Tampere University

JANI MÄKINEN

Coherent and Holographic Imaging Methods for Immersive Near-Eye Displays

Tampere University Dissertations 915

Tampere University Dissertations 915

JANI MÄKINEN

Coherent and Holographic Imaging Methods for Immersive Near-Eye Displays

ACADEMIC DISSERTATION
To be presented, with the permission of
the Faculty of Information Technology and Communication Sciences
of Tampere University,
for public discussion in the auditorium SA203
of the Sähkötalo, Korkeakoulunkatu 3, Tampere,
on 21 December 2023, at 12 o’clock.

ACADEMIC DISSERTATION
Tampere University, Faculty of Information Technology and Communication Sciences
Finland

Responsible supervisor and Custos	Professor Atanas Gotchev Tampere University Finland	
Supervisor	Dr. Erdem Sahin Tampere University Finland	
Pre-examiners	Dr. Claas Falldorf Bremen Institut für angewandte Strahltechnik GmbH Germany	Senior Research Associate, Dr. Ali Özgür Yöntem University of Cambridge United Kingdom
Opponents	Professor Tomasz Kozacki Warsaw University of Technology Poland	Senior Research Associate, Dr. Ali Özgür Yöntem University of Cambridge United Kingdom

Cover design: Roihu Inc.

ISBN 978-952-03-3212-9 (print)
ISBN 978-952-03-3213-6 (pdf)
ISSN 2489-9860 (print)
ISSN 2490-0028 (pdf)
http://urn.fi/URN:ISBN:978-952-03-3213-6

HIILINEUTRAALI PAINOTUOTE
ClimateCalc CC-000025/FI
PunaMusta Printing

Carbon dioxide emissions from printing Tampere University dissertations have been compensated.

PunaMusta Oy – Yliopistopaino
Joensuu 2023

PREFACE

This dissertation represents the research in holographic and coherent imaging conducted during my doctoral studies in Tampere University. I am honored that I have had the opportunity to share this research to the scientific community, and I hope that the contributions put forth in this dissertation will bring immersive holographic near-eye displays one step closer to practical reality.

I want to thank my pre-examiners Dr. Claas Falldorf and Dr. Ali Özgür Yöntem for their valuable and positive feedback, as well as Prof. Tomaz Kozacki and Dr. Ali Özgür Yöntem for agreeing to serve as the opponents for the public examination of my dissertation.

The seeds for this dissertation were planted already back during the summer of 2015, when Prof. Atanas Gotchev and Dr. Erdem Sahin gave me the opportunity to work on an exciting and, at the time, a completely new-to-me topic of holography. While at times it has been a daunting task to learn the necessary theoretical background and the inner workings of various approaches within the field, the journey until this point has been enjoyable largely thanks to the supervision skills of Prof. Gotchev and Dr. Sahin. Not only has Prof. Gotchev provided an excellent environment for me to work in, giving me the support and freedom to work independently as I have needed it, but his guidance and feedback has always been prompt and valuable. The vast knowledge of Dr. Sahin has been instrumental to the success of this work, along with his patience with me whenever I have needed the extra time to properly learn something new.

While conducting my research over the past years, I have had the privilege of working in the 3D Media Group of Tampere University, surrounded by amazing people. I want to thank my co-authors Ugur Akpinar, Ugur Gudelek, Lauri Varjo from the 3D Media Group, and Prof. Elena Stoykova from the Bulgarian Academy of Sciences for their valuable contributions to this dissertation. I would also like to express my appreciation to my colleagues: Yuta Miyanishi, for his help with all things human vision related; Filipe Da Graca Gama, for his friendship and support; Ihtisham Ali, for being a great office mate; Olli Suominen and Dr. Robert Bregovic, for helping out with various practical matters. Additionally, my sincerest gratitude goes out to all the lovely group members who have brought tremendous joy into my life, whether it has been by joining our floorball games, coffee breaks or other social events, or by keeping me company during workshop and conference visits.

Finally, words cannot express how grateful I am to my family for their never-ending love, support, and encouragement to always reach higher. Without them, I simply would not be where I am today.

ABSTRACT

Near-eye displays have been designed to provide realistic 3D viewing experience, strongly demanded in applications, such as remote machine operation, entertainment, and 3D design. However, contemporary near-eye displays still generate conflicting visual cues which degrade the immersive experience and hinders their comfortable use. Approaches using coherent, e.g., laser light for display illumination have been considered prominent for tackling the current near-eye display deficiencies. Coherent illumination enables holographic imaging whereas holographic displays are expected to accurately recreate the true light waves of a desired 3D scene. However, the use of coherent light for driving displays introduces additional high-contrast noise in the form of speckle patterns, which has to be taken care of. Furthermore, imaging methods for holographic displays are computationally demanding and impose new challenges in analysis, speckle noise and light modelling.

This thesis examines computational methods for near-eye displays in the coherent imaging regime using signal processing, machine learning, and geometrical (ray) and physical (wave) optics modeling. In the first part of the thesis, we concentrate on analysis of holographic imaging modalities and develop corresponding computational methods. To tackle the high computational demands of holography, we adopt holographic stereograms as an approximative holographic data representation. We address the visual correctness of such representation by developing a framework for analyzing the accuracy of accommodation visual cues provided by a holographic stereogram in relation to its design parameters. Additionally, we propose a signal processing solution for speckle noise reduction to overcome existing issues in light modelling causing visual artefacts. We also develop a novel holographic imaging method to accurately model lighting effects in challenging conditions, such as mirror reflections.

In the second part of the thesis, we approach the computational complexity aspects of coherent display imaging through deep learning. We develop a coherent accommodation-invariant near-eye display framework to jointly optimize static display optics and a display image pre-processing network. Finally, we accelerate the corresponding novel holographic imaging method via deep learning aimed at real-time applications. This includes developing an efficient procedure for generating functional random 3D scenes for forming a large synthetic data set of multiperspective images, and training a neural network to approximate the holographic imaging method under the real-time processing constraints.

Altogether, the methods developed in this thesis are shown to be highly competitive with the state-of-the-art computational methods for coherent-light near-eye displays. The results of the work demonstrate two alternative approaches for resolving the existing near-eye display problems of conflicting visual cues using either static or dynamic optics and computational methods suitable for real-time use. The presented results are therefore instrumental for the next-generation immersive near-eye displays.

TIIVISTELMÄ

Lähinäytöt on suunniteltu tarjoamaan realistisia kolmiulotteisia katselukokemuksia, joille on merkittävää tarvetta esimerkiksi työkoneiden etäkäytössä ja 3D-suunnittelussa. Nykyaikaiset lähinäytöt tuottavat kuitenkin edelleen ristiriitaisia visuaalisia vihjeitä, jotka heikentävät immersiivistä kokemusta ja haittaavat niiden miellyttävää käyttöä. Merkittävänä ratkaisuvaihtoehtona pidetään koherentin valon, kuten laservalon, käyttöä näytön valaistukseen, millä voidaan korjata nykyisten lähinäyttöjen puutteita. Erityisesti koherentti valaistus mahdollistaa holografisen kuvantamisen, jota käyttävät holografiset näytöt voivat tarkasti jäljitellä kolmiulotteisten mallien todellisia valoaaltoja. Koherentin valon käyttäminen näyttöjen valaisemiseen aiheuttaa kuitenkin huomiota vaativaa korkean kontrastin häiriötä pilkkukuvioiden muodossa. Lisäksi holografisten näyttöjen laskentamenetelmät ovat laskennallisesti vaativia ja asettavat uusia haasteita analyysin, pilkkuhäiriön ja valon mallintamisen suhteen.

Tässä väitöskirjassa tutkitaan laskennallisia menetelmiä lähinäytöille koherentissa kuvantamisjärjestelmässä käyttäen signaalinkäsittelyä, koneoppimista sekä geometrista (säde) ja fysikaalista (aalto) optiikan mallintamista. Työn ensimmäisessä osassa keskitytään holografisten kuvantamismuotojen analysointiin sekä kehitetään hologrammien laskennallisia menetelmiä. Holografian korkeiden laskentavaatimusten ratkaisemiseksi otamme käyttöön holografiset stereogrammit holografisen datan likimääräisenä esitysmuotona. Tarkastelemme kyseisen esitysmuodon visuaalista oikeellisuutta kehittämällä analyysikehyksen holografisen stereogrammin tarjoamien visuaalisten vihjeiden tarkkuudelle akkommodaatiota varten suhteessa sen suunnitteluparametreihin. Lisäksi ehdotamme signaalinkäsittelyratkaisua pilkkuhäiriön vähentämiseksi, ratkaistaksamme nykyisten menetelmien valon mallintamiseen liittyvät visuaalisia artefakteja aiheuttavat ongelmat. Kehitämme myös uudenlaisen holografisen kuvantamismenetelmän, jolla voidaan mallintaa tarkasti valon käyttäytymistä haastavissa olosuhteissa, kuten peiliheijastuksissa.

Väitöskirjan toisessa osassa lähestytään koherentin näyttökuvantamisen laskennallista taakkaa koneoppimisen avulla. Kehitämme koherentin akkommodaatioinvariantin lähinäytön suunnittelukehyksen, jossa optimoidaan yhtäaikaisesti näytön staattista optiikka ja näytön kuvan esikäsittelyverkkoa. Lopuksi nopeutamme ehdottamaamme uutta holografista kuvantamismenetelmää koneoppimisen avulla reaaliaikaisia sovelluksia varten. Kyseiseen ratkaisuun sisältyy myös tehokkaan menettelyn kehittäminen funktionaalisten satunnais-3D-ympäristöjen tuottamiseksi. Kehittämämme menetelmä mahdollistaa suurten synteettisten moninäkökulmaisten kuvien datasettien tuottamisen, joilla voidaan kouluttaa sopivia neuroverkkoja mallintamaan holografista kuvantamismenetelmäämme reaaliajassa.

