論文 / 著書情報
Article / Book Information

題目(和文)	光線の計測と変調による光学シースルー頭部搭載型ディスプレイにおける現実感の高い質感再現
Title(English)	Realistic Appearance Reproduction by Optical See-Through Head-Mounted Display based on Light Measurement and Modulation
著者(和文)	廣井裕一
Author(English)	Yuichi Hiroi
出典(和文)	学位:博士（工学）, 学位授与機関:東京工業大学, 報告番号:甲第11803号, 授与年月日:2022年3月26日, 学位の種別:課程博士, 審査員:小野 峻佑,小池 英樹,篠田 浩一,井上 中順,大上 雅史,伊藤 勇太
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第11803号, Conferred date:2022/3/26, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

REALISTIC APPEARANCE REPRODUCTION BY OPTICAL SEE-THROUGH HEAD-MOUNTED DISPLAY BASED ON LIGHT MEASUREMENT AND MODULATION

Yuichi Hiroi

Submitted in partial fulfilment of the requirements
for the degree of Doctor of Engineering

March 2022

School of Computing
Tokyo Institute of Technology

Abstract

Optical See-Through Head-Mounted Displays (OST-HMDs) directly overlay images on the wearer's field of view (FoV), which facilitates applications in Augmented Reality (AR) and vision augmentation. Reproducing realistic appearances in OST-HMDs encourages the wearer to make correct judgments and actions in these applications. Conventional OST-HMD technologies are, however, tend to focus on reproducing the objects' appearance, such as their contrast and depth of field (DoF).

Visual appearance is perceived by a combination of the various light property reaching the eye and the wearer's vision. Hence, reproducing a realistic visual appearance requires two factors: a display system that can reproduce rich light information and an eye measurement system that estimates ocular characteristics. In contrast, conventional OST-HMD technologies utilize devices based on digital image processing, such as cameras and displays. We focus on that the limitations of conventional OST-HMDs arise from the imaging procedure ignoring the rich information of light entering and leaving the eye.

In this dissertation, we propose an OST-HMD system that addresses the above limitation by directly measuring or modulating the light entering and leaving th eye. My technology basis includes: (1) an OST-HMD with improved contrast and DoF fidelity by direct modulation of incoming light, (2) contrast enhancement in the first-person view via the OST-HMD based on the measurement of light entering the eye using co-axial optics, and (3) vision augmentation by fusion of the proposed OST-HMD and machine vision using deep learning.

Our first contribution is focal surface occlusion, reproducing DoF for multiple virtual objects at continuous depths in an Occlusion-Capable OST-HMD (OCOST-HMD). An OCOST-HMD realizes high dynamic range (HDR) presentation by inserting an occlusion mask into the optical path. An amplitude-only spatial light modulator (ASLM) is commonly used as the occlusion mask to selectively attenuate the incoming light. Although reproducing the DoF on objects on this occlusion mask greatly facilitates depth perception, there is a trade-off among the fidelity of DoF, the display's FoV, and spatial resolution.

Our focal surface occlusion additionally inserts a phase-only SLM (PSLM) that acts as if a dynamic free-form lens into an OCOST-HMD. By approximating the focal length

of the PSLM with a curved surface to align the depth of the virtual scene, our concept can reproduce multiple and continuous focal blurs while maintaining the spatial resolution and an FoV. To reduce the distortion of the see-through view while reproducing DoF, I established an optical design based on afocal optics and edge-based optimization to exploit the property of the occlusion mask. The prototype with the PSLM and transmissive ASLM can reproduce DoF on the occluded objects at multiple and continuous depths with an FoV of 14.6°.

To reproduce realistic appearances according to the user's ocular characteristics, we need to measure the light perceived by the eyes from the user's eye position. Our second contribution is AdaptiVisor, a vision augmentation system that assists the brightness adaptation of the eye. This system uses a high-dynamic-range (HDR) camera to acquire the light information entering the eye from the viewpoint using a coaxial optical system. Based on the acquired images, the OCOST-HMD selectively darkens or brightens a part of the FoV so that the user does not feel the sudden change in brightness. Experiments show that the proof-of-concept system reduces the overexposed portions of the scene by 1/15 and the underexposed portions by half with a delay of 0.5 seconds.

If we can reproduce various appearances by combining light information measurement on the viewpoint and high-contrast OST-HMDs, then we can freely redesign the vision according to the situation. As such a system, our third contribution is DehazeGlasses, a vision augmentation system to remove haze from the first-person view using an OCOST-HMD. The system converts the viewpoint image to HDR through a dehazing algorithm, and selectively modulates the intensity of light entering the eye. The proof-of-concept system can dehaze the scene closer to the ground truth under a perceptual metric.

In summary, this dissertation aims to take a step toward realizing AR that freely manipulates the personal vision for a future in which OST-HMDs are worn daily like prescription glasses. We investigate an OST-HMD with directly measuring or modulating the light in terms of the display, eye measurement, and potential applications.

Acknowledgments

I would like to express my sincere gratitude to my supervisor, Project Associate Professor Yuta Itoh, University of Tokyo, for guiding me through my doctoral research. His infinite curiosity for new technologies and his ideas supported by his deep insight and knowledge have always motivated me to pursue my research. I would also like to thank him for his continuous support, which allowed me the freedom to explore my research.

I am also grateful to Associate Professor Shunsuke Ono, Tokyo Institute of Technology, Japan, for supporting me for a year after my supervisor moved to the University of Tokyo and cooperating with my research as an expert in mathematical optimization. Without his help, I would not have accomplished my research.

I would also like to thank my supervisor during my master's period, Professor Maki Sugimoto, Keio University, for teaching me the joy of research. Even after I entered the doctoral program, he provided me with meaningful advice and funding as a member of JST CREST, Collaborative Visual Sensing for Understanding Group Attention and Behaviors Project.

I would like to thank Associate Professor Tobias Langlotz and Assistant Professor Alexander Plopski, University of Otago, New Zealand, for hosting me as a visiting researcher and providing research guidance. The base of the idea for this doctoral dissertation was developed during my stay in their laboratory.

I am grateful to Associate Professor Daisuke Iwai, Osaka University, for his many discussions and advice as an expert in projection mapping and display technology. I am also grateful to Professor Jun Rekimoto, University of Tokyo, for his many discussions and advice from the standpoint of human augmentation technology.

I would like to thank my colleagues, all the members of Itoh Lab. In particular, I would like to thank T. Kaminokado, A. Mori, M. Kaneko for their many collaborations as co-researchers. I would also like to thank Assistant Professor Masahito Ohue, Tokyo Institute of Technology, for his support in all aspects of my doctoral life and in revising my funding applications.

This work was supported by the Japan Society for the Promotion of Science (JSPS),

Reserch Fellow (DC2) grant. In addition, this work is partially supported by JST FOREST Grant Number JPMJFR206E, JST PRESTO Grant Number JPMJPR17J2, and JSPS KAKENHI Grant Number JP20J14971 and JP20H04222, Japan.

Last but not the least, I am infinitely indebted to my parents for accepting and supporting my choice to pursue Ph.D., and my wife for accepting my precarious position as a doctoral student, marrying me, and supporting me on a daily basis.

Contents

Abstract iii

Acknowledgments v

I Introduction 1

1 Introduction for Augmented Reality Displays and their Applications 5

1.1 Augmented Reality . . . 5

1.1.1 What is Augmented Reality? . . . 5

1.1.2 Category of AR Displays . . . 6

1.2 Head-Mounted Displays . . . 9

1.2.1 Video See-Through HMDs . . . 9

1.2.2 Optical See-Through HMDs . . . 10

1.3 Vision Augmentation . . . 10

1.3.1 Comparison of HMD Types for VA Systems . . . 11

1.3.2 Overview of VA systems . . . 11

1.4 Summary . . . 12

2 Realistic Appearance Reproduction by OST-HMDs 13

2.1 Appearance Reproduction in AR . . . 13

2.2 Reproducing Photo-Realistic Appearance by OST-HMDs . . . 14

2.2.1 Accommodation-Supporting Displays . . . 15

2.2.2 Occlusion-Capable OST-HMDs . . . 17

2.3 Reproducing Appearance According to Eye Perception . . . 20

2.3.1 Appearance Reproduction with Incident Light Measurement . . . 21

2.3.2 Appearance Reproduction with Ocular Measurement . . . 22

2.4 Challenges of Appearance Reproduction in OST-HMDs . . . 22

2.5 Our Approach . . . 24

2.6 Summary . . . 25

3 Contributions of the Dissertation **27**

II High-Dynamic-Range OST-HMDs with DoF Reproduction **29**

4 Focal Surface Occlusion **31**

4.1 Introduction . . . 31
4.2 Related Work . . . 32
4.2.1 Occlusion-capable AR Displays . . . 32
4.2.2 Appllications of PSLMs in Displays . . . 33
4.3 Methods . . . 33
4.3.1 System Overview . . . 34
4.3.2 Focal Surface Formation for the Occlusion Mask . . . 35
4.3.3 Synthesizing Focal Surfaces with the PSLM . . . 38
4.3.4 Edge-Based Optimization for Minimizing the Distortion of the See-Through View . . . 41
4.4 Designing Focal Surface Occlusion . . . 42
4.5 Implementation . . . 44
4.5.1 Optical Design . . . 45
4.5.2 Hardware Setup . . . 45
4.5.3 Software Setup . . . 46
4.6 Experiments . . . 47
4.6.1 Comparison of Focal Blur Reproduction . . . 47
4.6.2 Resolution Analysis for Edge Profiles . . . 48
4.6.3 Focal Blur Reproduction in Various Scenes . . . 52
4.7 Limitation and Future Work . . . 53
4.7.1 Pixel-Wise Calibration between PSLMs and Devices . . . 53
4.7.2 Miniaturization of Form Factor . . . 55
4.7.3 Real-Time Phase Image Generation . . . 56
4.7.4 Combination with additive display . . . 56
4.8 Summary . . . 57

III High Dynamic Range Appearance Reproduction with Light Measurement on the Viewpoint 59

5 Adaptation Assistance with OCOST-HMDs 63

5.1 Introduction . . . 63
5.2 Related works . . . 65
5.2.1 Systems for Adjusting Scene Brightness . . . 65
5.2.2 Computational Photography . . . 65
5.3 Method . . . 66
5.3.1 System Overview . . . 66
5.3.2 Computation of Displaying Images . . . 67
5.4 Technical Setup . . . 69
5.4.1 Hardware and Software Setup . . . 69
5.4.2 System Calibration . . . 71
5.5 Experiment . . . 74
5.5.1 Experiment Setup . . . 74
5.5.2 Static Capture Result . . . 74
5.5.3 Dynamic Capture Result . . . 76
5.5.4 Preliminary Trial by Real Users . . . 77
5.6 Discussion . . . 77
5.6.1 Spatial Calibration with Eyes . . . 77
5.6.2 Low-latency rendering . . . 78
5.6.3 Simulating Eye Adaptation . . . 78
5.6.4 Occlusion Mask Optimization . . . 79
5.6.5 Long-term Use in Daily Life . . . 79
5.7 Summarys . . . 80

6 First-person Dehazing with OCOST-HMDs 81

6.1 Introduction . . . 81
6.2 Related Work . . . 83
6.2.1 Haze Removal . . . 83
6.2.2 Vision Augmentation to Improve the Visibility . . . 84

6.3 Method . . . 84
6.3.1 Problem formulation and Notations . . . 85
6.3.2 Basic computation for display images . . . 86
6.3.3 Alleviation of the complete occlusion issue by the occlusion layer . 87
6.4 Technical Setup . . . 91
6.4.1 System Overview . . . 91
6.4.2 Scene Camera . . . 92
6.4.3 Hardware and Software setup . . . 92
6.4.4 System Calibration . . . 93
6.5 Experiments . . . 94
6.5.1 Experiment setup . . . 94
6.5.2 Experiment procedure . . . 95
6.5.3 Experiment Result . . . 96
6.6 Discussions and Limitations . . . 99
6.6.1 Aligning image screens . . . 99
6.6.2 Color occlusion capability . . . 100
6.6.3 Depth of field . . . 100
6.6.4 Color Correction . . . 100
6.6.5 Scene Camera . . . 101
6.6.6 Hardware minimization . . . 101
6.7 Summary . . . 101

IV Conclusions and Future Work 103

7 Conclusions and Future Work 105
7.1 Conclusion . . . 105
7.2 Future Work . . . 106
7.2.1 Light Measurement and Modulation with Feedback Control . . . 106
7.2.2 Using Computational Optics with Embedded Optimization . . . 107
7.2.3 Image Presentation with Correction for Eye Aberration . . . 108
7.2.4 Low-latency VA System . . . 109

7.3 Closing Remark 109

Bibliography **115**

Publications related to this dissertation **133**

Other publications **135**

Part I

Introduction

This part serves as a general introduction to the theme of th dissertation for realistic appearance reproduction by Optical See-Through Head-Mounted Displays (OST-HMDs). The realistic appearance reproduction by OST-HMDs not only enhances the realism of Augmented Reality (AR) experiences but can be applied to Vision Augmentation (VA) technology, which re-designs individual vision ability by manipulating the real-world view.

First, we introduce a brief overview of AR, OST-HMD, and VA (Chapter 1), the area in which this dissertation is based. Then, we describe the importance of reproducing a realistic appearance with OST-HMDs and the challenges of existing technologies (Chapter 2). Finally, we summarize the contributions of this work (Chapter 3).

1

Introduction for Augmented Reality Displays and their Applications

In this chapter, we first introduce the concept of AR and its vision-based displays (Sec. 1.1). Then, we introduce Head-Mounted Displays (HMDs) in depth as the core resarch topic (Sec. 1.2). Finally, we overview the VA technologies, a major application of HMDs (Sec. 1.3).

1.1 Augmented Reality

In this chapter, we will first introduce the concept of AR (Sec. 1.1.1), and then describe the displays for vision-based AR using an established display classification [1] (Sec. 1.1.2).

1.1.1 What is Augmented Reality?

In a broad sense, AR is a technology that modifies a part of the real world by modulating sensory information that should be obtained from the real world in some way. This sensory information modulation includes various ways such as deletion, emphasis, attenuation, as well as addition. Modulating visual information is currently dominant in AR, with auditory information coming in a close second. In a broader sense, AR includes modulating any sensory information such as olfactory [2], gustatory [3, 4], haptics [5, 6], and vestibular stimulation [7, 8], etc.

Virtual Reality (VR) is similar to AR, yet modulates all sensory information. VR ultimately aims to create all sensory information at will. VR is, however, fundamentally different from AR because it completely blocks sensory information obtained from the real world and replaces it with artificially reproduced sensory information. Augmented Virtuality (AV) is another concept similar to AR. In AR, the base is the real world and is augmented with virtual information; in AV, the user is immersed in a virtual environment, where some of the virtual information is augmented with real world information. In this way, AR and AV can be viewed as essentially a continuum. The only difference is that the mixing ratio of real and virtual information changes between the real world and VR. Milgrams et al. proposed such a continuum in 1994 as the Reality-Virtuality Continuum [9]. The broader concept that includes both AR and AV is called Mixed Reality (MR).

The progenitor of today's AR and VR is Ultimate Display, proposed by Sutherland in 1965 [10]. In 1968, he created an optical see-through Head Mounted Display (HMD) based on this concept, which provided real-time 3D computer graphics that matched the user's movements [11]. In 1990, Caudell et al. coined the word *Augmented Reality* as the technology of presenting images according to the user's position in a see-through HMD, and discussed its advantages and applications in manufacturing [12]. Today, AR is widely utilized in various applications such as manufacturing [13], and medicine [14], marketing [15], entertainment [16], and education [17].

1.1.2 Category of AR Displays

Vision-based AR is the most common type of AR, which modulates the Field of View (FoV) by overlaying a computationally-generated image on the real-view. Bimber and Raskar promote that our AR experience changes depending on the spatial relationship of the eye and the display/screen and classified AR Displays as shown in Figure 1.1 [18]. In this section, we overlook the characteristics of each AR display based on Fig. 1.1. Note that, in this study we mainly focus on the category of head-attached displays in Fig 1.1.

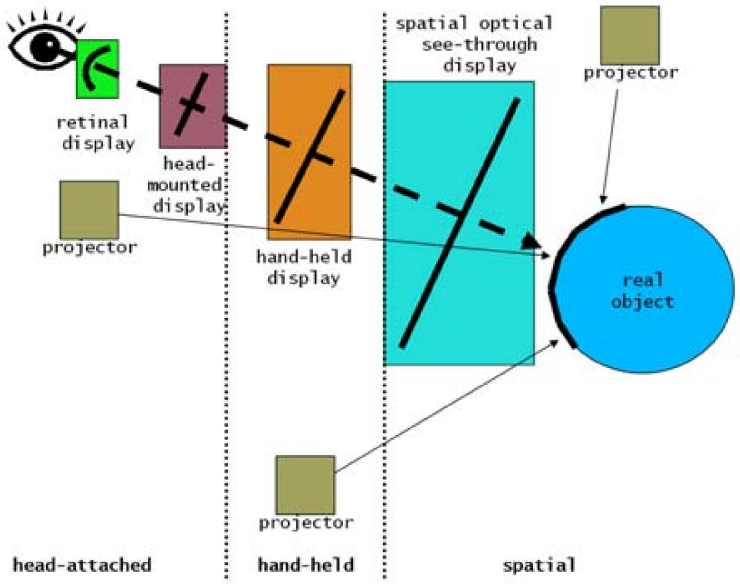

Figure 1.1: AR display categories according to the distance between the eye and the display. This figure is taken from Bimber and Rasker [18].

Spatial Displays

Spatial Augmented Reality (SAR), proposed by Raskeret al., is a category of AR that uses projection-type displays [19]. In SAR, images are projected onto the surface of real objects to manipulate the physical objects' color, texture, and apparent shape. Recently, SAR has been well-known as projection mapping and has found numerous applications in many fields, such as entertainment [20], education [21], and industrial design [22].

SAR has the advantage of working directly with the real environment. However, since the projected light is diffusive and visible from all viewpoints, it is not easy to correctly represent viewpoint-dependent appearances such as glossy texture from multiple perspectives. In addition, the dynamic range and the resolution of the projected image are limited by the display's hardware. Also, the projected image is affected by the physical object's appearance.

Spatial OST Displays [18] use a bench-top wide optical combiner (e.g., half-mirror) to overlay virtual images onto a real object. While it can improve the contrast and color representation of SAR, the form factor tends to be large, and their FoV is smaller than SAR.

Hand-Held Displays

Hand-held displays are another category of AR displays that utilize displays on mobile devices, which have developed rapidly in recent years with the spread of smartphones and tablets. Although these displays have the advantage of allowing users to interact with a familiar touch panel, the need to hold the device in one's hand interferes with activities in the real environment.

Head-Attached Displays

Head-Attached Displays overlays images directly on the user's field of view, which is inherently compatible with visual AR. Head-attached displays can provide customized AR contents for each user; thus, they can be rendered stereo images and viewpoint-depending appearances such as glossy materials. Most importantly, since head-attached displays are wearable and hands-free, they can be used in various situations with few restrictions. On the other hand, various issues remain in spatial calibration between the eye and the display, optical fidelity, and low latency rendering. We are thus far from the ultimate goal of realizing all of these simultaneously. This dissertation focuses on improving the optical fidelity of NEDs to reproduce realistic appearances, which will be discussed in detail in Sec. 2.

The classification in Fig. 1.1 defines two types of Head-Attached Displays: HMDs and retinal (projection) displays. HMDs overlay virtual images onto real-view through a display in front of the user's eyes. Video See-Through (VST) HMDs synthesize AR images by post-processing the real-view images captured from image sensors. Like goggles worn with a smartphone, VST-HMD can be defined as an extension of the handheld displays, and the user sees reality indirectly through the video. In contrast to this indirect approach, OST-HMDs overlay images onto the user's FoV while observing the real-view directly. The details of these HMDs are described in Sec. 1.2.

Retinal Projection Displays (RPDs) [23] directly form images on the retina based on Maxwellian view optics. The display illuminates the observer's eye pupil with a bundle of narrow light rays from each direction to illuminate the retina evenly [24]. The projection results in deep Depth-of-Field (DoF), creating a virtually always-in-focus (focus-free) view. In practice, RPDs are often adopted as a form of OST-HMD because they present images

directly to the user's visual field. Some commercial OST-HMDs (e.g. QD Laser's Retissa 3D) adopts the RPD as a displaying module. The major drawback of RPDs is that the periphery of the image is lost when the eye shifts from the normal viewing direction. Recent studies [25,26] have introduced Holographic Optical Elements (HOE) into RPDs to expand the range of eye position that can present a clear image (i.e., *eyebox*).

1.2 Head-Mounted Displays

As mentioned in the previous chapter, although HMDs are compatible with visual AR, the current technology is not perfect. Thus, we need to select its specification according to the target AR application. In this section, we make a detailed comparison of the HMD designs: VST (Sec. 1.2.1), and OST (Sec. 1.2.2), including trends in commercial HMDs and suitable applications.

1.2.1 Video See-Through HMDs

VST-HMDs provide an indirect view of the real world on opaque screens. VST-HMD uses a camera to capture the scenery and then displays that images together with virtual contents overlaid on top of the images. Since the display is opaque, the user cannot see the real world directly while wearing the device.

VST-HMDs are suitable for AR applications with image processing because the result of video composition is uniform, and the application can utilize real-view as a digitized image. Recent commercial VR HMDs such as Oculus Quest and HTC VIVE Pro equip a built-in scene camera on the front and record a live stream of the actual scene, thus they can be used as VST-AR HMDs.

The major concern with VST-HMDs is the discomfort caused by the divergence of the real view from the natural human perception, leading to motion sickness. Examples include reduced resolution, rendering delay, and misalignment between the camera position and the actual eye position. Emerging high-end commercial VST-HMDs, such as Varjo XR-3, partially address these issues. They simultaneously achieve a resolution of 30/30 vision, a

wide field of view of over 115°, and ultra-low latency (less than 20 milliseconds). However, in principle, VST-HMDs are difficult to apply in situations where time constraints, workability, and safety are important, such as driving, medical care, and working at heights

1.2.2 Optical See-Through HMDs

In contrast to VST-HMDs, OST-HMDs use optical combiners such as half-mirrors or prisms to overlay virtual images translucently while maintaining the visibility of the real environment. Therefore, the user can see the real-view world directly through an OST-HMD. In general, OST-HMDs use a microdisplay, so the image perceived by the user is a two-dimensional flat image.

OST-HMDs do not significantly interfere with reality, thus they are more suitable for AR experiences in daily use. Recently, OST-HMDs have been penetrating the market, as represented by EPSON BT-300, Microsoft HoloLens 2, Magic Leap One, and so on. While VST-HMDs can easily blend virtual contents into reality through image processing, OST-HMDs have many issues regarding the consistency between reality and virtual, such as narrow field of view, color and contrast issues, and misalignment of coordinates. In particular, aligning visual appearances between real and virtual has a significant impact on users' value judgments and behavioral choices.

1.3 Vision Augmentation

As AR HMDs become more sophisticated and widespread, there are a growing interest in VA technologies that can freely modulate vision to assist visual capability and to support daily life. VA is a branch of Augmented Human (AH) [27] technology that augments human abilities with technology. Unlike classical AR and MR, VA is not concerned with rendering independent virtual objects in real-view. Rather, VA is concerned with providing some convenience by directly controlling how we see the real world. Hence, it is easy to imagine VA as a direct implementation of machine vision, such as image filtering, to human vision. In this section, we first compare the features of VST and OST-HMD in terms of implementing VA systems (Sec. 1.3.1), then provide an overview of the research on VA

systems (Sec. 1.3.2).

1.3.1 Comparison of HMD Types for VA Systems

As a precursor to VA, Mann et al. proposed Mediated Reality in 1994, which aimed to modulate the real view with computation [28]. They also implemented the proof-of-concept VA system to deform the field of view and magnify interpupillary distance (i.e. turn into a Titan) on VST-HMDs. As seen in this example, early VAs were mostly based on VST-HMDs that could directly access a digital copy (image) of the actual scene shown to the user. On the other hand, when using VST-HMDs for VA implementation, the rich real-world information is lost due to the limited resolution and dynamic range of scene cameras and displays. To make matters worse, VST-HMDs risk blacking out the user's display due to accidental shutdown.

OST-HMDs are preferred to VA applications to retain real-world light and do not block the user's view in case of display failures. In recent years, with the increasing performance of OST-HMDs and the emergence of commercial products, VAs using OST-HMDs have also emerged [29]. While this maintains a natural view of reality, the quality of the overlaid virtual image is still limited by the specifications of the OST-HMD. Thus, there is a constant need to improve the reality of appearance reproduction with OST-HMDs.

1.3.2 Overview of VA systems

The basic VA system is used for visual assistance to solve various eye problems. There has been some vision assistance systems with VST-HMDs, such as visual acuity [30] and tunnel vision [31, 32] testing and treatment, eye movement disorder testing [33], strabismus [34], amblyopia [35] testing and treatment, and simulating various visual disorders [36, 37].

Since OST-HMD overlays the image with translucency, the degree of freedom in modulating the real-view is narrower than that of VST. Still, various VA systems with OST-HMDs have been proposed, such as low vision aids based on enhancing edge informatation [38], contrast [39] and correcting chroma for color blindness [40] by modulating colors on the field of view.

Furthermore, some VA systems can redesign vision ability at will. Examples include

augmentation of static [41–45] or dynamic [46] visual acuity, a user's field of view [47], shifting viewpoint [48–52], future forecasting [53], visible wavelength [54] (e.g. seeing infra-red images), etc.

1.4 Summary

In this chapter, we first introduced the core roll of OST-HMDs in the concept of AR, and their advantages and disadvantages. We then introduced VA, the prominent applications of OST-HMDs. The discussion from this chapter leads to the prospect that reproducing highly realistic appearances in OST-HMDs will improve the quality of AR experiences. Moreover, VA applications can potentially realize social welfare appropriate for an advanced information society by assisting or augmenting user's vision ability. In next chapter (Chapter 2), we will discuss how realistic appearance can be reproduced on OST-HMDs.

2

Realistic Appearance Reproduction by OST-HMDs

The previous chapter briefly introduced the importance of realistic appearance reproduction in OST-HMD for visual AR and VA. In this chapter, we will first describe the word *appearance* discussed in this dissertation in detail, and then we break down what is necessary to reproduce appearance in AR (Sec. 2.1). Next, we introduce related works on appearance reproduction in OST-HMDs in two aspects: reproduction of optical properties (Sec. 2.2) and measurement of how the eye perceives incident light (Sec. 2.3). After that, to clarify the scope of this study, we describe the challenges of the previous approach (Sec.2.4), then introduce our approach (Sec.2.5).

2.1 Appearance Reproduction in AR

Human beings perceive a variety of appearances through their senses. Here, *appearance* is defined as the human ability to estimate the subjective value (beautiful, yummy) of things and events in addition to the senses of physical properties (glossy, translucent), physical states (wet, dusty), material category (metal, ceramic). The perception of appearance is closely related to the recognition of reality; it plays an important role in basic human abilities such as understanding the environment, selecting actions, and controlling human body movements.

Among appearances, visual appearance accounts for most of the information that humans receive from the outside world. Visual appearance plays a major role in our perception of materials, determining their value, and making decisions and behavioral choices. Therefore, reproducing appearances in AR that is visually indistinguishable from reality is very important in encouraging users to make correct value judgments and action. Moreover, the visual appearances indistinguishable from reality can be appropriately modulated and presented to humans, which can also be considered to actively change human recognition of reality as a context of VA (Sec. 1.3).

Visual appearance is perceived when light from an object's surface reaches the retina. More specifically, the perception of visual appearance contains the estimation of the optical properties (color, material, state, and shade) of an object's surface from a two-dimensional image formed on the retina through the visual properties of the individual eye (lens, pupil, and cornea). Hence, reproducing a realistic visual appearance requires two factors: a display system that can reproduce rich light information and an eye measurement system that estimates advanced ocular characteristics.

2.2 Reproducing Photo-Realistic Appearance by OST-HMDs

Here, we introduce the designs of OST-HMD to reproduce photo-realistic appearances. Components for reproducing photo-realistic appearance with OST-HMDs include environmental lighting [55, 56], color reproduction [57–60], focal blur, and contrast reproduction.

In particular, the reproduction of contrast and focal blur is a problem that arises from the inherent limitations of common OST-HMD designs. As described in Sec. 1.2.2, common OST-HMDs provide a two-dimensional planar image to the eye through an optical combiner. Imaging with optical combiners leads to a significant contrast degradation caused by the translucent overlay of virtual images. In addition, since OST-HMDs produce a planar image with a fixed depth in the air, the virtual object is reproduced with the wrong depth of field (DoF), i.e., the optical depth of the image presented does not match the accommodation of our eyes.

In this section, we first introduce accommodation-supporting displays that reproduce

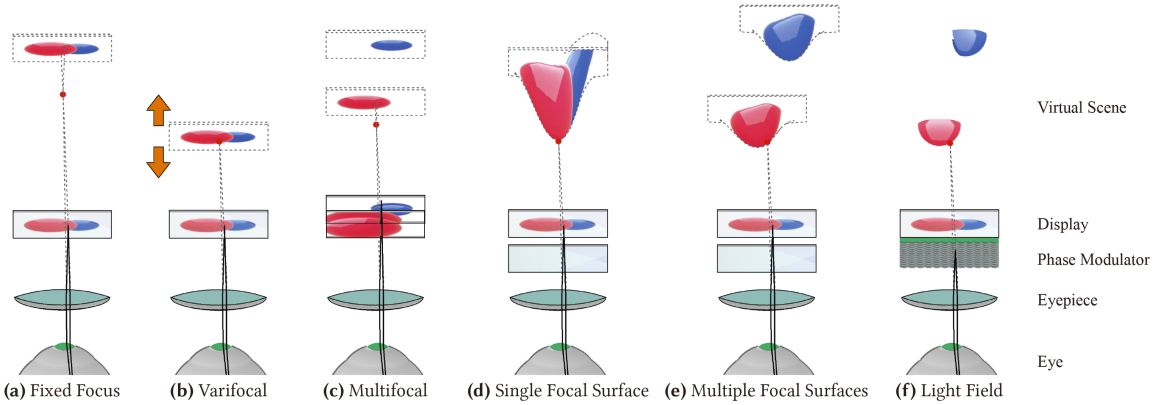

Figure 2.1: Classification of Accommodation-Supporting Displays. This figure is taken from Matsuda et al. [61]

the correct DoF (Sec. 2.2.1), and then introduce occlusion-capable OST-HMDs (OCOST-HMDs) that support optical occlusion to block real-world light to improve contrast of virtual images (Sec. 2.2.2).

2.2.1 Accommodation-Supporting Displays

Accommodation-supporting displays have been extensively studied in both VST and OST-HMD as a solution to reduce the vergence-accommodation conflict (VAC), which is caused by a mismatch between the binocular disparity of a stereoscopic image and the optical focus cues of the virtual objects. The essential solution for the mismatch is to render virtual objects at correct depths so that the light rays from the object appear as if they are coming from the proper distances.

Figure 2.1 shows the classification of accommodation support displays. Typical HMDs always fix the focal length of the display (Fig. 2.1, a). In the following, we describe an overview of each accommodation-supporting display category.

Varifocal Displays (Fig. 2.1, b) Varifocal displays are a simple solution for solving the VAC by mimicking continuous focal planes in a time-multiplexing manner. This time-multiplexing can be achieved by using focus tunable lenses (FTLs) [62–66] or modulating beam splitters, such as membrane mirrors [67, 68] and holographic optical elements [69].

Although varifocal displays can reduce VAC, they must measure the wearer's accommodation and render the depth of field (DoF) according to the measurements.

