3D w/ LF vs 3D w/ RGB-D, p <0.05: 3D w/ LF vs 2.5D) except the 2.5D vs 3D w/ RGB-D (p=0.36). In case of *Vignetted*, significant differences were present in the paired conditions (p <0.001: 4D vs the other cases, 3D w/ LF vs 3D w/ RGB-D, p <0.05: 3D w/ LF vs 2.5D) except the paired condition of 2.5D vs 3D w/ RGB-D (p=0.38). Lastly, in the case of *w/ head movement*, the JOD scores of paired conditions were significantly different (p <0.001: 4D vs the other cases, 3D w/ LF vs 3D w/ RGB-D, 3D w/ LF vs 2.5D) except the paired condition of 2.5D vs 3D w/ RGB-D (p=0.85). The results demonstrate significant differences in 3D perceptual realism in every viewing condition when the parallax cues were taken into

account in CGH supervision. The 4D-supervised CGH outperformed all other cases by considerable margins, and even the 3D w/ LF CGH outperformed the other cases with strong evidence of significance.

Throughout the experiment, the positions of the pupil were monitored and recorded. Figure 4.14(b) presents the measured data for one representative subject in a single session, with different colors indicating different viewing conditions (red: *Center*, green: *Decentered*, blue: *Vignetted*, and black: *w/ head movement*). The errorbars represent the 95% confidence interval of the pupil displacement. The measured data unequivocally demonstrate that the experiments were carried out under varying positions. Furthermore, it is noteworthy that the eye exhibited substantial movement even when head movement was restricted. This visual behavior can be also noticed with the extensive measurements of eye-tracked data as shown in Fig. 4.15. However, it is important to acknowledge that some degree of noise may have been introduced due to detection errors of the eye tracker.

It is worth noting that, even in the experiment conducted at the central position within the eyebox, the performance of the *4D* approach surpasses that of the *3D w/ LF* approach in terms of the perceived quality of 3D visuals. Note that the focal stacks of *3D w/ LF* case are generated with 25×25 views while *4D* employs 9×9 orthographic views for CGH supervision. Interestingly, the FovVideoVDP metric yields contradictory results when simulating images with the pupil placed at the center.

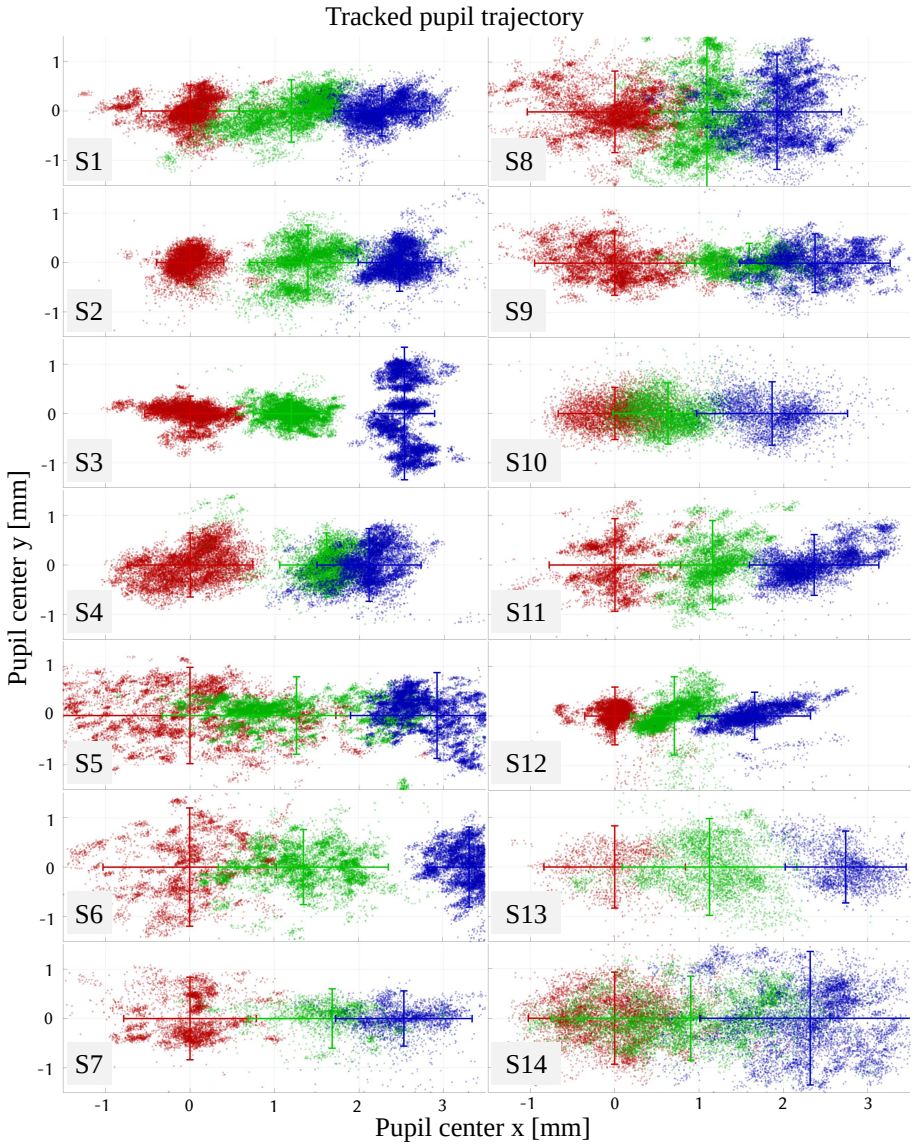


Figure 4.15 Tracked eye trajectory of fourteen subjects in a single session.

Moreover, when the viewer's eye is decentered or experiences maximum ocular movements, the disparities between CGH supervision approaches based on center views and those based on multiple views tend to escalate. This implies

that the influence of parallax on 3D realism will be noticeable in holographic near-eye displays with an expanded eyebox [76].

4.4 Number of view requirements in light field-based CGH

The previous section confirmed that the utilization of 4D CGH supervision, which provides complete holographic parallax support, significantly enhances the perceived 3D realism compared to other CGH supervision approaches. However, it is wise to know the perceptual threshold for the number of views to alleviate the computation load.

4.4.1 User validation

Stimuli and Conditions

CGHs are acquired with a distinct number of views for supervision; 3×3, 5×5, 7×7, and 9×9. Due to the optical configuration using a side-band filter, the effective number of views provided by the near-eye display prototype was 3 (H)$\times$1 (V), 5×2, 7×3, and 9×4 while maintaining the gap between adjacent views. Three scenes employed in the first experiment were utilized as stimuli.