Kaiken kaikkiaan tässä työssä kehitettyjen menetelmien osoitetaan olevan erittäin kilpailukykyisiä uusimpien koherentin valon lähinäyttöjen laskentamenetelmien kanssa. Työn tuloksena nähdään kaksi vaihtoehtoista lähestymistapaa ristiriitaisten visuaalisten vihjeiden aiheuttamien nykyisten lähinäyttöongelmien ratkaisemiseksi joko staattisella tai dynaamisella optiikalla ja reaaliaikaiseen käyttöön soveltuvilla laskentamenetelmillä. Esitetyt tulokset ovat näin ollen tärkeitä seuraavan sukupolven immersiivisille lähinäytöille.

CONTENTS

List of Tables

ABBREVIATIONS

ASM	angular spectrum method
CGH	computer generated hologram
CNN	convolutional neural network
DOE	diffractive optical element
DoF	depth-of-field
DSCP	diffraction-specific coherent panoramagram
DSLF	densely sampled light field
FCNN	fully convolutional neural network
FFT	fast Fourier transform
Fourier-CNN	fully convolutional neural network with spectral connections
hogel	holographic element
HS	holographic stereogram
HVS	human visual system
LCA	longitudinal chromatic aberration
LF	light field
ML	machine learning
MTF	modulation transfer function
NED	near-eye display
OLAS	overlap-add stereogram
PAS	phase-added stereogram
PBM	point-based method

PSF	point spread function
PSNR	peak signal-to-noise ratio
ReLU	rectified linear unit
RS	Rayleigh-Sommerfeld
SLM	spatial light modulator
SSIM	structural similarity index measure
STFT	short-time Fourier transform
VAC	vergence-accommodation conflict

ORIGINAL PUBLICATIONS

Publication I	E. Sahin, E. Stoykova, J. Mäkinen, and A. Gotchev, "Computer-Generated Holograms for 3D Imaging: A Survey", *ACM Computing Surveys*, vol. 53, no. 2, pp. 1–35, 2020. DOI: 10.1145/3378444.
Publication II	J. Mäkinen, E. Sahin, and A. Gotchev, "Speckle reduction method for image-based coherent stereogram generation", *Optics Express*, vol. 26, no. 5, pp. 5381–5394, 2018. DOI: 10.1364/OE.26.005381.
Publication III	——, "Analysis of accommodation cues in holographic stereograms", in *2018 - 3DTV-Conference*, IEEE, 2018, pp. 1–4. DOI: 10.1109/3DTV.2018.8478586.
Publication IV	J. Mäkinen, E. Sahin, U. Akpinar, and A. Gotchev, "Computational coherent imaging for accommodation-invariant near-eye displays", in *2021 IEEE International Conference on Image Processing*, IEEE, 2021, pp. 3433–3437. DOI: 10.1109/ICIP42928.2021.9506773.
Publication V	J. Mäkinen, U. Gudelek, L. Varjo, E. Sahin, and A. Gotchev, "Computer-generated holography using hybrid planar-spherical wave primitives", *ACM Transactions on Graphics*, submitted.

Author's contribution

Publication I	This publication was designed as an overview of state-of-the-art methods in computer-generated holography. Due to the wide scope of the study, the different topics were assigned to co-authors

	in the following manner. Co-authors Dr. Stoykova and Dr. Sahin proposed the article structure and prepared the content for Sections 1–3 and 6. The thesis author overviewed the relevant methods in Sections 4.1 (Holographic Stereogram) and 5 (Speckle Suppression in CGH) and drafted the section contents for the publication, with feedback from Dr. Sahin. Additionally, the thesis author designed and prepared the publication figures, and conducted the numerical view simulation experiments in the appendix. Dr. Stoykova, Dr. Sahin and Prof. Gotchev refined the draft version of the manuscript for publication.
Publication II	The thesis author implemented the method, conducted the numerical simulations and drafted the manuscript. Co-author Dr. Sahin provided valuable feedback and suggestions while the details of the method were refined. Dr. Sahin and Prof. Gotchev both provided feedback for revising and finalizing the manuscript.
Publication III	The thesis author formulated the analysis framework based on discussions with Dr. Sahin, as well as conducted the numerical simulations and drafted the manuscript. Co-authors Dr. Sahin and Prof. Gotchev contributed to the manuscript as in Publication II.
Publication IV	The original incoherent imaging implementation of the end-to-end system for accommodation-invariant displays was developed and provided by co-author Ugur Akpinar. The thesis author converted the system to utilize coherent imaging, added support for partially coherent light sources and speckle suppression. Additionally, he conducted the experiments and subsequent analysis. Dr. Sahin provided feedback and suggestions during the development phase, and together with Prof. Gotchev helped to polish the manuscript.
Publication V	The overall framework was designed by the thesis author. He also developed the novel analytical hologram synthesis method. Co-author Ugur Gudelek was in charge of the machine learning

implementation. The random scene generator was implemented by co-author Lauri Varjo with guidance from the thesis author and Dr. Sahin. The thesis author drafted the manuscript. Dr. Sahin and Prof. Gotchev contributed similarly to Publication IV.

1 INTRODUCTION

In this chapter, the dissertation topic is introduced by providing a concise overview of goals for achieving a truly immersive 3D viewing experience in a near-eye display (NED). The research objective is followed by the main research questions guiding the research work during this thesis. The final section summarises the general structure of the thesis.

1.1 Thesis objective

Whether developing methods for photography, televisions or modern 3D displays, researchers and engineers have strived for a realistic viewing experience as if seeing an alternate reality through a window. As the development of computing platforms and electronics has progressed to the stage where mobile, head-worn displays are readily available for the consumer market, we are closer than ever to immersing ourselves in realistic appearing 3D spaces, both synthetic and real-world ones as recorded by advanced capturing devices. Such a visual experience is beneficial, not only for entertainment like video games, but also for various design tasks, remote machine/vehicle operation and complex data visualization. Generally speaking, wherever it is vital that the user can see the environment as realistically as possible, 3D display systems capable of producing *all* relevant visual cues and their combinations accurately, termed as *immersive* displays, are applicable. Such displays are of particular importance as the realistic recreation of 3D scenes empowers the viewer to infer structural information intuitively and to focus on the actual task, regardless of whether it is remotely operating heavy work machinery or enjoying a 3D movie. One of the more interesting segments of potential immersive displays are wearable form-factor displays or NEDs, as they allow the user to freely look at different directions.

A significant issue impeding the comfortable use and immersive experience of conventional NEDs is that they provide conflicting visual cues for the viewer in terms of accommodation and vergence, resulting in the well-known phenomenon referred to as *vergence-accommodation conflict* (VAC). The conflict can be resolved either by decoupling accommodation from defocus blur, in an *accommodation-invariant* display, or by providing all visual cues accurately and consistently relative to each other, in an *accommodation-enabling* display. The design problem of accommodation-invariant NEDs is to engineer the perceived image to be sharp (i.e. defocus-free) in a large depth range behind the display, otherwise known as *extended depth-of-field* (DoF) imaging. An attractive, yet seldom used, approach for this task is to limit the display illumination to be coherent (e.g. laser-generated). This inherently extends the DoF at the cost of introducing high-contrast speckle noise, a fundamental coherent imaging issue. If the speckle noise is effectively diminished, coherent accommodation-invariant display becomes an appealing low computational cost solution for solving VAC. Alternatively, under coherent display illumination, one can utilize dynamic optics to modulate both the amplitude and phase of the incoming illumination and recreate light waves as emitted by a desired 3D scene. Such a display is referred to as holographic display and provides an accommodation-enabling solution for VAC. The holographic imaging methods, or hologram synthesis, convert an intermediate representation of a 3D scene such as point cloud or a collection of multiperspective images into a complex-valued image or wave field to drive a holographic display to recreate the desired light waves through diffraction and interference.

In this dissertation, we approach the overarching NED problem of VAC through coherent and holographic imaging methods. For accommodation-invariant displays, we examine whether the implicitly larger DoF from coherent illumination makes the extended DoF problem easier, and if the resulting speckle noise can be managed while maintaining the accommodation-invariant response. For accommodation-enabling displays, our interest is in holographic imaging from multiperspective views, where we identify three major research topics: accurate hologram synthesis, speckle noise reduction, and the trade-off between computational complexity and hologram accuracy. Accurately synthesizing holograms here entails also modelling challenging light behaviour correctly, such as reflection and refraction, for providing accurate visual cues in realistic 3D scenes, which is not fully solved yet. Regarding speckle noise reduction, existing approaches reduce the perceived image quality either by introducing added defocus blur or visual artefacts. Finally, the issue of computational complexity is that the existing hologram synthesis methods have to make approximations in their wave modelling in order to reduce their computational burden. Holographic stereograms are well-known approximative holographic data representation enabling fast hologram generation. They achieve this by trading-off the accurate modelling of the recorded 3D scene content being further away from the hologram. However, holographic stereograms are yet to be systematically characterized in terms of accommodation visual cues. Quantifying their accuracy in this aspect can provide useful insight about how suitable a holographic stereogram is for immersive display experience. Alternatively, the trade-off between hologram quality and computational burden can be circumvented in some cases using deep learning to train a sufficiently simple neural network to approximate the holographic imaging process at fast inference rates. This has been typically approached using intermediate 3D scene representations different from multiperspective images, which are however the closest to 3D scene capture and creation. This leaves a research gap for us to explore.

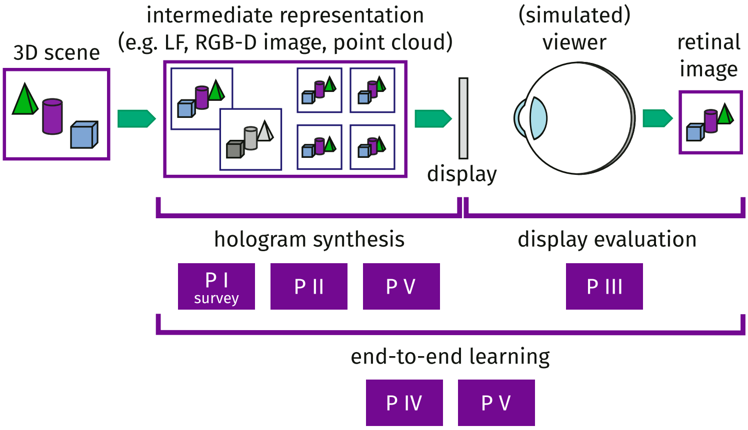


Figure 1.1 Overview of the general research topics covered in the thesis.