Multifocal Displays (Fig. 2.1, c) Instead of dynamically moving single focal planes, some OST-HMDs use multiple displays to form multiple focal planes in advance. This mutifocal display is initially achieved by a stack of SLMs [70–72], and recently such displays have been revisited with improved scene decomposition [73] and gaze-contingent capabilities [74]. For example, for a simple LCD-stacked multifocal display, achieving a 2.0 D (diopter, m^{-1}) DOF requires up to 27 planes [70]. Other approaches for achieving a multifocal display include using FTLs [75, 76], deformable mirrors [77], a fast polarization rotator and the Savart plate [78], and a geometric phase holographic lens [79].

Focal Surface Displays (Fig. 2.1, d, e) Focal surface displays [61] is an another approach to generalize the multifocal displays. Focal surface displays utilizes a PSLM which acts as a dynamic free-form lens, which can display time-multiplexing multiple focal plane images as curved continuous focal surfaces. Compared to multifocal displays, the focal plane provides a better approximation of the focus cue performance and can be achieved in a smaller form factor. Although the focal plane display was first proposed in VST-HMD, recently research has been conducted in OST-HMD as well, such as the reproduction of static focal plane using arbitrarily shaped 3D printed curved lenses [80].

Accommodation-Invariant Displays A completely different approach to achieving accommodation supporting HMDs is a display that provide the invariant focus cue for any depth using Maxwellian view. Pinlight displays [81] is taking inspiration from a pinhole camera, which captures a photo where everything is in focus thanks to the small aperture, to achieve always-in-focus virtual images onto OST-HMDs. Retinal scanning displays [26, 82] are another approach to achieve Maxwellian view by sweeping a point of light across the viewer's retina. These accommodation-invariant displays need to blur the images in the render step if out of focus effects are required for the displayed image.

Light-Field and Holographic Displays (Fig. 2.1, f) Light-field displays [83] and holographic displays [84] offer, at least in theory, simultaneously all depth cues by representing 4D light fields. However, these displays still have limitations, including trade-off between small eyebox and limited resolution. Recently, a scattering-based holographic see-through display [85] has been proposed to improve the trade-off between resolution and eyebox.

2.2.2 Occlusion-Capable OST-HMDs

The accurate occlusion of the incoming light is essential to improve the contrast of OST-HMDs. Without occlusion, the OST-HMD can only semi-transparently overlay virtual information, causing a lack of reality in the virtual information. Furthermore, occlusion plays an essential role as a depth cue [86]; a virtual object close to the wearer's point of view occludes a real object in the distance, allowing the wearer to perceive the depth between objects.

The most common approach to achieve an occlusion-capable OST-HMD (OCOST-HMD) is to insert a spatial light modulator (SLM), typically a transmissive liquid crystal display (LCD), into the optical path of the OST-HMD to selectively block light rays entering each pixel of the SLM. In general, SLM-based OCOST-HMDs usually assume that the occlusion mask is placed at a single fixed depth [87]. Thus, it cannot provide accurate focal blur when the user's accommodation changes. In recent years, OCOST-HMDs that support DoF reproduction are emerged by applying the approaches of accommodation-supporting displays described in the previous section.

A comparison between this work and other OCOST-HMDs is shown in Table 2.1, which is an extension from the detailed review in [88]. In Table 2.1, "Occlusion Generation" column shows the devices that consist the occlusion mask. "Configuration" column indicates the configuration of the focus of the occlusion. "Adaptive Depth" column indicates that whether the occlusion images correspond to variable focus or not. "The 3D occlusion" column shows whether the systems can reproduce the 4D ray field of the occlusion object. Moderate optical complexity designs use 3-4 optical elements (including the SLM) and moderate FoV is 40-80°.

This section introduce related works on OCOST-HMDs based on Table 2.1, focusing

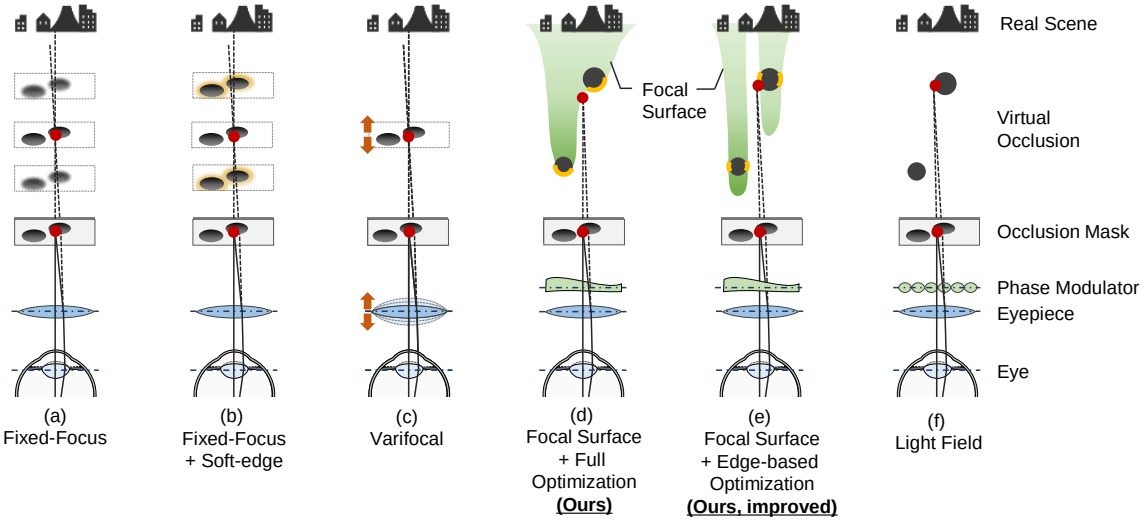

Figure 2.2: Classification of OCOST-HMD designs based on differences in accommodation support.

on DoF reproducibility on the occlusion mask as shown in Figure 2.2.

Hard-edge Occlusion (Fig. 2.2, a) In an early study of OCOST-HMDs, Kiyokawa et al. used relay optics to focus the incident light on a transmissive LCD to achieve a fixed-focus, sharp occlusion [87, 100]. Another approaches to achieve sharp edge occlusion has also been proposed by using a digital micro-mirror device (DMD) to switch between the background image and the AR image pixel by pixel [89, 90]. Krajancich et al. enhanced this DMD-based approach with a simpler optical design by switching between light from LEDs and light from the real world, making the DMD itself a display screen [91]. LCoS-based occlusion-capable OST-HMDs have also been proposed in combination with x-cube prisms [95], free-form lenses [93, 94], and affordable lens components [92, 101]. They generally provide sharp occlusion in fixed focus while achieving a thin form factor.

Soft-edge Occlusion (Fig. 2.2, b) Pinlight displays [81] and light-field displays with stacked LCD arrays [96] directly insert an occlusion mask in front of the eye without focusing. Although these approaches allow thinner form factors, simpler optical configurations, and wider viewing angles (FoV), they result in a blurred edge in the mask due to a depth

Table 2.1: Comparison of occlusion methods for OST-HMDs adapted from [88]. *The items with apostrophes were estimated by us from their paper.

.

Methods	Occlusion Generation	Configuration	Adaptive Depth	3D Occlusion	Mask Appearance	Computational Demand	Optical Complexity	Form Factor	FoV
Kiyokawa et al. ('03) [87]	LCD	Fixed-Focus	no	no	sharp	low	complex	bulky	30°
Uchida et al. ('02) [89]	DMD	Fixed-Focus	no	no	sharp	low	complex	moderate	narrow*
Kim et al. ('19) [90]	DMD	Fixed-Focus	no	no	sharp	low	complex	moderate	narrow*
Krajancich et al. ('20) [91]	DMD	Fixed-Focus	no	no	sharp	high	complex	moderate	8.7°
Wilson and Hua ('21) [92]	LCoS	Fixed-Focus	no	no	sharp	low	complex	thin	40°
Gao et al. ('12) [93,94]	LCoS	Fixed-Focus	no	no	sharp	low	complex	thin	40°
Cakmakci et al. ('04) [95]	LCoS	Fixed-Focus	no	no	sharp	low	moderate	thin	40°
Maimone and Fuchs ('13) [96]	Stacked LCD Array	Light Fields	no	no	dull	high	moderate	thin	moderate*
Maimone et al. ('14) [81]	Pinholes with LCD	Light Fields	no	no	dull	moderate	simple	moderate	moderate*
Itoh et al. ('17) [97]	LCD	Fixed-Focus Compensated	partially	no	sharp	low	simple	thin	70-80°
Yamaguchi and Takaki ('16) [98]	LCD with Microlenses	Light Fields	yes	yes	sharp	high	moderate	thin	4.3°
Hamasaki and Itoh ('19) [88]	LCD on linear stage	Varifocal	yes	no	sharp	low	moderate	moderate	narrow*
Rathinavel et al. ('19) [99]	LCD with FTLs	Varifocal	yes	no	sharp	moderate	complex	moderate	15.3°
This Work (Chap. 4)	LCoS and LCD	Focal Surface	yes	yes	sharp	moderate	moderate	moderate	14.6°

mismatch between the LCD and the virtual object overlaid in the scene. To resolve this blurry occlusion while maintaining the simple optics and wide FoV, Itoh et al. proposed a method to compensate for the blurry occlusion by fusing the scene image through an OST-HMD [97]. However, this system requires complex calibration between devices and continuous eye tracking and accommodations.

Varifocal Occlusion (Fig. 2.2, c) In order to resolve VAC in the occlusion mask, some varifocal OCOST-HMDs have been proposed. Hamasaki and Itoh [88] proposed a varifocal OCOST-HMD by mounting a transmissive LCD on a linear stage and mechanically shifting the depth of the occlusion mask according to the accommodation. Rathinavel et al. proposed another varifocal OCOST-HMD approach where they place two FTLs in the see-through path [99]. By controlling the focus of the lenses optimally, the system can optically tune the focus depth of the reflective LCD layer, which is placed between the lenses.

These methods can adaptively adjust the depth of the occlusion layer. However, because such techniques change the depth of only a single mask, only one discrete depth can be represent at a given time. Furthermore, even if the depth of a single mask is changed in a time-multiplexing manner, it is impossible to simultaneously occlude all depths; thus, the 3D occlusion mask appears translucent in such cases.

Light-Field Occlusion (Fig. 2.2, f) A light field occlusion approach using micro-lens array [98] has been proposed as a way to completely reproduce the depth of the virtual object including occlusion and provide a completely accurate focal blur. However, discrete lens arrays and liquid crystal structures involve resolution limitations and a trade-off between the 3D representation and FoV owing to diffraction and compression artifacts.

2.3 Reproducing Appearance According to Eye Perception

The previous section introduced accommodation-supporting displays reproducing correct DoFs and OCOST-HMDs reproducing high-dynamic range images. To reproduce a more

realistic appearance, we need to measure both the field of view from viewpoint and eye characteristics to estimate how they perceive the incoming light. For example, suppose that the intensity of light entering the eye from reality is measured directly; then we measure the pupil diameter to estimate the image perceived on the retina. By modulating the amplitude of the incident light based on this estimated image, we can enhance the brightness perception of the user's eyes. Similarly, if we can measure the phase of the incident light and the eye aberration, we can reproduce the appropriate DoF by phase modulation.

In this section, we first introduce an appearance reproduction method on an HMD based on measurements of incident light into eyes (Sec. 2.3.1). Then, we introduce an appearance reproduction according to measured ocular characteristics (Sec. 2.3.2), mainly focused on eye accommodation measurement and DoF reproduction.

2.3.1 Appearance Reproduction with Incident Light Measurement

As a method of measuring incident light and improve the brightness of the light entering the eye, Aiteanu et al. developed a video see-through welding helmet that provides contrast-enhanced images [102]. As an OST-HMD system, Bhagavathula et al. developed a low-power CMOS glare sensor, and combined it with an LCD shutter glasses to form a wearable glare reduction glasses. Similarly, Mori et al. proposed a system to attach LC shutters and Illuminometer to OST-HMD [103]. In a system similar to ours, Tochimoto et al. [C3]. realized adaptive blue light-cutting glasses by using a spectrometer to measure the spectrum of ambient light and then using an LC shutter to attenuate the light entering the eye. These system measure the light from the scene only in terms of (circadian) illumination. In addition, since the sensor is placed in a different location than the viewpoint, it does not measure the exact illuminance that the eye receives.

Few studies on color correction in OST-HMDs measure the incident light from the same optical position as the eye. Our study is inspired by Langlotz et al. [60], placing the scene camera in the same optical position as the eye using a beam splitter. While their research utilizes light information entering the eye for real-time radiometric color compensation, we combine their approach with OCOST-HMDs, and consider both color and contrast correction to implement advanced machine vision into human eyes.

2.3.2 Appearance Reproduction with Ocular Measurement

In AR-HMDs, display manufacturers simplify the measured accommodation depth to be accurate to within ± 0.3 D (Diopter, i.e., m^{-1}), the range where the user cannot perceive focus blur [104]. Note that this range translates to the vergence angle of 0.541° for an adult male with normal vision [105].

Some VST-HMD designs reproduce DoF by using FTLs after measuring eye accommodation. Konrad et al. combined a commercial autorefractor with a custom focus-tunable binocular display [64]. They use the measured ocular aberration information, but the use was limited to control their FTLs to adjust the image planes to be in focus to the user so that the visual appears focus-invariant. Similarly, Mercier et al. combined both an eye tracker and SHWFSs [74], to drive three adjustable-focus displays per eye with FTLs. Also in OST-HMDs, Rompapas et al. proposed EyeAR [45], a system for rendering refocusable AR content with correct DoF. This system combines a microdisplay with an autorefractor to measure an individual's focal length and pupil size, reproducing DoF-reproducing AR content through software.

In addition, some OST systems have been proposed to utilize the measured accommodations for correcting the real-view distortion. As a method that uses pre-measured prescriptions, Chakravarthula et al. proposed FocusAR [42], an auto-focus OST-HMD that can properly focus on real-view and virtual content simultaneously, using a varifocal membrane beam combiner design. Recently, Wu et al. [44] proposed Prescription AR, which covers hyperopia, myopia, astigmatism, and presbyopia by using prescription lenses as waveguides for AR displays. As a vision correction system using continuously measuring accommodations, Padmanaman et al. [43] used FTL to develop a real-view auto-focus system from accommodations obtained from binocular eye-tracking.

2.4 Challenges of Appearance Reproduction in OST-HMDs

In the previous sections, we introduced related works on appearance reproduction on OST-HMD, both in terms of reproducing photo-realistic and ocular characteristics. Based on these studies, we summarize what can be done and challenges in the current appearance

reproduction as follows:

- As for the photo-realistic appearance reproduction (Sec.2.2), some methods have been proposed to reproduce DoF on OCOST-HMDs. However, the varifocal OCOST-HMD cannot avoid displaying semi-transparent images in practical use, while the light field-based OCOST-HMD remains limited in FoV and image quality.

- As for appearance reproduction based on eye perception (Sec.2.3), it is mainly based on adaptive contrast adjustment based on the average brightness of the scene. There is an example of appearance reproduction based on measuring the user's field of view from the viewpoint; however, due to the dynamic range that OST-HMD can represent, this approach is limited in compensating both color and contrast.

As can be seen from the Section 2.2 and 2.3, current research trends independently deal with the eye measurement and the display system in general. Some studies utilize eye measurements to present a wide field of view and high resolution [25, 26]. However, these studies are limited to the use of gaze measurement and have not yet utilized the accommodation measurement or advanced ocular estimation.

One of the reasons the measurement and presentation systems are researched independently is that companies are leading the research, so it is assumed that their priority is to improve the performance of practical elemental technologies directly related to product specification. For example, OCOST-HMDs (Sec. 2.2.2) have been researched for almost 20 years, but it has not been implemented in consumer OST-HMDs.

More importantly, in the technological background, it was difficult to combine both the measurement and display systems while miniaturizing their form factor. For example, to implement a varifocal VST-HMD on both eyes while performing gaze and eye focus measurements, it was necessary to design a large benchtop system [74].

On the other hand, due to recent technological developments, the fundamentals for OST-HMDs that can integrate both measurement and presentation, is beginning to mature. The technological developments include the miniaturization of hardware and advanced machine vision with embedded system implementation. Regarding the miniaturization of hardware, sensing devices are becoming smaller, not only RGB cameras but also wavefront sensors [106], polarization cameras (e.g., SONY Polarsens CMOS image sensor),

and multi-spectral cameras [107, 108]. Also, micro LEDs and micro OLEDs (Organic LED) have emerged for miniature display devices. Furthermore, with the development of deep learning and the emergence of TPUs and GPUs for embedded systems, such as Google Coral and NVIDIA Jetson, it has become possible to implement advanced machine vision on HMDs.

2.5 Our Approach

This dissertation explores an approach to reproduce realistic appearance by handling the measurement system and the presentation system as a single unit on OST-HMDs. The advantage of this approach is that it uses the measured field of view as input and optimizes the compensating image using the properties of the presentation system, thereby enabling expressions that cannot be achieved by the presentation system or the measurement system independently.

Moreover, we focused on an approach based on direct modulation of light rays that goes beyond image processing, taking advantage of the characteristics of OST-HMD where the wearer perceives the light rays directly entering the eye. Conventional OST-HMDs use a conventional imaging camera to measure ocular or real-view information and a display to overlay virtual information. However, these camera and display devices lose part of the ray information in the imaging process, such as light wavefront and spectral distribution. We thought that by utilizing the rich information of light entering and leaving the eyes on OST-HMDs, we could reproduce a realistic appearance beyond conventional limitations.

Close to our work, Itoh et al. [109] realizes computer-modulated see-through glasses by directly modulating the phase of incoming light using PSLMs as free-form lenses. With this system, they implemented VA applications such as field of view magnification, viewpoint shift and re-focus. Similarly, Arias et al. [110] simulate the view of cataracts by PSLM and compensate for edge information of the virtual image with its inverse simulation. While these studies propose new VA applications by directly modulating the phase of light entering the eye, they have limitations for reproducing photo-realistic fidelity, especially in contrast. Our work extends these studies, aiming to integrate more flexible machine vision into human vision by measuring the incident light to the eye and modulating both the phase

and amplitude of the light.

2.6 Summary

In the above, we introduced the components for reproducing realistic appearances on OST-HMDs, what has been achieved to date, and the remaining challenges. Then, we hypothesized that an approach based on direct modulation of light could utilize the rich light information lost in the conventional image processing and break through the trade-offs of the existing methods. The following chapters will elaborate on our proposed OST-HMD based on the direct modulation of light ray information and VA applications based on this approach.

3

Contributions of the Dissertation

In this dissertation, we propose an OST-HMD system that addresses the limitation for reproducing realistic appearances by directly measuring or modulating the light entering and leaving the eye. Our system directly measures the amplitude of light entering the eye by placing the sensors at the same optical position as the eye using a co-axial optical system. Using this measured light information, we can directly modulate the amplitude of the light entering the eye by OCOST-HMDs and the phase by PSLM to increase the fidelity of appearance reproduction of OST-HMDs. Furthermore, we explore the VA applications of the our concept to realize a realistic appearance modulation for real-view depending on the individual vision characteristics under various situations.

We summarize the specific contributions of this work below:

Contribution 1: Realistic Depth-of-Field Reproduction for High-Dynamic Range OST-HMDs (Chap. 4) We propose *Focal Surface Occlusion*, reproducing DoF for multiple virtual objects at continuous depths in an OCOST-HMD. Our focal surface occlusion additionally inserts PSLM that acts as if a dynamic free-form lens into an OCOST-HMD. By approximating the focal length of the PSLM with a curved surface to align the depth of the virtual scene, our concept can reproduce multiple and continuous focal blurs while maintaining the spatial resolution and an FoV. To reduce the distortion of the see-through view while reproducing DoF, we established an optical design based on afocal optics and edge-based optimization to exploit the property of the occlusion mask. The prototype with

the PSLM and transmissive ASLM can reproduce DoF on the occluded objects at multiple and continuous depths with an FoV of $14.6°$.

Contribution 2: Adaptation Assistance System with OCOST-HMDs (Chap. 5) We propose *AdaptiVisor*, VA system that assists the brightness adaptation of the eye. Our system selectively modulates the intensity of the light coming into the eyes via OCOST-HMD. An integrated camera captures highlights and brightness in the environment via high-dynamic range capture. Our display system selectively dims or enhances the part of the field of view so that the user would not perceive rapid brightness changes. We build a proof-of-concept system to evaluate the feasibility of the adaptation assistance by combining a transmissive LCD panel and an OST-HMD, and evaluate it with a user-perspective camera. Our evaluation shows that the system decreases the overexposed area in a scene to 1/15 and enhances the color by reducing underexposed area to half.

Contribution 3: Dehazing System with OCOST-HMDs (Chap. 6) We propose *DehazeGlasses*, VA system that dehazes the user's field of vision. Our system selectively modulates the intensity of the light entering the eyes via an OCOST-HMD. We built a proof-of-concept system to evaluate the feasibility of our haze removal method by combining a DMD and an OCOST-HMD, and evaluated it with a user-perspective camera. A quantitative evaluation with 80 scenes from a haze removal dataset shows that our system realizes a dehazed view that is significantly closer to the ground truth scene compared to the native view under a perceptual image similarity metric.

The dissertation is organized as follows: In Part II, we introduce the reproduction of DoF on OST-HMD using phase modulation (Contribution 1). In Part III, we introduce the high-dynamic-range appearance reproduction on OST-HMDs by applying field-of-view measurement for system design (Contribution 2) and its VA application (Contribution 3).

Part II

High-Dynamic-Range OST-HMDs with DoF Reproduction

4

Focal Surface Occlusion

This chapter introduces Focal Surface Occlusion, a method to reproduce high-fidelity DoF on high dynamic range appearance reproduction with an OCOST-HMD. This section is based on the work that the author published in Optics Express, Vol.29, Issue 22, 2021 [J1].

4.1 Introduction

In practical AR applications, virtual objects are rendered at multiple and continuous depths. When applying occlusion to virtual objects with such complex virtual depths, it is necessary to ensure accurate focal blur for the entire virtual geometry. For example, consider an FoV including three objects in a line: a real object at the back, and two virtual objects in the middle and front. When the user's focus is on the virtual object in the middle, the virtual object in the front must be blurred similar to the real object at the back. Reproducing this focal blur based on the eye accommodation in virtual objects is of significance in facilitating accurate depth perception. However, as discussed in Sec. 2.2.2, it remains challenging to develop OCOST-HMDs that can reproduce the focal blur for multiple objects placed at different depths or continuous objects that extend from the front to the back.

To reproduce the light field while maintaining the resolution and FoV, we propose a novel OCOST-HMD concept, *focal surface occlusion* based on the focal surface approach [61], which relaxes the light field constraint by using a continuously curved lens. This focal surface approach uses a PSLM, such as a reflective LCoS. The PSLM acts as a dynamic

free-form lens; by approximating the focal surfaces to be closer to the shape of the virtual scene, the PSLM can make it appear as if the rays are emerging from the surface of the virtual scene. We formulate an optical design to apply a focal surface to the occlusion mask of OCOST-HMD by incorporating the PSLM into the afocal optical system that views the real image at equal magnification.

The focal surface approach has been initially applied in VR-HMDs [61]. When a focal surface is applied to an OCOST-HMD, the challenge is to minimize the optical distortion of the see-through view. In particular, in an OCOST-HMD, the light from both the real scene and occlusion mask passes through the same optical path. Therefore, optimizing the lens to reproduce only the light field from the occlusion mask may result in undesired scaling, distortion, and contrast loss in the see-through view. To alleviate this see-through view degradation, we adopt an edge-based lens optimization method that exploits the occlusion property.

Moreover, we develop a proof-of-concept system based on an optical design using PSLM and a transmissive LCD occlusion mask and experimentally confirm that the focal blur for multiple objects with continuous depths can be expressed while suppressing the see-through distortion. Furthermore, we evaluate the spatial resolution of the prototype and confirm that the proposed system can present a 3D focal blur with a higher resolution and FoV than those associated with conventional light field occlusion approaches.

4.2 Related Work

4.2.1 Occlusion-capable AR Displays

We described related works of OCOST-HMDs in Sec. 2.2.2, and compared them with our focal surface occlusion is shown in Table 2.1. Here, we briefly introduce the differences between the existing OST-HMD design and our Focal Surface Occlusion.

The varifocal OCOST-HMD [88, 99] (Fig. 2.2, c) requires accommodation measurement and hardware control systems, and cannot avoid displaying semi-transparent images to simultaneously occlude all depths. In contrast, our focal surface occlusion approximates the light field of the occlusion object and thus does not require additional equipments to

reproduce DoF. Furthermore, the occlusion mask can be displayed an occlusion layer as opaque because it does not use time-multiplexing.

In light field OCOST-HMDs [98] (Fig. 2.2, f), the pitch of the lens array and PSLMs limits the resolution and FoV due to diffraction. Our focal surface occlusion approximately reproduces the 4D light field while maintaining resolution by displaying continuous, differential phase function on the PSLM.

4.2.2 Appllications of PSLMs in Displays

Holographic displays [84, 85] is the major application of PSLMs. In addition, PSLMs are used to spatially change the direction of light to reproduce arbitrary caustics patterns [111] or create a passive view-point sensitive display [112]. The light-attenuation display [J2] realizes color filters by sandwiching PSLMs between two polarizers arranged in a cross-Nicol configuration. Recently, to reduce the defocus blur leakage of each transparent display in a multifocal VR-HMD, Chang et al.. propose a method to attach the PSLM to the display and tilt the light cone with a free-form lens [113].

Itoh et al. [109] realize computer-modulated glasses by modulating the see-through view using PSLMs as free-form lenses. We extend their optical design to provide proper focal cues for the occlusion mask while maintaining real-world visibility.

4.3 Methods

The focal surface occlusion approach uses a PSLM as a programmable free-form lens to spatially focus each pixel of a fixed-plane occlusion mask onto a different depth. By ensuring that the focal surface of the occlusion mask approaches the depth of the virtual object observed from the viewpoint, our system yields a perceptually correct focal blur for multiple and continuous depths. To this end, we optimize the phase function displayed on the PSLM to ensure that the virtual focal length of the light rays from the occlusion mask approaches the input target depth.

Sec. 4.3.1 presents an overview of the proposed system, along with the notations and

prerequisite information. Sec. 4.3.2 describes the simplification of the PSLM as a thin concave lens and the method to realize varifocal occlusion (Fig. 2.2, c) through the proposed system. In Sec. 4.3.3, after extending the PSLM focal length to a spatially varying function, we introduce the existing method that optimizes the focal surface using all the target depth map pixels (Fig. 2.2, d). Finally, Sec. 4.3.4 describes the edge-based focal surface optimization method to reduce the distortion of the real scene (Fig. 2.2, e).

4.3.1 System Overview

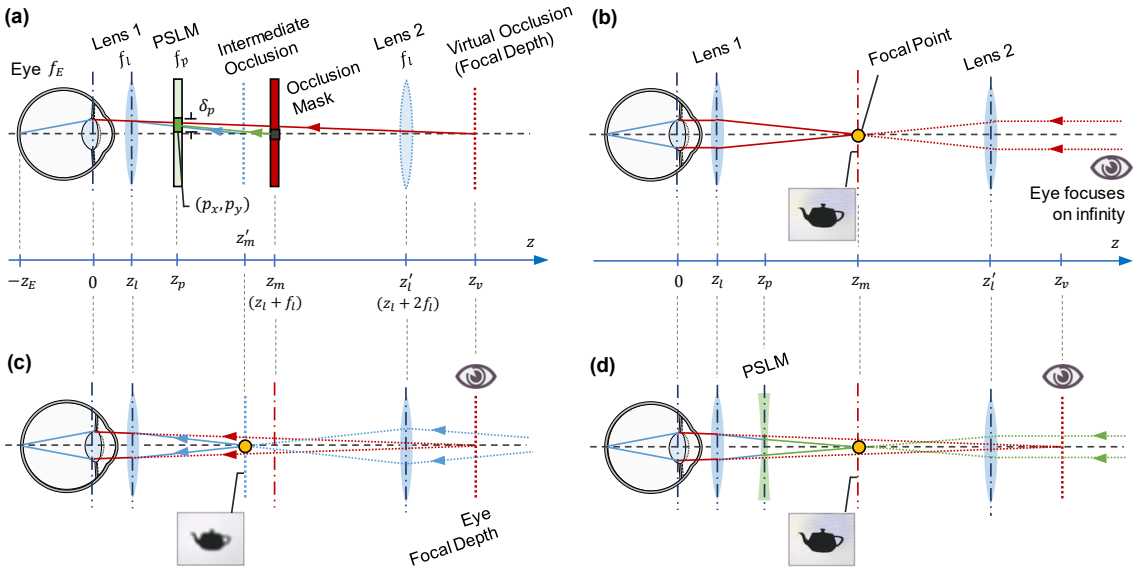

Figure 4.1: Schematic of the light path in focal surface occlusion. (a) Optical configuration of the proposed system. (b) If the PSLM is ignored, the occlusion mask appears sharp when the eye focuses on infinity. (c) As the focus of the eye moves forward, the occlusion mask appears blurred. (d) We adjust the focal length of the PSLM according to the eye focus to ensure that the occlusion mask is sharp.

Figure 4.1 shows the schematic of the optical design for the focal surface occlusion. The proposed system includes two lenses, Lenses 1 and 2, having the same focal length, and an occlusion mask placed on the focal point of these lenses. In other words, we place an occlusion mask on the focus plane of an afocal optical system that does not change the magnification of the see-through view. A PSLM is placed between the occlusion mask and

Lens 1. In this design, the occlusion mask can be an arbitrary amplitude-only SLM, such as a DMD or a transmissive LCD.

Notations and Prerequisite Information

Displayed Images and Device Specifications. $\phi(p_x, p_y)$ is the phase function displayed on the PSLM, which is to be optimized. (p_x, p_y) is the physical pixel position [m] on the PSLM. δ_p and r_p denote the pixel pitch and diagonal radius of the PSLM, respectively.

Focal Lengths of Lenses. $f_p(p_x, p_y)$ is the focal length of the PSLM at (p_x, p_y). As described in Sec. 4.3.3, $f_p(p_x, p_y)$ depends on the phase function $\phi(p_x, p_y)$. f_l is the focal length of Lenses 1 and 2.

Distances of Each Component. The distance of each optical component along the optical axis of the eye lens is denoted with reference to the position of the eye lens ($z = 0$). Along this axis, $z_l, z_p, z_m = z_l + f_l$, and $z'_l = z_l + 2f_l$ denote the depth of Lens 1, PSLM, occlusion mask, and Lens 2, respectively.

Input Depth Map. $\hat{d}(\theta_x, \theta_y)$ is the depth of the target virtual scene along each angle $(\theta_x, \theta_y) \in \Omega_\theta$ from the center of the pupil as the origin, where Ω_θ is a discrete set of retinal image samples. Similarly, $d(\theta_x, \theta_y)$ is the depth of the focal surface for each viewing angle, which is a smooth approximation of the true depth.

Virtual Focal Depths. z_v is the depth of the virtual occlusion mask, as perceived by the eye through the system. z'_m is the intermediate depth at which the eye perceives the occlusion if the refraction by the PSLM is ignored.

Eye Lens System. f_E is the focal length of the eye lens, and z_E is the distance from the eye lens to the retina.

4.3.2 Focal Surface Formation for the Occlusion Mask

This section describes the method to calculate the focal length of the free-form lens in the PSLM to ensure that the occlusion mask is reproduced at the desired depth in the afocal optical system. For simplicity, we assume that the PSLM is a thin concave lens and explain

how to achieve varifocal occlusion by moving the single mask to an arbitrary depth. As described in Sec. 4.3.3, we later extend this varifocal setup to the focal surface setup.

In Sec. 4.3.2, we schematically describe how the optical design shifts the focal depth of the occlusion mask. This explanation is formulated in Sec. 4.3.2 by using the ray transfer matrix to calculate the focal length to be displayed on the PSLM.