This work investigates the number of views required for 4D CGH supervision through both camera-incorporated experiments and a user study. Each condition is easily understood with the simulated images of various scenes provided in Fig. 4.16-4.17.

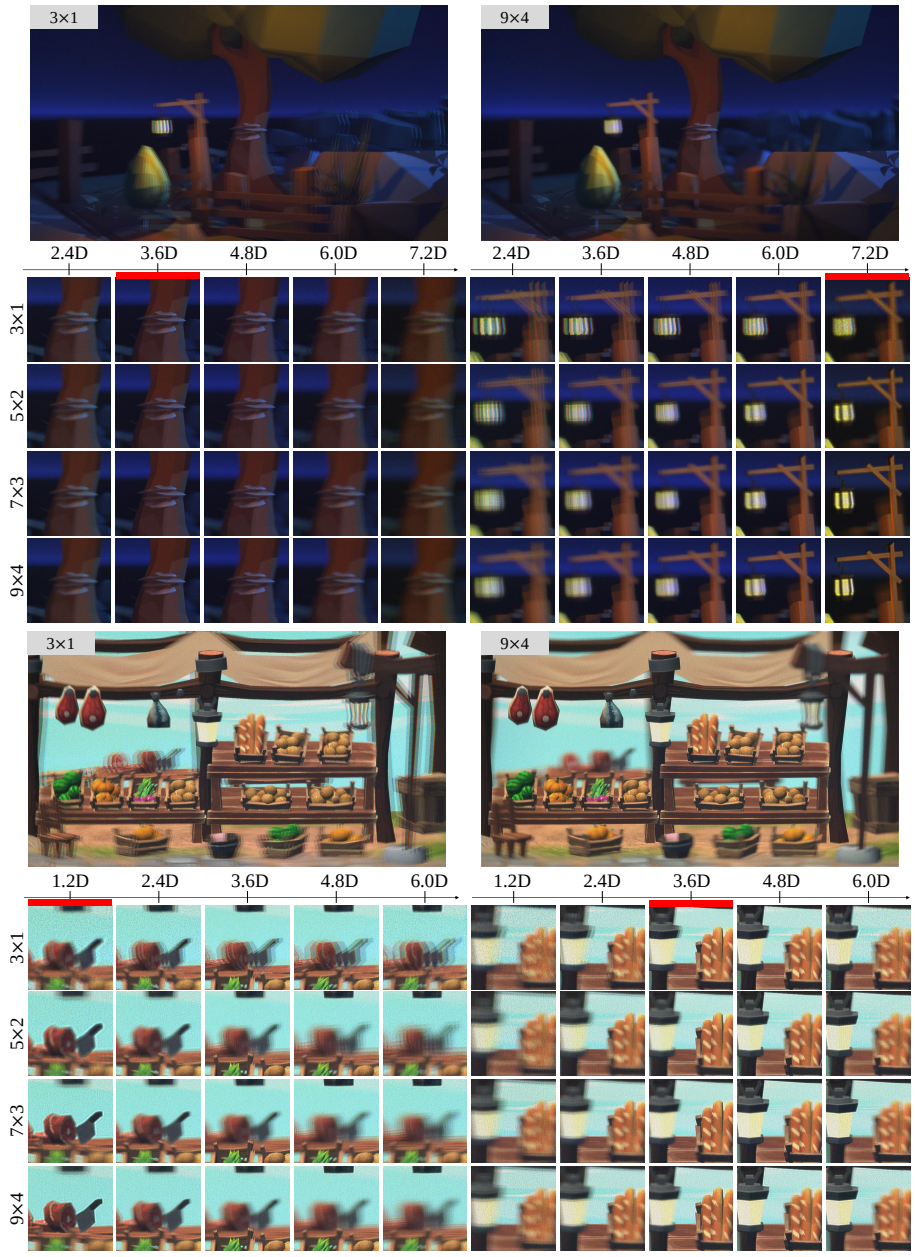


Figure 4.16 Reconstruction results of landscape_night, village scene.