1.2 Research questions

Following the previously introduced research gaps, we formulate the following research questions to guide this thesis work:

RQ1 How to quantify the accuracy of accommodation cues present in holographic stereograms in a systematic manner?

RQ2 What is the most accurate way of converting multiperspective views and corresponding depth maps into a hologram regardless of computational complexity?

RQ3 In hologram generation from multiperspective views, how can speckle noise be best avoided with minimal loss of perceived quality?

RQ4 How well can a neural network approximate the hologram conversion found from RQ2 under a real-time inference constraint?

RQ5 As a low computational complexity solution to VAC, how well do coherent imaging methods compare against incoherent in accommodation-invariant displays, and does the inherently larger DoF of coherent imaging outweigh the downside of speckle noise?

1.3 Thesis structure

The thesis is organized in five chapters. Chapter 2 provides context and positions our work within the research landscape. This is achieved by discussing how 3D information is observed by humans, what are the core concepts of 3D vision relative to displays, followed by an overview of competing solutions in the realm of 3D near-eye displays. Furthermore, we look at different approaches to hologram synthesis, summarize state-of-the-art research, similar to Publication I, and identify gaps in current research landscape.

The main contributions of the thesis are divided between the next two chapters; Chapter 3, addresses the accuracy of hologram synthesis, while Chapter 4 discusses what can be done about the computational efficiency in order to bring the technology closer to practical relevance. In Chapter 3, we discuss the solution proposed in Publication II towards reducing speckle noise while accurately modelling the light propagation between different locations inside the recorded 3D scene. Furthermore, the chapter discusses the problem of analyzing the accuracy of accommodation provided by holographic displays that was investigated in more detail in Publication III, and how to synthesize a hologram that most accurately recreates a 3D scene from multiperspective images and corresponding depth maps. Chapter 4 examines utilizing learning-based approaches for achieving efficient and immersive 3D experience; first by removing defocus blur entirely to mitigate conflicting visual cues typically present in conventional NEDs, as proposed in Publication IV, then finally by accelerating the accurate-yet-slow hologram synthesis method presented in Chapter 3, as elaborated in Publication V.

Finally, Chapter 5 concludes the dissertation and discusses the potential future work within the topic. Fig. 1.1 visualizes the general research topics of the thesis and places the publications according to their main contributions.

2 BACKGROUND

This chapter discusses the background information needed to understand the core concepts presented later in the thesis. The first section discusses how humans see in 3D, particularly concentrating on visual cues most relevant for displays. The next section overviews different approaches for displaying 3D information in NEDs, giving some examples of widely available displays, as well as recently proposed ones yet to be commercialized. Finally, the state-of-the-art methods and gaps in current research in hologram synthesis are highlighted, providing further context to Publications II–V.

2.1 Human 3D perception

As the main draw of holographic displays is their ability to accurately recreate 3D scenes, it is essential to understand the key factors for human 3D perception, to contextualize how such displays and other competing 3D display approaches are capable of delivering these factors. Though there exist multiple psychological aspects and learned contextual clues which drive our spatial perception, here we will focus on factors which are understandable through concepts of imaging and optics.

2.1.1 Binocular disparity

Scene objects live in 3D space. When projected on multi-perspective planes, referred to as 2D views, a 3D object point appears on different locations on the corresponding views. *Disparity* is defined as the relative difference between two such locations. Disparity is depth-dependent, i.e. object points at different scene depths generate different disparities. In the context of human vision, i.e. binocular viewing, this corresponds to the different relative positions of object points on the retinae based on where the eyes are verging towards, as in Fig. 2.1. When the eyes fixate at a specific point in the scene, like the purple circle in Fig. 2.1, it is imaged on the same position on the two retinae and the disparity between the two views is zero. Objects closer or further than the fixation point are imaged at non-identical locations and have non-zero disparity: the larger the difference between the depth of the fixation point and the object in question, the larger the magnitude of the disparity.

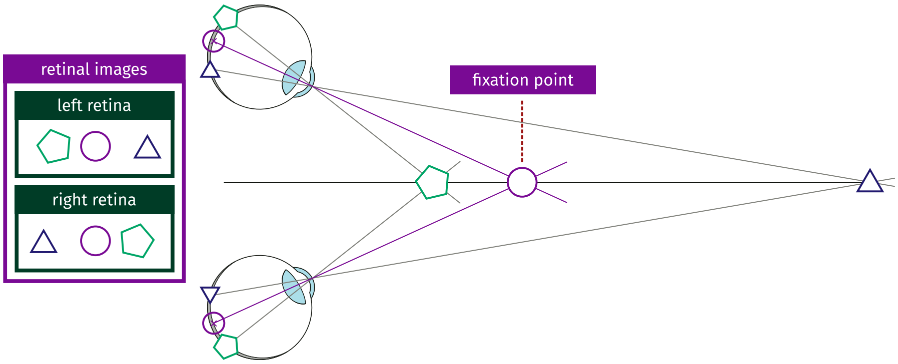


Figure 2.1 Two eyes fixated (converged) on the purple circle in the middle, which is imaged to the center of the retinae. The other two objects are imaged to different locations on the retinae resulting in non-zero disparity.

In addition to providing information during static viewing, binocular disparity also drives the eyes to either converge or diverge when attempting to refocus the eyes towards another point in the scene. This is due to the fact that disparity contains depth-sign information, i.e. whether the imaged object is in front or behind the fixation point. Altogether, binocular disparity drives the oculomotor function of *vergence* [1]: the rotation of eyes towards a fixation point while focusing at a specific object or depth in the 3D scene.

2.1.2 Motion parallax

When viewing a 3D scene while moving horizontally and/or vertically, the relative motion of objects depending on their depth is known as *motion parallax*: objects near the viewer move faster than objects farther in the scene, as illustrated in Fig. 2.2. Unlike binocular disparity, motion parallax is a monocular depth cue, i.e. it can be observed with a single eye. It is, however, closely related to disparity and resulting vergence: combined they are the two strongest cues for inferring 3D structure from retinal images [2].

Motion parallax, particularly in terms of 3D displays, can be further categorized

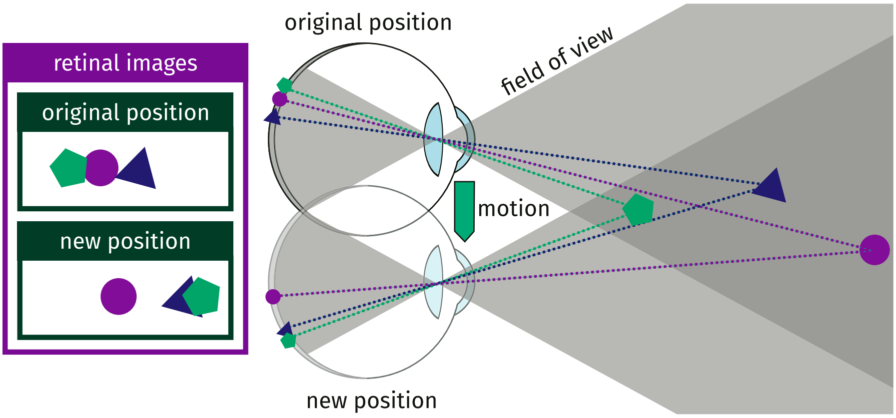


Figure 2.2 An illustration of motion parallax: the relative motion of objects at various depths when the viewer moves. Objects near the viewer (e.g. the pentagon) move a larger distance than objects farther away.

into horizontal or vertical parallax when restricting movement to one axis, and full parallax when movement in both directions is supported. In certain cases, displays provide only horizontal parallax to relax the display design or computational requirements [3], [4], especially when the viewer is assumed to have limited vertical motion, e.g. when sitting in front of a conventional display. The perception of motion parallax can also be considered from two different movement sources: head and eye. Current NEDs can easily provide head motion parallax via positional tracking of the display device. However, eye motion parallax is more cumbersome, due to the need for either accurate eye tracking or capability of providing multiple views across the expected eye positions. In a study by Nadler et. al. [5], it was discovered that smooth eye movement, rather than head movement, provides the critical information for the human visual system (HVS) to compute depth-sign from motion parallax. Therefore, it is advantageous for a NED to support motion parallax for even small viewpoint changes like these within the extent of eye movement for a truly immersive 3D experience.

2.1.3 Accommodation

Accommodation is a key part of the action of focusing in human vision. It refers to the oculomotor refractive power adjustment of the crystalline lens in the eye, such that a sharp image of the target object is formed on the retina (see Fig. 2.3). Objects at and within a nearby region of the accommodated depth produce sharply imaged objects in the retinal images, or the DoF, which for human eyes is typically around ±0.3 diopters[1] [6], [7]. Objects outside the DoF are imaged with retinal defocus blur, of which size increases when the objects get further away from the accommodated depth. Unlike disparity, defocus blur in ideal optical systems does not provide by itself information about the direction of depth: objects both in front and behind the accommodated depth get blurred identically relative to dioptric distances according to the imaging equation. In practice, the optical aberrations of human eyes (such as chromatic aberration) transform the blur to form differently depending on the direction of depth, thus providing additional information for the visual system [8].