Light Path in Focal Surface Occlusion

Figure 4.1 (a) shows the light path from the occlusion mask to the retina on the eye. The *negative* light ray from the occlusion mask is refracted by the PSLM and Lens 1. This refracted light reaches the eye (green arrow). If the PSLM is ignored, Lens 1 produces an intermediate occlusion image at z'_m between the occlusion mask and PSLM (marked in blue). Furthermore, if Lens 1 is ignored, the virtual occlusion mask is perceived at distance z_v (marked in red).

The relationship between the light path of the virtual object and real environment is clarified considering Figs. 4.1 (b–d).

Without the PSLM, the real-world light from infinity is focused at the position of the occlusion mask (Fig. 4.1, b). In other words, in this optical design, the light from the real world is sharply occluded when the eye accommodation is at infinity. As the focal depth of the eye approaches the front, the focus position between the lenses moves to the intermediate z'_m (Fig. 4.1, c), and thus, the occlusion mask appears to be blurred (blue arrow). As shown in Fig. 4.1 (d), assuming that the PSLM is a thin concave lens, the PSLM adjusts the focal length $f_p < 0$ to ensure that the focus lies on the occlusion mask (green arrow). Therefore, this system achieves a varifocal occlusion that can present a single sharp occlusion mask depending on the eye accommodation. The real view is focused at infinity at this time, which may be unnatural; however, this light is occluded and cannot be perceived (dotted green arrow).

Formulating the Focal length of the PSLM

We formulate the relationship between z_v and f_p. First, we introduce a ray transfer matrix. When a light ray propagates from a certain point and reaches a point at a distance d on the

optical axis from the initial position, the ray propagation matrix $\mathbf{P}(d)$ is

$$\begin{bmatrix} x' \\ u' \end{bmatrix} = \mathbf{P}(d) \begin{bmatrix} x \\ u \end{bmatrix}, \quad \mathbf{P}(d) \coloneqq \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix}. \tag{4.3.1}$$

where x is the distance between the point and the axis, and u is the ray angle from the axis. Herein, the propagation axis is defined as the optical axis z of the eye lens. Similarly, when the rays pass through a thin lens with focal length f, the lens matrix can be defined as

$$\begin{bmatrix} x' \\ u' \end{bmatrix} = \mathbf{L}(f) \begin{bmatrix} x \\ u \end{bmatrix}, \quad \mathbf{L}(f) \coloneqq \begin{bmatrix} 1 & 0 \\ -1/f & 1 \end{bmatrix}. \tag{4.3.2}$$

According to Fig. 4.1, the ray transfer matrix $\mathbf{M} \coloneqq [m_{ij}]$ from the retina to Lens 1 can be expressed as follows:

$$\mathbf{M} = \mathbf{P}(z_E)\mathbf{L}(f_E)\mathbf{P}(z_l)\mathbf{L}(f_l) \tag{4.3.3}$$

$$= \begin{bmatrix} \left(1 - \dfrac{z_E}{f_E}\right)\left(1 - \dfrac{z_l}{f_l}\right) - \dfrac{z_E}{f_l} & \left(1 - \dfrac{z_E}{f_E}\right) z_l + z_E \\ -\left(1 - \dfrac{z_l}{f_l}\right)\dfrac{1}{f_E} - \dfrac{1}{f_l} & -\dfrac{z_l}{f_E} + 1 \end{bmatrix} \tag{4.3.4}$$

Assuming that the eye lens is a thin lens, the eye focal depth z_v is

$$z_v^{-1} = f_E^{-1} - z_E^{-1}. \tag{4.3.5}$$

Without the PSLM, the intermediate occlusion position z'_m can be formulated using $\mathbf{M}$ for $\forall u \in \mathbb{R}$ (Fig. 4.1, c):

$$\begin{aligned}
\mathbf{MP}(z'_m - z_l) \begin{bmatrix} 0 \\ u \end{bmatrix} &= \begin{bmatrix} 0 \\ u' \end{bmatrix} \\
\Leftrightarrow \quad z'_m &= z_l - \frac{m_{12}}{m_{11}} \\
&= \frac{z_l^2 - z_l z_v - f_l z_v}{z_l - f_l - z_v} \quad (\because \text{Eq. } (4.3.5)).
\end{aligned} \tag{4.3.6}$$

Similarly, the focal length of the PSLM f_p can be calculated as follows (Fig. 4.1, d):

$$\begin{aligned}
&\mathbf{MP}(z_p - z_l)\mathbf{L}(f_p)\mathbf{P}(z_m - z_p) \begin{bmatrix} 0 \\ u \end{bmatrix} = \begin{bmatrix} 0 \\ u' \end{bmatrix} \\
&\Leftrightarrow f_p = \frac{\{m_{11}(z_p - z_l) + m_{12}\}(z_m - z_p)}{m_{11} f_l + m_{12}}.
\end{aligned} \tag{4.3.7}$$

We denote $z_m - z_p = r f_l$ $(0 < r < 1)$. In this case, $z_p - z_l = (1 - r) f_l$. Substituting these values and Eq. (4.3.5) into Eq. (4.3.7) yields

$$\begin{aligned}
f_p &= \frac{\{m_{11}(1 - r) f_l + m_{12}\} r f_l}{m_{11} f_l + m_{12}} \\
&= r\{f_l - r(f_l + z_v - z_l)\}.
\end{aligned} \tag{4.3.8}$$

Note that the focal length of the PSLM f_p is bounded as $|f_p| \geq 2\delta_p r_p / \lambda$ [114]. This constraint depends on the hardware of the PSLM. Hence, by varying parameters such as the PSLM position z_p, the range of z_v that can be modulated can be varied.

4.3.3 Synthesizing Focal Surfaces with the PSLM

In the previous section, we assumed that f_p is constant and formulated the relationship between f_p and the eye accommodation z_v. In practice, the focal length of the PSLM is

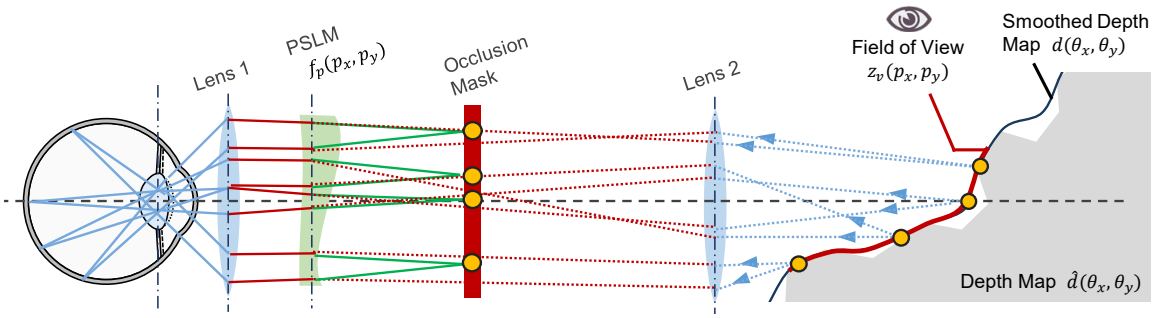

Figure 4.2: Optimization of PSLM as a free-form lens for the given depth map. While focusing the light rays onto the occlusion mask (green arrows), the phase image $\phi(p_x, p_y)$ on the PSLM is optimized to ensure that the light is perceived to be emanating from the smoothed virtual depth map when the PSLM is ignored (dotted blue arrows).

spatially varied, $f_p(p_x, p_y)$, to ensure that the depth of the virtual object perceived from the eye, $z_v(p_x, p_y)$, approaches an arbitrary input depth map $\hat{d}(\theta_x, \theta_y)$ (Fig. 4.2). In this section, we summarize [61] the process of optimizing the free-form lens from a given depth map.

Approximating Depth Maps with Focal Surfaces

Before optimizing the phase map, we approximate the input depth map $\hat{d}(\theta_x, \theta_y)$ to a smooth focal surface $d(\theta_x, \theta_y)$. In VR-HMDs, multiple focal surfaces can be generated in a time-multiplexing manner. However, because the OST-HMD occludes light from the real environment, time multiplexing can cause the occlusion to appear semi-transparent, thereby also leading to an unwanted flicker in the see-through view. Thus, our system approximates the depth map to a single focal surface. If the PSLM can reproduce an arbitrary topology, the depth map does not need to be optimized. However, the desired depth should be smooth and continuous because the PSLM represents the lens as a locally quadratic function. Thus, we optimize the single focal surface $d(\theta_x, \theta_y)$ using the objective function that adds a smoothing term to the input depth $\hat{d}(\theta_x, \theta_y)$:

$$\min_{d} \sum_{(\theta_x, \theta_y) \in \Omega_\theta} \|\hat{d}(\theta_x, \theta_y) - d(\theta_x, \theta_y)\|^2 + \gamma \sum_{(\theta_x, \theta_y)} \|\partial^2 d(\theta_x, \theta_y)\|^2, \tag{4.3.9}$$

where $\partial^2 d(\theta_x, \theta_y) = \left[\frac{\partial^2 d}{\partial x^2}, \frac{\partial^2 d}{\partial y^2}\right]^\top$ is the vector of the second partial derivatives of $d(\theta_x, \theta_y)$, and γ is a weighting parameter.

Synthesizing Focal Surfaces with PSLMs

Next, we optimize the phase image $\phi(p_x, p_y)$ from a smoothed focal surface $d(\theta_x, \theta_y)$. As the FoV of the system is constrained by the PSLM dimension, we truncate the given depth map in the range in which the viewing angle Ω_θ intersects the PSLM dimension, i.e. $d(\theta_x, \theta_y) = z_v(p_x, p_y)$. Subsequently, we calculate the focal length of each pixel on the PSLM $f_p(p_x, p_y)$ from the given $z_v(p_x, p_y)$ by using Eq. (4.3.8).

We consider that the PSLM functions as a local thin lens in the neighborhood of each pixel. In other words, the focal length $f_p(p_x, p_y)$ of the ϵ-sized bundle of rays passing through the pixel (p_x, p_y) is determined by the curvature of the phase function in the neighborhood of the pixel. This property can be expressed with the Hessian of the phase function $\mathbf{H}_\phi$ [114]:

$$\mathbf{H}_\phi(p_x, p_y) = -\frac{2\pi}{\lambda f_p(p_x, p_y)}\mathbf{I}. \tag{4.3.10}$$

In practice, the values of ϕ that satisfy the above equation are limited to those for which $f_p(p_x, p_y)$ is constant and $\phi(p_x, p_y)$ is quadratic, i.e. the PSLM acts as a uniform lens. Instead, we obtain $\phi(p_x, p_y)$ with a Hessian as close as possible to Eq. (4.3.10) by solving the following linear least-squares problem:

$$\min_{\phi} \sum_{(p_x, p_y)} \left\| \hat{\mathbf{H}}_{[\phi]}(p_x, p_y) - \frac{-2\pi}{\lambda f_p(p_x, p_y)}\mathbf{I} \right\|_F^2, \tag{4.3.11}$$

where $\| \cdot \|_F^2$ is the Frobenius norm, and $\hat{\mathbf{H}}[\cdot]$ is the discrete Hessian operator, specified by the finite differences of ϕ. Because the sum of ϕ and any linear function $a + bx + cy$ has the same Hessian $\mathbf{H}$, we add two more constraints: $\phi(0, 0) = 0$ and $\nabla\phi(0, 0) = 0$.

4.3.4 Edge-Based Optimization for Minimizing the Distortion of the See-Through View

In the existing study pertaining to VR-HMDs [61], Eq. (4.3.11) was optimized considering all pixels in the entire FoV to achieve a focal surface. However, when reproducing the focal surface on an OCOST-HMD, it is necessary to consider the visibility of the see-through view and virtual geometry. As confirmed in the experiments described later, a simple optimization based on all the pixels of the virtual depth map by using Eq. (4.3.11) may result in over-fitting of the intense changes in the virtual depth, resulting in large gradient changes in the see-through view. To minimize the distortion of the see-through view, it is desirable to minimize the effect of the PSLM in the area without virtual objects.

To alleviate this degradation, we perform an edge-based focal surface optimization that exploits the properties of occlusion masks. When SLMs that display occlusion masks in binary (black or transparent) are used, the observer cannot perceive the depth inside the black mask, and depth cues are perceived only by blurring at the mask edges. Using this occlusion depth perception, the distortion of the see-through view is reduced as the phase image is placed as close to the target focal length as possible only for pixels around the edge of the mask, and the gradient change of the phase image for other pixels is suppressed as much as possible. These conditions can be expressed as the following optimization problem:

$$\min_{\phi} \left\{ \sum_{(p_x,p_y)\in E} \left\| \hat{\mathbf{H}}_{[\phi]}(p_x,p_y) - \frac{-2\pi}{\lambda f_p(p_x,p_y)}\mathbf{I} \right\|_F^2 + \sum_{(p_x,p_y)\in \bar{E}} \|\partial\phi(p_x,p_y)\|^2 \right\}, \quad (4.3.12)$$

where $\partial\phi(p_x,p_y) = \left[\frac{\partial\phi}{\partial x}, \frac{\partial\phi}{\partial y}\right]^\top$, and E represents a set of edge pixels detected using the Canny edge detector [115] and dilated with a kernel of $N_e \times N_e$. In our implementation, we manually set $N_e = 21$.

Because the Hessian and partial derivatives for Eq. (4.3.12) are discrete, the expression can be rewritten as a least-squares problem. We denote $\mathbf{x}$ as the vectorized phase image $\phi(p_x,p_y)$, and $\mathbf{A}, \mathbf{b}$ as a sparse constant matrix and a vector. If a PSLM pixel (u,v) located at an edge on the depth map, i.e. $(u,v) \in E$, three rows of 2nd-order differential constraints

corresponding to $\hat{\mathbf{H}}_{[\phi]}$ are added to $\mathbf{A}, \mathbf{b}$. Similarly, if $(u', v') \in \bar{E}$, two rows of 1st-order differential constraints are added to $\mathbf{A}, \mathbf{b}$. These constraints can be expressed as follows:

$$\overbrace{\begin{bmatrix} & & & & & & & & & & & & & & \\ & & & & & \vdots & & & & & & & & & \\ \cdots & 0 & 0 & 0 & 1 & -2 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots \\ \cdots & 0 & 1 & 0 & 0 & -2 & 0 & 0 & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots \\ \cdots & -1 & 1 & 0 & 1 & -2 & 1 & 0 & 1 & -1 & \cdots & 0 & 0 & 0 & \cdots \\ & & & & & & \vdots & & & & & & & & \\ \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & -1 & 1 & 0 & \cdots \\ \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & -1 & 0 & 1 & \cdots \\ & & & & & & \vdots & & & & & & & & \end{bmatrix}}^{\mathbf{A}} \overbrace{\begin{bmatrix} \vdots \\ \phi(u-1, v-1) \\ \phi(u, v-1) \\ \phi(u+1, v-1) \\ \phi(u-1, v) \\ \phi(u, v) \\ \phi(u+1, v) \\ \phi(u-1, v+1) \\ \phi(u, v+1) \\ \phi(u+1, v+1) \\ \vdots \\ \phi(u', v') \\ \phi(u'+1, v') \\ \phi(u', v'+1) \\ \vdots \end{bmatrix}}^{\mathbf{x}} = \overbrace{\begin{bmatrix} \vdots \\ \frac{-2\pi}{\lambda f_p(u, v)} \\ \frac{-2\pi}{\lambda f_p(u, v)} \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \end{bmatrix}}^{\mathbf{b}} \tag{4.3.13}$$

The solution of Eq. (4.3.13) equals to the solution of a large linear equation:

$$\mathbf{A}^\top \mathbf{A}\mathbf{x} = 2\mathbf{A}^\top \mathbf{b}. \tag{4.3.14}$$

4.4 Designing Focal Surface Occlusion

Similar to the focal surface display on VR-HMDs, in OST-HMD, there are trade-offs in FoV, the depth of field (DoF, the supported accommodation range), and optical aberration caused by the PSLM. In this section, we discuss these trade-offs and give suggestions for the display design.

The element with the smallest dimension limits the FoV of the entire optical system.

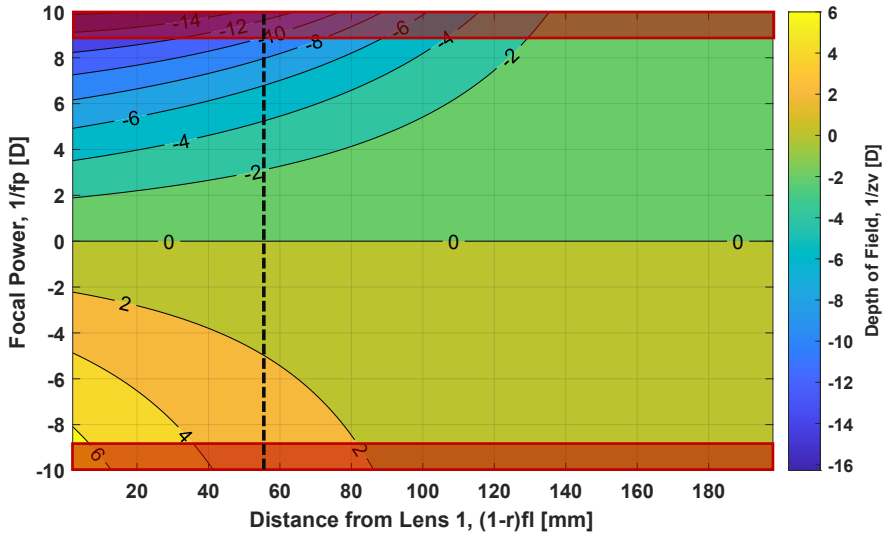

Figure 4.3: The achievable DoF of our focal surface occlusion. The labeled plot contours and the fill color represent the virtual image distances z_v^{-1} [D] achieved with the PSLM, when used to represent a lens of focal length f_p and distance between Lens 1 and PSLM $(1-r)f_l$. Red areas indicate focal lengths beyond the dynamic range of the SLM, and the vertical black dashed line is the distance between the Lens 1 and the PSLM as set by our hardware (Sec. 4.5.2).

In the current case, since the dimension of the PSLM constrains the FoV, the FoV is given by $2\arctan\frac{r_p}{z_p - z_l}$. To increase the FoV, we need to move the PSLM as close as possible to Lens 1.

In addition, the closer the PSLM is to the eye, the more extensive DoF the PSLM can support. Figure 4.3 shows the relationship between the distance between the eye and the PSLM $z_p - z_l$, the focal power of the PSLM f_p, and the supported range of z_v, derived from Eqs. (4.3.4) and (4.3.8). Following in Sec. 4.3.2, the range of f_p that the PSLM can modulate is constrained by $|f_p| \geq 2\delta_p r_p/\lambda$. Given this constraint (the red box in Fig. 4.3), the modulatable DoF range becomes wider when the PSLM is closer to the Lens 1.

On the other hand, to reduce the optical aberration of the PSLM and increase the resolution of the occlusion mask, it is desirable to concentrate the ray bundle refracted by the PSLM to a single point on the occlusion mask, i.e., to reduce the circle of confusion. The circle of confusion can be reduced by making the phase image on the PSLM more quadratic or moving the occlusion mask closer to the PSLM.

In the implementation of the focal surface occlusion, it is necessary to determine the position of the PSLM while considering the trade-off between FoV, DoF, and optical aberration. In our prototype, we designed the PSLM to be as close to the lens as possible, because we use a binary occlusion mask in the prototype. Using the binary mask, when aberrations occur in the mask, the blur inside the occlusion is filled with black; thus, we observe aberrations only at the periphery of the occlusion. Based on this, we prioritize FoV and the modulatable DoF.

In our prototype, the position where we place the PSLM is limited by the volume of the beam splitter. Also, the light ray is folded by the beam splitter to maintain a see-through view, which increases the focal length f_l of the lens and limits the DoF range. We later discuss the prospects for miniaturization of the form factor to improve all the trade-offs in Sec. 4.7.2. However, even with these constraints, our system can support a wide range of DoF and achieve continuous focal cues.

4.5 Implementation

This section describes the proof-of-concept prototype of the focal surface occlusion.

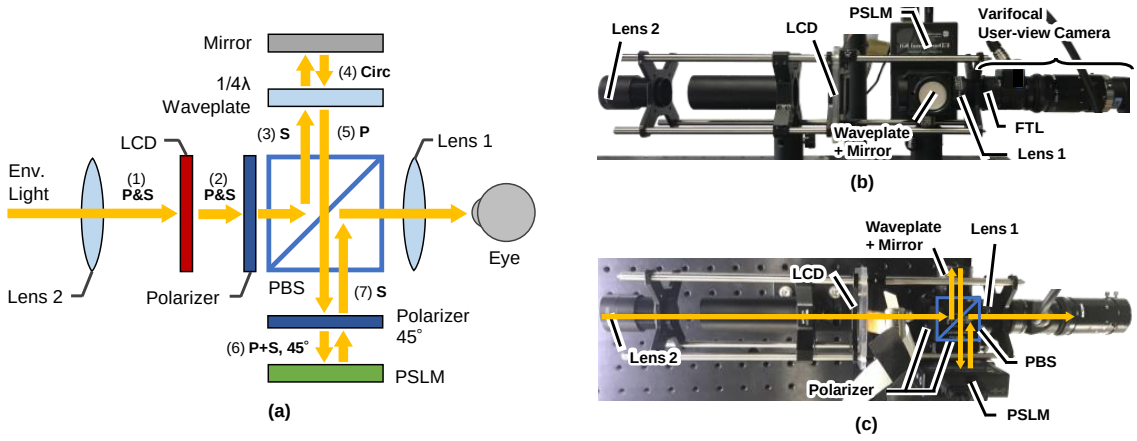

Figure 4.4: (a) Schematic of the optical design of the focal surface occlusion system. The numbers and P & S represent the incident path and polarization state of the light, respectively. (b) Top view and (c) front view of our proof-of-concept system. The schematic for the front view is identical to that shown in (a).

4.5.1 Optical Design

For the proposed system, an optical design that allows only the focal surface of the occlusion to be modulated by the PSLM while maintaining the see-through view must be used. Figure 4.4 (a) shows a schematic of the optical design and light paths passing through the system. The focal surface occlusion framework is designed based on a polarized beam splitter (PBS). The see-through phase modulation design is based on the PSLM [109], and we realize focal surface occlusion by placing a transmissive LCD in front of this design, sandwiching the ends with lenses having the same focal length. In this design, the light entering the eye is upside down. We invert the captured images in sequential experiments. Extending the system to $4f$ optics can prevent inversion.

4.5.2 Hardware Setup

Figure 4.4 (b) shows our proof-of-concept prototype. We implemented a benchtop lab prototype using commercially available optical elements.

The PSLM was a Jasper Display EDK 9554 A+ (1920 $\times$ 1080 pixels, 60 Hz). The pixel pitch of the PSLM $\delta_p = 6.4$ μm, and its active area is 12.5 mm $\times$ 7.1 mm, i.e. the diagonal radius $r_p = 7.0$ mm. The focal length that can be modulated by the PSLM is bounded as $|f_p| \geq 2\delta_p r_p/\lambda = 0.236$ m $\sim$ 0.115 m ($\lambda = 380$ nm $\sim$ 780 nm).

The transmissive LCD was a SONY LCD 0178 (1024 $\times$ 768 pixels, 60 Hz). This LCD was connected to the computer via the controller made by bbs Bild- und Lichtsysteme. Certain fixtures for the LCD were prepared using a 3D printer.

The other optical components were obtained from Thorlabs: LPVISE100-A linear polarizers, AQWP10M-580 achromatic $\lambda/4$ waveplate, PBS251 PBS, BB1-E02 mirror, and AC254-200-A lenses (Lenses 1 and 2).

To obtain images from a viewpoint through the OST-HMD, we installed a user-perspective camera, specifically, a C-mount camera, FLIR BFS-U3-122S6C (4096 $\times$ 3000 pixel), instead of a real eye. To capture the images, we mounted a C-mount lens (RICOH FL-CC7528-2M) on the user-perspective camera, and the focal length of the C-mount lens was fixed at infinity. In addition, an FTL (Optotune EL-10-30-Ci-VIS-LD-MV) was placed in front of the C-mount lens to simulate the eye accommodation changes and capture images

by varying the focal length z_v^{-1} in the range from -1.5 D (m^{-1}) to 3.5 D.

In this hardware configuration, the distance of the optical components according to the manufacturer specifications was as follows: $z_l = 38.6$ mm, $z_p = 93.2$ mm, $z_m = 238.6$ mm, and $f_l = 200.0$ mm. Given these parameters and the constraint on f_p, the DoF that could be achieved, according to Eq. (4.3.8), was $z_v^{-1} = -2.88$ D ~ 1.83 D for $\lambda = 380$ nm and $z_v^{-1} = -8.46$ D ~ 3.15 D for $\lambda = 780$ nm. The FoV of our prototype was 14.6 °.

4.5.3 Software Setup

Optimization To obtain a smooth depth map, we solved Eq. (4.3.9) by assuming the circulant boundary condition and using the fast Fourier transform, which is a well-known technique for image deblurring with a unique point spread function [116]. The average run time was 0.23 s. Next, we evaluated the phase images by solving Eq. (4.3.12). The framework was implemented via a sparse linear solver using the LDL factorization in MATLAB, with an average run time of 3.4 s.

Calibration between the LCD and PSLM Pixel-wise calibration was required to be performed between the PSLM and LCD. We obtained the relationship via the viewpoint camera. In the calibration process, we switched on a surface light source (Viltrox L116T) to provide a white background in front of the system.

First, we determined the pixel-wise correspondence between the LCD and user-perspective camera by displaying a gray code pattern on the LCD. Next, we obtained the correspondence between the PSLM and user-perspective camera. Notably, an accurate calibration between these entities has not yet been reported, which is a key reason for the lack of satisfactory results yielded by the current prototypes, as discussed in Section 4.7.1. In this study, we manually calibrated the position of the PSLM; after increasing the exposure of the user-perspective camera to capture the upper and lower frames of the PSLM, we specifies the area of the PSLM to ensure an aspect ratio of 16:9.

4.6 Experiments

Using the proof-of-concept prototype, we evaluated the performance of the focal surface occlusion scheme.

4.6.1 Comparison of Focal Blur Reproduction

First, we compared our focal surface occlusion with a fixed-focus OCOST-HMD and simulated the focal blur on edges in an object with continuous depth.

Experiment Setup

A virtual object in the form of a continuous bar from the back to the front was used. To evaluate the results, three edges of interest were considered: the edge of the real scene that exists at infinity (Far, 0D), edge of the virtual object that exists at the back the viewpoint (Middle, 0.25D), and the corresponding edge at the front side (Near, 0.75D), as shown in Fig. 4.5 (a).

After obtaining a target depth map from the object, we generated phase images by performing optimization based on the entire depth map (Eq. (4.3.11)) and edge-based optimization (Eq. (4.3.12)). These optimized phase images were displayed on the PSLM. In the fixed-focus setup, we displayed a black image without a gradient, i.e., the occlusion mask was located as infinity. We binarized the input depth map to create an occlusion image. This occlusion image was transformed from the PSLM coordinates to the LCD coordinates on the pixel-wise map obtained as described in Section 4.5.3.

We focused on each region of interest by changing the focal length of the FTL and captured the occlusion mask with the user-perspective camera. The F-number of the camera was set as 4 to approximate the human eye, the F-number for which ranges between 2.8 and 8.

We captured the occlusion mask in two environments: one with a see-through view, in which the edge of the building at infinity was the edge of interest at 0 D (Far), and the other including a surface light, as described in Section 4.5.3. These settings were used to eliminate the influence of the background and evaluate only the blur of the edge of interest

of the virtual object in the Middle and Near configurations. Additionally, the prototype corresponded to the same virtual object and red plane placed at infinity on Unity 5 by using the same focal length and F-number, and we treated the rendered images as the simulation results.

Results

Figure 4.5 (b) shows the results of the experiment. The fixed-focus setup (column 1) focuses on all edges of the virtual object at infinity and is blurred as the accommodation moves closer. Moreover, the phase image with full optimization (column 2) overfits the depth of the Near edge and is blurry as the accommodation moves farther. In contrast, the edge-based optimization (column 3) can reproduce the focus blur at the edges of the virtual object according to the accommodation as well as the simulation (column 4). Moreover, the edge-based optimization does not significantly degrade the see-through view, and the see-through view blurs as the accommodation changes, as in the fixed-focus case.

4.6.2 Resolution Analysis for Edge Profiles

By quantitatively analyzing the edge profiles obtained as described in the previous section, we evaluated the effectiveness and reproducible spatial resolution of the proposed method. The edge modulation transfer function (MTF) [117] analysis was performed for the edge profiles in the Middle and Near configurations against a surface light source background, obtained as described in the previous section.

Results

Comparison with Fixed-Focus Setup and Simulation. Figure 4.6 shows the all normalized MTF curves. To compare the spatial resolution among all experiment conditions, we calculated the spatial resolution that can be represented with a half contrast, which results in Figure 4.7.

In our simulation (Fig. 4.7, d), the spatial resolutions of the Middle edge decrease in the order of the camera focus: Middle (42.4 cycles per degree, cpd) $>$ Far (23.2 cpd) $>$ Near (13.1 cpd). Moreover, the spatial resolutions of the Near edge decrease in the order of the

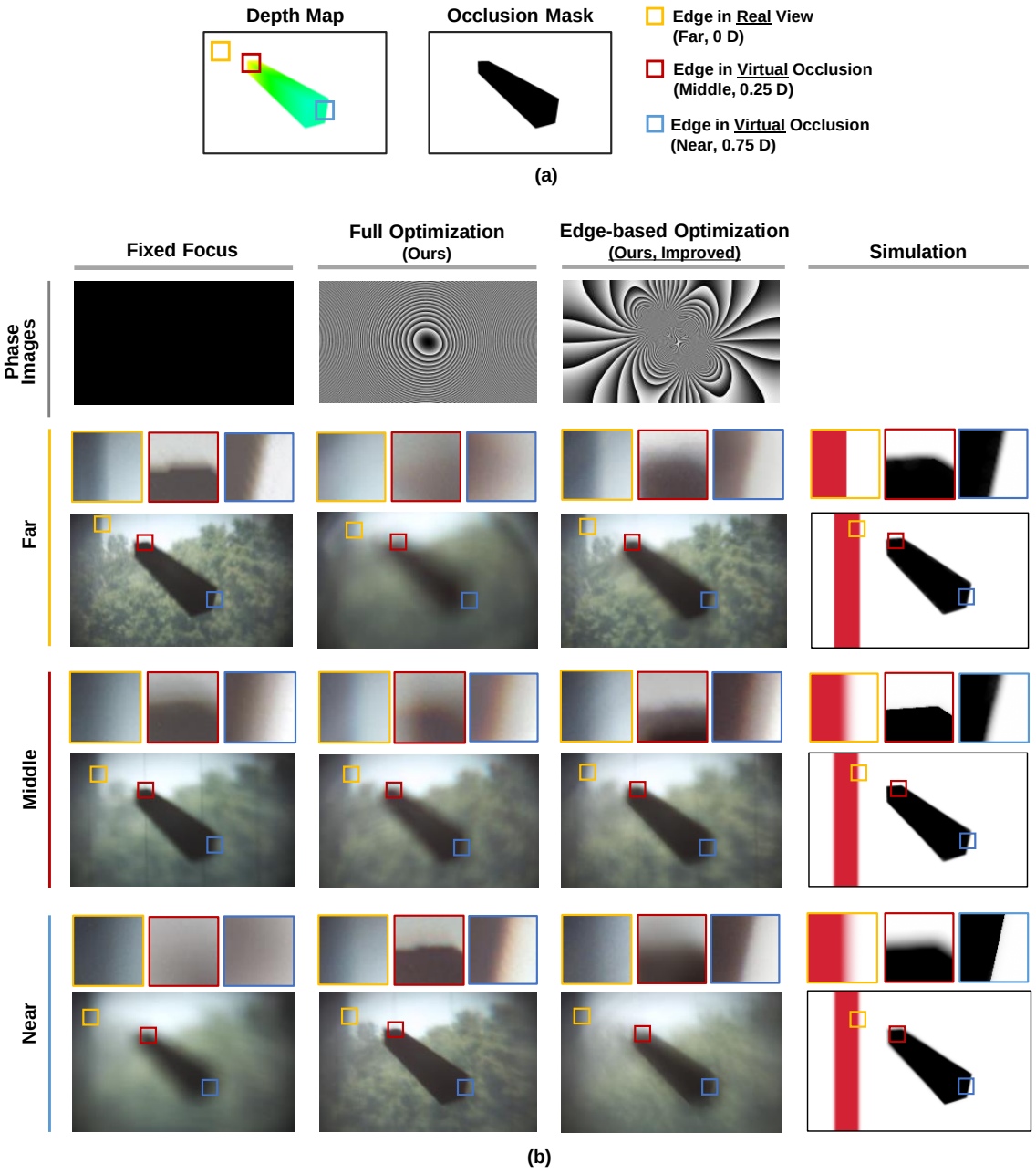

Figure 4.5: (a) Depth map and occlusion mask of the virtual object used in the comparison experiment. The yellow, red, and blue boxes indicate the edges of interest in the Far (0 D), Middle (0.25 D), and Near (0.75 D) configurations, respectively. (b) Comparison of the focal blur reproduced by each method for a virtual occlusion at continuous depth. (row 1) Phase images obtained using each method. (rows 2–4) Images captured using the user-perspective camera with the FTL set to Far, Middle, and Near. In addition to the overall view, we present a detailed profile of the edge of interest for each patch. Among the edges displayed in each patch, Far corresponds to the edge of a real object at infinity, captured with a see-through view, whereas Middle and Near correspond to the edges of virtual objects, captured with a surface light source to remove the effect of the background.