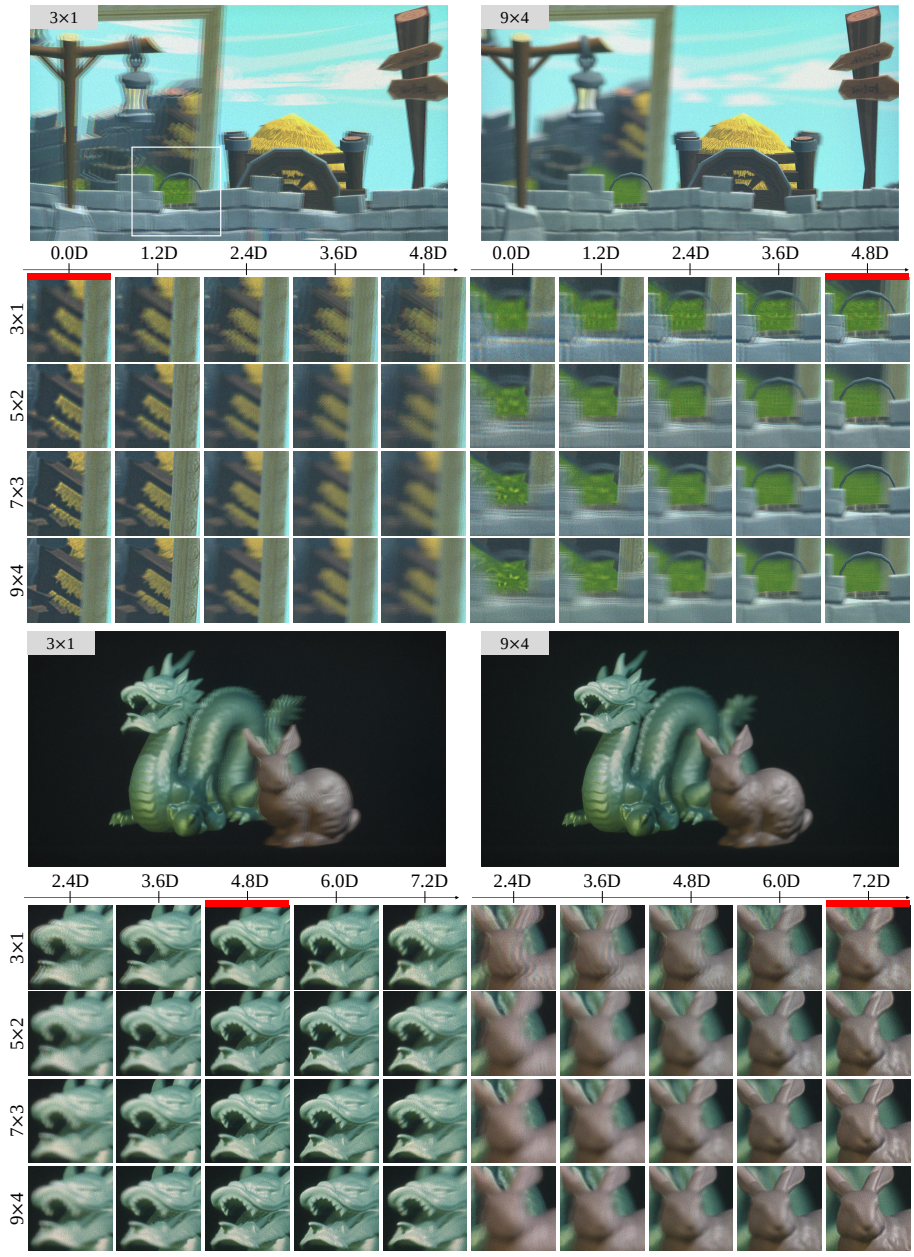


Figure 4.17 Reconstruction results of village_mirror, dragon_bunny scene.

By examining these images, it becomes apparent that the overall quality of

the reconstructed 3D scene improves as more views are incorporated into the 4D CGH supervision. However, the objects placed near the WRP are reconstructed with a high resolution (see the enlarged images of tree, baguette in Fig. 4.16, and handle of basket, head of dragon in Fig. 4.17) and do not demonstrate the noticeable difference when the scenes are rendered with denser views. On the other hand, the image quality suffers for the objects lying at planes that deviated from the WRP (see the enlarged images of street light, fruit in Fig. 4.16 and wagon reflected by a mirror, head of bunny in Fig. 4.17) especially when the views are sparsely sampled. Therefore, the reconstructed results visually demonstrate that the placement of objects relative to the WRP and the number of views directly affect the resolution of the reconstructed scene.

Fig. 4.18 presents the experimental results captured with the benchtop prototype of a holographic near-eye display, illustrating the impact of the number of views used in 4D CGH supervision on 3D visualization and serving as an assessment of the stimuli used in the test. The captured results of four different scenes with number of views of 3×1, 5×2, 7×3, and 9×4 are provided. Insets show the images captured with three different focuses randomly chosen for each scene and the images in gray boxes are enlarged. Among the enlarged images, the red boxes indicate that the object is focused.

The whole test has six pairs and three scenes, and the complete pairwise comparison was repeated five times for each pair as we aim to estimate the threshold. Thus, the number of trials was 90 and it took around thirty minutes. The test was initialized after checking the subject's pupil located at the center of the eyebox, but the eye-tracking data was not recorded for this test.

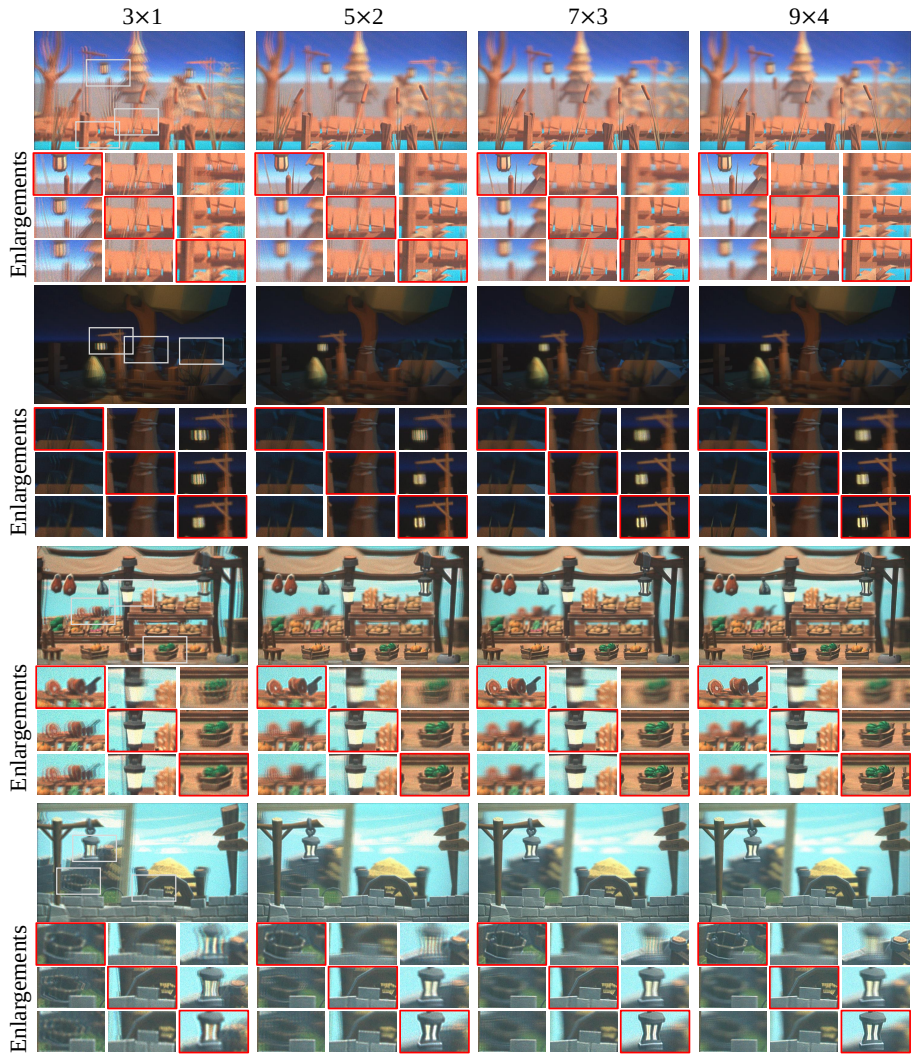


Figure 4.18 Experimental results with different number of views (3×1, 5×2, 7×3, and 9×4) used in 4D CGH supervision.

Subjects

All subjects that performed the user experiment in Sec. 4.3.3 participated in the test.