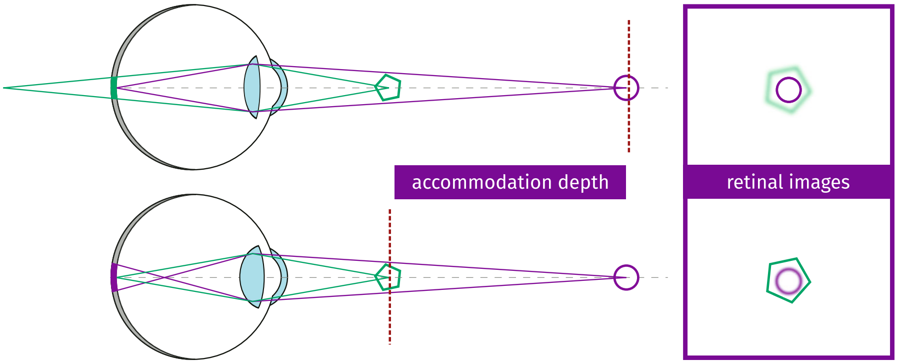


Figure 2.3 Accommodation in human eye when looking at two objects at different depths. The top row corresponds to the eye accommodating at the far object (circle), the bottom row at the near object (pentagon).

Closely related to retinal blur and its perception are two concepts: blur detection and blur discrimination. Blur detection refers to the smallest change in retinal defocus allowed before the blur of a clear test target is first detected, whereas blur discrimination refers to the smallest change in defocus level when an already blurred object is noticed to be blurrier, i.e. how finely changes in blur are observed [9]. Both are typically measured in dioptric differences in accommodation depth relative to the original accommodation state, which is accommodating on the target object in blur detection and the initial accommodation depth when testing for blur discrimination. Subjective studies have shown that the human blur discrimination threshold is approximately 0.05 D and the blur detection threshold is at best 0.10 D, making blur

---

[1] Diopter (D) is a commonly utilized measure for optical distances, expressed as the inverse of metric distance.

discrimination more sensitive in human vision [10]. Defocus blur changes significantly in a linear dioptric distance scale, which corresponds to short distances from the eye optics. Therefore, accommodation is typically instrumental for estimating depth to distances shorter than those supported by binocular visual cues.

Accommodation and vergence do not operate completely independently. Accommodation is partially driven by disparity and vergence is partially driven by retinal blur [11], [12]. Breaking the synchrony between visual cues results in VAC, in which the visual cues from retinal blur and disparity produce conflicting signals, causing visual discomfort for the viewer [13]–[15]. Thus, significant research efforts have been directed toward minimizing VAC while designing 3D NEDs.

2.2 Near-eye 3D displays

After discussing the main aspects of human 3D perception, we move forward to discuss how 3D content can be shown to a human using electronic displays. This section overviews the main categories of 3D NEDs. The aim of these displays is to visualize 3D content in an immersive fashion using wearable, head-mounted configurations. The different head-mounted 3D displays and the content they visualize are sometimes referred to as virtual reality, augmented reality or mixed reality, depending on to what extent the user can see their surroundings.

2.2.1 Stereoscopic

Stereoscopic NEDs constitute the vast majority of commercially available products. In such displays, one 2D image is projected for each eye, which in turn provides the visual cue of disparity to drive the perception of 3D. While the disparity is provided relative to the scene content, the display plane is fixed at a specific distance from the viewer (typically imaged by a magnifying lens in the NED). The former guides the vergence, while the latter is the focal depth of the display where the viewer typically accommodates their eyes to [16]. Apart from very specific situations, these are not matched, thus leading to VAC, which is a significant source of discomfort when using such displays [13], [14].

In addition to the conventional, commercially commonplace NEDs, various display design have been proposed particularly to alleviate or completely eradicate VAC. A design approach has aimed at making the defocus blur depth-invariant. For example, in Maxwellian-view displays the display image is focused directly on the retina through an aperture effectively smaller than the pupil [17]. However, the limited aperture size restricts light throughput and the viewing area. Alternatively, the displayed images can be projected to multiple different depths faster than the human temporal resolution using dynamic, varifocal optics [18]. A highly competitive approach to both, avoiding their major drawbacks, was presented in [19], where a jointly optimized pre-processing neural network and (static) display optics result in an extended DoF to achieve the depth-invariant defocus response.

In stereoscopic displays, only one view is provided per eye, thus limiting the viewing region or *eye box*. Even small movement of the viewer requires displaying (i.e. rendering) a new pair of views for the altered position. This further extends for providing motion parallax: the stereo views must be constantly refreshed based on the viewer position relative to the 3D scene. This facilitates the need for rapid, latency-free rendering in order to avoid noticeable lag.

2.2.2 Light field

There are multiple variations of *light field* (LF) displays, though regardless of the physical implementation, they recreate the 3D scene as a LF, i.e. a collection of light rays. Typically such LFs are parametrized on two parallel planes [20], and LF displays provide multiple views, or angular samples[2], within the extent of the pupil using incoherent imaging. The two-plane parametrization is visualized in Fig. 2.4. Depending on the exact display characteristics, LF NEDs avoid the need for constant refreshing of views, as multiple views are readily provided within the eye box, for both eyes separately. This is commonly achieved using a spatially-multiplexing design, where multiple display panel pixels are utilized for different angular LF samples (u, v) at the same spatial LF sample location (s, t). Such displays provide information about disparity and parallax also within the pupil, and thus, more accurate information for the eye to drive accommodation. Therefore, LF displays are considered potentially accommodation-enabling in terms of avoiding VAC.

The major downside of spatial multiplexing LF displays is the spatio-angular res-

---

[2] In a two-plane parametrization of LFs, typically one is related to spatial and the other to angular samples [21]. When captured using multiperspective images, we consider here the camera positions (or views) as the angular samples and the image plane (or sensor conjugate plane) samples as the spatial ones.

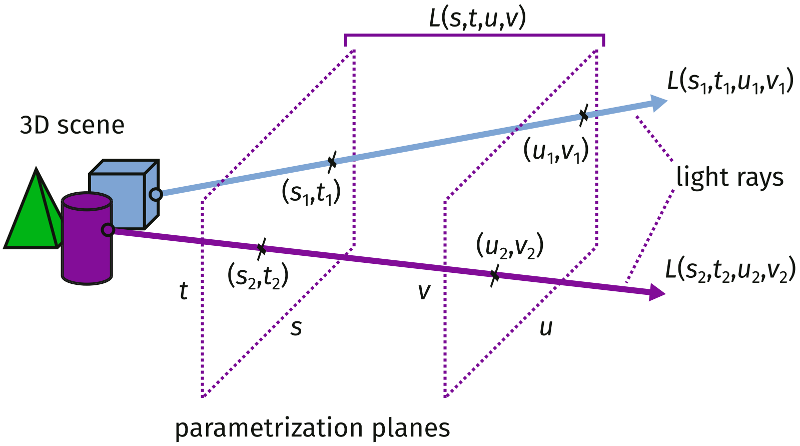


Figure 2.4 Two-plane parametrization of a light field. Illustrating two light rays (originating from two different objects within the scene) intersecting the (s, t) and (u, v) planes at different locations.

olution tradeoff; in order to provide multiple views within the eye box, the spatial samples of the display plane have to be redirected such that only a partial set is utilized for each different view. Thus, increasing the angular resolution (number of views) reduces the spatial resolution (number of pixels per view). This can be mitigated partially by suitable optical designs in the case of horizontal-parallax only LF displays, essentially by sacrificing vertical resolution instead of only horizontal, to allow more views [22]. Furthermore, directing spatial samples to angular samples typically requires additional optical elements, such as lenticular sheets (single direction parallax) [23], microlens arrays (full parallax) [24] or parallax barriers [25], increasing the complexity and cost of such display designs.

Due to the rather general definition of LF displays, a wide variety of different display designs have been researched and proposed to tackle the issues highlighted in the previous paragraph while maintaining the notable benefit of avoiding VAC. One such design for high spatial and angular resolution LF recreation consists of several point-like light sources illuminating the display panel sequentially to image multiple full-parallax viewpoints (up to 21, one view per light source) for the viewer [26]. This design avoids the harsh spatio-angular resolution tradeoff in more conventional spatial multiplexing designs, at the cost of reducing effective temporal resolution. In [27], a see-through design for augmented-reality use was achieved using a discrete lens array with gaps in between lenses to allow real world light to pass through. Recent advances in novel optics have accelerated different original LF display designs. For example, switchable optics, as in [28], allow for instance displaying views at multiple focal planes rapidly. Furthermore, metaoptics have enabled controlling optical properties with polarization, such as the integral imaging style LF display in [29]: the proposed design utilizes a metalens array where the polarization state tunes the focal length and pitch of the lens array.

2.2.3 Holographic

Often considered the ultimate method for displaying 3D content, *holographic* NEDs recreate the (complex-valued) wave field of the recorded scene utilizing coherent imaging and wave optics, where the amplitude of the wave encodes the intensity of light and the phase encodes the directionality [30]. The exact level of detail and accuracy is largely dependent on how the holographic data was formed, however, holographic displays can potentially provide smooth, continuous motion parallax including realistic object surface material properties. Like LF displays, holographic displays avoid VAC via accommodation-enabling design [31], [32]. Moreover, holograms can typically encode directional information of the 3D scene efficiently, thus avoiding the need for multiplexing spatial samples and compromising between angular and spatial resolution.

The highly accurate 3D imaging comes with substantial drawbacks. As a coherent imaging technique, it requires coherent light sources to illuminate the display, which in turn introduces speckle noise in certain situations (explained further in Sec. 2.3.3) [33]. Moreover, the full holographic information is complex-valued, which would require modulating both the amplitude and phase of the light; however, currently available spatial light modulators (SLM) can only modulate one at a time on a pixel level, not both simultaneously. Although full complex modulation can be achieved using a single phase-only SLM, it requires optically combining multiple adjacent pixels, effectively reducing the resolution of the display [34]. As such, the full information needs to be encoded either only as amplitude or phase values [35], [36], thus reducing efficiency and accuracy. Alternatively, both SLM types need to be incorporated into the display construction and optically combined [37], thus increasing the design complexity. Lastly, the computational complexity is a critical issue, particularly for real-time use, which will be further elaborated in the next section.