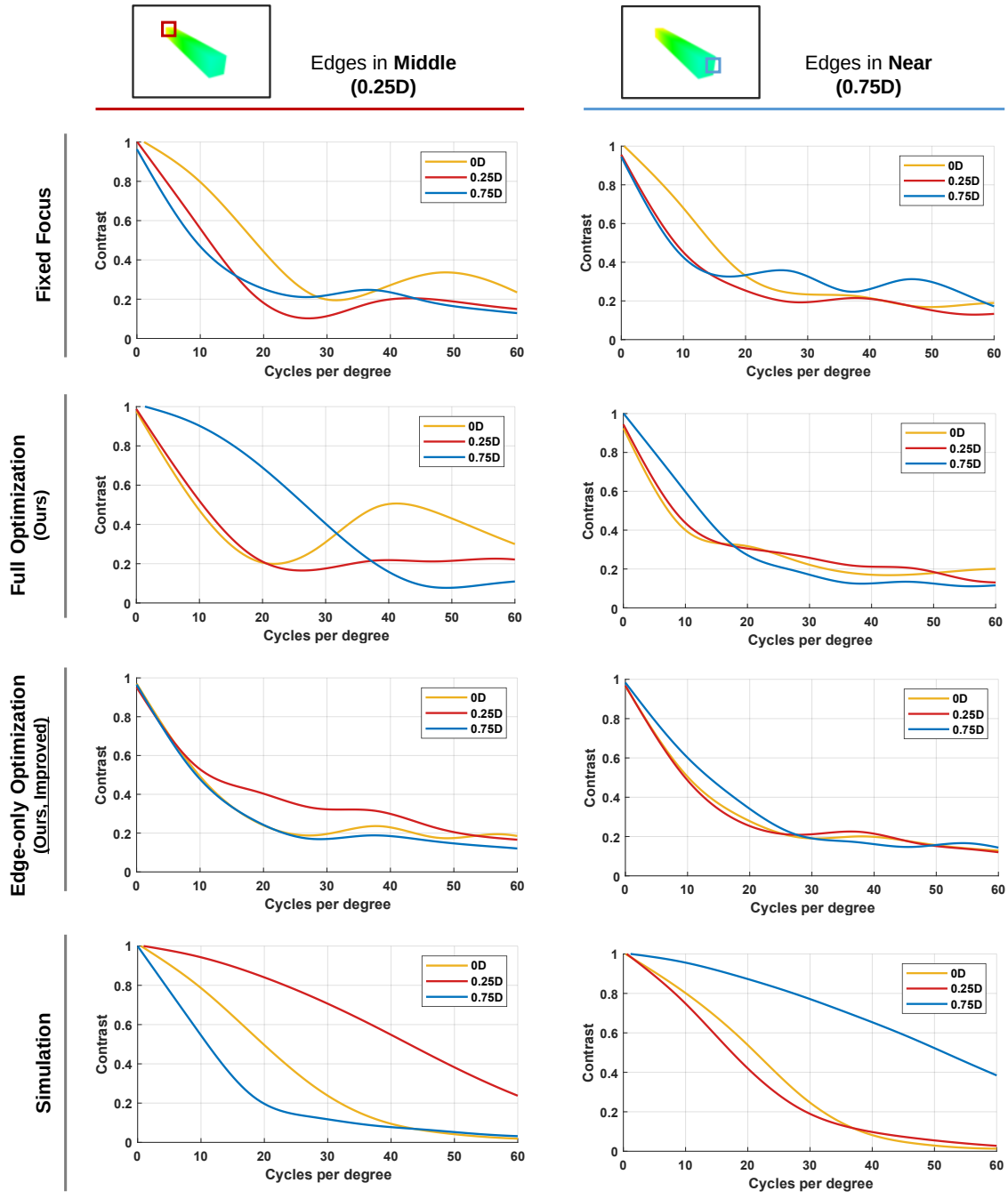

Figure 4.6: Normalized MTF curves in all experiment conditions in Sec. 4.6.2. In this graph, the original MTFs are interpolated by 3rd-order smoothing spline curves.

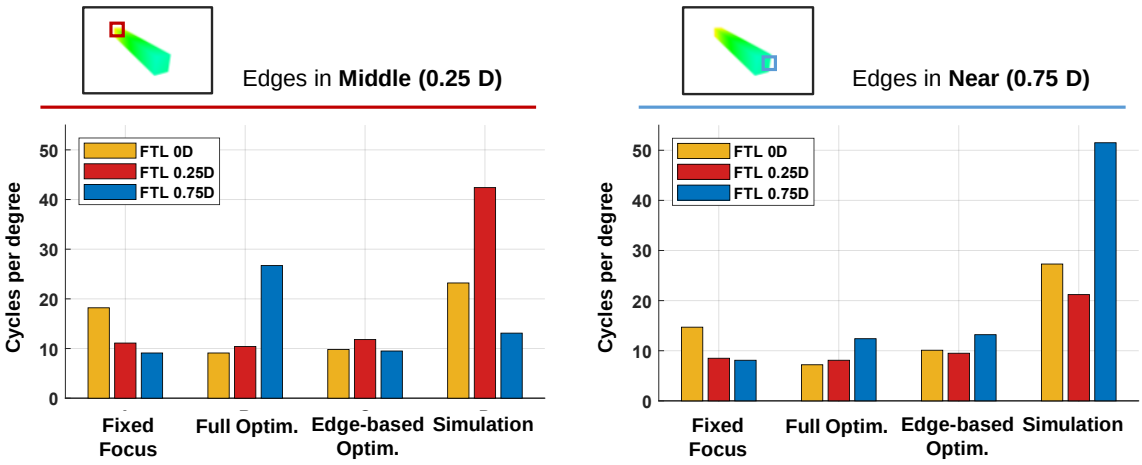

Figure 4.7: Spatial resolution at which half-contrast can be reproduced in the edge profile of each method, captured with the FTL set as Far, Middle, and Near. We used the edge profile shown in the patches in Fig. 4.5.

camera focus: Near (51.5 cpd) > Far (27.3 cpd) > Middle (21.2 cpd). As mentioned in the previous section, for the fixed-focus setup, all the spatial resolutions decrease in the order Far > Middle > Near, regardless of the edge depth. Similarly, for the full-optimization setup, all the spatial resolutions decrease in the order Near > Middle > Far. In contrast, in the case of the edge-based optimization, the spatial resolution decreases in the same order as in the simulation: The spatial resolutions of the Middle edge decrease in the order Middle (11.8 cpd) > Far (9.8 cpd) > Near (9.5 cpd), and those of the Near edge decrease in the order Near (13.2 cpd) > Far (10.1 cpd) > Middle (9.5 cpd). Hence, it can be quantitatively confirmed that the edge-based optimization can reproduce the focal blur of the edges according to eye accommodation.

Although our proof-of-concept system can provide a focal blur that is more in accordance with the accommodation compared with that attained using fixed and varifocal OCOST-HMDs, sufficiently sharp edges are not obtained compared to those in the simulations. This phenomenon can likely be attributed to the insufficient calibration between the PSLM and occlusion mask. Section 4.7.1 further discusses the causes for this phenomenon and possible remedies.

Comparison with Consumer-Level HMDs. The 20/20 vision in humans corresponds to 30 cpd, and the resolution of current consumer-level near-eye displays is typically approximately $5 \sim 15$ cpd. Our edge-based optimization scheme achieves a resolution of $11 \sim 14$ cpd at the sharpest edges. Thus, the proof-of-concept prototype with edge-based optimization can reproduce sharp edges comparable to those of consumer HMDs even with phase modulation according to the content.

Comparison with Light-Field Occlusions. In principle, the proposed focal surface occlusion scheme can yield a 3D occlusion mask with an improved FoV and resolution to light-field occlusion with a micro-structure phase modulator. The FoV of the proof-of-concept prototype is 14.6°, which is larger than that (4.7°) of the existing light field occlusion scheme [98]. Because the diagonal length of the PSLM limits the FoV of the current prototype, the FoV can be expanded by using a wider-angle reflective PSLM or transmissive PSLM (Section 4.7.2).

In the lens-array-based light field occlusion scheme [98], a single display is spatially divided according to the number of lens arrays, resulting in a resolution that is typically one order of magnitude lower than the original display resolution. Furthermore, the pixel pitch cannot be reduced to less than the diffraction limit imposed by individual lenses. The stacked display [96] can alleviate this limitation, with values of 3.9 cpd reported in simulations, although the value of the consumer level cannot be reached,. In contrast, although the focal cues are not as perfect as those of the light field, our focal surface occlusion scheme achieves a high spatial resolution.

4.6.3 Focal Blur Reproduction in Various Scenes

We set a target for multiple objects and a complex and natural scene and attempted to reproduced the focal blur by using the phase image generated by edge-based optimization. We captured the focal blur under the experimental settings described in Section 4.6.1, except for several minor changes described in the following text.

Spheres. A scene in which two virtual spheres are placed in a line (front, 0.25 D and back, 0.75 D) was considered. Because of the misalignment between the PSLM and LCD, as described in Section 4.7.1, it was difficult to precisely align the phase image with the

edge of the mask. Thus, we used the optimized phase image with $N_e = 51$ for only this scene.

Figure 4.8 (a) shows the results. The edges of each sphere are sharpened when the ideal depth and accommodation match, although a large N_e slightly reduces the sharpness. Moreover, the see-through view between the objects tends to be blurred owing to the increased phase image folding between the multiple objects at different depths.

City. A complex and challenging scene, a 3D city model, was examined to evaluate the future practical use of the proposed method. Depending on the scene geometry, we selected 0.1 D, 0.2 D, and 0.5 D edges on the virtual geometry as the regions of interest.

Figure 4.8 (b) shows the results. The proposed edge-based optimization ensures that the see-through view is visible even in complex scenes. In addition, the Far edge profile changes according to the accommodation. The Near edge is not considerably affected because an edge is located at another depth around the edge of interest. Moreover, the focus blur at the edge of the bridge is overfit relative to the surrounding depth.

These results show that edge-based optimization is not entirely effective in scenes in which edges of different depths are close to one another. Alleviating the misalignment between the PSLM and LCD is the first step in overcoming this problem. Adding a penalty term to Eq. (4.3.12), which strongly affects cases involving many edges in a certain region, may also enhance the results.

4.7 Limitation and Future Work

Although our proof-of-concept prototype of focal surface occlusion can accurately reproduce the focus blur in natural scenes, it has not yet reached the level necessary for a practical OCOST-HMD. In this section, we discuss the possible limitations of our prototype and highlight the scope for future work.

4.7.1 Pixel-Wise Calibration between PSLMs and Devices

To achieve higher quality 3D occlusion, spatial calibration that fully accounts for the misalignment of the LCD and PSLM optical planes, including affine (scaling and translation)

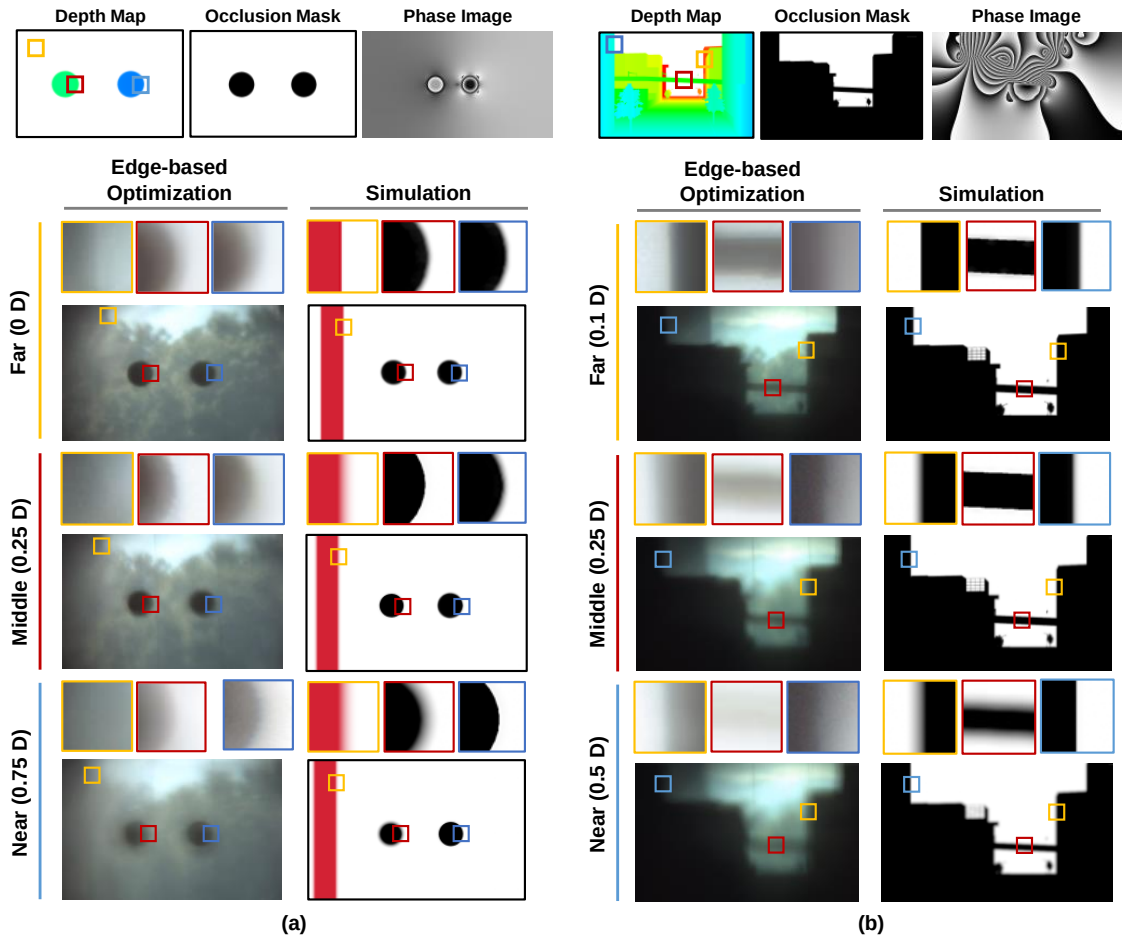

Figure 4.8: Examples of applying phase images with edge-based optimization to (a) multiple objects (spheres) and (b) a natural, complex scene (city). The images and notations in the figure are the same as those in Fig. 4.5.

and non-affine transformations, must be realized. In our prototype, owing to the misalignment between the PSLM and LCD, the focus distribution may not be perfect at the edges, and several details may not be adequately sharp. In addition, owing to the misalignment of the scale, the focal surface is not aligned with the actual depth, which makes the depth-of-field appear narrow.

To the best of our knowledge, no method has been reported to realize pixel-wise calibration between PSLMs and displays including an occlusion mask. A possible method to ensure the calibration between a viewpoint camera and PSLM is to place the resolution evaluation chart in the real world, display the phase pattern on the PSLM to blur the chart, estimate a blur map, and match the blur map to the phase pattern. Because the study of specific calibration methods is beyond the scope of this research, it must be addressed in future research.

4.7.2 Miniaturization of Form Factor

The current prototype resembles a telescope and has a large form factor. The optical path is folded three times through the PBS to ensure modulation with the reflective LCoS while maintaining a see-through view, resulting in an extremely long distance between Lens 1 and the occlusion mask $z_m - z_l$. Owing to the characteristics of the afocal optics, the distance between the mask and Lens 2 must be equal to this folded optical path length (i.e. $z_l' - z_m = z_m - z_l = f_l$), leading to the telescope-like appearance of the prototype. Furthermore, because the light passes through the 45° polarizer twice, the final brightness decreases to 1/8 of the original brightness.

A straightforward approach to solve these issues is to adopt a transmissive LCoS as a PSLM. Unfortunately, the current transmissive LCoS has a large pixel pitch δ_p, which causes diffraction artifacts of the see-through view when inserted directly in front of the eye. Noably, to achieve a wide FoV through this approach, the diagonal radius r_p of the PSLM must be increased. Increasing δ_p and r_p significantly reduces the range of focal lengths that the PSLM can present as $|f_p| \geq 2\delta_p r_p/\lambda$. For reference, in HOLOEYE LC2012 (1024 $\times$ 768 pixel, δ_p = 36 μm and $r_p = 23.0$ mm), a transmissive PSLM currently commercially available, $|f_p| > 3.12$ m ($= 0.32$ D) at $\lambda = 532$ nm. The adoption of transmissive

LCoS thus relies on future hardware improvements.

4.7.3 Real-Time Phase Image Generation

The optimization calculations for the prototype require approximately 3.4 s (Eq. (4.3.12)), which is not adequate to achieve the high frame rate required for real-time applications ($\sim$ 0.03 s). The first step to ensure a high frame rate calculation is to reduce the number of pixels used in the linear calculation. In particular, because non-edge pixels have an extremely small gradient change and account for a large portion of the image, our edge-based optimization method can significantly reduce the computation time by pre-complementing the gradient in the non-edge pixels. Another method to accelerate the computation is to convert the linear equations into a form that can be used for parallel computation or convert them into a form that the fast Fourier transform can solve, as in Eq. (4.3.9).

4.7.4 Combination with additive display

While our scope in this paper is reproducing the focal surface on the occlusion mask in OCOST-HMDs, it is important in AR-HMDs to present a virtual image in addition to the mask. To overlay the fixed-focus virtual images on the mask in our system, one can just insert a beam splitter in front of the eye. Furthermore, by introducing an additional PSLM, focal surfaces can be applied to virtual images as well.

Figure 4.9 shows a schematic diagram of the system when the current prototype is extended to a virtual image with the focal surface. To simplify the configuration in Fig. 4.9 (a), we assume the occlusion mask as a reflective spatial amplitude modulator (e.g. reflective LCoS or DMD). We can use Lens 2 as a common eyepiece by placing the occlusion mask and the virtual image display at the same optical depth, as shown in Fig. 4.9 (b).

Since the virtual image is introduced by PBS independently from the occluded light path of real scene, we can apply time-multiplexing phase modulation to virtual images [61], which we could not apply to occlusion masks due to the distortion of the see-through view (Sec. 4.3.3). Thus, the MTF of the virtual image is expected to be better than the MTF of the occlusion mask by increasing the number of time division for the multiplexing.

The addition of optical elements corresponds to increased complexity. While rendering

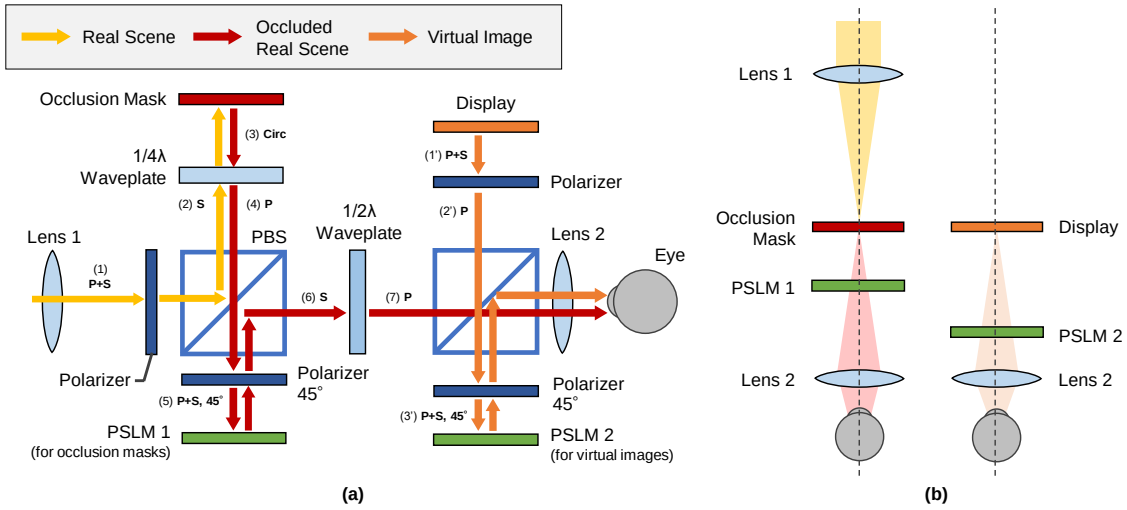

Figure 4.9: (a) Schematic of the optical design when adding a virtual image to the current prototype. (b) An optically equivalent schematics for (left) the occlusion mask and (right) the display. The optical depth of the mask and the display need to be matched.

the focal blur in software on the virtual image can reduce the hardware complexity, it is a trade-off with computational cost. The OCOST-HMD that used a single DMD for the occlusion and the virtual image display simultaneously [118] would be promising to reduce the complexity of optical systems. When applying focal surface to this single DMD configuration, a PSLM is inserted between the DMD and the illumination (e.g., RGB LED). Fortunately, in this configuration, the occlusion mask is also independent of the real optical path, which allows time-multiplexing phase modulation for the occlusion mask. However, since the refresh rate of PSLM is relatively low compared to the control speed of DMD, it is difficult to increase the number of time divisions. While future improvement for the refresh rate of the PSLM will essentially solve the problem, for the occlusion mask, our edge-only optimization would reproduce the blur with a small number of time divisions.

4.8 Summary

A focal surface occlusion scheme is proposed to facilitate the depth representation of OCOST-HMDs. Focal length calculations for a PSLM based on afocal optics are performed to ensure that the rays from the mask are perceived to be emitted from an arbitrary

depth while maintaining the magnification of the see-through view. Furthermore, we formulate an edge-based optimization scheme of PSLMs as dynamic free-form lenses based on the occlusion properties to suppress degradation in the see-through view. Results of experiments show that the proposed method can reproduce the focus blur of continuous depth objects at a resolution comparable to that of current consumer-level near-eye displays while suppressing the degradation of the see-through view. Nevertheless, certain limitations of the focal-surface occlusion concept remain to addressed to ensure its practical use. The presented findings and recommendations are expected to promote innovative research on realistic depth representations in OCOST-HMDs.

Part III

High Dynamic Range Appearance Reproduction with Light Measurement on the Viewpoint

This part introduces a method for HDR appearance reproduction by adaptive amplitude modulation of incident light using OCOST-HMD based on the light measurement on the viewpoint. We introduce AdaptiVisor, a VA system that assists eye adaptation by implementing HDR appearance representation in the first-person view (Chapter 5). Furthermore, suppose various appearances can be reproduced by combining the adaptive measurement and the HDR presentation on OST-HMDs; in that case, it will be possible to freely redesign vision according to the purpose. As an example of such system, we introduce DehazeGlasses, a method for implementing haze removal in the first-person perspective based on light measurements on the viewpoint and analysis by machine vision (Chapter 6).

5

Adaptation Assistance with OCOST-HMDs

This section is based on the work that the author published in Augmented Human 2017 Conference [C2].

5.1 Introduction

Brightness adaptation of the eye is a fundamental ability in human visual system. By sensing the surrounding brightness, our eyes adjust the amount of light passing through the lens by widening the pupil [119]. Moreover, our eyes adjust the sensitivity of the retina by switching between rod and cone cells in the photoreceptors [120].

However, once the switching occurs in the both types, it takes a certain amount of time to return to the state where the cells can react again. In other words, the brightness adaptation of the eye cannot keep up with the rapid change of the brightness [120]. When the environment lighting shifts from bright to dark, the adaptation already takes 20-40 seconds to adapt to the dark environment [121]. Even worse, when the shift is from dark to bright, the adaptation takes more than 30 minutes to completely adapt to the bright environment [122].

In our daily life, there are various situations where such adaptation cannot complete in time, which may lead to decreased task performance, or even serious accidents. For example, when a light from an oncoming vehicle gets into the eyes during night driving, the eye first dazzles since the intense beam light irradiates the eyes that are adapted to a

dark environment, and then the bright adaption is quickly occurred. As the result, it would take time for the eyes to adapt back to the dark environments.

To avoid such risks caused by the adaptation, this paper conducts a feasibility study of VA system in brightness adaptation assistance via OST-HMDs. For the study, we build AdaptiVisor, an OCOST-HMD system that assists the brightness adaptation. Our approach combines a transmissive LCD as an occlusion layer with an OST-HMD as an additive layer to dynamically block or add the light entering into the eyes, which allows controlling the brightness of the scene at a viewpoint.

In order to realize such smart dimming glasses, it is necessary to measure how the eye perceives light and then calibrate between the eye and the display. Our system uses a scene camera that detects both the change of the brightness and the position of particularly bright or dark areas in the scene. The scene camera is optically aligned to the position of the viewpoint for the OST-HMD by a beam splitter. Combined with a spatially-calibrated transmissive LCD and HMD, our system can selectively control the light entering into the eye up to the pixel resolution of each occlusion and display layers in real-time.

Contributions Our main contributions include

- implementing a proof-of-concept OST-HMD system that selectively suppress and enhance the amount of the light entering into the user's vision up to the pixel resolution of the LCD mask and HMD screen in realtime,

- conducting quantitative and qualitative analyses of the system with a user-perspective camera, which shows that the system achieves to decrease the overexposed area in a scene to 1/15 th, and enhances the color by reducing majorly underexposed area to half,

- providing a thorough analysis of the current setup including limitations and possible research directions.

5.2 Related works

In Section 1.3, we introduced the existing VA systems on OST-HMDs. Also, in Sec. 2.3.1, we describe a method for dynamically controlling the light entering the eye using an HMD. Here, we focus on introducing other VA systems for adjusting the scene brightness, and computional photography.

5.2.1 Systems for Adjusting Scene Brightness

Rekimoto proposed Squama, a programmable window that can dynamically and selectively adjust light-blocking properties using smart window panels that can switch transparency [123]. Their system detects the positions of external light and indoor objects, and can adjust which panel to turn on so that the light does not hit the object.

Hara et al. demonstrated a glare removal system for transparent surfaces [124]. Their LCD-based system removes the glare of water droplets on a transparent surface by detecting them with a camera and by tuning the transmittance of the surface on which the droplets stay. While their system can remove such disturbing effects caused by obstacles on the lens, it does not consider controlling the brightness of the light from the scene. Santos et al. developed an occlusion-capable see-through binocular display that uses LCD panels [125]. Although their LCD mask is controllable, the system only creates masks for the position of the virtual contents to be rendered, and does not analyze scene brightness.

5.2.2 Computational Photography

In computational photography, there are rich researches on appearance control. Nayar and Branzoi proposed adaptive dynamic range imaging that adapts the exposure of each pixel on a camera image, based on the brightness of the corresponding scene point [126]. Their system attaches an LCD mask in front of a camera and uses the mask to attenuate light rays for each pixel. Amano et al. constructed a system that dynamically controls the appearance of real objects using a projector-camera system [127], realizing various visual-aid effects such as color vision correction, color removal, and contrast enhancement.

Tamburo et al. developed a smart headlight system for automobiles [128]. Their system

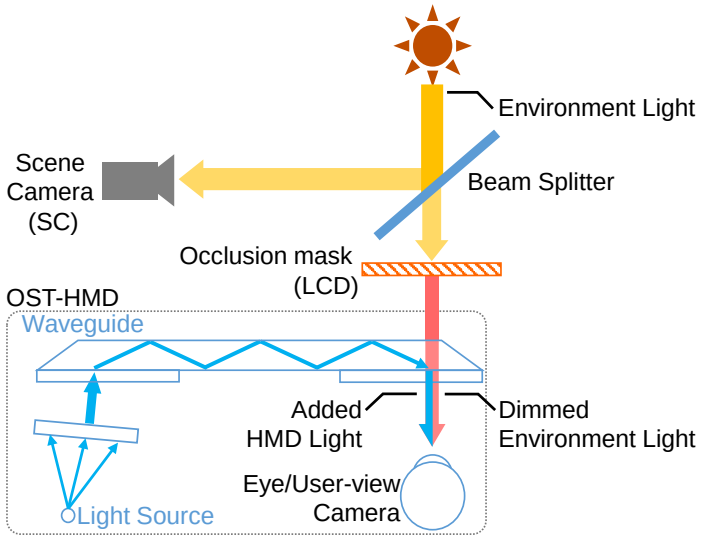

Figure 5.1: Schematic diagram of our adaptation assistance system with OCOST-HMDs.

combines a DLP projector based on a high-resolution digital micro-mirror device with a high-speed camera system. The authors demonstrated that their system with 1 to 2.5 ms latency could illuminate an outdoor scene while avoiding the light to illuminate falling snow flakes that would cause flickering.

5.3 Method

5.3.1 System Overview

Figure 5.1 shows a schematic diagram of our adaptation-assistance system using an OST-HMD and an occlusion mask. Our approach combines a transmissive LCD panel and an OST-HMD to dynamically dim or add the environmental light entering into the eyes, which allows to control the brightness of the scene at a viewpoint. The scene camera (SC) captures the information of the environment light. The user-perspective camera (UC) represents the user's eye for our test to obtain images from the viewpoint. The coordinates of the cameras and the displays are calibrated beforehand, as mentioned in Sec. 5.4.2.

Firstly, to specify the position in which the light needs to be controlled at the viewpoint, the SC is set at the optically same position as the viewpoint by using a beam splitter as

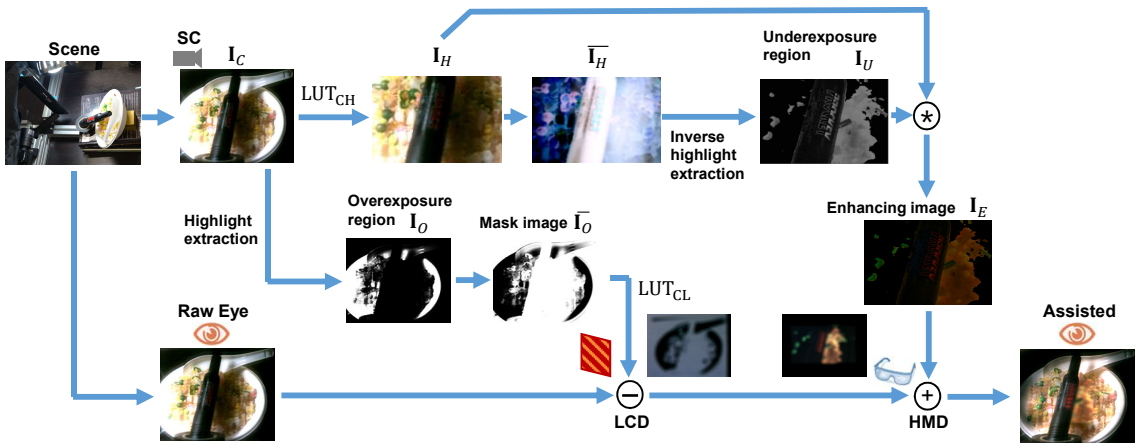

Figure 5.2: Algorithm flow of the computation of the mask and HMD images.

with [60]. Then, the controlled occlusion mask dynamically dims the environment light passing coming through the beam splitter toward the user's eye. The occlusion mask blocks light from a scene area where the highlight is strong at the viewpoint. After that, the OST-HMD layer provides the brightness and the color to the positions that are not clearly visible from the viewpoint. As a result, the amount of the light entering into the user's eye is dynamically controlled.