Procedure

The overall procedure is identical to the experiment described in Sec. 4.3.3.

Results

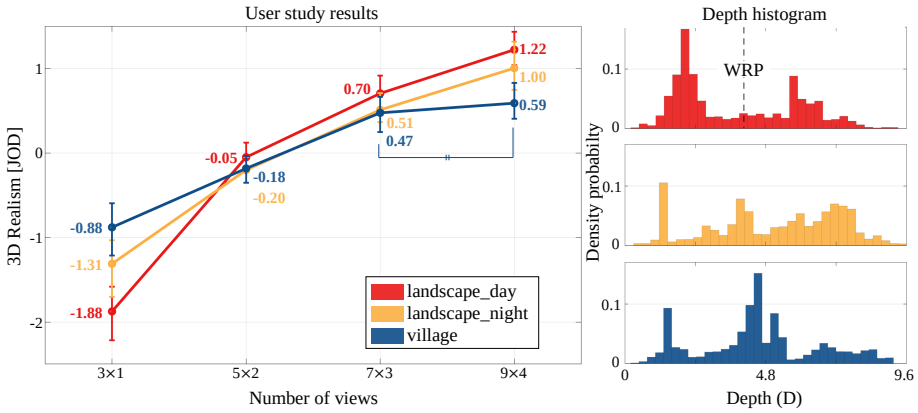


Figure 4.19 Scene-dependent user experiment results with a different number of views (3×1, 5×2, 7×3, and 9×4) used in 4D CGH supervision.

An evaluation was conducted comparing the CGHs that were supervised using 4D targets, with variations in the number of views. The author subjected to the response of 24 subjects excluding the responses of four participants after outlier analysis with JOD scores estimated with the vote counts accumulated over the scenes. The results, presented in Fig. 4.19(a), were dependent on the number of views. The difference in 3D realism of the individual scene between the neighboring view number conditions was evaluated by a two-tailed z-test with the scaled JOD scores. Most of the paired options elicited very strong statistical significance on the difference (p <0.001). The paired condition of 7×3 vs 9×4 in the landscape_night scene showed strong evidence (p <0.01), while the paired option of 7×3 vs 9×4 in the village scene showed no evidence of the

significance (p=0.45). For the village scene, the objects are relatively concentrated near the depth of the wavefront recording plane (WRP) compared to other scenes as shown in Fig. 4.19(b). These scene-specific findings regarding 3D perception can be better understood by the light field sampling theorem, which will be discussed in the next analysis section.

4.4.2 Analysis

The analysis on the number of views required for 4D CGH supervision is conducted based on light field sampling theorem [69, 73]. In the theorem, the depth range covered by the light field is proportional to the angular resolution.

Let's assume the situation when the WRP is placed at a certain distance, and the FCP is located at the focal length of the eyepiece lens as Fig. 4.4. The metric distance between FCP and NCP is $2z_o$ which makes the depth coverage from 0 D to D_{max} and the single-sided distance from WRP to the clipping plane is as follows:

$$z_o = \frac{f}{2} - \frac{1}{2\left(D_{\max} + \frac{1}{f}\right)} = \frac{f^2 D_{max}}{2\left(fD_{\max} + 1\right)}. \tag{4.9}$$

Here, the angular resolution (N_u) of the light field required to reconstruct the image with spatial bandwidth (B_x) at the distance of z_o can be obtained as:

$$N_u \geqslant \lambda z_o B_x^2. \tag{4.10}$$

The expressible depth range supported by the near-eye display depends on the scene's spatial bandwidth and the angular resolution as

$$D_{max} = \frac{2N_u}{f(f\lambda B_x^2 - 2N_u)}. \tag{4.11}$$

If the spatial bandwidth is bounded within a certain range, the low-pass filtered spatial bandwidth ($B_{x,v}$) can be simply acquired with the ratio of spatial frequency (v) and cut-off frequency (v_{cutoff}) as $B_{x,v} = B_x \frac{v}{v_{cutoff}}$. The cut-off

spatial frequency can be acquired with the optical configuration of the near-eye display.

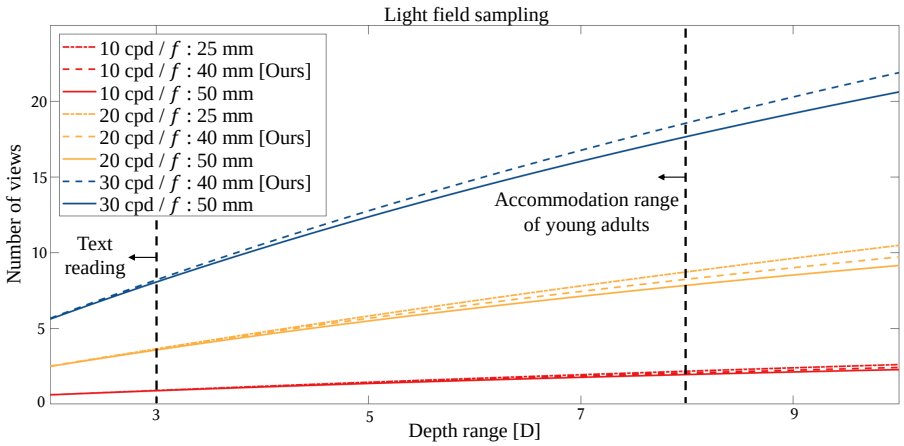


Figure 4.20 Required number of views in horizontal for 4D CGH supervision based on light field sampling theorem.

The graph labeled as Fig. 4.20 demonstrates the number of horizontal perspectives necessary for 4D-supervised CGH based on LF sampling analysis, contingent upon the depth range. We explored three different spatial bandwidth regions (10 cpd, 20 cpd, and 30 cpd), with the results illustrated using three distinct eyepiece lenses (f=25 mm, f=40 mm [prototype], and f=50 mm). The simulation was conducted with a 532 nm light source and an SLM having specifications identical to the prototype. Despite the eye relief of the near-eye display, assumed to be identical to the focal length of the eyepiece lens, being down to the conventional range, the estimated number of required views is similar to the parameter obtained with the system's configuration.

Additionally, if the total depth range is reduced to 3 D, a distance typically used for reading text, approximately 8 horizontal views are required, supporting a resolution of 30 cpd. Conversely, if the 3D content is designed to cover a

full depth range of 8 D at high resolution, over 15 views become necessary. It's crucial to mention that this analysis doesn't consider factors like diffraction from the eye pupil or individual eye aberration, which can alter the point spread functions and consequently affect the results.