Despite the aforementioned drawbacks, a large body of research has been developed around holographic display schemes. In terms of holographic NEDs, some of the less critical issues have been approached specifically via optical means; for example, effective extension of the viewing region was realized with eye-tracking and pupil shifting optics to adjust the eye box position [38]. Furthermore, the majority of holographic display hardware prototypes have been demonstrated in a large optical table benchtop form factor, however, considerable efforts have been made to minimize their size to be more suitable for compact wearable (head-mounted) use [39], [40]. In [39], the small form-factor was achieved via backlight steering and waveguide usage, whereas the construction of Maimone et. al. relies on highly off-axis illumination and a miniaturized holographic projector without 4f optics [40].

2.3 Hologram synthesis

The computational synthesis of holograms, often referred to as *computer-generated holography* (CGH)[3] [41], [42], is a vital part of holographic NEDs; it defines how the hologram, or the complex-valued interference pattern, is calculated from a specific representation of a 3D scene. Here we will overview different hologram synthesis methods, in two broad categories: analytical and machine learning. Furthermore, the formation of speckle noise and how to minimize its effects is discussed.

2.3.1 Analytical methods

Analytical hologram synthesis methods cover the algorithms which apply a closed-form solution or a numerical approximation for transferring a 3D scene into a complex wave field on the display plane. As a part of the thesis work, we conducted a thorough survey in Publication I of analytical CGH methods, which we summarize in this section. The wide scope and long history of CGH is reflected in the number of different methods, as well as varying terminology and categorization of methods. However, one simple categorization can be formed based on the 3D scene representation used as the input data. Thus, we differentiate the methods into *ray-based*, *ray and depth-based*, and *model-based*. In general, each hologram can be represented in the form $ae^{j\phi}$, where a is the amplitude, carrying the intensity information and ϕ is the phase, containing the geometric information of the recorded scene. The relative

---

[3] The abbreviation CGH is used interchangeably for the area of computer-generated holography, the hologram generation methods and the resulting holograms.

magnitude of a for different wavelengths determines the color of the wave. Here we formulate the hologram synthesis methods without the initial phase term ϕ_0 for brevity. When needed, it can be included by simply adding ϕ_0 inside the exponential function ($ae^{j(\phi+\phi_0)}$) per relevant input primitive (e.g. point source or light ray).

2.3.1.1 Ray-based

In ray-based methods, the basic primitive of the 3D scene is a light ray, as captured in a set of images. As such, the directional ray information is implicitly known from the capture geometry. From this data, the hologram can be formed in a segmented manner, where each segment contains the angular information from different views sampled on the image plane position equivalent to the segment center. These segments, referred to as *holographic elements* (hogels), form a *holographic stereogram* (HS) [43], [44]: a non-overlapping tiling formation encoding the light rays of the input LF as waves. As illustrated in Fig. 2.5, each light ray corresponds to a windowed plane wave emitted from the center of the ray-sampled location on the hologram plane at (x_m, y_n) (discrete indices $[m, n]$) towards the camera plane sampling location at (s_p, t_q) (discrete indices $[p, q]$) with intensity $L(x_m, y_n, s_p, t_q)$ (or using discrete notation $L[m, n, p, q]$). The frequency of such a wave is a constant f_x^{mnpq} within the hogel, and is related to the geometric angle θ_x^{mnpq} through the grating equation:

$$f_x^{mnpq} = \frac{\sin\theta_x^{mnpq}}{\lambda},\ f_y^{mnpq} = \frac{\sin\theta_y^{mnpq}}{\lambda}. \tag{2.1}$$

The HS in then defined as

$$\begin{aligned} H_{\mathrm{HS}}(x, y) = \sum_{m=1}^{M}\sum_{n=1}^{N} \Pi\left(\frac{x - mh_x}{h_x}\right)\Pi\left(\frac{y - nh_y}{h_y}\right) \\ \sum_{p=1}^{P}\sum_{q=1}^{Q}\sqrt{L[m, n, p, q]}\exp[j2\pi(f_x^{mnpq}x + f_y^{mnpq}y)], \end{aligned} \tag{2.2}$$

where h_x and h_y are the hogel sizes horizontally and vertically. The hogels correspond to a spatial windowing by the rectangular function $\Pi(x)$

$$\Pi(x) = \begin{cases} 1 & \text{if } |x| \le \frac{1}{2} \\ 0 & \text{otherwise} \end{cases} \tag{2.3}$$

scaled and shifted accordingly. The sinusoidal fringes in Eq. 2.2 encode the directional information in their frequencies according to Eq. 2.1 [44], [45]. The inner summations in Eq. 2.2 over p and q are similar form to a discrete Fourier transform, and can be therefore efficiently obtained via fast Fourier transforms (FFTs) of the LF angular samples.

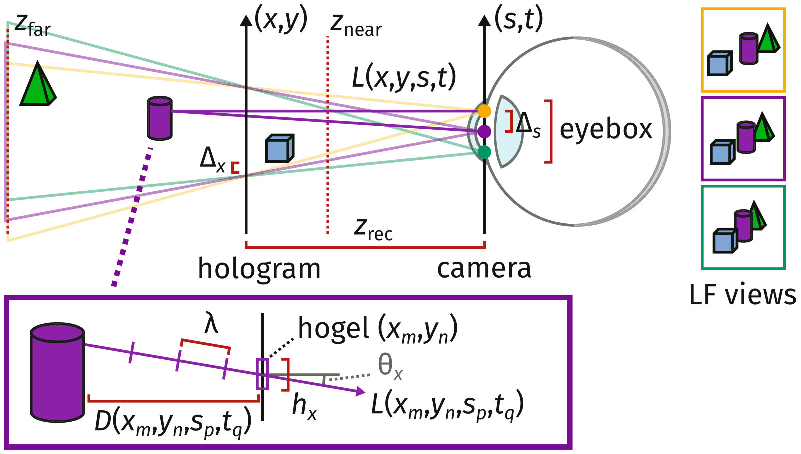


Figure 2.5 Correspondence between a typical two-plane parametrized LF and a plane wave CGH. The hologram plane (x, y) corresponds to the recentered image plane and the camera plane (s, t) to the assumed viewer location. Each light ray passes through the hologram plane at an angle θ relative to the hologram plane normal, which corresponds to a spatial frequency for a plane wave according to the grating equation.

For a self-emitting point source, in terms of wave optics, the HS corresponds to a discontinuous, piecewise linear approximation of a spherical wavefront. That is, the phases of each hologram segment do not align with each other creating discontinuities across hogel borders. As a result, such waves from a single point in the scene will intersect the location of the point correctly, albeit at different phases, therefore not summing coherently as an ideal wave would and thus causing minor errors. However, as the input scene representation (LF) intrinsically contains view-dependent properties, such as reflections and occlusions, the HS also encodes (and reconstructs when displayed) them. Thus, HSs can be considered as a holographic analog to LF displays, accompanied with similar drawbacks in terms of recreating 3D scenes and their visual cues.

2.3.1.2 Ray and depth -based

As an extension of the previous category, ray and depth -based methods augment the directional ray information with depth values. Typically such methods are considered in the context of computer graphics, where the depth values can be simply obtained from the Z-buffer of a 3D rendering engine, though depth sensors (such as time-of-flight [46]) can be utilized to acquire the same information during real world capture. The simplest extension is the phase-added stereogram (PAS) and its later variants [47], [48], which forms the hologram similarly to HSs, while aligning the planar waves to better match the approximation of a spherical wave:

$$\begin{aligned} H_{\mathrm{PAS}}(x, y) = &\sum_{m=1}^{M}\sum_{n=1}^{N} \Pi\left(\frac{x - mh_x}{h_x}\right) \Pi\left(\frac{y - nh_y}{h_y}\right) \\ &\sum_{p=1}^{P}\sum_{q=1}^{Q} \sqrt{L[m, n, p, q]} \exp\left[j2\pi\left(f_x^{mnpq} x + f_y^{mnpq} y + \frac{r^{mnpq}}{\lambda}\right)\right], \end{aligned} \tag{2.4}$$

where $r^{mnpq}(mh_x, nh_y)$ is the Euclidean distance along the ray from the hogel center (mh_x, nh_y) towards the origin of the ray, obtained from the depth map value $D[m, n, p, q]$ (defined here relative to the hologram plane[4]) as

$$x^{mnpq} = D[m, n, p, q] \tan \theta_x^{mnpq}, \quad y^{mnpq} = D[m, n, p, q] \tan \theta_y^{mnpq}, \tag{2.5}$$

$$r^{mnpq}(x, y) = \sqrt{(x - x^{mnpq})^2 + (y - y^{mnpq})^2 + D[m, n, p, q]^2}. \tag{2.6}$$

Like HSs, PASs can be computed using FFTs.

Alternatively, the plane wave approximation can be forfeited altogether, placing windowed spherical wave segments in place of hogels. In this approach, each LF sample then corresponds essentially to a bandlimited fringe based on the spatial and angular sampling of the LF. This hologram synthesis method is known as *diffraction-*

---

[4] Often the depth map obtained from computer graphics renderers corresponds to the orthographic image projection distance between the rendering (camera) location and the first intersection point in the 3D scene. In our notation, we use the depth map as a version which has been subtracted from the distance between the hologram plane and the capture device to get the z coordinate of each light ray relative to the hologram plane.

specific coherent panoramagram (DSCP), and is defined as [49]

$$H_{\mathrm{DSCP}}(x,y) = \sum_{m=1}^{M}\sum_{n=1}^{N} \Pi\left(\frac{x-mh_x}{h_x}\right)\Pi\left(\frac{y-nh_y}{h_y}\right) \sum_{p=1}^{P}\sum_{q=1}^{Q} \sqrt{L[m,n,p,q]}\exp\left[jk\left(r^{mnpq}(x,y)-D[m,n,p,q]\right)\right], \quad (2.7)$$

where $k = 2\pi/\lambda$ is the wave number for wavelength λ, and $r^{mnpq}(x,y)$ is the Euclidean distance between the hologram coordinate (x,y) and the origin of the discrete ray $[m,n,p,q]$ according to Eq. 2.6. Unlike the previously mentioned stereograms, DSCP cannot be computed using Fourier transforms, thus making it significantly more expensive in term of computation time. The computational complexity can be partially reduced by altering the capture geometry, placing the cameras between the hologram and the viewer, and limiting the LF rendering to only samples within each hologram segment [50]. Regardless, the resulting hologram is a better approximation of the continuous wave field due to the curved wavefronts, and thus more suited for accurate reconstruction of larger scene depth ranges.