5.3.2 Computation of Displaying Images

In our system, the LCD attenuates the overexposure region of the light entering into the eyes and the OST-HMD compensates the underexposure region. By adding or subtracting the light with displays pixel-wise, our system can selectively change the amount of the light entering the eyes up to the pixel resolution of the displays. Figure 5.2 shows an overview of our algorithm flow. Our system has two key computation steps: To compute the mask image displayed on the LCD and the enhancing image displayed on the HMD. $\mathbf{I}$ represents a 2D image. Both the overexposure region and the underexposure region are detected from the SC image $\mathbf{I}_S$.

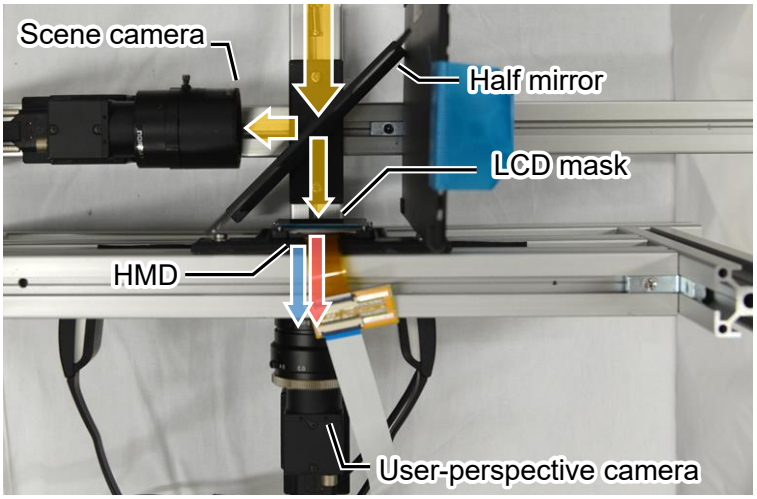

Figure 5.3: Hardware setup with schematic visualization of light paths in Fig. 5.1. The image is vertically flipped for the visualization purpose.

Compute the mask image

To compute the mask image, we first obtain the overexposure region $\mathbf{I}_O$ from $\mathbf{I}_S$. In our current setup, we apply the real-time highlight removal method proposed by Shen et al [129] to $\mathbf{I}_S$, then we treat the removed highlight image as $\mathbf{I}_O$. Since $\mathbf{I}_O$ shows the pixel-wise intensity of lights on the SC, the mask image $\overline{\mathbf{I}_O}$ is obtained by the negative/positive conversion of $\mathbf{I}_O$ to compensate the intensity. Finally, LCD displays the image $\mathrm{LUT}_{\mathrm{CL}}(\overline{\mathbf{I}_O})$, where $\mathrm{LUT}_{\mathrm{CL}}(\cdot)$ is coordinate transformation by the look-up table (LUT) between the SC and the LCD (described in Sec. 5.4.2).

Compute the enhancing image

To obtain the image displayed on HMD, we first transform $\mathbf{I}_S$ to $\mathbf{I}_H = \mathrm{LUT}_{\mathrm{CH}}(\mathbf{I}_S)$ by the LUT between SC and the HMD (Sec. 5.4.2). Due to the limited FoV of the OST-HMD, the area of captured OST-HMD screen is smaller than the entire camera image. We thus transform the camera image beforehand to reduce the computation area in the image.

We calculate $\overline{\mathbf{I}_H}$ by the negative/positive conversion of $\mathbf{I}_H$. Since the underexposure region such as blocked up shadows in $\mathbf{I}_H$ is converted to (inverse-)highlight in $\overline{\mathbf{I}_H}$, we obtain $\mathbf{I}_U$ as a gray-scale image by applying the same highlight removal to $\overline{\mathbf{I}_H}$.

While the LCD can display only the gray-scale image, the HMD can compensate the RGB color of the environment. For this reason, we convert the gray-scale image $\mathbf{I}_U$ into the color enhancing image $\mathbf{I}_E$ by using the original image $\mathbf{I}_H$. The RGB color $\mathbf{I}_E(x)$ at pixel x is calculated from $\mathbf{I}_H(x) = [I_H^r(x), I_H^g(x), I_H^b(x)]^\top$ and the gray-scale value $I_U(x)$:

$$\mathbf{I}_E(x) = \alpha I_U(x)\mathbf{I}_H(x) \tag{5.3.1}$$

where α is the enhancing ratio. In Eq. (5.3.1), $\mathbf{I}_U$ is applied as a mask to $\mathbf{I}_H$ to obtain the image $\mathbf{I}_E$ displayed on the OST-HMD. By this masking process, the black background is kept black and only the region where the color and brightness is faded can be extracted. Through the experiment, we notice that just displaying $\mathbf{I}_E$ (i.e. $\alpha = 1.0$) is not enough to fully compensate the color and brightness. For this reason, we manually set $\alpha = 2.5$ and trimmed the exceeded elements $I_E^{r,g,b}(x) > 255$ to 255. The proper value of α depended on the brightness of HMD, therefore beyond the scope of this paper. We discuss the color and brightness calibration on Sec. 5.6.3.

5.4 Technical Setup

We describe our hardware and software setup and elaborate calibration steps.

5.4.1 Hardware and Software Setup

Figure 5.3 shows our hardware setup. Our system combines an Epson Moverio BT-200 (23° field of view) as an OST-HMD with a LCD panel (SONY LCX017, 1024×768 pixel) for the opaque layer. We use an acrylic beam splitter with 30% transmittance and 70% reflectance. Some fixtures were self-made parts using a 3D printer. All devices are connected to the same Windows 10 desktop machine (Intel Core i7-6700K CPU 4.00GHz, 16GB RAM). The LCD panel is connected to the computer via the controller made by bbs Bild- und Lichtssysteme.

Figure 5.4 shows the transmittance property of the LCD panel measured by a gamma-corrected camera. We displayed uniform values, i.e. uniform occlusion masks, on the LCD while capturing the scene through the panel by the camera. The figure shows the profile

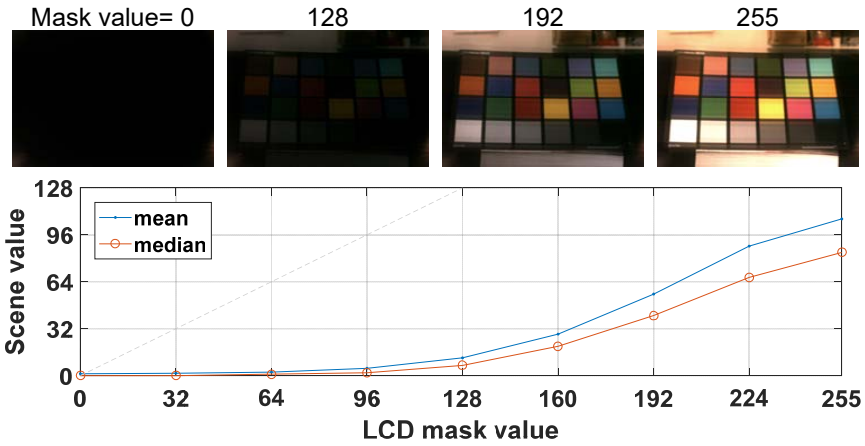

Figure 5.4: Transmittance profile of the LCD panel used in our setup. Given uniform values on the entire LCD mask from 0 to 255, the transmittance changes nonlinearly when seen by a gamma-corrected camera.

of the captured images in terms of the measured brightness. Note that, in this preliminary test, we did not place the beam splitter.

To output video signal from a PC to the BT200, we used a DM484CS DVI-D (HDMI) Interface for Moverio BT-200 from Colorado Video. We set the virtual video signal to 1280×720 pixel for the display, which is larger than the actual display resolution, yet yields the same aspect ratio of 16:9.

To obtain images from a viewpoint through the OST-HMD, we installed a user-perspective camera placed behind the left optical element of the display. For both the user-perspective and the scene camera, we use PointGrey Flea3 FL3-U3-13E4C-C with a 1280×960 resolution. The model uses 1/1.8" CMOS sensor. For the two cameras, we set the frame rate to 30Hz.

For the scene camera, we used a C-mount lens, Kowa LM12HC (focal length 12.5mm). For the user-perspective camera, we used a varifocal C-mount lens, Tamron M12VM412, and set its focal distance to 12 mm.

For calibration procedure and image processing, we used OpenCV and a calibration framework, Ubitrack [130], to handle PointGray cameras, and Matlab 2016 to compute calibration data.

5.4.2 System Calibration

To control the brightness of each light rays, it is crucial to create a pixel-wise mapping from an observed camera pixel in SC to the corresponding display pixel and a user view, i.e. UC in our case. Figure 5.5 summarizes each calibration steps described in the following.

Calibration between the User-Perspective and the Scene Camera (Fig. 5.5 A)

We first compute a homography between SC and UC. Assuming that the two cameras are optically almost aligned, we decided to approximate the mapping between the two cameras by homography. We let both SC and UC capture a physical checkerboard simultaneously, and then we applied a checker board detection to automatically detect grid corners. Finally, we computed a 3×3 homography matrix out of the 2D point correspondences. Since both cameras used lenses with rather long focal distances, we left the cameras uncalibrated and did not undistort.

Calibration between the HMD Screen and the User-Perspective Camera (Fig. 5.5 B1)

We compute $\mathrm{LUT}_{\mathrm{CH}}(\cdot)$, the LCT from the UC image plane to HMD screen, by showing gray-code patterns on the HMD screen while capturing them by the UC. By applying a standard LUT calculation, we can compute an LUT from the camera to the screen. Note that the direction of the LUT obtained by the above procedure is opposite of what we require, i.e. we need to transfer the pixel position of the screen to that of the UC image.

In our implementation, we applied Gaussian kernel regression between the 2D correspondences to build a continuous mapping function between the image planes. We then built an integer LUT from the screen to the camera.

Calibration between the LCD Screen and the User-Perspective Camera (Fig. 5.5 B2)

In principal, $\mathrm{LUT}_{\mathrm{CL}}(\cdot)$, the LUT from the UC image plane to LCD mask, can be computed in the same manner as the above step. Note that, however, the occlusion mask is out of focus for the UC and the mask itself is just a transparent layer without background illumination.

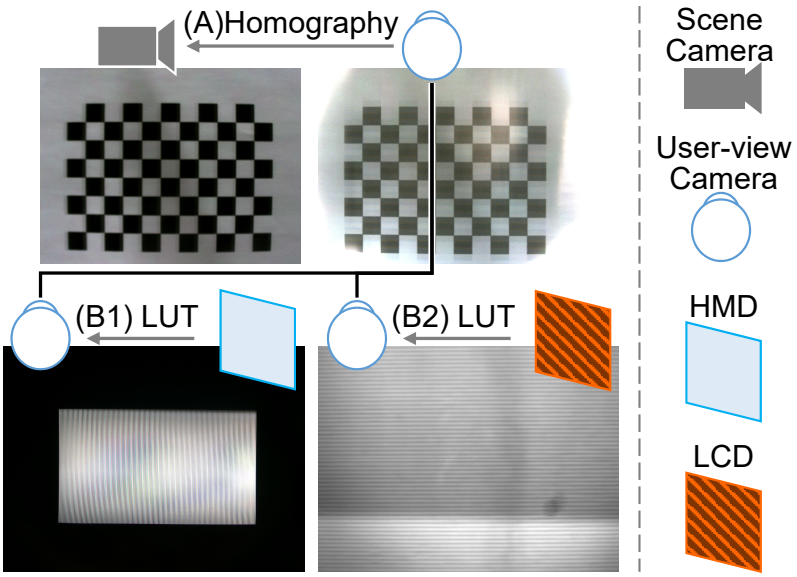

Figure 5.5: System calibration overview. Combining mapping functions A and B1 or B2 creates desired mappings for mask and HMD image visualization.

We therefore let the UC focus on the LCD to capture sharp gray-code patterns. We also placed uniform white diffuse illumination at the background to make the patterns visible.

We then let the UC focus to the final background scene after calibration. Thus, the LCD image is not perfectly calibrated against UC in pixel-wise. This misalignment is, however, has less impact than the blur effect induced by the LCD, and both the spatial calibration and the mask blur are beyond the scope of this paper.

Nevertheless, we discuss this issues and possible solutions later in Secs. 5.6.1 and 5.6.4 to provide future research questions.

In this calibration, gray-code patterns gave smaller number of reliable 2D correspondence pairs compared to Sec. 5.4.2. To properly estimate the continuous mapping function, we set a wider kernel width for the regression thus the LUT. Since the LCD is larger than the camera lens, the camera captured a part of the LCD panel.

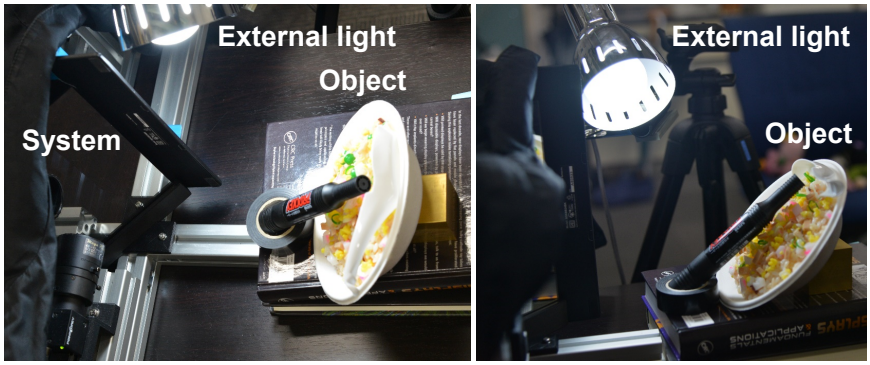

Figure 5.6: Experiment setup. A target object is placed at about 30cm away from the system, and an extra lighting is set above the object.

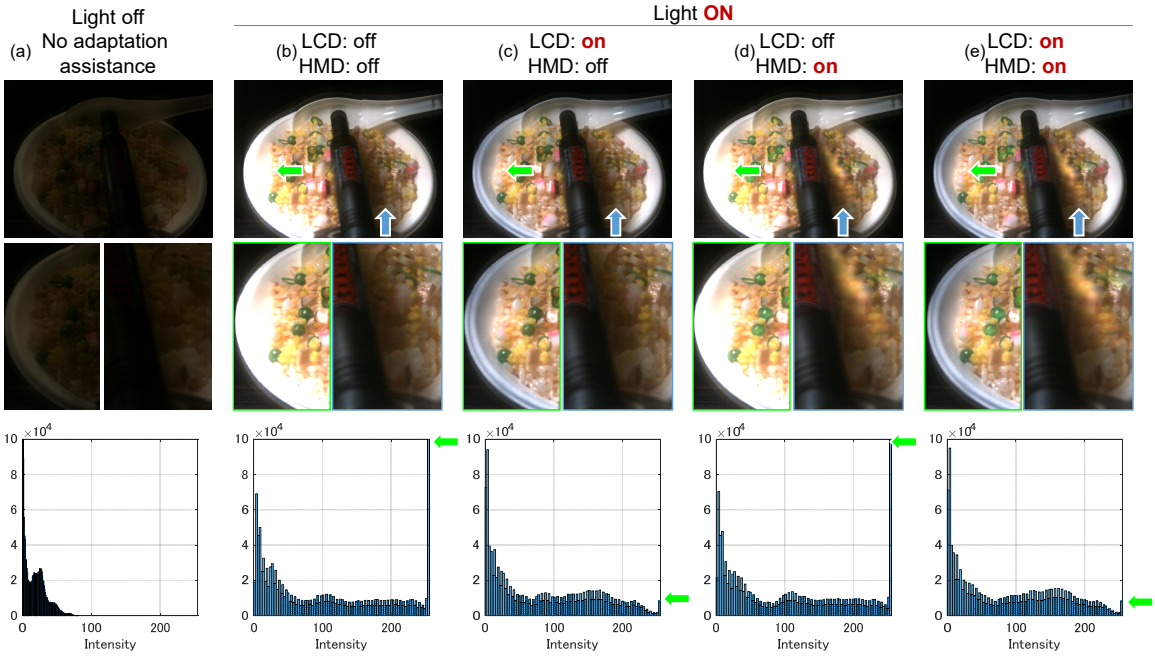

Figure 5.7: Overview of the static-capture experiment in Sec. 5.5.2.

5.5 Experiment

We conduct two experiments with our proof-of-concept system. The static capture experiment (Sec. 5.5.2) quantitatively evaluates the adaptation ability of our system under a controlled environment. The dynamic capture experiment (Sec. 5.5.3) clarifies issues of the current system under more practical use where the system runs in real-time under a dynamic-lighting condition. We then discuss possible solutions for realizing a practical system in the discussion section. Finally, to see actual impression of the users, we include preliminary feedback from real users by letting them try the system.

5.5.1 Experiment Setup

For evaluation, we built a workbench that consists of a target object to be captured and an external light source to be switched (Fig. 5.6). The target is a plastic food model and a black marker pen. The target object is placed at about 30cm away from UC. We then let UC capture the target object while the external lighting is switched on and off. We evaluate different adaptation assistance conditions by choosing if the HMD is used or not and if the occlusion mask by the LCD is used or not. We thus have four conditions: C1 {LCD:off,HMD:off} (i.e. raw view), C2 {LCD:on,HMD:off}, C3 {LCD:off,HMD:on}, and C4 {LCD:on,HMD:on}. We keep the exposure of UC fixed through the experiment so that the comparison among different conditions becomes fair.

5.5.2 Static Capture Result

We first evaluate a static condition where we capture the scene with the four conditions offline. Although our system runs realtime, capture and display latency induce a constant delay from the lighting change to the system starts assistance. This affects the visualization quality when the lighting condition changes. To see the static performance of our proof-of-concept system, we first create in offline an occlusion mask and a brightness compensation image by capturing the static target under the extra lighting, we then capture the view with these mask and image shown on the LCD and the HMD respectively.

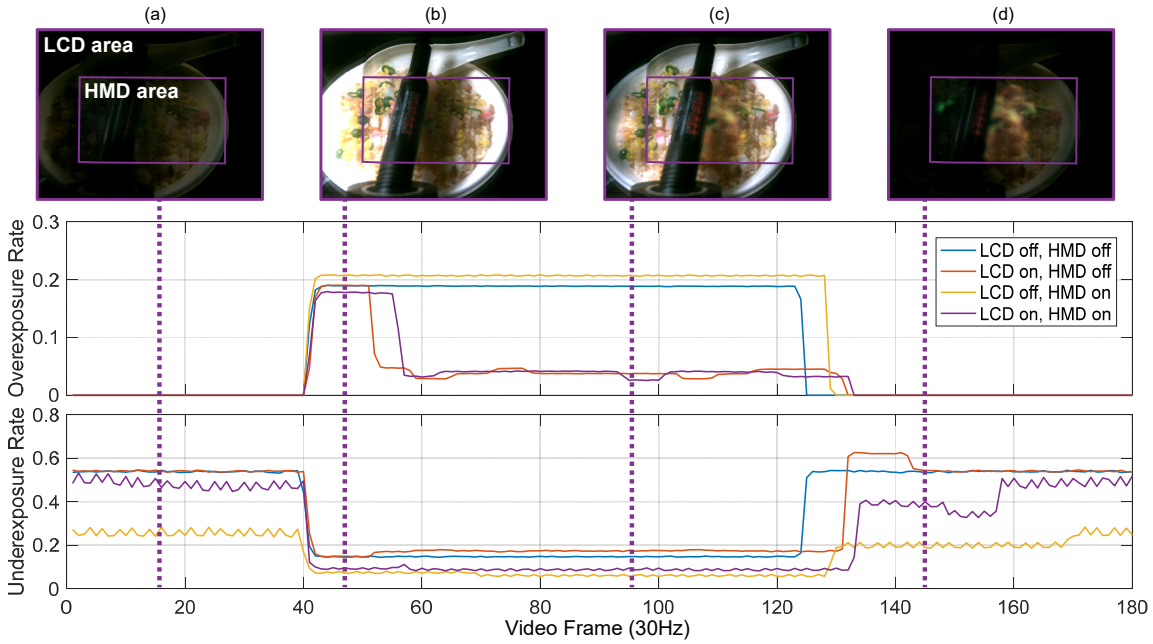

Figure 5.8: Overview of the realtime experiment in Sec. 5.5.3.

Figure 5.7 shows the summary of the experiment. Without the extra light, the scene appears dark to UC (Fig. 5.7 a). Once the lighting is turned on, the view suffers overexposure and partially darker areas due to shadows (Fig. 5.7b).

If we only activate the occlusion mask (Fig. 5.7 c), we can see that overexposed area is reduced and can see the detail of the edge of the dish. By comparing the histogram with that of 5.7 b, we can quantitatively confirms that the overexposed area in a scene is shrunk to 1/15th. This clearly shows that the LCD layer successfully dimmed overexposed area selectively.

On the other hand, when we only activate the HMD screen (Fig. 5.7 d), the histograms do not tell much about the effect. We can, however, qualitatively see that the shadow area, which is at right to the marker pen, is compensated by an additive image overlaid by the HMD. This allows us to see some fake yellow grains brighter. Unfortunately, the overlaid image lost detail due to the different focal lengths of the camera (30cm) and the HMD screen (2m). We discuss this accommodation issue in the discussion section.

Finally, if we combine the LCD and the HMD (5.7 e), we can see both effect from the LCD-only and HMD-only conditions above.

5.5.3 Dynamic Capture Result

We now evaluate a realtime response of the current proof-of-concept system. In this experiment, we recorded UC image sequences with the four conditions. For each recording sequence, we start without the extra lighting, then we turn on the external lighting, and finally turn off. The lighting is controlled by hand.

In this setup, we defined the overexposure value as that radiance larger than 235 and underexposure lower than 20. This is to see more general tendency of our system.

Figure 5.8 shows a time-domain analysis of our adaptation assistance system in terms of the over-/under-exposure area ratios. The images in the figure are from the condition C4 {LCD:on, HMD:on}. To align each recording on the timeline in the figure, we detected the timing when the light was turned on by checking a jump of overexposed-pixel count in each sequences.

In the cases with the HMD turned on (i.e. C2 and C4), the graph lines are jittering at the figure of the under-exposure rate in Fig. 5.8. As well, the enhanced visuals seemed to be jittering in the recorded UC image sequences. The jittering is caused by the difference of refresh rates between HMD and UC, thus the jittering will not be occurred when we see the HMD with the naked eye.

At Fig. 5.8 (b), the system is not yet able to provide the mask and the compensation image, thus the view is not optimized. This causes a flickering effect since the system will soon start assisting adaptation.

At Fig. 5.8 (c), our adaptation visualization is shown and we can see the overexposure is suppressed to about 25% in terms of its area size. The underexposure area also decreases about 50% between conditions with the same HMD state, i.e. C1 vs C3, and C2 and C4.

Through this dynamic capture setup, we can see that the proof-of-concept system can handle both over-/under-exposure properly and can assist our brightness adaptation. On the other hand, we observed that a low-latency system is necessary to reduce flicking effects (Fig. 5.8 b) caused by a delay to suppress overexposure. We further discuss this latency issue in the discussion section.

5.5.4 Preliminary Trial by Real Users

After removing the UC, two real users tried out the proof-of-concept system, and we let them mention the impression of the system. In the feedback, the users first realized that the HMD part clearly enhances darker areas on the object. We further decreased the base brightness setting of the HMD by the requests of the users, and one user mentioned, “The color on the HMD blended seamlessly into the background (dark) object”.

For the LCD mask, on the other hand, the users initially could not clearly recognize that the LCD was dimming the environmental light. This is because the mask got stronger blur when seeing with the real eyes than with the UC. The users, however, easily recognized the brightness reduction when the LCD was switched on and off. We speculate that the change of the blur size stems from the difference of that of the aperture of the eye, i.e. pupil size, and of the UC. This question follows to our discussion sections (Sec. 5.6.4.

5.6 Discussion

In our current system, there are various assumptions and simplifications. In order to encourage the development of a more practical adaptation assistance system in future, we organize these issues and discuss on future research directions below.

5.6.1 Spatial Calibration with Eyes

Analogous to the spatial calibration of OST-HMD in AR [131,132], a practical system must track the current 3D position of the eyes with respect to the display system. In our current proof-of-concept implementation, we achieved the pixel-wise mappings between the image layers and the user-perspective camera by LUTs since we can directly get user-perspective images. This is not true for real users.

Fortunately, we can still model the mappings as perspective projections by treating the system as off-axis pinhole cameras where the layers are image planes and the eye is the camera center. We can thus compute the projection by tracking the eye position. Such automated display calibration with wearable eye trackers has been a hot topic in recent AR research areas [133, 134]. Since the LCD and the HMD’s waveguide causes optical

distortion, we would also need to calibrate optical aberration to make a predistorted images [135].

5.6.2 Low-latency rendering

Our current system has a delay roughly about 500 msec. This is due to the latency of the screens and the computation loop to calculate $\mathbf{I}_E$ and $\overline{\mathbf{I}_O}$. Although we leave this latency issue to other dedicated works, this is a must-be-solved issue to realize a practical AdaptiVisor system. As we introduced in the related work section, hardware systems optimized for low-latency rendering in the computational photography field [127, 128] are worth exploring.

5.6.3 Simulating Eye Adaptation

Even if we have a calibrated user-perspective camera, at the end of the day, a practical system is used by a real user. We thus require a way to estimate the current state of the eye's brightness adaptation in real-time so that the system can determine the optimal mask darkness.

Furthermore, the system relies on a scene camera to analyze the scene brightness and thus to determine the mask shapes and darkness. This requires estimating color distortion between the user and the scene camera. In other words, we have to correct the color of an image captured by the scene camera to be consistent with the visual stimulus that the user receives from the scene. The same problem occurs to the HMD screen color, which has been explored in the HMD color calibration context [59, 60]. It is also desirable to use a high dynamic-range camera for the SC so that the system can handle various lighting conditions.

A fundamental difficulty in these issues is that the sensitivity of rod and cone cells of the eyes are different from that of image sensors. We thus would need a perceptually correct color rendering [136] and a way to estimate the current state of the eye.

Another potential problem is the optical property of the transmissive LCDs. Due to their transparent TFT circuits, such LCDs have diffraction and make the see-through images

blurred as we can see in figures in our paper. Integral imaging displays might be one solution to reduce this optical degradation while keeping the system compact [98].

5.6.4 Occlusion Mask Optimization

For realizing a consistent occlusion effect, we have to determine the shape of occlusion patterns so that they only and completely cover undesired light rays to be blocked. Our single-layer occlusion LCD, however, is too close to the user's eyes, and the occlusion patterns normally appear blurred to the eyes that focus at a distant object in a scene. The aperture of the eye, i.e. the pupil, predominantly causes this focus blur. The larger the pupil diameter is, the stronger the blur effect will be.

The blur can also be modeled as a point spread function or its Fourier transform, optical transfer function (OTF). Watson et al. proposed a model of OTF as its mean real function, i.e. radial modulation transfer function for the best-corrected human eye as a function of pupil diameter [137]. Bursky et al. explored vision-realistic rendering by using wavefront data of an actual human [138]. Given such function, we can apply image deconvolution on occlusion masks so that their projected retinal images appear sharp [139].

The same blur effect could occur for OST-HMDs since most existing displays only have a single focus plane. Recent light-field displays can mitigate this issue by displaying volumetric focusible images. Retinal scanning displays can also solve this phenomenon since such display ignores the crystalline lens and can directly project image onto the retina.

Furthermore, as shown in Fig. 5.4, non-linearity of the transmittance property of the LCD mask is also needed to be calibrated.

5.6.5 Long-term Use in Daily Life

The proposed system has concerns that the artificial adaptation might have bad effect to the natural adaptation of the eyes. We consider that the system might not degrade the natural adaptation by the photoreceptors, because the system only assists the excessive environment light to fit within the original dynamic range of the eyes. However, as a long-term goal, we would need to evaluate the actual influence of the system on the eyes quantitatively.

5.7 Summarys

We presented AdaptiVisor, a proof-of-concept vision augmentation system for brightness adaptation assistance. Our system consists of an occlusive optical see-through HMD and a scene camera which is optically aligned to the viewpoint. For occlusion, we employ a transmissive LCD rigidly attached in front of an OST-HMD.

Our proof-of-concept system with a user-perspective camera as a viewer demonstrates that it enables to selectively and dynamically suppress overexposed areas in the scene by the LCD mask. At the same time, the system can enhance underexposed areas by showing additive image on the screen. We use a highlight detection method to determine which area in the FoV to be masked and which area of the FoV to be enhanced by the HMD. For displaying the mask and the image correctly in the user-perspective camera's FoV, we spatially calibrated the LCD, HMD, and the user-perspective camera by using homography and gray-code patterns.

Our evaluations based on the static and dynamic scene capture show that our system was able to suppress the overexposed area to 1/15th compared to raw FoV. It also improved the underexposed area by correctly overlaying the image. Preliminary user trial with two real users indicates that the system also works for real eyes for the HMD part and to some extent for the LCD, and suggests that optimizing occlusion mask shape for a given user's eye is a key issue.

Discussion followed then gave thorough discussion on steps towards practical adaptation assistance systems. The discussion includes eye simulation and occlusion mask optimization.

6

First-person Dehazing with OCOST-HMDs

This section is based on the work that the author published in Augmented Human 2020 Conference [C1].

6.1 Introduction

Human vision often suffers from the haze, small particles floating in the air. The floating particles such as smoke, dust, fumes, and mist scatter the light in the atmosphere and attenuate the direct scene irradiance in correspondence with the scene depth. Furthermore, the particles reflect the ambient light and veils the scene uniformly as an opaque white layer called airlight [140]. Haze thus leads to low contrast, blurring, color distortion, and other visual degradation, which significantly affects human perception and behavior. The visual degradation may even cause severe accidents in some situations, e.g., driving in a fog or snow environment, and evacuation and rescue activities in fire or other disasters.

Removing haze to provide a clear view from a user's viewpoint helps the user to make correct decisions and actions. In the computer vision community, haze removal also attracts interest as a preprocessing step to aid high-level visual tasks such as object recognition, person re-identification, and visual tracking.

The early dehazing methods stand on the atmospheric scattering model [141, 142], which is a simple approximation of the haze effect, and formulate the haze-free image assuming some priors [143]. Recent rising-up of deep learning pursues robust performance

of haze removal without any prior [144, 145].

In this work, we consider *re-importing* these machinery vision of haze removal into the human vision by presenting a first-person system. Our future goal is to enhance the high-level human recognition in hazy situations and to decrease the potential misjudgment.

The most straightforward approach is VST systems where a system feeds a video stream from a scene camera to the user's view through a monitor or an opaque head-mounted display. While such a VST system can easily apply existing dehaze methods, the approach may not be suitable for visual assistance in a wearable system [146]. A VST system loses rich light information of the real world consisting of the real light field, which has the resolution and dynamic range higher than existing displays. More importantly, VST may have a risk of blackout the user's view under accidental shutdown or running out of power. We think these disadvantages make VST systems not be suitable for dehaze applications that may be used in critical tasks [146, 147].