The resolution of CGHs shows depth-dependent characteristics due to the reconstruction of individual view images in 4D-supervised CGHs. However, with CGHs created using a plane-to-plane model, the highest resolution remains consistent irrespective of depth changes. Interestingly, the user study results present differing findings, even though the experiments involve scenes with extensive depth distribution. These results indirectly suggest that the evaluation of 3D perceptual realism isn't solely determined by the resolution of objects at various distances from the viewer.

4.5 Discussion

User study stimuli

The validation could have been performed with low-level psychophysics methodologies [77] to identify the concrete discrimination thresholds for the number of views depending on the depth distribution of scenes. However, these methods require densely sampled evaluation sets, necessitating the rendering of various light field sets and CGHs for individual cases. While the work did not extensively address the computational load of rendering 4D CGH supervision, the joint optimization of binary CGHs for temporal multiplexing and 9×9 light field views in this study required clever strategies to overcome memory limitations. Nevertheless, it took 12 hours to acquire full frames for three color channels, even with the high-end GPU (Nvidia DGX A100) having an 80 GB memory. Despite these challenges, the stimuli, comprising complex scenes with contin-

uous depth distribution, effectively elicited natural eye movements mimicking real-world visual behavior during navigation.

Perceptual quality metric of 3D contents

A thorough understanding of the human visual system and careful design of display mechanisms have led to the development of perceptual quality metrics for evaluating images displayed on commercial devices. However, there have been discrepancies between the realism predicted by these metrics [3] and user study results with 3D content. This can be attributed to the fact that conventional displays do not typically incorporate 3D visualization and the actual ocular response to 3D stimuli may increase sensitivity. This presents an intriguing opportunity for researchers in optics, graphics, and vision science to explore perceptual metrics specifically tailored for evaluating 3D visual stimuli produced by modern displays.

4.6 Conclusion

In this chapter, the author questions the optimal target of CGH supervision needed to realize high-quality 3D holographic scenes under natural viewing conditions. This is because camera-based experiments can often approximate the 3D scene and neglect the parallax cues. To address this issue, a range of simulations and experiments have been carried out. Notably, a study conducted with the perceptual testbed of a holographic near-eye display confirmed that certain types of 3D targets fall short in terms of perceptual realism. Furthermore, the inclusion of parallax cues significantly improves 3D perceptual realism even in viewing conditions with limited head movement. This study carried out the first user evaluation on 3D holographic content. The findings from this work

offer key insights into the effectiveness and 3D realism of CGH algorithms and these will be instrumental in guiding the community towards passing the visual Turing test of displays using future holographic light-field near-eye displays.

Chapter 5. Conclusion

This dissertation presents perceptual studies on 2D and 3D holographic near-eye displays. These displays feature unique characteristics due to their configuration, including 3D visualization, speckle noise, and the limited size of the eyebox. These differences, compared to conventional near-eye displays, significantly impact the overall visual experience. However, much of the existing research on holographic near-eye displays is primarily focused on enhancing visuals through camera-based experiments, often neglecting certain shortcomings that may deteriorate the user experience. Therefore, this dissertation introduces perceptual findings obtained through visual experiences with holographic near-eye displays that produce both 2D and 3D visuals.

At the start of the dissertation, the principles of holographic displays are explained, including unique visual features of holographic near-eye displays such as 3D visualization, speckle noise, and a limited eyebox size. These features significantly influence the user experience with holographic near-eye displays. The author provides a brief introduction to human visual perception, divided into two main categories: image perception and depth perception. The chapter introduces research on quantifying image quality, ranging from traditional image-based metrics to recently introduced perceptual metrics, with a succinct summary of the procedures used to quantify preferences. It also explains depth perception mechanisms and the potential impact of depth cues on 3D perceptual realism.

In Chapter 3, the author introduces perceptual findings related to 2D holographic near-eye displays along with solutions to improve the accommodation response. Recently introduced high-quality 2D CGHs extend the depth of field,

thereby eliminating accommodation cues. Conversely, speckle noise, a unique feature of holographic displays, undermines the accommodation cues widely known to be provided by such displays. Methods of speckle reduction and CGH optimization with a contrast ratio regularizer were found to significantly enhance the accommodation response.

Chapter 4 presents perceptual findings regarding 3D holographic near-eye displays, questioning the feasibility of conventional 3D scene approximations to ideal viewpoints. User studies were conducted using a perceptual testbed with a holographic near-eye display that produces high-quality 3D visuals. The study found a significant improvement in 3D perceptual realism when CGHs were supervised to reconstruct parallax, even under viewing conditions with limited head movement.

For future research, various approaches can be adopted to quantify the 3D perceptual quality of holographic content, which would enhance our chances of passing the visual Turing test with holographic near-eye displays. The author found that actual evaluations of 3D content yield different results compared to predictions made using perceptual metrics established for 2D content. Furthermore, evaluating 3D holographic content could be integrated into the differentiable 3D CGH optimization framework, enabling us to generate perceptually realistic 3D CGHs.

In conclusion, the author presents perceptual findings on 2D and 3D holographic near-eye displays, marking initial strides towards understanding the perceptual impacts of holographic stimuli. The dissertation sets out milestones to be achieved in order to produce perceptually realistic holographic near-eye displays, thereby increasing the likelihood of passing the visual Turing test.

Bibliography

[1] G. Wetzstein and D. Lanman, "Factored displays: improving resolution, dynamic range, color reproduction, and light field characteristics with advanced signal processing," *IEEE Signal Processing Magazine*, vol. 33, no. 5, pp. 119–129, 2016.

[2] A. Patney, M. Salvi, J. Kim, A. Kaplanyan, C. Wyman, N. Benty, D. Luebke, and A. Lefohn, "Towards foveated rendering for gaze-tracked virtual reality," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 6, pp. 1–12, 2016.

[3] R. K. Mantiuk, G. Denes, A. Chapiro, A. Kaplanyan, G. Rufo, R. Bachy, T. Lian, and A. Patney, "Fovvideovdp: A visible difference predictor for wide field-of-view video," *ACM Transactions on Graphics (TOG)*, vol. 40, no. 4, 2021.

[4] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *ACM SIGGRAPH 2008 classes*, pp. 1–10, 2008.

[5] A. B. Watson, "A formula for the mean human optical modulation transfer function as a function of pupil size," *Journal of Vision*, vol. 13, no. 6, pp. 18–18, 2013.