One of the state-of-the-art approaches in this category is the overlap-add stereogram (OLAS) [51]. It is similar to HSs and PASs, depending on the configuration, with the notable exception to the placement of hogels as they are allowed to overlap (rather than be tiled). The method takes advantage of the link between holograms and (observable) LFs via short-time Fourier-transform (STFT): the squared magnitude of the forward STFT of a hologram results in a LF [52]. The inverse problem, i.e. LF to hologram, is ill-posed due to the LF on its own not containing phase information. However, the problem can be solved by adding depth information as a phase to the LF rays and by utilizing the sliding window overlap-add method to invert the STFT. Consequently, the OLAS hologram from a LF $L(x,y,s,t)$ and corresponding ray distances $r(x, y, s, t)$ (to each hogel center as in PAS) is defined as

$$\begin{aligned} H_{\mathrm{OLAS}}(x,y) = \iint\limits_{-\infty}^{\infty} w(x-x', y-y') \iint\limits_{-\infty}^{\infty} \sqrt{L(x', y', s, t)} \\ \exp\left[j2\pi \left(f_x(s,t)x' + f_y(s,t)y' + \frac{r(x', y', s, t)}{\lambda} \right) \right] \mathrm{d}s\, \mathrm{d}t\, \mathrm{d}x'\, \mathrm{d}y' \\ = \mathrm{STFT}^{-1} \left\{ \sqrt{L(x,y,s,t)} \exp[jkr(x,y,s,t)] \right\}, \end{aligned} \tag{2.8}$$

where $w(x, y)$ is the Hann synthesis window function of the inverse STFT, and f_x and f_y are the horizontal and vertical frequencies of the plane waves, as in Eq. 2.1.

The main advantage of OLAS is that it breaks the spatio-angular tradeoff present in other stereograms, while still maintaining the majority of their benefits, such as view-dependent effects and hogel generation via Fourier transforms. Compared to HSs and PASs, however, the computational burden is increased due to having only a single pixel shift between the overlapping hogels, and thus a larger number of hogels (and Fourier transforms). As such, OLAS is not suitable for real-time applications.

2.3.1.3 Model-based

Model-based hologram synthesis relies on 3D scene representations utilizing geometric primitives. This can be for example a point cloud [53] or a polygon mesh [54], from which waves or partial wave fields are emitted from and superposed to obtain the hologram. Due to the scene representation requirement, such methods are typically restricted to synthetic (computer graphics) 3D scenes, although they have some overlap with ray and depth -based methods as ray-casting (or sometimes rendered images) is often utilized to determine occlusions [55].

The simplest hologram synthesis method considers the scene as formed by a 3D point cloud and the wave field is generated by independent self-emitting point sources of light. The total hologram from the entire set of point sources (P points) is obtained as a superposition of spherical waves [53]:

$$H_{\mathrm{PBM}}(x,y) = \sum_{p=1}^{P} \frac{a_p}{r_p} \exp(jkr_p), \tag{2.9}$$

where a_p is the amplitude of point p, r_p is the Euclidean distance between point p

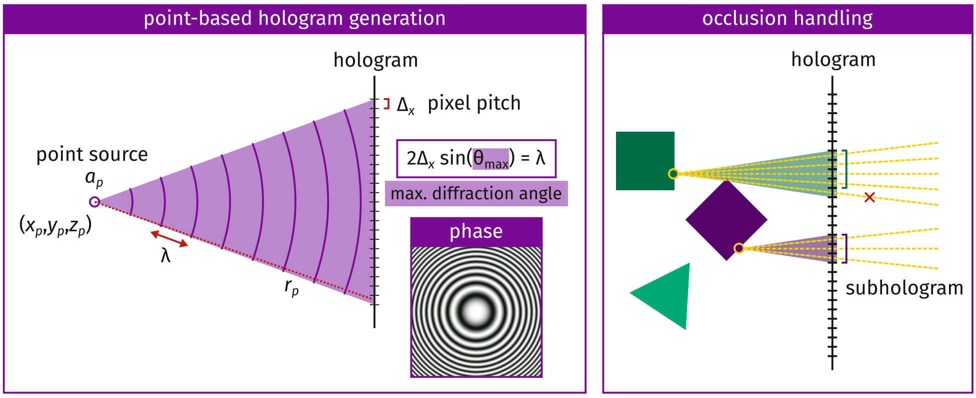


Figure 2.6 Basic principle behind point-based hologram synthesis (left). Each point in the scene corresponds to a spherical wave fringe on the hologram plane. When considering discretely sampled holograms, the maximum fringe (or subhologram) size is limited by the maximum diffraction angle to avoid aliasing the signal. The right block illustrates one option for handling occlusion boundaries along different point contributions. A ray is cast from the point towards each subhologram sample. If it intersects a surface within the scene (usually estimated via simple meshing techniques), the sample is deemed occluded and thus not included in the subhologram.

and (x, y). An example of a single point source hologram is shown in Fig. 2.6. The point cloud information can be explicitly provided as coordinate and color triplets, however, the same data can be inferred from a single color image and depth map to inherently manage most occlusions within the scene. Although the method is simple to implement, it has several critical drawbacks; most notably, point clouds do not contain any directional amplitude information, thus lacking realistic material reconstruction for glossy and other non-Lambertian surfaces. The methods also require careful handling of occlusions not to superpose hidden scene parts with visible ones [56] (also shown in Fig. 2.6), and quickly become more computationally expensive as the number of points P increases. The computational complexity can be partially relieved by precomputing several holographic patterns (corresponding to points at different distances from the hologram) and storing them in a look-up table at the cost of increased data storage requirements [57].

2.3.2 Machine learning methods

Machine learning (ML) has been steadily becoming more common in CGH to alleviate the computational burden typically present in analytical methods. Such ML-based approaches utilize large data sets of pre-calculated reference holograms and their corresponding input data to train the hologram generation to be performed by an efficient neural network structure, which in turn allows real-time hologram synthesis at the cost of scene reconstruction accuracy. Naturally, the ML-based methods are heavily affected by the input data properties, the network structure, the reference hologram generation method, as well as how the network output and reference hologram are compared against each other (i.e. the loss metric). Fig. 2.7 visualizes the general structure of how ML-based methods are typically trained. While certain specialized approaches have utilized ML as a tool for holographic imaging and could be included in this category, they are of limited interest in the context of this thesis. These include examples such as encoding flat 2D images into phase-only holograms optimized via a display-observing camera setup [58], learning a non-iterative alternative for converting fully complex holograms to phase-only [59], [60] and obtaining holograms under strict hardware limitations [61].

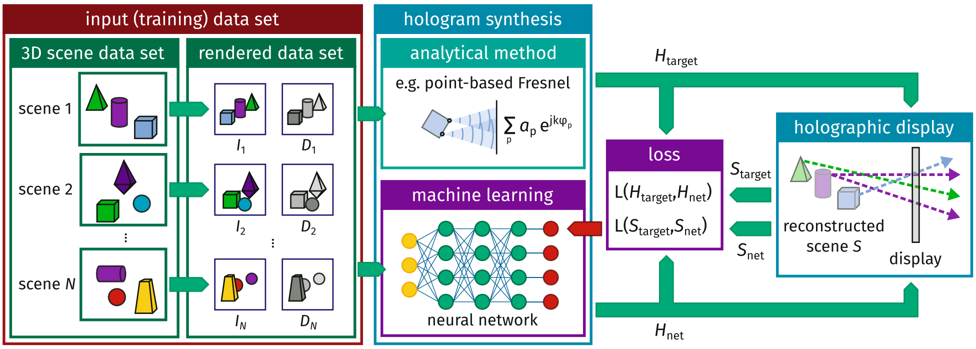


Figure 2.7 General pipeline for ML-based hologram synthesis. During training, the system receives scene information (e.g. RGB-D image) as an input, the network transforms it into a hologram, which is compared against a reference hologram from the same scene/input via a loss function driving the network weights to be changed.

In the context of this thesis, the most relevant approach in machine learning methods is Tensor Holography (or shortened TensorHolo) [62]. In TensorHolo, a fully-connected convolutional neural network (CNN) is trained using Fresnel-based reference holograms from single 2D RGB views and their equivalent depth maps. The resulting CNN is capable of reproducing single viewpoint Fresnel-based holograms at real-time speeds. Subsequently, the approach was updated to utilize layered depth images and a layer-type hologram synthesis method to better handle occluded scene regions [63], [64]. Compared to analytical methods, TensorHolo has a noticeable advantage in terms of synthesis speed. However, due to utilizing only a single viewpoint, the reference holograms (and the subsequently trained network) implicitly model each scene point as a diffusely emitting light source and lack support for view-dependent properties within the eye box. It should also be noted that the loss component, particularly the perceptual loss, lacks the effects of finite aperture (i.e. diffraction-limited imaging), non-defocus imaging effects (e.g. spherical lens aberration or other higher order Zernike polynomials) and support for other human vision imperfections (e.g. chromatic aberrations).

2.3.3 Speckle noise

The final image of an imaging system, in general, can be modelled as a superposition of point source contributions on the sensor, each one represented by a point spread function (PSF). If the points are laterally near each other, the PSFs overlap, and under coherent illumination, interfere with each other. For coherent imaging systems, such as hologram synthesis and holographic displays, this means that neighboring content in the recorded 3D scene (such as point sources or other similar primitives) interfere on the retina of a human viewer. Moreover, if the initial phases of such points ϕ_0 are assigned randomly to better diffuse light on the display[5], the resulting interference pattern is also random, causing high-contrast intensity variations on the perceived view known as *speckle noise* [33].