OST approaches, where a system minimally augments the user's first-person view, have advantages over the VST approaches. Firstly, the OST approach is less intrusive to the user's view compared to VST systems. It can maintain the real-world light and the modification of the view is minimally done by OST-HMD. Secondly, the OST approach is safer. The malfunction of an OST-HMD system does not block the user's view and it merely results in the user sees the actual view without augmentation.

Lin et al. recently adapted a dehaze method to a commodity OST display without occlusion capability [148]. Their system calculates the difference between the real scene and a dehazed image, then computed an auxiliary image that enhances the visual image by applying the just-noticeable difference decomposition [149]. Since the OST display can only add light in the scene, their method essentially loses the contrast of the scene, and the resulting view rather appears as a brightness-enhanced view instead of truly removing the airlight layer.

To realize a see-through dehaze system that optically removes haze, we built an occlusion capable see-through display. Our system combines an OST-HMD with a DMD, which can switch the direction of light rays for each pixel by micro-electro-mechanical mirrors to occlude the incoming rays. By adopting a dehazing method to the occlusion-capable OST-HMD system, we can optically subtract the scene light to dehaze the hazy view.

To the best of our knowledge, this is the first attempt to realize an occlusion-capable see-through dehazing system to enhance the first-person view of the user.

Our main contributions include:

- providing a computation method to realize haze removal using occlusion-capable OST-HMD,
- implementing a proof-of-concept OST AR display that can optically dehaze the first-person view of the user,
- conducting quantitative and qualitative analyses of the system with a user-perspective camera, which shows that the system achieves perceptually natural haze removal while keeping the see-through view of actual scenes, and
- providing a thorough analysis of the current setup including limitations and possible research directions.

6.2 Related Work

We described the overview of VA systems in Sec. 1.3 and OCOST-HMDs in Sec. 2.2.2. Here, we first introduce related research on haze removal. Then, we focus on VA systems to improve visibility, which utilizes SLMs to manipulate the contrast or brightness of real-view.

6.2.1 Haze Removal

Existing dehazing methods rely on the physical scattering model [141,142]. The model employs the atmospheric transmission map and the atmospheric illumination. Most of methods based on this model use prior-based methods [143]. Schechner et al. took the airlight polarization into account on top of the scattering model [150, 151]. Recent advancement in deep neural networks brought works using these networks [144, 145].

All these methods require image processing on the input image to reconstruct a dehazed image. Haze adds up gray colors on top of the clear image, the processing inevitably

requires the subtraction of the colors from the input image. Since our goal is to dehaze the user's view via a see-through system, the system must be capable of *occlusion*—selectively attenuating the scene light in the user's view.

6.2.2 Vision Augmentation to Improve the Visibility

Tamburo et al. developed a DMD-based smart automotive headlight to improve visibility in a driving scene, especially in risky weather conditions such as rain, fog and snowstorms [128]. Hara et al. demonstrated an LCD-based system that can remove the glare of water droplets on a transparent surface by tuning the transmittance where droplets are positioned [124]. These systems can improve the visibility by modulating the incoming rays in the driving scene or computational photography. Similarly, we aim to directly modulate the incoming light into eye and improve the visibility from the first-person view.

6.3 Method

In the following, we first formulate the see-through dehaze problem and provide mathematical notations used through the rest of the paper in Sec. 6.3.1. Secondly, we describe our base computation approach in Sec. 6.3.2. Finally, in Sec. 6.3.3, we further provide a visually less-intrusive method based on the base approach, which aims to reduce the amount of illumination from the OST-HMD image layer.

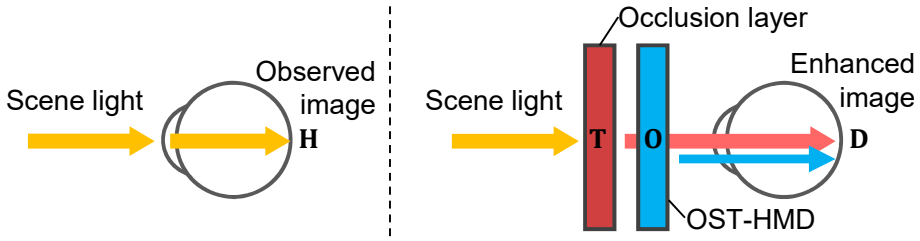

Figure 6.1: A schematic diagram of a generic hardware formation of our see-through setup defined in Sec. 6.3.1. This generic model lacks some implementation details (Sec. 6.4), yet just sufficient to formulate our problem.

6.3.1 Problem formulation and Notations

Our goal is to model a see-through haze removal filter on a glasses system equipped with an occlusion layer (e.g., with a DMD) and the additive image layer in an OST-HMD. Since the occlusion layer can only control the transparency of the light passing through it, the layer can only rescale the brightness of the scene image instead of directly subtracting its colors. This limitation prevents us from simply using existing haze removal pipelines. Instead, we have to build a dehaze approach with attenuation and addition operations only.

For simplicity, we start with a simplified setup where each image layer is already spatially aligned with each other and has the same image size. In practice, we need to calibrate each image layer to associate each corresponding pixel as elaborated in Sec. 6.4.4.

Our common notations and variables are as follows:

- $\mathbf{H} = \{\mathbf{h}_{ij}\} = \{h_{ijc}\} \in \mathbb{R}_{+}^{W \times H \times 3}$ is the observed (hazy) color scene image,
- $\mathbf{D} = \{\mathbf{d}_{ij}\} = \{d_{ijc}\} \in \mathbb{R}_{+}^{W \times H \times 3}$ is the dehazed color image from the viewpoint,
- $\mathbf{T} = \{t_{ij}\} \in \mathbb{R}^{W \times H}$ $(0 \leq t_{ij} \leq 1)$ is the $W \times H$ transmittance matrix for the occlusion layer, and
- $\mathbf{O} = \{\mathbf{o}_{ij}\} = \{o_{ijc}\} \in \mathbb{R}_{+}^{W \times H \times 3}$ is the color image displayed on the OST-HMD,

where i, j are image pixel indices, W is the image width, H is the image height, and $c \in \{r, g, b\}$ is the color channel index. In the following, we may put the index c as a subscript to image matrices and color vectors to represent their channel-wise slices and elements, respectively.

Based on these notations, we formulate our problem as shown in Fig. 6.1. At each color channel c, the occlusion layer attenuates incoming light of a hazey image $\mathbf{H}_c$ by $\mathbf{T}$, and the OST-HMD adds an image $\mathbf{O}_c$. We want the resulting image $\mathbf{D}_c$ to be obtained by:

$$\mathbf{D}_c = \mathbf{T} \odot \mathbf{H}_c + \mathbf{O}_c \tag{6.3.1}$$

where $\odot$ is the Hadamard (element-wise) product of two matrices of the same size. Note that the attenuation matrix $\mathbf{T}$ is shared over the color channels. Figure 6.2(top) shows an overview of the computation model of $\mathbf{T}$ and $\mathbf{O}$ with an example scene.

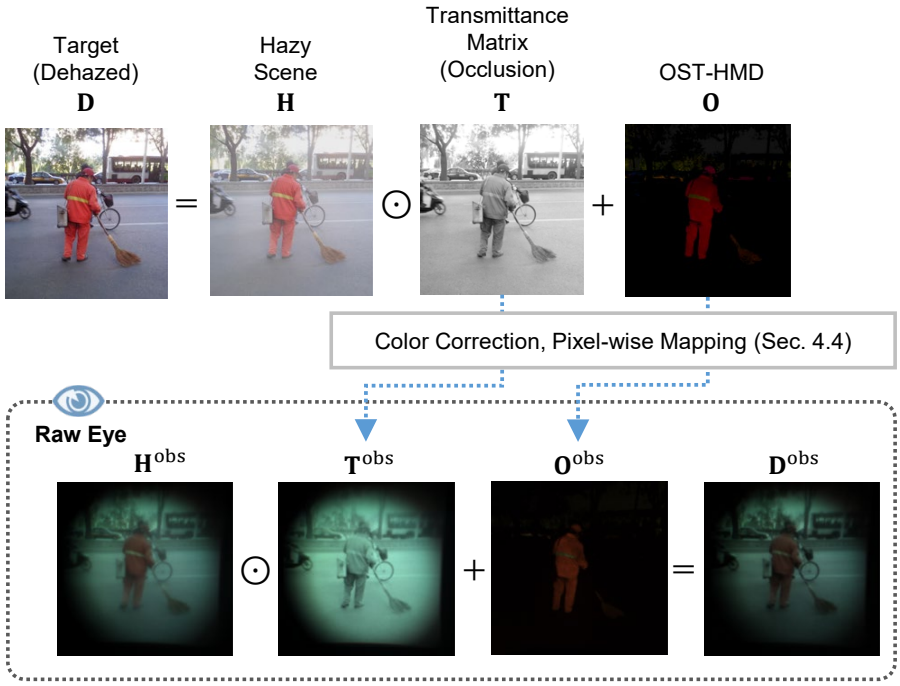

Figure 6.2: Overview of the computation of the transmittance matrix $\mathbf{T}$ and the displayed image $\mathbf{O}$. (top) Computation in the captured image layer. (bottom) Observed scenes and the display images from the user view. Since the observed OST-HMD image $\mathbf{O}^{\text{obs}}$ is not clearly visible in the original image, the brightness of $\mathbf{O}^{\text{obs}}$ in this figure is multiplied by 1.6.

Our problem is now to find $\mathbf{T}$ and $\mathbf{O}$ given $\mathbf{H}$ and $\mathbf{D}$. There are, however, two points we need to pay attention to in practice. Firstly, to capture the $\mathbf{H}$, we need an extra scene camera aligned with the user's view [C2]. Since the camera and the OST-HMD have different color profiles, their color spaces need to be calibrated [40, 59]. Secondly, $\mathbf{D}$ is unknown and needs to be estimated from a hazy scene image $\mathbf{H}$. The dehaze algorithm itself is not the focus of this paper, so we simply use one of the most promising dehaze algorithms, FFA-Net [145], in the community and treat the obtained image as a true dehazed image $\mathbf{D}$.

6.3.2 Basic computation for display images

The trivial solution for Eq. (6.3.1) is to set $t_{ij} = 0$ and $o_{ijc} = d_{ijc}$, i.e. the system completely cuts the see-through view and use the OST-HMD as a VST system. Using complete VST

view leads the issues compared to the OST view [146]. We reduce the displayed colors on OST-HMD as minimal as possible while preserving the maximum use of the scene light.

We thus need to solve the following optimization problem for each pixel:

$$\text{minimize}\,||\mathbf{o}_{ij}||, \quad \text{s.t.}\,||\mathbf{d}_{ij} - (t_{ij}\mathbf{h}_{ij} + \mathbf{o}_{ij})|| = 0. \tag{6.3.2}$$

Given $\mathbf{o}_{ij} = \mathbf{d}_{ij} - t_{ij}\mathbf{h}_{ij}$ from Eq. (6.3.1) and $\mathbf{o}_{ij} \geq 0$ by definition, the above problem is reformulated as:

$$\underset{t_{ij}}{\text{argmin}}\,||\mathbf{d}_{ij} - t_{ij}\mathbf{h}_{ij}||, \quad \text{s.t.}(\mathbf{d}_{ij} - t_{ij}\mathbf{h}_{ij}) \geq 0. \tag{6.3.3}$$

Given all variables are non-negative, we solve the above in the following. We start from $t_{ij} = 0$, which meets the constraint. Gradually increasing t_{ij} from 0 only decrease the cost function, so we can safely continue increasing it until one of the color channels gives $d_{ijc} - t_{ij}h_{ijc} = 0$. Since other color channels hold $d_{ijc} - t_{ij}h_{ijc} > 0$ and we can not increase t_{ij} anymore due to the constraint, we found the solution:

$$t_{ij} = \min_{c \in (r,g,b)} t'_c, \quad t'_c = \frac{d_{ijc}}{h_{ijc}} \tag{6.3.4}$$

6.3.3 Alleviation of the complete occlusion issue by the occlusion layer

In the basic computation, the pixel value t_{ij} in the transmittance matrix is derived from the division between the target and the input from Eq. (6.3.4). Thus, the t_{ij} tends to be close to 0 when a color channel of the target d_{ijc} has a much smaller value than the other color channels according to Eq. (6.3.4).

The lower value of the t_{ij} particularly affects the color compensation in the pixel with high saturation, as shown in Fig. 6.3. When the pixel value in the scene image $\mathbf{h}_{ij}$ has low saturation (Fig. 6.3 a), the transmissive ratio t_{ij} becomes higher, and the light reflected by the occlusion layer still keeps the actual scene light. Otherwise, the values of the other color channels are suppressed by the lower transmittance t_{ij} (Fig. 6.3 b). In such pixels, the occlusion layer almost completely occludes the input, and the OST-HMD needs to add almost the same value as the target. As a result, Our system works as the VST system at

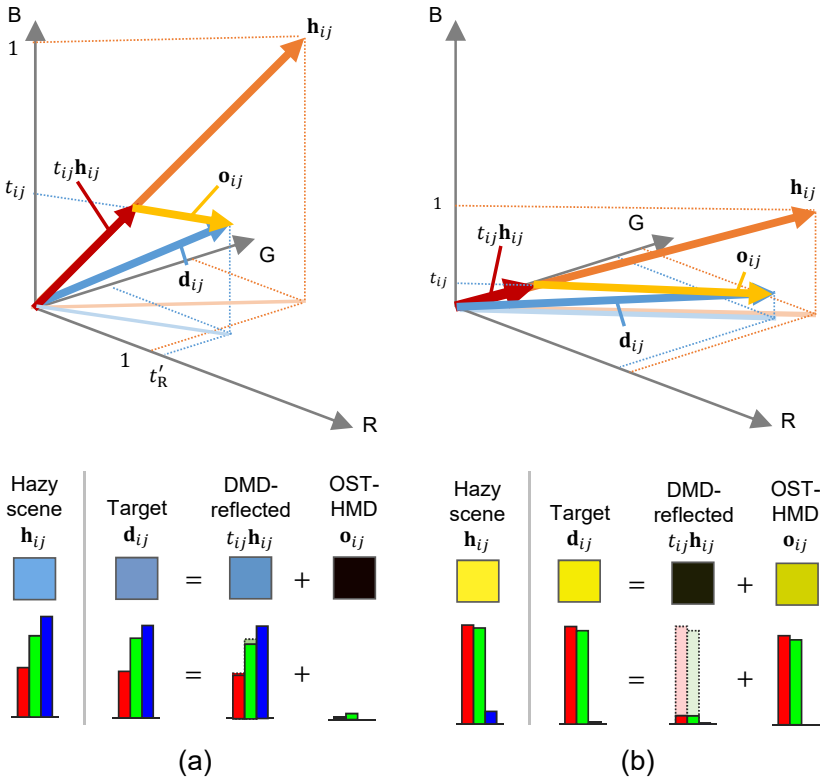

Figure 6.3: Visualization of the complete occlusion issue explained in Sec. 6.3.3. (top) The figures visualize the Eq. (6.3.1) for cases when the intensity ratio t'_c is (a) not much saturated and (b) highly saturated. The pixel value in the scene image $\mathbf{h}_{ij}$ is colored in orange, the target image $\mathbf{d}_{ij}$ in blue, the scene image reflected by the DMD $t_{ij}\mathbf{h}_{ij}$ in red, and the image displayed by the OST-HMD $\mathbf{o}_{ij}$ in yellow. (bottom) RGB colors and components in each image pixel corresponding to models shown above. Note that t'_B is selected as the transmittance ratio t_{ij} in both saturation cases.

that pixel.

To keep the actual view even for highly saturated pixels, we introduce the alleviation term, s_{ij}, corresponding to the saturation at each pixel that can be denoted as:

$$s_{ij} = \frac{M - t_{ij}}{M} \tag{6.3.5}$$

$$M = \max_{c \in (r,g,b)} t'_c \tag{6.3.6}$$

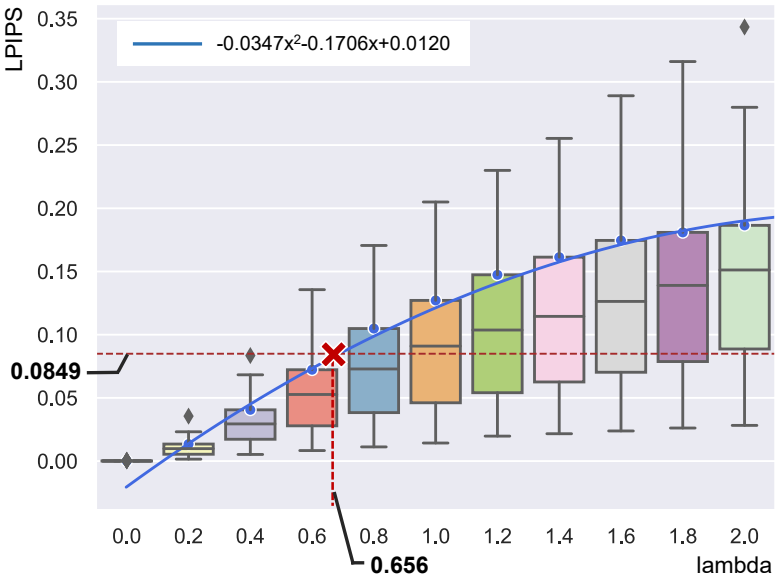

Figure 6.4: The determination of the penalty coefficient λ. The curve of quadratic regression is colored in blue. The red cross point is the intersection of the desired λ (=0.656) and the perceptually indistinguishable threshold of LPIPS (=0.0849).

Then we update the t_{ij} using the s_{ij}:

$$t_{ij} \leftarrow t_{ij} + \lambda s_{ij} \tag{6.3.7}$$

where λ is the penalty coefficient.

Determination of the penalty coefficient λ

By applying the alleviation method, we compromise the replication quality of the dehaze image while suppressing the total brightness of the image shown on the OST-HMD. Since the method includes the hyper-parameter λ, we seek an objective way to choose a proper one with reasonable justification. Note that $\lambda = 0$ is equivalent to the base method.

In general, a larger λ degrades the resulting dehaze view while reducing the total value of the OST-HMD pixels. If the degraded view is not perceptually visible for the humans, we may accept that result and use the corresponding λ.

To determine λ that does not degrade the dehaze view, one possible way is to increase

λ for a given hazy scene unless the perceptual difference between the resulting dehaze view and the view given by the base method (i.e., when $\lambda = 0$) is unnoticeable. Since the approach is impractical as it increases the computation time excessively, instead, we apply the same principle on an image dataset.

We use a public hazy image dataset (RESIDE [152][1] Synthetic Objective Testing Set). The dataset is well populated in the community and used for developing various dehaze methods [145,153]. The dataset contains 1000 image pairs of ground truth images, and their corresponding hazy images that are synthesized by their depth images and the scattering model [141].

Figure 6.4 shows our determination process of λ. Using their hazy images and the approach we mentioned above, we find the best λ by measuring the perceptual similarity between the base dehazed image computed with $\lambda = 0$ and that with different λ. We searched λ from 0 to 2.0 with the step 0.2, and computed the third quartile (75%) of the perceptual similarity of 100 hazy images randomly chosen from the dataset. We then fit the curve to the 75% values, and found the λ which gives the perceptual similarity value that humans cannot distinguish.

For the perceptual similarity, we actually used Learned Perceptual Image Patch Similarity (LPIPS) metric, version 0.1 [154]. LPIPS is a metric based on deep learning and outperforms compared to other major image metrics in terms of perception. Since LPIPS does not provide the value that humans consider two images are the same, we evaluated the LPIPS and decided the value for our application as the following. From the LPIPS's dataset, we took the just noticeable difference (JND) image dataset where 4800 pairs of original and visually-distorted small image patches (64×64). Each pair has a label showing that if three human observers saw each pair in a short amount of time (250 msec) could see differences. On the JND dataset, we computed LPIPS values over image pairs of which the all observers could not see the difference. As the result, we get the threshold LPIPS value 0.0849 as the median of the values.

By using the LPIPS and its perceptual threshold value, we obtained the threshold λ as 0.656, which we use in the evaluation in Sec. 6.5.

[1] https://sites.google.com/view/reside-dehaze-datasets/

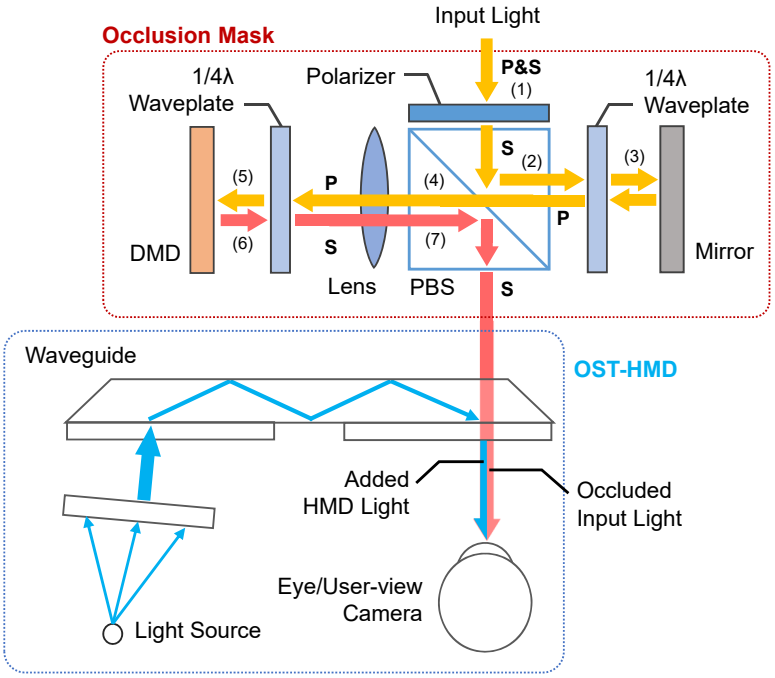

Figure 6.5: Schematic diagrams of our optical designs and prototype systems.

6.4 Technical Setup

We describe our hardware/software setups and calibration steps required for the experiment.

6.4.1 System Overview

Figure 6.5 shows a schematic diagram of our optical dehazing system. We combine a DMD-based occlusion mask with a consumer OST-HMD to selectively modulate the incoming light by both additive and subtractive manner.

The optical design of the occlusion mask consists of a DMD, a polarized beam splitter (PBS), and $\lambda/4$-waveplate. The DMD controls the reflection intensity of the incoming light by rotating several hundred thousand micro-mirrors individually, which achieves higher contrast and provides brighter images than the transmissive LCD.

To occlude the incoming light while keeping the see-through view from the eye, we

implemented the optical design for the DMD similar to [155]. PBS is made of a birefringent material and splits the light depending on the polarization states: transmits p-polarized light and reflects s-polarized light at an angle of $90°$.

By leveraging the property of PBS, our setup realizes the following optical paths: (1) the incoming light is firstly s-polarized by a linear polarizer, and (2) reflected to right by the PBS. Then (3) the s-polarized light is converted into p-polarized by reflected on a mirror and passing through a $\lambda/4$ waveplate twice. (4) The p-polarized light passes through the PBS from right to left, and also passes through the lens. After that, (5) another $\lambda/4$ waveplate rotates the p-polarized light into circularly polarized, then (6) the DMD modulates the amplitude of the light and the waveplate again rotates the light into s-polarized. Finally, (7) the s-polarized light is again reflected by the PBS and exits in the same direction as the input light. As a result, the occlusion mask optically aligns with the view direction of the user's eye.

Note that the distance between the DMD and the lens determines where the occlusion mask is virtually located. We set this distance as same as the focal length of the lens, which means the occlusion mask is located at the infinite distance from the eye. Using the DMD for this optical design has an issue. The tilt angles of the micromirrors we used can only take ±12° with the tilt axis 45° diagonal to each mirror pixel. This led us to slightly tilt the DMD so that the mirrors face perpendicular to the incoming light at the path (5).

6.4.2 Scene Camera

In addition to the above setup, we also need a scene camera which captures the hazy view $\mathbf{H}$. Ideally, the camera should be aligned with the viewpoint of the user's eye via a half mirror [C2]. In our proof-of-concept system, we leave this design factor. Instead, we focused on evaluating the full potential of our see-through dehaze system by placing a user-perspective camera at the viewpoint, which serves $\mathbf{H}$.

6.4.3 Hardware and Software setup

Figure 6.6 shows our hardware setups. For the DMD and its controller, we used DLP LightCrafter 6500 from Texus Instruments (1920×1080 pixels). The DMD is connected

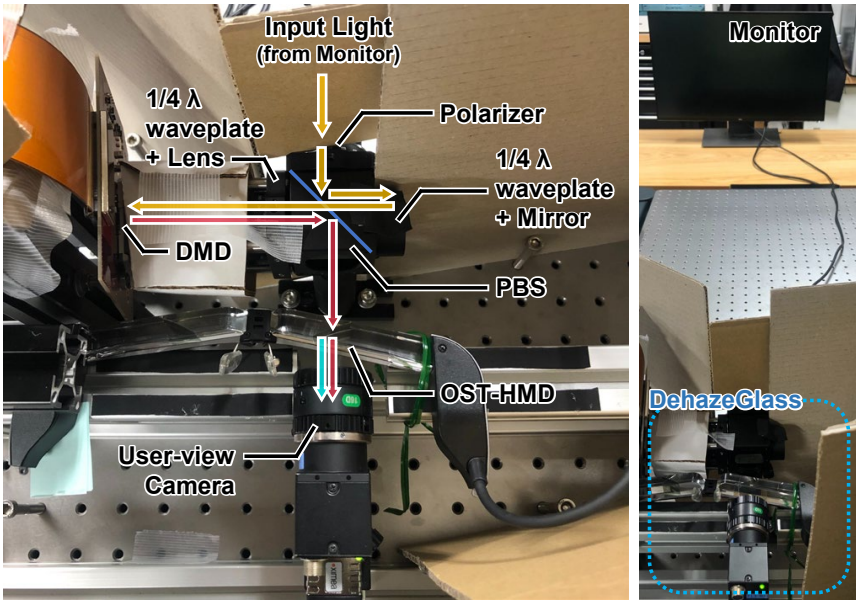

Figure 6.6: (left) Hardware setup with schematic visualization of light paths corresponding to Figure 6.5. (right) Input light (i.e. hazy scene) is coming from the monitor placed in front of our system.

to a laptop by the HDMI interface via the controller. For the OST-HMD, we used Epson Moverio BT-30E (23° field of view diagonally, 1920 × 1080 pixels). Other components were from Thorlabs: LPVISE100-A for linear polarizers, WPQ10ME-546 for the $\lambda/4$ waveplates, PBS251 for the PBS, BB1-E02 for the mirror, and AC254-045-A for the lens. We installed Point Grey FL3-U3-88S2C-C (4096 × 2160 pixels) as the user-perspective camera and set its gamma value as 1.0. A monitor displaying the hazy images (3840 × 2160 pixels) was placed 1.6 m away from our system. We set the camera focus and the position of the display according to the focal plane of the OST-HMD.

6.4.4 System Calibration

To realize dehaze functionality that is equivalent to the formulation we denoted in Sec. 6.3.1, spatial and color calibration is essential in our system. Therefore, we created some look-up tables (LUTs) which reflect our system properties in advance and correct input images based on them.

Pixel-wise mapping between displays

In the spatial correction, we need to warp images to display on the DMD and OST-HMD so that they exactly overlap an image on the monitor. To achieve this, we used a gray code pattern and obtained pixel correspondences between the user-perspective camera and each display. We then computed approximate projection functions with the neural net fitting of MATLAB. This function is a simple two-layer neural network (Input (2D) $\Rightarrow$ Affine (10D) $\Rightarrow$ Sigmoid $\Rightarrow$ Output (2D)) that takes the 2D coordinates of the OST-HMD or DMD as input and the 2D coordinates of the viewpoint camera image as output. We optimize this network by Levenberg-Marquadt algorithm. After that, we obtained the pixel-wise mappings from the DMD and OST-HMD to the monitor as LUTs by sampling from composite functions of them.

Color and Gamma Correction on each display

In the color correction, we need to correct input images to the DMD, OST-HMD, and monitor, so that the camera response to each color channel becomes linear and color balance is unified. For the OST-HMD and monitor, we displayed RGB colors one by one in 256 levels on each display and measured by the camera. And for the DMD, we input monotone images in 256 levels to the DMD and measured by the camera while displaying a white background on the monitor. We then computed approximate power functions and created LUTs by sampling from composite functions of them.

6.5 Experiments

We evaluated our method through an objective evaluation with quantitative measures.

6.5.1 Experiment setup

To conduct a fair quantitative evaluation, we introduced two conditions in the setup.

Firstly, we placed a user-perspective camera behind the glasses as mentioned in Sec. 6.4.2. We used the camera to capture hazy images $\mathbf{H}$ and dehazed images $\mathbf{D}$. Secondly, we used

hazy images instead of using the real scenes. Since it is hard to collect ground truth dehaze view of different real scenes, we substituted the real view with a monitor screen and display various hazy scenes. For the scenes, we again used the dataset introduced in Sec. 6.3.3. The images displayed on the monitor were scaled so that their short side is 2000 pixels, which is close to the diameter of the visible area of the monitor from the user-perspective camera. Given the setup, we conducted our evaluation as shown below.

6.5.2 Experiment procedure

Our goal is to evaluate if the observed dehaze image is indistinguishable from the true dehaze view. For clarity of the evaluation, we (re)define several dehaze images:

- $\mathbf{D}$ is a ground truth dehaze image from the dataset, which will be displayed on the monitor
- $\widehat{\mathbf{D}}$ is an estimated dehaze image computed from a hazy image $\mathbf{H}$ using the existing dehaze algorithm,
- $\mathbf{D}^{\mathrm{obs}}$ is an observed ground truth dehaze image captured by the user-perspective camera while the monitor is displaying $\mathbf{D}$, and
- $\widehat{\mathbf{D}}^{\mathrm{obs}}$ is an observed dehaze image captured by the user-perspective camera while the system displaying $\mathbf{T}$ on the DMD and $\mathbf{O}$ on the OST-HMD and the monitor displaying $\mathbf{H}$.

Note that there are two options for the choice of $\mathbf{D}$. The first option is the ground-truth dehazed scene images provided in the dataset. The second option is dehazed scene images we obtain by applying the dehaze method (FFA-Net) to the hazy scene images. In the experiment, we consider both options and provide their results for comparisons.

Finally, our goal is to measure the performance of $\widehat{\mathbf{D}}^{\mathrm{obs}}$ against the $\mathbf{D}^{\mathrm{obs}}$ for given sample scenes. Since the scene can be arbitrary, we used the image dataset in Sec. 6.3.3 and capture 80 different scenes and obtain a set of $\{(\mathbf{D}_k^{\mathrm{obs}}, \widehat{\mathbf{D}}_k^{\mathrm{obs}})\}_k$ for $\mathbf{D}$ from the ground truth images and the dehazed images via FFA-Net.

To compare $\mathbf{D}_k^{\mathrm{obs}}$ and $\widehat{\mathbf{D}}_k^{\mathrm{obs}}$, we used both the L2 matrix norm and the LPIPS metric. As mentioned in Sec. 6.3.3, we considered the LPIPS metrics as a more suitable perceptual metric to compare image pairs.