[6] L. N. Thibos, X. Hong, A. Bradley, and X. Cheng, "Statistical variation of aberration structure and image quality in a normal population of healthy eyes," *Journal of the Optical Society of America A*, vol. 19, no. 12, pp. 2329–2348, 2002.

[7] R. Navarro, P. Artal, and D. R. Williams, "Modulation transfer of the human eye as a function of retinal eccentricity," *Journal of the Optical Society of America A*, vol. 10, no. 2, pp. 201–212, 1993.

[8] M. S. Banks, W. S. Geisler, and P. J. Bennett, "The physical limits of grating visibility," *Vision research*, vol. 27, no. 11, pp. 1915–1924, 1987.

[9] D. H. Kelly, "Motion and vision. ii. stabilized spatio-temporal threshold surface," *Journal of the Optical Society of America*, vol. 69, no. 10, pp. 1340–1349, 1979.

[10] S. J. Daly, "Visible differences predictor: an algorithm for the assessment of image fidelity," in *Human Vision, Visual Processing, and Digital Display III*, vol. 1666, pp. 2–15, International Society for Optics and Photonics, 1992.

[11] P. G. Barten, *Contrast sensitivity of the human eye and its effects on image quality*. SPIE press, 1999.

[12] B. Guenter, M. Finch, S. Drucker, D. Tan, and J. Snyder, "Foveated 3D graphics," *ACM Transactions on Graphics (TOG)*, vol. 31, no. 6, pp. 1–10, 2012.

[13] C. W. Tyler and R. D. Hamer, "Analysis of visual modulation sensitivity. iv. validity of the ferry–porter law," *Journal of the Optical Society of America A*, vol. 7, no. 4, pp. 743–758, 1990.

[14] B. Krajancich, P. Kellnhofer, and G. Wetzstein, "A perceptual model for eccentricity-dependent spatio-temporal flicker fusion and its applications to foveated graphics," *ACM Transactions on Graphics (TOG)*, vol. 40, 2021.

[15] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, "Hdr-vdp-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," *ACM Transactions on Graphics (TOG)*, vol. 30, no. 4, 2011.

[16] M. Perez-Ortiz and R. K. Mantiuk, "A practical guide and software for analysing pairwise comparison experiments," *arXiv preprint arXiv:1712.03686*, 2017.

[17] R. Bogacz, E. Brown, J. Moehlis, P. Holmes, and J. D. Cohen, "The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks.," *Psychological Review*, vol. 113, no. 4, p. 700, 2006.

[18] P. G. Engeldrum, "Psychometric scaling: a toolkit for imaging systems development," *(No Title)*, 2000.

[19] J. E. Cutting and P. M. Vishton, "Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth," in *Perception of space and motion*, pp. 69–117, Elsevier, 1995.

[20] D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, "Vergence–accommodation conflicts hinder visual performance and cause visual fatigue," *Journal of vision*, vol. 8, no. 3, pp. 33–33, 2008.

[21] T. G. Martens and K. N. Ogle, "Observations on accommodative convergence: Especially its nonlinear relationships," *American Journal of Ophthalmology*, vol. 47, no. 1, pp. 455–463, 1959.

[22] B. G. Cumming and S. J. Judge, "Disparity-induced and blur-induced convergence eye movement and accommodation in the monkey," *Journal of Neurophysiology*, vol. 55, no. 5, pp. 896–914, 1986.

[23] E. F. Fincham, "The accommodation reflex and its stimulus," *The British Journal of Ophthalmology*, vol. 35, no. 7, p. 381, 1951.

[24] P. B. Kruger, S. Mathews, K. R. Aggarwala, and N. Sanchez, "Chromatic aberration and ocular focus: Fincham revisited," *Vision Research*, vol. 33, no. 10, pp. 1397–1411, 1993.

[25] S. Reichelt, R. Häussler, G. Fütterer, and N. Leister, "Depth cues in human visual perception and their realization in 3D displays," in *Three-Dimensional Imaging, Visualization, and Display 2010 and Display Technologies and Applications for Defense, Security, and Avionics IV*, vol. 7690, p. 76900B, International Society for Optics and Photonics, 2010.

[26] E. Peli and G. McCORMACK, "Dynamics of cover test eye movements.," *American journal of optometry and physiological optics*, vol. 60, no. 8, pp. 712–724, 1983.

[27] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments: A presence questionnaire," *Presence*, vol. 7, no. 3, pp. 225–240, 1998.

[28] F. Zhong, A. Jindal, Ö. Yöntem, P. Hanji, S. Watt, and R. Mantiuk, "Reproducing reality with a high-dynamic-range multi-focal stereo display," *ACM Transactions on Graphics (TOG)*, vol. 40, no. 6, p. 241, 2021.

[29] J. March, A. Krishnan, S. Watt, M. Wernikowski, H. Gao, A. Ö. Yöntem, and R. Mantiuk, "Impact of correct and simulated focus cues on perceived realism," in *SIGGRAPH Asia 2022 Conference Papers*, pp. 1–9, 2022.

[30] B. Lee, D. Yoo, J. Jeong, S. Lee, D. Lee, and B. Lee, "Wide-angle speckle-less dmd holographic display using structured illumination with temporal multiplexing," *Optics Letters*, vol. 45, no. 8, pp. 2148–2151, 2020.

[31] S. Choi, M. Gopakumar, Y. Peng, J. Kim, M. O'Toole, and G. Wetzstein, "Time-multiplexed neural holography: a flexible framework for holographic near-eye displays with fast heavily-quantized spatial light modulators," in *ACM SIGGRAPH 2022 Conference Proceedings*, pp. 1–9, 2022.

[32] B. Lee, D. Kim, S. Lee, C. Chen, and B. Lee, "High-contrast, speckle-free, true 3d holography via binary cgh optimization," *arXiv preprint arXiv:2201.02619*, 2022.

[33] V. R. Curtis, N. W. Caira, J. Xu, A. G. Sata, and N. C. Pégard, "Dcgh: dynamic computer generated holography for speckle-free, high fidelity 3d displays," in *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 1–9, IEEE, 2021.

[34] Y. Peng, S. Choi, N. Padmanaban, and G. Wetzstein, "Neural holography with camera-in-the-loop training," *ACM Transactions on Graphics (TOG)*, vol. 39, no. 6, pp. 1–14, 2020.