The intensity of speckle noise can be diminished by reducing the coherence of the display illumination source; reducing spatial or temporal coherence of the light source have both been shown to suppress speckle noise in practice, the former being achieved with a diffuser [66] and the latter by utilizing light-emitting diodes as the illumination source [67]. However, the reduction in coherence comes at the cost of additional spatial blur in the reconstructed images which can hinder the 3D reconstruction of large depth range scenes [67]. Alternatively, the problem can be approached via computational means. For example, by generating several holograms with different random phases and displaying them faster than the HVS temporal resolution (i.e. time-multiplexing), the random noise patterns are effectively averaged, and thus, reduced in intensity [68]. This random averaging approach, however,

---

[5] Concentration of light on the display is especially problematic when utilizing phase-only modulation for the display. In such cases, it is common to add phases (per primitive) from a random uniform distribution [65].

achieves speckle contrast reduction proportional to the square root of the number of frames, therefore necessitating a large number of frames to be displayed in sequence for effective speckle suppression. This cuts down the usable frame rate significantly and requires a high framerate SLM in practice.

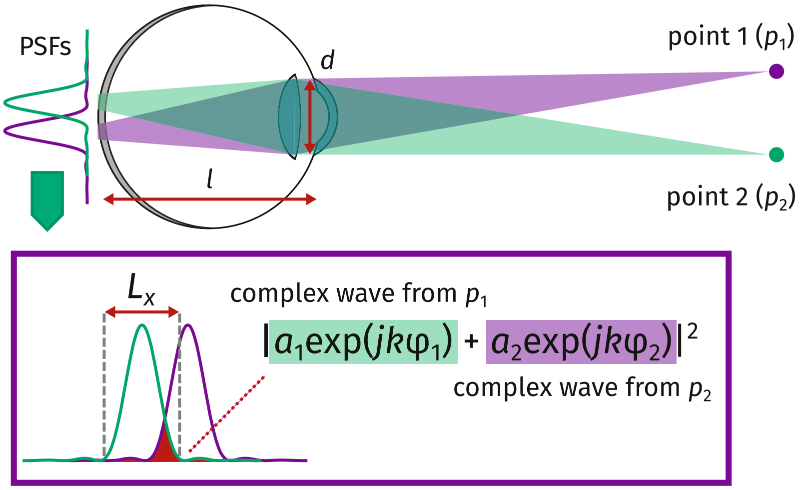


Figure 2.8 Interference between two observed points close to each other. Each point is imaged as a PSF on the sensor (retina), estimated and visualized here as an Airy disk. The degree of overlap between the points (highlighted in red) is dependent on the Airy disk main lobe width and the distance between the two points when imaged on the sensor.

The strongest effect of the PSF on speckle formation is quantified by the main lobe of the Airy disk, under the assumption of a circular-aperture imaging system and an aberration-free thin lens. Such imaging system can be considered as a simplified eye model, in which case the average size of speckle(s) on the sensor (retina) for a diffused scene surface can be estimated via the Airy disk main lobe width [33]:

$$L_x \approx 2.44\frac{\lambda l}{d}, \tag{2.10}$$

where d is the circular aperture diameter and l is the distance between the aperture and the sensor.

If the overlapping region between adjacent PSFs on the retina, as shown in red in Fig. 2.8, is minimized, the resulting speckle patterns from the randomly interfering phases are also suppressed. As such, the speckle noise can be reduced by increasing the distance between scene points that are superposed at a given frame. If frames from such sparsely populated scenes are then later combined, e.g. via time-multiplexing, the severity of the speckle noise is greatly reduced as less overlap between adjacent points occurs. This is the basic principle of various separation-based speckle suppression methods applied for model-based holograms [69] and ray and depth -based holograms [70].

2.4 Optical modelling of coherent displays

In order to computationally evaluate the performance and visual quality of various coherently illuminated displays, we must model light diffraction and the propagation of light in free space. The Huygens-Fresnel principle defines the diffraction of monochromatic scalar wave fields [71]: for a reference (input) wave field $U_0(\xi, \eta) \in \mathbb{C}$, the diffraction field at distance z is obtained via the convolution

$$U_z(x, y) = \iint_{\Sigma} U_0(\xi, \eta) h_z(x - \xi, x - \eta) \,\mathrm{d}\xi \,\mathrm{d}\eta, \tag{2.11}$$

where $h_z(x, y)$ is the convolution kernel and Σ is the nonzero region of $U_0(\xi, \eta)$. Assuming that $z \gg \lambda$, the principle in Eq. 2.11 can be solved using the Rayleigh-Sommerfeld (RS) kernel for h_z:

$$h_z(x, y) = \frac{z}{j\lambda} \frac{\exp\left(jk\sqrt{x^2 + y^2 + z^2}\right)}{x^2 + y^2 + z^2}. \tag{2.12}$$

The convolution kernel h_z is also referred to as a diffraction kernel.

In practice, angular spectrum method (ASM) is frequently utilized to implement RS diffraction. ASM essentially represents the propagation as a set of plane waves in the spectral domain, meaning that the convolution from Eq. 2.11 is replaced with forward and inverse Fourier transforms (denoted here as $\mathcal{F}\{\cdot\}$ and $\mathcal{F}^{-1}\{\cdot\}$) and multiplication as

$$U_z(x, y) = \mathcal{F}^{-1}\{H_z(f_x, f_y)\mathcal{F}\{U_0(\xi, \eta)\}|_{(f_x, f_y)}\}, \tag{2.13}$$

where

$$H_z(f_x, f_y) = \begin{cases} \exp\left(j2\pi z\sqrt{\frac{1}{\lambda^2} - f_x^2 - f_y^2}\right) & \text{if } \sqrt{f_x^2 + f_y^2} < \frac{1}{\lambda} \\ 0 & \text{otherwise} \end{cases} \tag{2.14}$$

is the frequency response of the RS kernel. If the field propagation is assumed to follow the small-angle, or paraxial approximation, the spherical wavefronts of Eq. 2.12 can be replaced with parabolic ones. This results in the so-called Fresnel diffraction kernel, which can also be implemented via ASM using the appropriate frequency response:

$$H_z(f_x,f_y) = \exp(jkz)\left[-j\pi\lambda z\left(f_x^2 - f_y^2\right)\right] \tag{2.15}$$

The discrete implementation of ASM assumes periodic signals at all times (i.e. input, during propagation and output). As the propagated plane waves during ASM are confined within $\pm\arcsin\left(\frac{\lambda}{2\Delta_\xi}\right)$ (assuming field sampling at Δ_ξ), i.e. bandlimited by frequencies $\pm 1/\Delta_\xi$ and sampled according to the Nyquist theorem, each period of the signal contributes energy onto a fixed area on the output field as dictated by the propagation distance and plane wave bandwidth. Avoiding replica contributions to obtain accurate results requires padding the input field with zeroes such that the waves corresponding to the highest frequency component from the periodic replicas do not interfere with the output field (at the same horizontal and vertical extent as the input) after propagating by z [72]. For a field sampled at Δ_ξ, this means adding at least

$$n_z = \frac{z}{\Delta_\xi}\tan\left[\arcsin\left(\frac{\lambda}{2\Delta_\xi}\right)\right] \tag{2.16}$$

zero samples around the input field U_0.

2.4.1 Optical elements

Optical elements along the light path have varying effects on the wave field. For example, dynamic optoelectronic devices such as SLMs modulate either the amplitude or the phase of the incoming light, and can be simply modelled by multiplying the field with the amplitude or phase modulation effect. A commonly occurring optical element in various imaging setups is a lens which operates on the phase of the passing light. Essentially, an ideal lens delays the phase of light relative to the size and focal power of the lens. These can be modelled via a phase function known as lens transmittance[6]:

$$Q_f(x,y) = A_d(x,y)\exp\left[-\frac{j\pi}{\lambda f}\left(x^2+y^2\right)\right], \tag{2.17}$$

---

[6] Also called pupil function or simply lens function, depending on the context.

where f is the focal length of the lens, and

$$A_d(x, y) = \begin{cases} 1 & \text{if } \sqrt{x^2 + y^2} < \frac{d}{2} \\ 0 & \text{otherwise} \end{cases} \tag{2.18}$$

is the ideal circular lens aperture function for a lens diameter d. In the case of computational simulations, the lens transmittance function must be discretely sampled such that the sampling step is at most $\lambda f / d$ to avoid aliasing.

2.4.2 Basic coherent imaging system

The introduced concepts on modelling wave optics can be brought together to form a model of a basic coherent imaging system. More specifically, for a given input wave field $U_0(\xi, \eta)$ at distance z from the imaging system, a single lens and a sensor separated by distance l, the imaged intensity $I(x, y)$ on the sensor plane is modeled as:

$$I(x, y) = |\mathcal{P}_l\{\mathcal{P}_z\{U_0(\xi, \eta)\}Q_f(s, t)\}|^2, \tag{2.19}$$

where $\mathcal{P}_z$ is a scalar field propagation operation for distance z and $\mathcal{P}_l$ for distance l, e.g. using Eqs. 2.13 and 2.14. In discrete simulations, the sampling grids on the (s, t) plane of the lens transmittance and the incoming field should match; either by choosing the lens sampling according to the field sampling, or by resampling the field to match the lens sampling.

The imaging system in Eq. 2.19 provides a simple method for simulating either a capture device, such as a camera, or more importantly, a human viewer. In their simplest form, the parts equivalent to a human eye are pupil for the lens and retina for the sensor. Although this simplification is rather crude, in many cases it is sufficient for analyzing the visual quality of different holograms in terms of focusing and parallax. Moreover, the imaging system model can be extended to incorporate further phase terms of the lens by modifying the phase term of Q_f or amplitude-modulation by modifying the aperture function A.