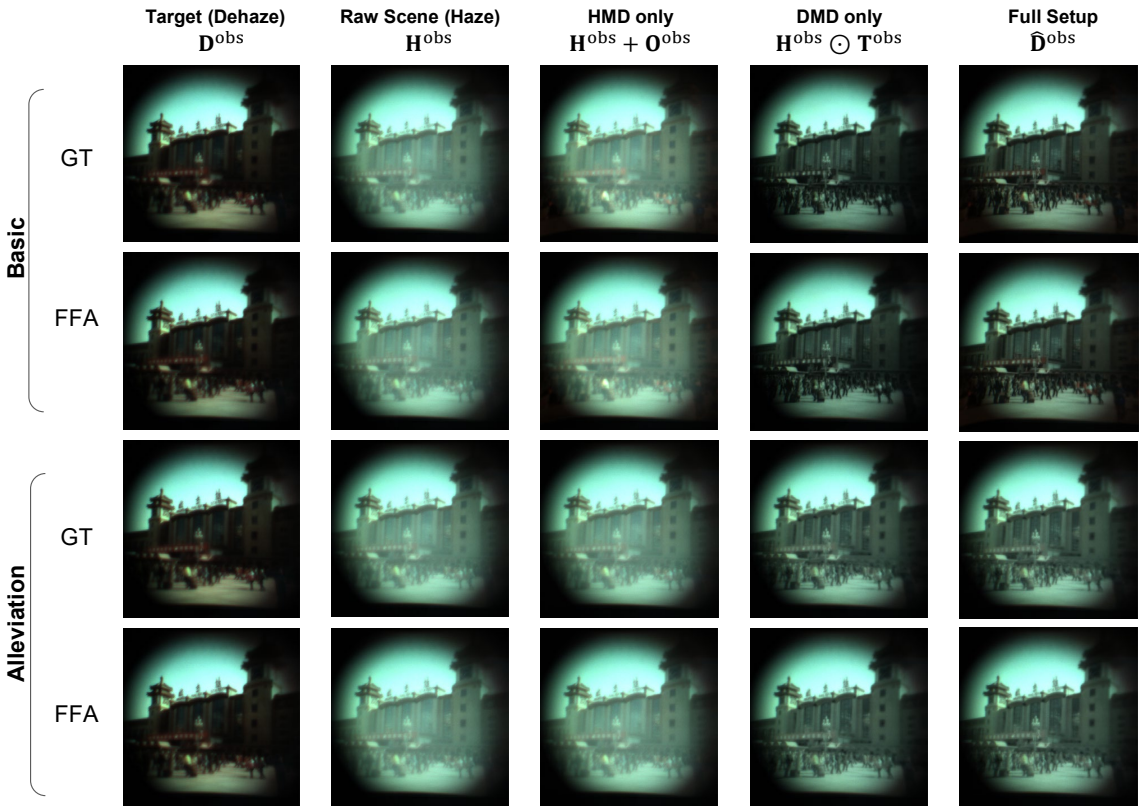

Figure 6.7: Results of haze removal by different combinations of the displays.

6.5.3 Experiment Result

Dehaze capability

Figure 6.7 shows the result images observed from our haze removal systems with different computation methods, target images, and combinations of the displays. Each row represents the experimental conditions: the basic computation (Sec. 6.3.2) or the alleviation of the complete occlsuion (Sec. 6.3.3), and the image regarded as the target in the computation: the ground truth dehaze image from the dataset (GT) or the estimated dehaze image using FFA-Net [145] (FFA). Each column represents the observed target scene ($\mathbf{D}^{\mathrm{obs}}$ or $\widehat{\mathbf{D}}^{\mathrm{obs}}$), the hazy scene, and the observed image with different combination of the displays. We selected the image where the difference of LPIPS between the Raw and the Scene+HMD+DMD is the median from the dataset.

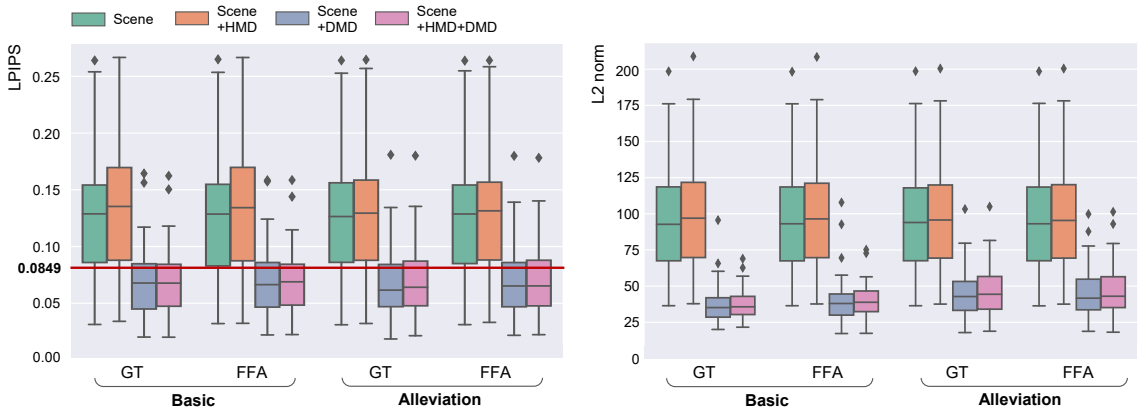

Figure 6.8: The overview of the dehaze results as described in Sec. 6.5.3. (left) With the LPIPS metric. (right) With the L2 matrix norm. Over the plots, our dehaze system (HMD+DMD) clearly reduced the similarity metric between the final view and the target dehazed view, which concludes our method successfully realized see-through view dehazing. The red line in the LPIPS plots refers to the line where our perceptual judgment analysis provided as a noticeable threshold.

Through the experiments, we observed that the choice of target images ($\mathbf{D}$ or $\widehat{\mathbf{D}}$) did not affect the experiment results and the conditions always brought similar results. For the difference of the calculation method, although we qualitatively confirmed that the final view of the alleviation method provides lower contrast, the overall image quality of the alleviation method was improved from the raw scene and perceptually acceptable.

Figure 6.8 visualizes the distribution of LPIPS and L2 matrix norm taken under different screen conditions. Applying either the base method or the alleviation method both clearly shows that our DehazeGlasses reduced the similarity metric between the final view and the target dehazed view, so we think our system successfully dehazed the user's see-through view. Our system also realized the LPIPS mean values lower than the noticeable threshold (red line) that our perceptual judgment analysis provided in Sec. 6.3.3 . In other words, it suggests that our system can dehaze the scene view which is perceptually as same as the original dehazed view.

Interestingly, applying DMD only also improved the see-through view whereas the change adding HMD (HMD+DMD) did not further improve the view. This suggests attenuating scene light correctly has a bigger impact on dehaze application than adding colors

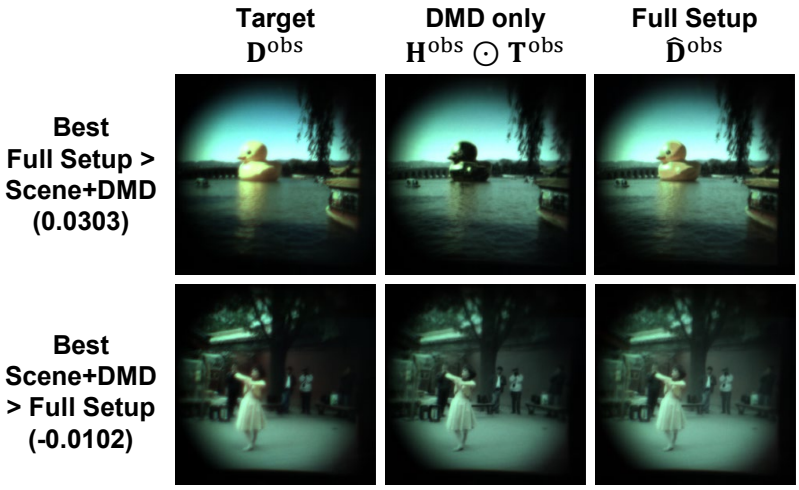

Figure 6.9: Results that the full setup is most superior to the DMD-only setup, and vise versa in the basic-GT condition. The values under the row legends are the differences of LPIPS between the scene+DMD and the full setup.

on top. Fig. 6.9 shows the scenes that the full setup is most effective than the DMD-only setup and vise versa. This result suggests the full-setup is more effective when the scene contains high-saturation pixels.

Alleviation result

The motivation in introducing the alleviation method was to let the OST-HMD display images as minimum as possible. To verify this, we draw a curve by calculating the ration of the pixels in an OST-HMD image $\mathbf{O}$ that is lower than a given threshold. By moving the threshold 0 to 255, we obtain a curve starting from (0,0) and ending at (255, 1.0). The area under the curve (AUC) represents how bright a given image is. When an image is completely black, then the AUC takes its maximum value. We compute this response curve for all input OST-HMD images used in the experiment and draw the mean curves for different conditions (Fig. 6.10).

The result shows that the AUC of the alleviation methods clearly shows a larger AUC than the base method, which confirms that the alleviation method was successfully reducing the required image brightness of the OST-HMD.

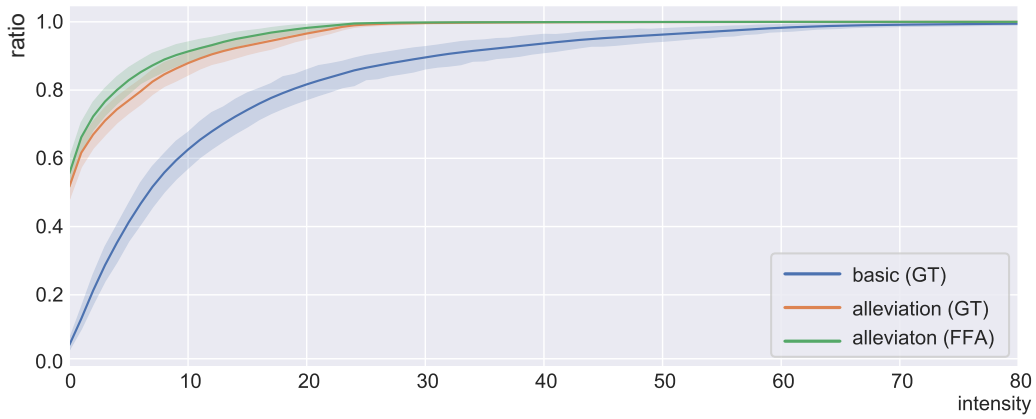

Figure 6.10: The intensity response curves with quartile areas for the comparison between the base method and the alleviation method. The alleviation method gives curves with larger AUCs than that of the basic method, which shows the alleviation method suppressed an excessive use of the OST-HMD. Note that we only show the results of the threshold 0 to 80 since the ratio can be regarded as 1.0 when the threshold is more than 80. We do not show the base (FFA) since it gave the almost identical result as the base (GT).

6.6 Discussions and Limitations

While our proof-of-concept system showed a promising result, there is still a large gap between the current system and a practical system used in a real life. Through the section, we describe issues of the current implementation and provide future research directions.

6.6.1 Aligning image screens

Our system defines several image screens (DMD, HMD and the scene camera). As we mentioned in the calibration section (Sec. 6.4.4), the ideal system requires real-time dynamic calibration that calibrates the system with respect to the current position of the user's eye. Such automated calibration has been a topic in OST-HMD community [133], yet not have been explored deeper in occlusion-capable OST-HMD scenarios with scene cameras.

6.6.2 Color occlusion capability

The current system uses a DMD for occlusion. Since a DMD is essentially a dynamic mirror, each micromirror of the DMD merely controls the transmittance of the user's see-through view regardless of the colors. In other words, our occlusion system cannot attenuate a specific color band of the scene view only. If a system can attenuate each color channel separately, the system less needs to display compensation colors on the OST display (Sec. 6.3.3). We think it is desirable that the OST layer renders as less intrusive images as possible following the discussion on the disadvantages of using a VST system in Sec. 1.3.1.

Existing color-occlusion displays use, for example, transmissive color LCD [156, 157] and reflective liquid crystal on silicon device [95, 155].

6.6.3 Depth of field

In practice, most of the hazy scenes at which the users see locate far enough to be treated as a 2D image captured from the infinite distance. Since our system renders a 2D image layer and occlusion layer, our dehaze approach works fine while the layers are optically focused at a far distance.

This 2D scene assumption is, however, invalid when the hazy scene is close to the user's view. For example, a garden scene with fog or smoke could be hazy and contain scene objects close enough to the user, which causes a deep depth of field in the user's view. Under such scenarios, relying on the 2D image model causes image blur or occlusion blur [97] due to the mismatch of the screens' depth and the user's focus depth.

To solve this issue, a possible solution is to make the system varifocal, i.e. making the image plane of the OST display and the occlusion layer are focus-tunable. Varifocal OST-HMDs have gained interest from the display community [158]. Researchers also investigated realizing varifocal occlusion-capable displays [88, 99].

6.6.4 Color Correction

Calibration the color property of each screen and the scene camera is an important issue in practice. In the evaluation, we calibrated the system in a global manner where we applied

the same color pre-distortion for each pixel of a screen. Since each pixel may have slightly different color changes, the local color calibration method which calibrates color pixel-wise would improve the system performance [60].

6.6.5 Scene Camera

As briefly mentioned in Sec. 6.4.2, the scene camera is a vital part of the system to provide the basis of the user's current first person view. There are all issues omitted in this paper: view alignment, latency, resolution, and so on.

6.6.6 Hardware minimization

The current DehazeGlaases is a proof-of-concept system built on an optical bench. For practical use, the system needs to be a wearable system. Although the development of a wearable OST-HMD system without occlusion is still a challenge, the community has shown wearable designs of occlusion-capable displays [101].

6.7 Summary

In this work, we proposed DehazeGlasses, a see-through view dehazing system using an occlusion-capable OST-HMD. We firstly formulated the computation for displaying images to realize the dehaze functionality for our system, and explored the parameter that provides perceptually reasonable dehazed scene while keeping the actual view. Then We implemented the proof-of-concept system of DehazeGlasses that consists of a DMD for occlusion layer and OST-HMD for image overlay. By using the proof-of-concept system, we verified that the result with dehaze image dataset shows that our system can clearly improve hazy views, and the perceived views from our system are difficult to perceive differences from the target dehazed view. Finally we provide the limitation of the current implementation and provide future research directions.

Part IV

Conclusions and Future Work

7

Conclusions and Future Work

We present the conclusion of this dissertation, and suggest future research directions.

7.1 Conclusion

This dissertation explored an approach to reproduce a highly realistic appearance by directly measuring and modulating the light rays entering the eye in an OST-HMD. The ultimate goal of this research is to reproduce an appearance that does not differ from reality with OST-HMD. In other words, the ideal is to present a virtual image indistinguishable from a real object when viewed through an OST-HMD. Although there is still a long way to reach this goal, we have achieved a previously unattainable appearance expression on OST-HMDs, mainly by improving dynamic range and reproducing DoF.

We proposed Focal Surface Occlusion (Chap. 4) as a method to reproduce a highly realistic DoF for OST-HMDs that can present images with HDR range. By directly modulating the phase of light entering the eye using PSLM, this method can reproduce DoF on the occlusion mask with wider FoV and resolution than conventional methods while maintaining real-world visibility. We confirmed that the proposed method can present a occlusion mask with a resolution comparable to that of commercial OST-HMDs, although it does not reach the resolution of the retina.

We also proposed AdaptiVisor (Chap. 5), a system that uses direct measurement and

modulation of light rays to change the appearance of reality and assist the eye in adaptation. This system acquires light entering the eye using coaxial optics and presents the acquired brightness-compensating image on the OCOST-HMD. The proof-of-concept prototype enabled high dynamic range appearance modulation at the viewpoint with a delay of about 0.5 seconds.

Finally, we proposed DehazeGlasses (Chap. 6) as a method to remove haze at the viewpoint by directly modulating the amplitude of light rays entering the eye using the OCOST-HMD. The system converts the image on the viewpoint to HDR by passing it through a haze removal algorithm and then selectively modulates the intensity of the light rays entering the eye based on the result. The implemented proof-of-concept system achieves a dehaze vision that is significantly closer to the ground-truth scene than the native view under a perceptual image similarity metric.

7.2 Future Work

We described the discussions and future research on individual systems in each chapter. In this section, we discuss the possible research direction from a mid- to long-term vision on the approach throughout this paper, realistic appearance reproduction on OST-HMD and its VA application based on the direct measurement and modulation of light rays.

7.2.1 Light Measurement and Modulation with Feedback Control

While our system can measure the user's view from a viewpoint, the measurement and modulation are executed only in a single, sequential pipeline. However, the environment and the displacement between devices changes from time to time. Therefore, we can expect to be able to reproduce more realistic appearance based on feedback control, e.g. re-measuring the modulated real-view and learning the difference between the target and the observation.

Recently, in computer-generated holograms, the framework using camera-in-the-loop optimization achieves unprecedented image fidelity and real-time framerates [159, 160]. In this method, the generated hologram is captured by a camera. Then, the neural network

architecture, which interprets the propagation of light waves, learns the difference between the observation and the target, and updates the phase image. Camera-in-the-loop training has the advantage of implicitly compensating for hardware-dependent errors such as device misalignment and response calibration of the PSLM, in addition to improving the quality of the appearance through an iterative update.

We believe that this camera-in-the-loop learning is highly compatible with our approach. For example, while we are currently using a geometrical optics formulation for Focal Surface Occlusion, we believe that learning with a neural network that takes wavefront propagation into account will allow us to reproduce more realistic DoF closer to the simulation.

7.2.2 Using Computational Optics with Embedded Optimization

In this dissertation, we uses the measured FoV as input to a specific optimized function (e.g., Dehazing), and the OST-HMD presents a computed compensation image. Currently, for this optimized function, we apply existing machine vision software. On the other hand, there are many cases where information cannot be estimated or complemented by software processing alone, or where software computation becomes a bottleneck in terms of computational resources and execution time on the embedded system.

The strength of our OST-HMD framework is that it can integrate both sensing hardware and software optimization. In such a framework, it is possible to process data by both hardware and software. For example, it is possible to use software to supplement information that is difficult to measure with sensors, or to create sensing hardware that focuses on information that cannot be supplemented by software.

One such approach is to use an optical system with an optimization calculation embedded in the sensor used to measure light rays. For example, some research implements a single layer of a convolutional neural network on diffraction optics by optimizing the grating [161], or an encoded lens that can simplify the algorithm for HDR image restoration [162]. Utilizing such computational optics is expected to improve the appearance reproduction.

7.2.3 Image Presentation with Correction for Eye Aberration

The combination of a PSLM with an OST-HMD can not only reproduce DoF on virtual objects but compensate for real-world vision like prescription glasses. For such applications, it is necessary to measure the aberration of the eye lens in real-time, such as the focal depth and visual acuity.

In ophthalmic autorefractors, we estimate such aberration by first irradiating the light into the eye lens center, then measuring the wavefront reflected by the retina and exiting through the lens. Some small autorefractors utilize a Shack-Hartmann wavefront sensor (SHWFS) [163], which consists of a microlens array and an image sensor placed at the focal length of the microlens array. Since the periodicity of the microlens array limits the measurable range of displacement, the dynamic range of SHWFS is narrow, reported as a range from -4.30 D to +4.30 D [106]. Due to the limited dynamic range and large form factor, an SHWFS has not been adopted as a measurement device in the context of HMDs.

A Diffusive Wavefront Sensor (DWFS) is a new, computational approach that aims to overcome the limitation of SHWFS [106]. A DWFS places a thin holographic diffuser close to the image sensor and detects the caustic pattern of light passing through the diffuser. The caustic pattern changes according to the passing light's wavefront, so we can estimate the wavefront by detecting its displacement. DWFS can greatly expand the dynamic range of measurable diopters, reported as a range from -12 D to +12 D, even when the size and spacing of sensors. Therefore, we believe that DWFS will be useful for AR displays that require miniaturization.

Based on the idea above, we proposed an accommodation-supporting display by measuring the eye accommodation with DWFS and displaying the image reproducing DoF on a RPD [C4]. Since DWFS does not yet support real-time measurements, extending the measurement algorithm to real-time is a promising direction. Also, we expect to realize OCOST-HMD that can compensate for the real-view by convolution of the phase image compensating for the measured aberration with the original phase image.

7.2.4 Low-latency VA System

Although this dissertation focused on improving appearance reproduction on OST-HMDs, we have not yet considered the delay in the presentation of visual stimuli.

In daily life, people process various information instantaneously to make decisions. VA encourages the wearer to make decisions by displaying images according to the information received by humans. However, if the augmentation is delayed, it conversely inhibits instantaneous decisions. For example, while driving, the surroundings change at a speed beyond human perception, and the eyes cannot keep up with the rapid changes in ambient light. On the other hand, the current AdaptiVisor has a delay of about 0.5 seconds between the detection of a change in ambient light and the presentation of a visual stimulus.

Suppose we realize a low-latency VA system that presents visual stimuli at speed beyond human perception. In that case, it will improve the discomfort of wearing VA systems and promote quick judgment in various daily situations such as driving, sports, and task support, expanding VA's applicability in the real world. Specifically, it is desirable to present a visual stimulus less than 16 ms from the timing of a sudden change in vision, which is the critical fusion frequency that indicates the limit of human discrimination of flicker [164].

7.3 Closing Remark

I have been conducting this research with the vision of *creating a technology that allows all people to freely share sensations and imagination*, and *create various experiences and impressions by freely rewriting reality with information technology*. The technology to re-design vision sensation partially updates the user's understanding of the world through information technology. If a world in which all people can freely share their senses and images is realized, we can achieve self-realization by transforming ourselves into the person we envision. Moreover, we expect to realize a social infrastructure in which we can understand others and co-exist in diversity and creativity by reproducing and sharing the experience of others.

In this research, we have consistently pursued the technology *to reproduce a virtual*

image that is optically indistinguishable from reality, and have come one step closer to our vision. On the other hand, this paper only shows that it is possible to rewrite the optically faithful *appearance* on an OST-HMD by measuring and modulating the light reaching the eye. To move from the reproduction of *appearance* to the reproduction of *experience*, we believe that we need to estimate how people feel when they see the appearance. In order to do so, a deep understanding of the world perception of each individual, including human visual characteristics, behavioral patterns, and cognitive psychology, is required. In that sense, this dissertation is a new step toward that distant vision.

List of Figures

1.1 AR display categories according to the distance between the eye and the display 7

2.1 Classification of Accommodation-Supporting Displays 15
2.2 Classification of OCOST-HMD designs 18

4.1 Schematic of the light path in focal surface occlusion 34
4.2 Optimization of PSLM as a free-form lens for the given depth map 39
4.3 The achievable DoF of our focal surface occlusion 43
4.4 Schematic of the optical design of the focal surface occlusion prototype . . 44
4.5 Comparison of the focal blur reproduced by each method for a virtual occlusion at continuous depth 49
4.6 Normalized MTF curves in all experiment conditions in Sec. 4.6.2 50
4.7 patial resolution at which half-contrast can be reproduced 51
4.8 Examples of applying phase images with edge-based optimization to complex scene 54
4.9 Schematic of the optical design when adding a virtual image to the current prototype 57

5.1 Schematic of the optical design when adding a virtual image to the current prototype 66
5.2 Algorithm flow of the computation of the mask and HMD images. 67
5.3 Hardware setup with schematic visualization of light paths in Fig. 5.1 . . . 68
5.4 Transmittance profile of the LCD panel used in our setup 70
5.5 System calibration overview of our AdaptiVisor system 72
5.6 Experiment setup for our AdaptiVisor prototype 73
5.7 Overview of the static-capture experiment 73
5.8 Overview of the realtime experiment 75

6.1 A schematic diagram of a generic hardware formation of our DehazeGlasses setup 84

6.2 Overview of the computation of the transmittance matrix $\mathbf{T}$ and the displayed image $\mathbf{O}$. 86
6.3 Visualization of the complete occlusion issue explained in Sec. 6.3.3. 88
6.4 The determination of the penalty coefficient λ 89
6.5 Schematic diagrams of our DehazeGlasses prototype 91
6.6 Hardware setup with schematic visualization and actual prototype 93
6.7 Results of haze removal by different combinations of the displays 96
6.8 The overview of the dehaze results as described in Sec. 6.5.3 97
6.9 Results that the full setup is most superior to the DMD-only setup, and vise versa in the basic-GT condition. 98
6.10 The intensity response curves with quartile areas for the comparison between the base and the alleviation method 99

List of Tables

2.1 Comparison of occlusion methods for OST-HMDs adapted from [88] 19

Bibliography

[1] O. Bimber and R. Raskar, *Spatial augmented reality: merging real and virtual worlds*. CRC Press, 2005.

[2] J. Erkoyuncu and S. Khan, "Olfactory-based augmented reality support for industrial maintenance," *IEEE Access*, vol. 8, pp. 30306–30321, 2020.

[3] Y. Hashimoto, N. Nagaya, M. Kojima, S. Miyajima, J. Ohtaki, A. Yamamoto, T. Mitani, and M. Inami, "Straw-like user interface: virtual experience of the sensation of drinking using a straw," in *Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology*, p. 50, ACM, 2006.

[4] H. Nakamura and H. Miyashita, "Augmented gustation using electricity," in *Proceedings of the 2nd Augmented Human International Conference*, AH '11, (New York, NY, USA), Association for Computing Machinery, 2011.

[5] J. Vallino and C. Brown, "Haptics in augmented reality," in *Multimedia Computing and Systems, 1999. IEEE International Conference on*, vol. 1, pp. 195–200, IEEE, 1999.

[6] T. Nojima, D. Sekiguchi, M. Inami, and S. Tachi, "The smarttool: a system for augmented reality of haptics," in *Proceedings of IEEE Virtual Reality 2002*, pp. 67–72, IEEE, 2002.

[7] M. Sra, X. Xu, and P. Maes, "Galvr: A novel collaboration interface using gvs," in *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, VRST '17, (New York, NY, USA), Association for Computing Machinery, 2017.

[8] C.-Y. Yen, K.-H. Lin, M.-H. Hu, R.-M. Wu, T.-W. Lu, and C.-H. Lin, "Effects of virtual reality–augmented balance training on sensory organization and attentional demand for postural control in people with parkinson disease: a randomized controlled trial," *Physical Therapy*, vol. 91, pp. 862–874, Apr 2011.

[9] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino, "Augmented reality: A class of displays on the reality-virtuality continuum," in *Photonics for Industrial Applications*, pp. 282–292, International Society for Optics and Photonics, 1995.

[10] I. E. Sutherland, "The ultimate display," in *Proceedings of the Congress of the Internation Federation of Information Processing (IFIP) 65*, vol. 2, pp. 506–508, 1965.

[11] I. E. Sutherland, "A head-mounted three dimensional display," in *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, pp. 757–764, 1968.

[12] T. P. Caudell and D. W. Mizell, "Augmented reality: An application of heads-up display technology to manual manufacturing processes," in *Hawaii Inter. Conf. on System Sciences*, vol. 2, pp. 659–669, 1992.

[13] J. Weidenhausen, C. Knöpfle, and D. Stricker, "Lessons learned on the way to industrial augmented reality applications, a retrospective on arvika," *Comput. Graph.*, vol. 27, pp. 887–891, Dec 2003.

[14] C. Bichlmeier, F. Wimmer, S. M. Heining, and N. Navab, "Contextual anatomic mimesis hybrid in-situ visualization method for improving multi-sensory depth perception in medical augmented reality," in *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 129–138, 2007.

[15] T. Olsson, E. Lagerstam, T. Kärkkäinen, and K. Väänänen-Vainio-Mattila, "Expected user experience of mobile augmented reality services: a user study in the context of shopping centres," *Personal and Ubiquitous Computing*, vol. 17, pp. 287–304, Feb 2013.

[16] W. Piekarski and B. Thomas, "Arquake: the outdoor augmented reality gaming system," *Communications of the ACM*, vol. 45, no. 1, pp. 36–38, 2002.

[17] M. Billinghurst and A. Duenser, "Augmented reality in the classroom," *Computer*, vol. 45, no. 7, pp. 56–63, 2012.

[18] O. Bimber and R. Raskar, "Modern approaches to augmented reality," in *ACM SIGGRAPH 2006 Courses*, p. 1, ACM, 2006.

[19] R. Raskar, G. Welch, and H. Fuchs, "Spatially augmented reality," in *Proceedings of First IEEE Workshop on Augmented Reality (IWAR'98)*, pp. 11–20, Citeseer, 1998.

[20] B. Jones, R. Sodhi, M. Murdock, R. Mehra, H. Benko, A. Wilson, E. Ofek, B. MacIntyre, N. Raghuvanshi, and L. Shapira, "Roomalive: Magical experiences enabled by scalable, adaptive projector-camera units," in *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, (New York, NY, USA), p. 637–644, Association for Computing Machinery, 2014.

[21] D. Iwai, R. Matsukage, S. Aoyama, T. Kikukawa, and K. Sato, "Geometrically consistent projection-based tabletop sharing for remote collaboration," *IEEE Access*, vol. 6, pp. 6293–6302, 2018.

[22] C. Menk and R. Koch, "Truthful color reproduction in spatial augmented reality applications," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, pp. 236–248, Feb 2013.

[23] J. Kollin, "A retinal display for virtual-environment applications," in *SID International Symposium Digest of Technical Papers*, vol. 24, pp. 827–827, SOCIETY FOR INFORMATION DISPLAY, 1993.

[24] G. Westheimer, "The maxwellian view," *Vision research*, vol. 6, no. 11-12, pp. 669–682, 1966.

[25] C. Jang, K. Bang, S. Moon, J. Kim, S. Lee, and B. Lee, "Retinal 3d: augmented reality near-eye display via pupil-tracked light field projection on retina," *ACM TOG*, vol. 36, pp. 190:1–190:13, Nov. 2017.

[26] J. Kim, Y. Jeong, M. Stengel, K. Akşit, R. Albert, B. Boudaoud, T. Greer, J. Kim, W. Lopes, Z. Majercik, P. Shirley, J. Spjut, M. McGuire, and D. Luebke, "Foveated ar: Dynamically-foveated augmented reality display," *ACM Trans. Graph.*, vol. 38, July 2019.

[27] N. Davies, M. Langheinrich, P. Maes, and J. Rekimoto, "Augmenting humans," *IEEE Pervasive Computing*, vol. 17, pp. 9–10, Apr 2018.

[28] S. Mann and S. M. Nnlf, "Mediated reality," 1994.

[29] J. Sutton, T. Langlotz, and Y. Itoh, "Computational glasses: Vision augmentations using computational near-eye optics and displays," in *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pp. 438–442, 2019.

[30] S. Shekar, P. R. Pesaladinne, S. A. Karre, and Y. R. Reddy, "Vreye: Exploring human visual acuity test using virtual reality," in *Virtual, Augmented and Mixed Reality. Industrial and Everyday Life Applications* (J. Y. C. Chen and G. Fragomeni, eds.), (Cham), pp. 415–429, Springer International Publishing, 2020.

[31] F. Vargas-Martin and E. Peli, "Augmented-view for restricted visual field: Multiple device implementations," *Optometry and vision science : official publication of the American Academy of Optometry*, vol. 79, pp. 715–723, Dec 2002.

[32] O. Younis, W. Al-Nuaimy, F. Rowe, and M. H. Alomari, "A smart context-aware hazard attention system to help people with peripheral vision loss," *Sensors*, vol. 19, Apr 2019.

[33] J. Orlosky, Y. Itoh, M. Ranchet, K. Kiyokawa, J. Morgan, and H. Devos, "Emulation of physician tasks in eye-tracked virtual reality for remote diagnosis of neurodegenerative disease," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, pp. 1302–1311, Apr 2017.

[34] Y. Miao, J. Y. Jeon, G. Park, S. W. Park, and H. Heo, "Virtual reality-based measurement of ocular deviation in strabismus," *Computer Methods and Programs in Biomedicine*, vol. 185, p. 105132, Mar 2020.

[35] A. Nowak, M. Wozniak, M. Pieprzowski, and A. Romanowski, "Towards amblyopia therapy using mixed reality technology," in *2018 Federated Conference on Computer Science and Information Systems (FedCSIS)*, pp. 279–282, 2018.

[36] H. C. Ates, A. Fiannaca, and E. Folmer, "Immersive simulation of visual impairments using a wearable see-through display," in *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '15, (New York, NY, USA), p. 225–228, Association for Computing Machinery, 2015.

[37] P. R. Jones, T. Somoskeöy, H. Chow-Wing-Bom, and D. P. Crabb, "Seeing other perspectives: evaluating the use of virtual and augmented reality to simulate visual impairments (openvissim)," *npj Digital Medicine*, vol. 3, p. 32, Mar 2020.