[35] P. Chakravarthula, E. Tseng, T. Srivastava, H. Fuchs, and F. Heide, "Learned hardware-in-the-loop phase retrieval for holographic near-eye displays," *ACM Transactions on Graphics (TOG)*, vol. 39, no. 6, pp. 1–18, 2020.

[36] J. W. Goodman, *Introduction to Fourier optics*. Roberts and Company Publishers, 2005.

[37] C.-K. Hsueh and A. A. Sawchuk, "Computer-generated double-phase holograms," *Applied Optics*, vol. 17, no. 24, pp. 3874–3883, 1978.

[38] W. H. Lee, "Sampled fourier transform hologram generated by computer," *Applied Optics*, vol. 9, no. 3, pp. 639–643, 1970.

[39] A. Maimone, A. Georgiou, and J. S. Kollin, "Holographic near-eye displays for virtual and augmented reality," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 4, pp. 1–16, 2017.

[40] L. Shi, B. Li, C. Kim, P. Kellnhofer, and W. Matusik, "Towards real-time photorealistic 3D holography with deep neural networks," *Nature*, vol. 591, no. 7849, pp. 234–239, 2021.

[41] R. W. Gerchberg, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, pp. 237–246, 1972.

[42] J. R. Fienup, "Phase retrieval algorithms: a comparison," *Applied Optics*, vol. 21, no. 15, pp. 2758–2769, 1982.

[43] P. Chakravarthula, Y. Peng, J. Kollin, H. Fuchs, and F. Heide, "Wirtinger holography for near-eye displays," vol. 38, no. 6, 2019.

[44] Y. Takaki and N. Okada, "Hologram generation by horizontal scanning of a high-speed spatial light modulator," *Applied Optics*, vol. 48, no. 17, pp. 3255–3260, 2009.

[45] S. Broomfield, M. Neil, E. Paige, and G. Yang, "Programmable binary phase-only optical device based on ferroelectric liquid crystal slm," *Electronics Letters*, vol. 28, no. 1, pp. 26–28, 1992.

[46] B. Lee, D. Kim, S. Lee, C. Chen, and B. Lee, "High-contrast, speckle-free, true 3d holography via binary cgh optimization," *Scientific Reports*, vol. 12, p. 2811, Feb 2022.

[47] Y. Takaki and M. Yokouchi, "Accommodation measurements of horizontally scanning holographic display," *Optics Express*, vol. 20, no. 4, pp. 3918–3931, 2012.

[48] R. Ohara, M. Kurita, T. Yoneyama, F. Okuyama, and Y. Sakamoto, "Response of accommodation and vergence to electro-holographic images," *Applied Optics*, vol. 54, no. 4, pp. 615–621, 2015.

[49] A. Nozaki, M. Mitobe, F. Okuyama, and Y. Sakamoto, "Dynamic visual responses of accommodation and vergence to electro-holographic images," *Optics Express*, vol. 25, no. 4, pp. 4542–4551, 2017.

[50] A. Mehrfard, J. Fotouhi, G. Taylor, T. Forster, N. Navab, and B. Fuerst, "A comparative analysis of virtual reality head-mounted display systems," *arXiv preprint arXiv:1912.02913*, 2019.

[51] D. Owens, "A comparison of accommodative responsiveness and contrast sensitivity for sinusoidal gratings," *Vision Research*, vol. 20, no. 2, pp. 159–167, 1980.

[52] S. Mathews and P. B. Kruger, "Spatiotemporal transfer function of human accommodation," *Vision Research*, vol. 34, no. 15, pp. 1965–1980, 1994.

[53] D. Kim, S.-W. Nam, B. Lee, J.-M. Seo, and B. Lee, "Accommodative holography: improving accommodation response for perceptually realistic holographic displays," *ACM Transactions on Graphics (TOG)*, vol. 41, no. 4, pp. 1–15, 2022.

[54] E. Agustsson and R. Timofte, "Ntire 2017 challenge on single image super-resolution: Dataset and study," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 126–135, 2017.

[55] S. A. Cholewiak, G. D. Love, P. P. Srinivasan, R. Ng, and M. S. Banks, "Chromablur: Rendering chromatic eye aberration improves accommodation and realism," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 6, 2017.

[56] I. C. on Non-Ionizing Radiation Protection *et al.*, "Guidelines on limits of exposure to laser radiation of wavelengths between 180 nm and 1,000 μm," *Health Physics*, vol. 71, no. 5, pp. 804–819, 1996.

[57] G.-A. Koulieris, B. Bui, M. S. Banks, and G. Drettakis, "Accommodation and comfort in head-mounted displays," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 4, 2017.

[58] K. J. MacKenzie, D. M. Hoffman, and S. J. Watt, "Accommodation to multiple-focal-plane displays: Implications for improving stereoscopic displays and for accommodation control," *Journal of Vision*, vol. 10, no. 8, pp. 22–22, 2010.

[59] N. Padmanaban, R. Konrad, T. Stramer, E. A. Cooper, and G. Wetzstein, "Optimizing virtual reality for all users through gaze-contingent and adap-

tive focus displays," *Proceedings of the National Academy of Sciences*, vol. 114, no. 9, pp. 2183–2188, 2017.

[60] P. Chakravarthula, Z. Zhang, O. Tursun, P. Didyk, Q. Sun, and H. Fuchs, "Gaze-contingent retinal speckle suppression for perceptually-matched foveated holographic displays," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 11, pp. 4194–4203, 2021.

[61] R. Horisaki, R. Takagi, and J. Tanida, "Deep-learning-generated holography," *Applied Optics*, vol. 57, no. 14, pp. 3859–3863, 2018.

[62] J. Lee, J. Jeong, J. Cho, D. Yoo, B. Lee, and B. Lee, "Deep neural network for multi-depth hologram generation and its training strategy," *Optics Express*, vol. 28, no. 18, pp. 27137–27154, 2020.

[63] S. Choi, M. Gopakumar, Y. Peng, J. Kim, and G. Wetzstein, "Neural 3D holography: Learning accurate wave propagation models for 3D holographic virtual and augmented reality displays," *ACM Transactions on Graphics (TOG)*, vol. 40, no. 6, 2021.

[64] A. T. Bahill, M. R. Clark, and L. Stark, "The main sequence, a tool for studying human eye movements," *Mathematical biosciences*, vol. 24, no. 3-4, pp. 191–204, 1975.