[38] A. D. Hwang and E. Peli, "An augmented-reality edge enhancement application for google glass," *Optometry and vision science: official publication of the American Academy of Optometry*, vol. 91, p. 1021, Aug 2014.

[39] X. Hu, Y. Zhang, N. Isoyama, N. Sakata, and K. Kiyokawa, "Design and prototyping of computational sunglasses for autism spectrum disorders," in *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 581–582, 2021.

[40] T. Langlotz, J. Sutton, S. Zollmann, Y. Itoh, and H. Regenbrecht, "Chromaglasses: Computational glasses for compensating colour blindness," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, (New York, NY, USA), pp. 390:1–390:12, ACM, 2018.

[41] Y. Itoh and G. Klinker, "Vision enhancement: defocus correction via optical see-through head-mounted displays," in *Proceedings of the 6th Augmented Human International Conference*, pp. 1–8, 2015.

[42] P. Chakravarthula, D. Dunn, K. Akşit, and H. Fuchs, "Focusar: Auto-focus augmented reality eyeglasses for both real world and virtual imagery," *IEEE TVCG*, vol. 24, pp. 2906–2916, Nov. 2018.

[43] N. Padmanaban, R. Konrad, and G. Wetzstein, "Autofocals: Evaluating gaze-contingent eyeglasses for presbyopes," *Science Advances*, vol. 5, p. eaav6187, Jun 2019.

[44] J.-Y. Wu and J. Kim, "Prescription ar: a fully-customized prescription-embedded augmented reality display," *Opt. Express*, vol. 28, pp. 6225–6241, Mar 2020.

[45] D. C. Rompapas, A. Rovira, S. Ikeda, A. Plopski, T. Taketomi, C. Sandor, and H. Kato, "Eyear: Refocusable augmented reality content through eye measurements," in *15th IEEE ISMAR*, pp. 334–335, 2016.

[46] T. Tao, P. Ratsamee, Y. Uranishi, K. Kiyokawa, T. Mashita, and H. Takemura, "An interactive 4d vision augmentation of rapid motion," in *Proceedings of the 9th Augmented Human International Conference*, AH '18, (New York, NY, USA), pp. 1–4, Association for Computing Machinery, 2018.

[47] J. Orlosky, T. Toyama, K. Kiyokawa, and D. Sonntag, "Modular: Eye-controlled vision augmentations for head mounted displays," *IEEE Trans. Vis. Comp. Graph.*, vol. 21, pp. 1259–1268, Nov 2015.

[48] J. Nishida, H. Takatori, K. Sato, and K. Suzuki, "Childhood: Wearable suit for augmented child experience," in *Proceedings of the 2015 Virtual Reality International Conference*, VRIC '15, (New York, NY, USA), Association for Computing Machinery, 2015.

[49] K. Higuchi, Y. Ishiguro, and J. Rekimoto, "Flying eyes: Free-space content creation using autonomous aerial vehicles," in *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, (New York, NY, USA), p. 561–570, Association for Computing Machinery, 2011.

[50] K. Higuchi, K. Fujii, and J. Rekimoto, "Flying head: A head-synchronization mechanism for flying telepresence," in *2013 23rd International Conference on Artificial Reality and Telexistence (ICAT)*, pp. 28–34, 2013.

[51] S. Kasahara, S. Nagai, and J. Rekimoto, "Jackin head: Immersive visual telepresence system with omnidirectional wearable camera," *IEEE Tran. Vis. Comp. Graph.*, vol. 23, pp. 1222–1234, Mar 2017.

[52] R. Komiyama, T. Miyaki, and J. Rekimoto, "Jackin space: designing a seamless transition between first and third person view for effective telepresence collaborations," in *Proceedings of the 8th Augmented Human International Conference*, pp. 1–9, 03 2017.

[53] Y. Itoh, J. Orlosky, K. Kiyokawa, and G. Klinker, "Laplacian vision: Augmenting motion prediction via optical see-through head-mounted displays," in *Proceedings*

of the 7th Augmented Human International Conference 2016, AH '16, (New York, NY, USA), Association for Computing Machinery, 2016.

[54] J. Orlosky, P. Kim, K. Kiyokawa, T. Mashita, P. Ratsamee, Y. Uranishi, and H. Takemura, "Vismerge: Light adaptive vision augmentation via spectral and temporal fusion of non-visible light," in *Proceedings of 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 22–31, 2017.

[55] O. Bimber, A. Grundhö, fer, G. Wetzstein, and d. S. Knö, "Consistent illumination within optical see-through augmented environments," in *2nd IEEE ISMAR*, International Symposium on Mixed and Augmented Reality '03, pp. 198–207, IEEE Computer Society, 2003.

[56] Y. Wang and D. Samaras, "Estimation of multiple directional light sources for synthesis of augmented reality images," *Graphical Models*, vol. 65, no. 4, pp. 185–205, 2003.

[57] S. K. Sridharan, J. D. Hincapié-Ramos, D. R. Flatla, and P. Irani, "Color correction for optical see-through displays using display color profiles," in *ACM VRST 2013*, pp. 231–240, 2013.

[58] J. David Hincapie-Ramos, L. Ivanchuk, S. K. Sridharan, and P. Irani, "Smartcolor: Real-time color correction and contrast for optical see-through head-mounted displays," in *IEEE ISMAR*, pp. 187–194, 2014.

[59] Y. Itoh, M. Dzitsiuk, T. Amano, and G. Klinker, "Semi-parametric color reproduction method for optical see-through head-mounted displays," *IEEE Trans. Vis. Compute. Graph*, vol. 21, pp. 1269–1278, Nov. 2015.

[60] T. Langlotz, M. Cook, and H. Regenbrecht, "Real-time radiometric compensation for optical see-through head-mounted displays," *IEEE Trans. on Visualization and Computer Graphics*, vol. 22, pp. 2385–2394, Nov. 2016.

[61] N. Matsuda, A. Fix, and D. Lanman, "Focal surface displays," *ACM Trans. Graph.*, vol. 36, pp. 86:1–86:14, Aug 2017.

[62] S. Suyama, M. Date, and H. Takada, "Three-dimensional display system with dual-frequency liquid-crystal varifocal lens," *Japanese Journal of Applied Physics*, vol. 39, no. 2R, p. 480, 2000.

[63] S. Liu, D. Cheng, and H. Hua, "An optical see-through head mounted display with addressable focal planes," in *Proceedings of 7th IEEE International Symposium for Mixed and Augmented Reality (ISMAR)*, pp. 33–42, 2008.

[64] R. Konrad, N. Padmanaban, K. Molner, E. A. Cooper, and G. Wetzstein, "Accommodation-invariant computational near-eye displays," *ACM TOG*, vol. 36, pp. 88:1–88:12, July 2017.

[65] K. Rathinavel, H. Wang, A. Blate, and H. Fuchs, "An extended depth-at-field volumetric near-eye augmented reality display," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 11, pp. 2857–2866, 2018.

[66] X. Xia, Y. Guan, A. State, P. Chakravarthula, K. Rathinavel, T.-J. Cham, and H. Fuchs, "Towards a switchable ar/vr near-eye display with accommodation-vergence and eyeglass prescription support," *IEEE Trans. Vis. Comput. Graph*, vol. 25, pp. 3114–3124, Nov 2019.

[67] D. Dunn, C. Tippets, K. Torell, P. Kellnhofer, K. Aksit, P. Didyk, K. Myszkowski, D. Luebke, and H. Fuchs, "Wide field of view varifocal near-eye display using see-through deformable membrane mirrors," *IEEE Trans. on Vis. Comput. Graph.*, vol. 23, no. 4, pp. 1322–1331, 2017.

[68] D. Dunn, P. Chakravarthula, Q. Dong, and H. Fuchs, "Mitigating vergence-accommodation conflict for near-eye displays via deformable beamsplitters," in *Digital Optics for Immersive Displays*, vol. 10676, p. 106760U, 2018.

[69] K. Akşit, W. Lopes, J. Kim, P. Shirley, and D. Luebke, "Near-eye varifocal augmented reality display using see-through screens," *ACM Trans. Graph.*, vol. 36, nov 2017.

[70] J. P. Rolland, M. W. Krueger, and A. Goon, "Multifocal planes head-mounted displays," *Appl. Opt.*, vol. 39, pp. 3209–3215, July 2000.

[71] K. Akeley, S. J. Watt, A. R. Girshick, and M. S. Banks, "A stereo display prototype with multiple focal distances," in *ACM Trans. Graph.*, vol. 23, pp. 804–813, 2004.

[72] Y.-C. Liu and M.-H. Wen, "Comparison of head-up display (hud) vs. head-down display (hdd): driving performance of commercial vehicle operators in taiwan," *International Journal of Human-Computer Studies*, vol. 61, no. 5, pp. 679–697, 2004.

[73] R. Narain, R. A. Albert, A. Bulbul, G. J. Ward, M. S. Banks, and J. F. O'Brien, "Optimal presentation of imagery with focus cues on multi-plane displays," *ACM Trans. Graph.*, vol. 34, July 2015.

[74] 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, pp. 237:1—-237:15, Dec 2017.

[75] S. Liu, H. Hua, and D. Cheng, "A novel prototype for an optical see-through head-mounted display with addressable focus cues," *IEEE Trans. Vis. Comput. Graph.*, vol. 16, pp. 381–393, May-Jun 2010.

[76] J.-H. R. Chang, B. V. K. V. Kumar, and A. C. Sankaranarayanan, "Towards multifocal displays with dense focal stacks," *ACM Trans. Graph.*, vol. 37, dec 2018.

[77] X. Hu and H. Hua, "High-resolution optical see-through multi-focal-plane head-mounted display using freeform optics," *Optics Express*, vol. 22, no. 11, pp. 13896–13903, 2014.

[78] C.-K. Lee, S. Moon, S. Lee, D. Yoo, J.-Y. Hong, and B. Lee, "Compact three-dimensional head-mounted display system with savart plate," *Optics Express*, vol. 24, pp. 19531–19544, Aug 2016.

[79] C. Yoo, K. Bang, C. Jang, D. Kim, C.-K. Lee, G. Sung, H.-S. Lee, and B. Lee, "Dual-focal waveguide see-through near-eye display with polarization-dependent lenses," *Optics Letters*, vol. 44, pp. 1920–1923, Apr 2019.

[80] K. Akşit, P. Chakravarthula, K. Rathinavel, Y. Jeong, R. Albert, H. Fuchs, and D. Luebke, "Manufacturing application-driven foveated near-eye displays," *IEEE Trans. Vis. Comput. Graph.*, vol. 25, pp. 1928–1939, May 2019.

[81] A. Maimone, D. Lanman, K. Rathinavel, K. Keller, D. Luebke, and H. Fuchs, "Pinlight displays: wide field of view augmented reality eyeglasses using defocused point light sources," in *ACM SIGGRAPH 2014 Emerging Technologies*, p. 20, 2014.

[82] R. S. Johnston and S. R. Willey, "Development of a commercial retinal scanning display," in *Helmet-and Head-Mounted Displays and Symbology Design Requirements II*, vol. 2465, pp. 2–14, 1995.

[83] D. Lanman and D. Luebke, "Near-eye light field displays," *ACM Trans. Graph.*, vol. 32, p. 220, Nov 2013.

[84] A. Maimone, A. Georgiou, and J. S. Kollin, "Holographic near-eye displays for virtual and augmented reality," *ACM Trans. Graph.*, vol. 36, pp. 85:1–85:16, July 2017.

[85] G. Kuo, L. Waller, R. Ng, and A. Maimone, "High resolution étendue expansion for holographic displays," *ACM Trans. Graph.*, vol. 39, July 2020.

[86] J. E. Cutting, "How the eye measures reality and virtual reality," *Behavior Research Methods*, vol. 29, pp. 27–36, Mar 1997.

[87] K. Kiyokawa, M. Billinghurst, B. Campbell, and E. Woods, "An occlusion-capable optical see-through head mount display for supporting co-located collaboration," in *2nd IEEE International Symposium of Mixed and Augmented Reality*, p. 133, 2003.

[88] T. Hamasaki and Y. Itoh, "Varifocal occlusion for optical see-through head-mounted displays using a slide occlusion mask," *IEEE Trans. Vis. Comput. Graph*, vol. 25, pp. 1961–1969, May 2019.

[89] T. Uchida, K. Sato, and S. Inokuchi, "An optical see-through mr display with digital micro-mirror device," *Trans. of the Virtual Reality Society of Japan*, vol. 7, no. 2, 2002.

[90] K. Kim., D. Heo., and J. Hahn., "Occlusion-capable head-mounted display," in *Proceedings of the 7th International Conference on Photonics, Optics and Laser Technology*, pp. 299–302, INSTICC, SciTePress, 2019.

[91] B. Krajancich, N. Padmanaban, and G. Wetzstein, "Factored occlusion: Single spatial light modulator occlusion-capable optical see-through augmented reality display," *IEEE Trans. Vis. Comput. Graph.*, vol. 26, pp. 1871–1879, May 2020.

[92] A. Wilson and H. Hua, "Design of a pupil-matched occlusion-capable optical see-through wearable display," *IEEE Trans. Vis. Comput. Graph.*, p. Early Access, 2021.

[93] C. Gao, Y. Lin, and H. Hua, "Occlusion capable optical see-through head-mounted display using freeform optics," in *11th IEEE International Symposium of Mixed and Augmented Reality*, pp. 281–282, 2012.

[94] C. Gao, Y. Lin, and H. Hua, "Optical see-through head-mounted display with occlusion capability," in *Proc. SPIE*, vol. 8735, pp. 87350F–1:9, 2013.

[95] O. Cakmakci, Y. Ha, and J. P. Rolland, "A compact optical see-through head-worn display with occlusion support," in *3rd IEEE International Symposium of Mixed and Augmented Reality*, pp. 16–25, 2004.

[96] A. Maimone and H. Fuchs, "Computational augmented reality eyeglasses," in *12th IEEE International Symposium of Mixed and Augmented Reality*, pp. 29–38, 2013.

[97] Y. Itoh, T. Hamasaki, and M. Sugimoto, "Occlusion leak compensation for optical see-through displays using a single-layer transmissive spatial light modulator," *IEEE Trans. Vis. Comput. Graph.*, vol. 23, pp. 2463–2473, Nov 2017.

[98] Y. Yamaguchi and Y. Takaki, "See-through integral imaging display with background occlusion capability," *Applied Optics*, vol. 55, pp. A144–A149, Jan 2016.

[99] K. Rathinavel, G. Wetzstein, and H. Fuchs, "Varifocal occlusion-capable optical see-through augmented reality display based on focus-tunable optics," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, pp. 3125–3134, Nov 2019.

[100] K. Kiyokawa, Y. Kurata, and H. Ohno, "An optical see-through display for mutual occlusion with a real-time stereovision system," *Computers and Graphics*, vol. 25, pp. 765–779, Oct 2001.

[101] A. Wilson and H. Hua, "Design and prototype of an augmented reality display with per-pixel mutual occlusion capability," *Opt. Express*, vol. 25, pp. 30539–30549, Nov 2017.

[102] D. Aiteanu, B. Hillers, and A. Graser, "A step forward in manual welding: demonstration of augmented reality helmet," in *Mixed and Augmented Reality, 2003. Proceedings. The Second IEEE and ACM International Symposium on*, pp. 309–310, IEEE, 2003.

[103] S. Mori, S. Ikeda, A. Plopski, and C. Sandor, "Brightview: Increasing perceived brightness of optical see-through head-mounted displays through unnoticeable incident light reduction," in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, (Los Alamitos, CA, USA), pp. 251–258, IEEE Computer Society, 2018.

[104] F. Campbell, "The depth of field of the human eye," *Optica Acta: International Journal of Optics*, vol. 4, pp. 157–164, Dec 1957.

[105] D. Dunn, "Required accuracy of gaze tracking for varifocal displays," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1838–1842, 2019.

[106] G. N. McKay, F. Mahmood, and N. J. Durr, "Large dynamic range autorefraction with a low-cost diffuser wavefront sensor," *Biomedical Optics Express*, vol. 10, pp. 1718–1735, Apr 2019.

[107] S. Ono, "Snapshot multispectral imaging using a pixel-wise polarization color image sensor," *Opt. Express*, vol. 28, pp. 34536–34573, Nov 2020.

[108] X. Dun, H. Ikoma, G. Wetzstein, Z. Wang, X. Cheng, and Y. Peng, "Learned rotationally symmetric diffractive achromat for full-spectrum computational imaging," *Optica*, vol. 7, pp. 913–922, Aug 2020.

[109] Y. Itoh, T. Langlotz, S. Zollmann, D. Iwai, K. Kiyokawa, and T. Amano, "Computational phase-modulated eyeglasses," *IEEE Trans. Vis. Comput. Graph.*, vol. 27, pp. 1916–1928, 10 2019.

[110] A. Arias and P. Artal, "Wavefront-shaping-based correction of optically simulated cataracts," *Optica*, vol. 7, pp. 22–27, Jan 2020.

[111] G. Damberg, J. Gregson, and W. Heidrich, "High brightness hdr projection using dynamic freeform lensing," *ACM Trans. Graph.*, vol. 35, May 2016.

[112] D. Glasner, T. Zickler, and A. Levin, "A reflectance display," *ACM Trans. Graph.*, vol. 33, July 2014.

[113] J.-H. R. Chang, A. Levin, B. V. K. V. Kumar, and A. C. Sankaranarayanan, "Towards occlusion-aware multifocal displays," *ACM Trans. Graph.*, vol. 39, July 2020.

[114] D. Voelz in *Computational Fourier Optics: A MATLAB tutorial*, Society of Photo-Optical Instrumentation Engineers, 2011.

[115] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, pp. 679–698, Nov 1986.

[116] P. C. Hansen, J. G. Nagy, and D. P. O'leary, *Deblurring images: matrices, spectra, and filtering*. SIAM, 2006.

[117] P. D. Burns *et al.*, "Slanted-edge mtf for digital camera and scanner analysis," in *Is and Ts Pics Conference*, pp. 135–138, SOCIETY FOR IMAGING SCIENCE & TECHNOLOGY, 2000.

[118] Y.-G. Ju, M.-H. Choi, P. Liu, B. Hellman, T. L. Lee, Y. Takashima, and J.-H. Park, "Occlusion-capable optical-see-through near-eye display using a single digital micromirror device," *Opt. Lett.*, vol. 45, pp. 3361–3364, Jul 2020.

[119] A. B. Watson and J. I. Yellott, "A unified formula for light-adapted pupil size," *Journal of Vision*, vol. 12, pp. 12–12, Sep 2012.

[120] J. E. Dowling, *The retina: an approachable part of the brain*. Harvard University Press, 1987.

[121] H. D. Baker, "The course of foveal light adaptation measured by the threshold intensity increment," *JOSA*, vol. 39, no. 2, pp. 172–179, 1949.

[122] C. Haig, "The course of rod dark adaptation as influenced by the intensity and duration of pre-adaptation to light," *The Journal of general physiology*, vol. 24, pp. 735–751, Jul 1941.

[123] J. Rekimoto, "Squama: modular visibility control of walls and windows for programmable physical architectures," in *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pp. 168–171, ACM, 2012.

[124] T. Hara, H. Saito, and T. Kanade, "Removal of glare caused by water droplets," in *Visual Media Production, 2009. CVMP'09. Conference for*, pp. 144–151, IEEE, 2009.

[125] Santos, Pedro, Gierlinger, Thomas, Machui, O. Stork, , and André, "The daylight blocking optical stereo see-through hmd," in *Workshop on Immersive projection technologies/Emerging display technologies*, p. 4, ACM, 2008.

[126] S. K. Nayar and V. Branzoi, "Adaptive dynamic range imaging: Optical control of pixel exposures over space and time," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pp. 1168–1175, IEEE, 2003.

[127] T. Amano and H. Kato, "Appearance control by projector camera feedback for visually impaired," in *CVPRW*, pp. 57–63, IEEE, 2010.

[128] R. Tamburo, E. Nurvitadhi, A. Chugh, M. Chen, A. Rowe, T. Kanade, and S. G. Narasimhan, "Programmable automotive headlights," in *European Conference on Computer Vision*, pp. 750–765, Springer, 2014.

[129] H.-L. Shen and Z.-H. Zheng, "Real-time highlight removal using intensity ratio," *Applied Optics*, vol. 52, pp. 4483–4493, Jul 2013.

[130] J. Newman, M. Wagner, M. Bauer, A. MacWilliams, T. Pintaric, D. Beyer, D. Pustka, F. Strasser, D. Schmalstieg, and G. Klinker, "Ubiquitous tracking for augmented reality," in *Proceedings of 3rd IEEE ISMAR*, pp. 192–201, 2004.

[131] A. L. Janin, D. W. Mizell, and T. P. Caudell, "Calibration of head-mounted displays for augmented reality applications," in *Proceedings of IEEE Virtual Reality Annual International Symposium, 1993.*, pp. 246–255, 1993.

[132] M. Tuceryan and N. Navab, "Single point active alignment method (spaam) for optical see-through hmd calibration for ar," in *Proceedings of ISAR*, pp. 149–158, IEEE, 2000.

[133] Y. Itoh and G. Klinker, "Interaction-free calibration for optical see-through head-mounted displays based on 3d eye localization," in *Proceedings of IEEE Symposium on 3D User Interfaces (3DUI), Minneapolis, MN, USA, March 29-30, 2014*, pp. 75–82, 2014.

[134] A. Plopski, Y. Itoh, C. Nitschke, K. Kiyokawa, G. Klinker, and H. Takemura, "Corneal-imaging calibration for optical see-through head-mounted displays," *IEEE Transactions on Visualization and Computer Graphics (Proceedings Virtual Reality 2015)*, vol. 21, pp. 481–490, April 2015.

[135] Y. Itoh and G. Klinker, "Light-field correction for spatial calibration of optical see-through head-mounted displays," *IEEE Transactions on Visualization and Computer Graphics (Proceedings Virtual Reality 2015)*, vol. 21, pp. 471–480, April 2015.

[136] P. Irawan, J. A. Ferwerda, and S. R. Marschner, "Perceptually based tone mapping of high dynamic range image streams.," in *Rendering Techniques*, pp. 231–242, 2005.

[137] A. B. Watson, "A formula for the mean human optical modulation transfer function as a function of pupil size," *Journal of Vision*, vol. 13, pp. 18–18, Jan 2013.

[138] B. A. Barsky, "Vision-realistic rendering: simulation of the scanned foveal image from wavefront data of human subjects," in *Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization*, pp. 73–81, ACM, 2004.

[139] K. R. Moser, D. C. Rompapas, J. E. Swan, S. Ikeda, G. Yamamoto, T. Taketomi, C. Sandor, H. Kato, *et al.*, "Sharpview: Improved clarity of defocused content on optical see-through head-mounted displays," in *IEEE 3DUI*, pp. 173–181, IEEE, 2016.

[140] H. Koschmiere, "Theorie der horizontalen sichtweite," *Beitrage zur Physik der freien Atmosphare*, pp. 33–53, 1924.

[141] E. J. Cartney, "Optics of the atmosphere: Scattering by molecules and particles," *Physics Bulletin*, vol. 28, pp. 521–521, Nov 1977.

[142] S. G. Narasimhan and S. K. Nayar, "Vision and the atmosphere," *International Journal of Computer Vision*, vol. 48, pp. 233–254, Jul 2002.

[143] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, pp. 2341–2353, Dec 2011.

[144] C. Bolun, X. Xiangmin, J. Kui, Q. Chunmei, and T. Dacheng, "Dehazenet: An end-to-end system for single image haze removal," *IEEE Transactions on Image Processing*, vol. 25, pp. 5187–5198, Jan 2016.

[145] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "Ffa-net: Feature fusion attention network for single image dehazing," 2019.

[146] J. P. Rolland, R. L. Holloway, and H. Fuchs, "Comparison of optical and video see-through, head-mounted displays," in *Photonics for Industrial Applications*, pp. 293–307, 1995.

[147] J. P. Rolland and H. Fuchs, "Optical versus video see-through head-mounted displays in medical visualization," *Presence: Teleoperators & Virtual Environments*, vol. 9, no. 3, pp. 287–309, 2000.

[148] K.-E. Lin, K.-T. Shih, and H. Chen, "Enhancing the perception of a hazy visual world using a see-through head-mounted device," in *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 4397–4401, Sep. 2017.

[149] R. Mantiuk, K. Myszkowski, and H.-P. Seidel, "A perceptual framework for contrast processing of high dynamic range images," *ACM Trans. Appl. Percept.*, vol. 3, p. 286–308, July 2006.

[150] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar, "Instant dehazing of images using polarization," in *CVPR (1)*, pp. 325–332, 2001.

[151] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar, "Polarization-based vision through haze," *Applied Optics*, vol. 42, pp. 511–525, Jan 2003.

[152] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 492–505, 2019.

[153] D. Chen, M. He, Q. Fan, J. Liao, L. Zhang, D. Hou, L. Yuan, and G. Hua, "Gated context aggregation network for image dehazing and deraining," *CoRR*, vol. abs/1811.08747, 2018.

[154] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 586–595, 2018.

[155] Y. Itoh, T. Langlotz, D. Iwai, K. Kiyokawa, and T. Amano, "Light attenuation display: Subtractive see-through near-eye display via spatial color filtering," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, pp. 1951–1960, May 2019.

[156] G. Wetzstein, W. Heidrich, and D. Luebke, "Optical image processing using light modulation displays," *Computer Graphics Forum*, vol. 29, pp. 1934–1944, Sep 2010.

[157] T. Rhodes, G. Miller, Q. Sun, D. Ito, and L.-Y. Wei, "A transparent display with per-pixel color and opacity control," in *ACM SIGGRAPH 2019 Emerging Technologies*, p. 5, ACM, 2019.

[158] G. A. Koulieris, K. Akşit, M. Stengel, R. K. Mantiuk, K. Mania, and C. Richardt, "Near-eye display and tracking technologies for virtual and augmented reality," *Computer Graphics Forum*, vol. 38, pp. 493–519, May 2019.

[159] Y. Peng, S. Choi, N. Padmanaban, and G. Wetzstein, "Neural holography with camera-in-the-loop training," *ACM Trans. Graph.*, vol. 39, pp. 1–14, Dec 2020.

[160] 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 Trans. Graph. (SIGGRAPH Asia)*, Dec 2021.

[161] J. Chang, V. Sitzmann, X. Dun, W. Heidrich, and G. Wetzstein, "Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification," *Scientific Reports*, vol. 8, p. 12324, Aug 2018.

[162] C. A. Metzler, H. Ikoma, Y. Peng, and G. Wetzstein, "Deep optics for single-shot high-dynamic-range imaging," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (Los Alamitos, CA, USA), pp. 1372–1382, IEEE Computer Society, jun 2020.

[163] J. Liang, B. Grimm, S. Goelz, and J. F. Bille, "Objective measurement of wave aberrations of the human eye with the use of a hartmann-shack wave-front sensor," *Journal of the Optical Society of America A*, vol. 11, no. 7, pp. 1949–1957, 1994.

[164] S. Hecht and S. Shlaer, "Intermittent stimulation by light : V. the relation between intensity and critical frequency for different parts of the spectrum ," *Journal of General Physiology*, vol. 19, pp. 965–977, 07 1936.

Publications related to this dissertation

Peer-Reviewed Articles in Journal Papers

[J1] Y. Hiroi, T. Kaminokado, S. Ono, Y. Itoh, “Focal Surface Occlusion”, *Optics Express*, Vol. 29, Issue 22, pp. 36581–36597, Oct. 2021.

Peer-Reviewed Articles in International Conferences

[C1] Y. Hiroi, T. Kaminokado, A. Mori, Y. Itoh, “DehazeGlasses: Optical Dehazing with an Occlusion Capable See-Through Display,” In *Proceedings of the Augmented Humans International Conference (AHs 2020)*, pp. 1–11, Kaiserslautern, Germany, Mar. 16–17, 2020.

[C2] Y. Hiroi, Y. Itoh, T. Hamasaki, M. Sugimoto, “AdaptiVisor: Assisting Eye Adaptation via Occlusive Optical See-Through Head-Mounted Displays,” In *Prceedings of The 8th Augmented Human International Conference (AH 2017)*, pp.1–9, Silicon Valley, USA, Mar. 16-18, 2017.

Other publications

Peer-Reviewed Articles in Journal Papers

[J2] T. Kaminokado, Y. Hiroi, Y. Itoh, "StainedView: Variable-Intensity Light Attenuation Display with Cascaded Spatial Color Filtering for Improved Color Fidelity," *IEEE Trans. Vis. Comput. Graph.*, vol. 26, num. 12, pp. 3576–3856, Dec, 2020.

[J3] T. Hamasaki, Y. Itoh, Y. Hiroi, D. Iwai, M. Sugimoto, "HySAR: Hybrid Material Rendering by an Optical See-Through Head-Mounted Display with Spatial Augmented Reality Projection," *IEEE Trans. Vis. Comput. Graph.*, vol. 24 Num. 4, pp. 1457–1466, Apr, 2018.

Peer-Reviewed Articles in International Conferences

[C3] T. Tochimoto, Y. Hiroi, Y. Itoh, "CircadianVisor: Image Presentation With an Optical See-Through Display in Consideration of Circadian Illuminance", In *Proceedings of the Augmented Humans International Conference (AHs 2021)*, pp. 1–11, Online, Feb. 22–24, 2021.

Peer-Reviewed Poster Papers in International Conferences

[C4] M. Kaneko, Y. Hiroi, Y. Itoh, "Depth-of-Focus-Simulated Retinal Projection Display with Diffusive Ocular Wavefront Sensing", *IEEE International Symposium for Mixed and Augmented Reality (ISMAR) Adjunct*, Bali, Italy, Oct. 4–8, 2021.

[C5] X. Zhang, J. Lundgren, Y. Mesaki, Y. Hiroi, Y. Itoh, "Stencil Marker: Designing Partially Transparent Markers for Stacking Augmented Reality Objects", In *Proceedings of IEEE ISMAR Adjunct 2020*, Recife, Brazil/Virtual, Nov. 9-13, 2020.

[C6] K. Someya, Y. Hiroi, M. Yamada, Y. Itoh, “OSTNet: Calibration Method for Optical See-Through Head-Mounted Displays via Non-Parametric Distortion Map Generation,” In *Proceedings of IEEE ISMAR Adjunct 2019*, Beijing, China, pp 259–260, Oct. 14–18, 2019.

[C7] Y. Hiroi, Y. Itoh, T. Hamasaki, D. Iwai, M. Sugimoto, “HySAR: Hybrid Material Rendering by an Optical See-Through Head-Mounted Display with Spatial Augmented Reality”, In *Proceedings of IEEE Virtual Reality 2017 (IEEE VR 2017)*, Los Angeles, USA, pp. 211–212, Mar. 18–22, 2017.

[C8] T. Kikuchi, Y. Hiroi, R. T. Smith, B. H. Thomas, M. Sugimoto, “MARCut: Marker-based Laser Cutting for Personal Fabrication on Existing Objects”, In *Proceedings of 9th International Conference on Tangible, Embedded, and Embodied Interaction (TEI 2016)*, Eindhoven, Netherlands, pp.468–474, Feb. 14–17, 2016.

[C9] Y. Hiroi, K. Obata, K. Suzuki, N. Ienaga, M. Sugimoto, H. Saito, T. Takamaru, “Remote Welding Robot Manipulation using Multi-View Images”, In *Proceedings of 2015 IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2015)*, Fukuoka, Japan, Sep. 29–Oct. 3 pp.128–131, 2015.

Peer-Reviewed Demo in International Conferences

[C10] Y. Itoh, Y. Hiroi, J. Otsuka, M. Sugimoto, J. Orlosky, K. Kiyokawa, G. Klinker, “Laplacian Vision: Augmenting Motion Prediction via Optical See-through Head-Mounted Displays and Projectors,” *ACM SIGGRAPH 2016 Emerging Technologies*. Anaheim, USA, Jul 24–28, 2016.