[65] O. Mercier, Y. Sulai, K. Mackenzie, M. Zannoli, J. Hillis, D. Nowrouzezahrai, and D. Lanman, "Fast gaze-contingent optimal decompositions for multifocal displays," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 6, pp. 1–15, 2017.

[66] P. Guan, O. Mercier, M. Shvartsman, and D. Lanman, "Perceptual requirements for eye-tracked distortion correction in vr," in *ACM SIGGRAPH 2022 Conference Proceedings*, pp. 1–8, 2022.

[67] J.-H. Park, "Efficient calculation scheme for high pixel resolution non-hogel-based computer generated hologram from light field," *Optics Express*, vol. 28, no. 5, pp. 6663–6683, 2020.

[68] J.-H. R. Chang, A. Levin, B. V. Kumar, and A. C. Sankaranarayanan, "Towards occlusion-aware multifocal displays," *ACM Transactions on Graphics (TOG)*, vol. 39, no. 4, pp. 68–1, 2020.

[69] J.-H. Park, "Recent progress in computer-generated holography for three-dimensional scenes," *Journal of Information Display*, vol. 18, no. 1, pp. 1–12, 2017.

[70] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.

[71] O. Bryngdahl and A. Lohmann, "Single-sideband holography," *Journal of the Optical Society of America*, vol. 58, no. 5, pp. 620–624, 1968.

[72] E. Jang, S. Gu, and B. Poole, "Categorical reparameterization with gumbel-softmax," *arXiv preprint arXiv:1611.01144*, 2016.

[73] Z. Zhang and M. Levoy, "Wigner distributions and how they relate to the light field," in *2009 IEEE International Conference on Computational Photography (ICCP)*, pp. 1–10, IEEE, 2009.

[74] N. Padmanaban, Y. Peng, and G. Wetzstein, "Holographic near-eye displays based on overlap-add stereograms," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 6, pp. 1–13, 2019.

[75] A. Duane, "Normal values of the accommodation at all ages," *Journal of the American Medical Association*, vol. 59, no. 12, pp. 1010–1013, 1912.

[76] C. Jang, K. Bang, G. Li, and B. Lee, "Holographic near-eye display with expanded eye-box," *ACM Transactions on Graphics (TOG)*, vol. 37, dec.

[77] A. B. Watson and D. G. Pelli, "Quest: A bayesian adaptive psychometric method," *Perception & psychophysics*, vol. 33, no. 2, pp. 113–120, 1983.

Appendix

Portions of the work discussed in this dissertation were also presented in the following publications:

[Chapter 3] D. Kim, S.-W. Nam, B. Lee, J.-M. Seo, and B. Lee, "Accommodative holography: improving accommodation response for perceptually realistic holographic displays," *ACM Transactions on Graphics* (SIGGRAPH 2022), vol. 41, no. 4, article 111, 2022.

[Chapter 4] D. Kim, S.-W. Nam, S. Choi, J.-M. Seo, G. Wetzstein, and Y. Jeong, "Full holographic parallax improves 3D perceptual realism," (submitted to SIGGRAPH Asia 2023).

초 록

홀로그래픽 근안 디스플레이는 증강 현실과 가상 현실 디스플레이 중 가장 유망한 후보로 부상하였고, 다른 차세대 디스플레이와 차별된 시각 인지적 요인들을 제공한다. 최근 컴퓨터 생성 홀로그래피 알고리즘 분야의 발전으로 홀로그래픽 이미지 화질이 급속도로 향상되었으며 이러한 발전은 실제 세계와 디스플레이를 통해 바라본 가상 현실이 구분 불가한 수준임을 평가하는 비주얼 튜링 테스트를 통과할 가능성을 높였다. 하지만, 대부분의 연구는 카메라 기반 평가에 의존하고 있기에 특정 지각적 영향을 간과할 수 있다.

본 논문은 홀로그래픽 근안 디스플레이를 사용한 시각 인지적 품질 향상을 위한 첫 시도이며, 이에는 2차원, 3차원 홀로그래픽 콘텐츠에 대한 지각 평가가 포함된다. 인간 시각 시스템 및 인간 지각 모델을 고려하여 2차원 홀로그래피 자극의 초점 깊이에 따른 명암비 시뮬레이션하여 홀로그래픽 근안 디스플레이가 제공하는 초점 요인을 향상하고 반응을 개선한다. 본 저자는 광학 솔루션과 계산 접근법을 모두 활용하여 기존 컴퓨터 연산 홀로그램의 제한된 초점 요인 제공 문제를 해결한다. 이 제안된 방안들은 사용자 연구를 통해 검증되었으며, 이는 주관적 이미지 품질 관점에서 눈에 띄지 않는 차이를 갖지만 초점 반응에서의 유의미한 개선을 나타낸다.

본 논문은 3차원 콘텐츠를 포함시키는 방향으로 연구 범위를 확장하여 홀로그래픽 근안 디스플레이를 통한 3차원 지각적 현실감을 향상한다. 실제로, 홀로그램 렌더링에 활용되는 3차원 장면은 홀로그래픽 근안 디스플레이의 아이박스가 상대적으로 작고 평가 또한 고정된 카메라를 사용하여 이루어지기 때문에 이상적인 시점에서의 3차원 근사가 정당화되곤 한다. 하지만 가상현실 디스플레이 사용시에는 끊임없는 눈의 움직임을 동반하기에, 홀로그래픽 근안 디스플레이에서 제공하는 이미지를 다양한 동공 상태에서 재구성하고 이를 평가한다. 고품질 홀로그래픽 근안 디스플레이 지각 테스트베드를 구축

하고 이를 활용하여 실시한 사용자 평가에서 컴퓨터 생성 홀로그램 최적화 시 시차 정보를 포함하는 것이 3차원 지각적 현실감에서 우수함을 사용자 평가를 통해 검증하였다. 이 결과는 머리 움직임이 제한된 환경에서도 3차원 지각적 현실감에 유의미한 개선을 나타내는 것이 검증되었으며 다양한 시청 조건 전반에 걸쳐 일관성을 유지하였다.

결과적으로, 본 논문은 홀로그래픽 근안 디스플레이를 통한 지각적 현실감 향상에 기여하며, 홀로그래픽 근안 디스플레이를 사용한 비쥬얼 튜링 테스트를 통과하는 목표에 중대한 이정표를 설정한다.

주요어: 홀로그래픽 디스플레이, 인지 연구, 증강 현실, 가상 현실
학번: 2017-29782

감사의 글

2023년 8월

김동연 올림