3 SYNTHESIS AND ANALYSIS OF NEAR-EYE HOLOGRAMS

As affirmed in the previous chapters, there is a clear need for hologram synthesis methods which can accurately recreate a desired 3D scene in a noise-free manner, as well as methods for quantifying the accuracy of the reconstructed visual scenery. Due to the complete nature of LFs and capabilities to provide all necessary information within a 3D scene in the context of NEDs, the contributions in this thesis have focused on using LFs as the input data, further augmented by depth values to help to precisely align the phases of waves from the same scene source and hence to enhance the hologram accuracy.

The first work presented here, as reported in Publication II tackles the issue of speckle noise in ray and depth-based holograms. Specifically, the proposed approach addresses a light intensity mapping issue in an existing ray separation method. Furthermore, we discuss Publication III, which summarises a simple HS analysis framework in terms of provided accommodation cue accuracy. Lastly, the final section of the chapter presents a novel method for generating 3D holograms from LF and depth, capable of producing correct visual cues even for non-Lambertian surfaces and large depth ranges. The hologram synthesis method was developed as a part of Publication V.

3.1 Speckle suppression for ray and depth -based holograms

As part of hologram synthesis, one critical aspect is the initialization of the constant phase term ϕ_0. Typically, the best approach is to utilize, if available, the positional information of the scene and assign the phase relative to the distance between the wave-originating primitive (e.g. point source) and the hologram plane [40]. Yet, in some cases it is beneficial to utilize randomly distributed initial phase to avoid concentration of light on the display and to better simulate the effects of diffuse lighting. As explained in Sec. 2.3.3, this results in highly disruptive speckle noise, perceived by the viewer. In ray and depth-based CGH methods, time-multiplexing holograms from spatially separated ray sets has been shown to effectively reduce the noise [70]. However, the existing solution did not accurately map intensity changes across points when quantizing their locations, thus decreasing the reconstruction quality. We aim at alleviating this issue through signal processing means and by leveraging the reconstruction properties of densely sampled light fields (DSLF) [73].

3.1.1 Quantization of scene points

Although there are no explicit point sources in image-based scene representations, such can be associated to light rays utilizing the available depth information as illustrated in Fig. 3.1. The point location (x_{mi},z_{mi}) for the light ray corresponding to the pixel m of captured image at view i is obtained as

$$x_{mi} = mh_x + \frac{z_{mi}(i\Delta_s - mh_x)}{z_{\text{rec}}}, \tag{3.1}$$

$$z_{mi} = D[m, i], \tag{3.2}$$

where $D[m, i]$ is the corresponding depth map value and z_{rec} is the distance between the camera and hologram planes. Please note that the following discussion will only consider a horizontal cross-section of the full 3D space for brevity, as the extension to vertical y coordinates is straightforward. For effective spatial separation of light rays, we must first quantize these point source locations to a regular grid. Assuming that the imaging system considered (e.g. HVS) is diffraction-limited, no spatial resolution information is lost if the grid quantization step is defined to be less than the Rayleigh resolution limit [71]. As such, the lateral quantization step can be chosen as [70]

$$\Delta_{\tilde{x}} \leq 1.22\frac{\lambda z_{\text{rec}}}{d} \tag{3.3}$$

for a pupil diameter d. In ray-based (or ray and depth) CGH methods where the segments do not overlap, this is also typically the targeted segmentation size limit. Consequently, the quantization grid can be aligned with the hologram segment centers laterally, also resulting in $\Delta_{\tilde{x}} = h_x$. On the other hand, the axial quantization step can be derived from the depth acuity of human vision. Although it has been mostly

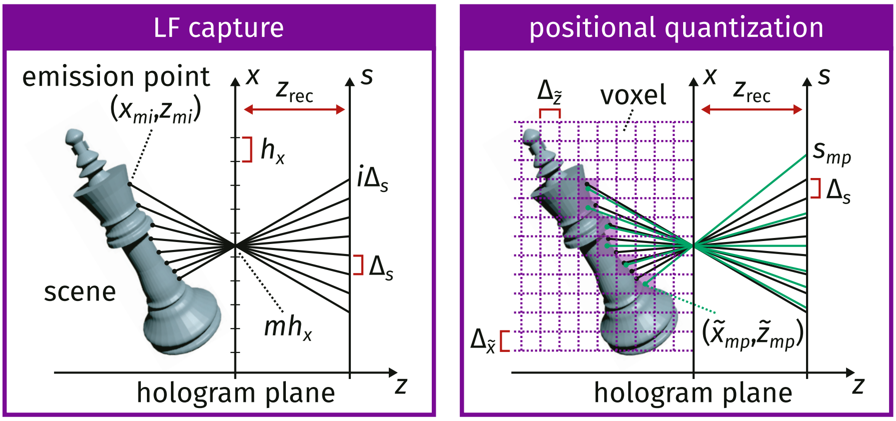


Figure 3.1 Obtaining the emission coordinates from a captured LF using depth information and knowledge of the capture geometry (left). These points are quantized in terms of their locations within a voxel grid, creating a new LF parametrization on the s-plane, s_{mp}, as indicated by the cyan-colored lines (right).

studied experimentally, we can utilize stereoacuity as a rough estimate in most of the scenarios: at depth z, the depth difference $\delta_z(z)$ just detectable by the HVS can be estimated as [74]

$$\delta_z(z) = \frac{z^2 \delta_\gamma}{cB}, \tag{3.4}$$

where δ_γ is the angular measure for the stereoacuity of the HVS, which is typically around 0.5 arcmin, B is the interpupillary distance (typically around 6.5 cm), and $c = 3437.75$ is a constant defining conversion from radian to arcmin. Therefore we can assume that little to no information is lost in terms of human vision by choosing the depth quantization step as $\Delta_{\tilde{z}} = \delta_z(z_{\text{rec}})$.

Having formed the quantization grid, the quantized point coordinates on the grid can be obtained by finding the nearest grid point for each point source as [70]

$$(\tilde{x}_{mi}, \tilde{z}_{mi}) = \underset{(\tilde{x},\tilde{z})\in S_q}{\arg\min} \left\{ \sqrt{(\tilde{x} - x_{mi})^2 + (\tilde{z} - z_{mi})^2} \right\}, \tag{3.5}$$

where S_q is the entire set of points on the quantization grid. As shown in Fig. 3.1, this divides the scene space into equal size quantization volumes (voxels) surrounding each quantization point.

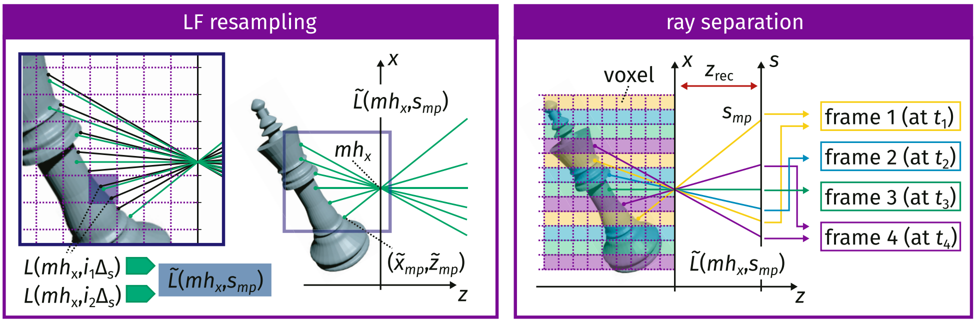


Figure 3.2 Resampling the LF intensities to match the quantized locations in the left figure. The speckle suppression is achieved by generating holograms from sparse voxel sets, as highlighted in different colors on the right figure, and time-multiplexing them.

3.1.2 Light field resampling

Once the quantized point source locations (or the equivalent voxels) have been found, their correct light intensities need to be accurately defined, particularly in relation to the hologram segments. Thus, let us denote the LF between the quantized points and the hologram segment m as $\tilde{L}(mh_x, s_{mp})$, where the corresponding rays intersect the camera plane at s_{mp}, $p = 1, 2, \ldots, P$; $P \leq N$ and N represents the total number angular samples per hogel. The new LF samples can be defined as done in [70] using the captured discrete LF $L[m, i]$:

$$\tilde{L}(mh_x, s_{mp}) = \sum_{i \in V^{mp}} L[m, i], \tag{3.6}$$

where V^{mp} denotes the set of indices i for which the emission points corresponding to captured rays for hogel m are inside the voxel p. The index sets V^{mp} can be found through

$$V^{mp} = \frac{1}{\Delta_s}\left[mh_x - \frac{z_{\mathrm{rec}}}{Z^{mp}}(X^{mp} - mh_x) \right], \tag{3.7}$$

where (X^{mp}, Z^{mp}) is the set of original ray emission coordinates that are quantized on to the point $(\tilde{x}_{mp}, \tilde{z}_{mp})$ representing the voxel p for hogel m. However, as pointed out in Publication II, this summation based mapping can create originally nonexistent intensity variations along the angular dimension (s plane), which the viewer observes as varying intensity regions along the surfaces of the 3D scene.

The proposed solution in Publication II addresses the mapping issue by more accurately obtaining the desired unknown samples $\tilde{L}(mh_x, s_{mp})$ from the known data samples $L[m, i]$. This is achieved through standard signal processing practices: we reconstruct the continuous function, i.e. the continuous LF between mh_x and the camera plane s, $L(mh_x, s)$, and resample it at the desired sample positions of $s = s_{mp}$. If the captured LF is sampled according to the DSLF criterion [73], the LF resampling can be achieved simply by utilizing linear interpolation as follows [1]; for the light ray emitted from the quantized point $(\tilde{x}_{mp}, \tilde{z}_{mp})$ corresponding to voxel p for hogel m, first the corresponding intersection point on the camera plane s_{mp} is obtained as

$$s_{mp} = mh_x - \frac{z_{\text{rec}}(\tilde{x}_{mp} - mh_x)}{\tilde{z}_{mp}}. \tag{3.8}$$

The surrounding ray indices i_1 and i_2 are then acquired from the two nearest camera plane coordinates of the captured set of rays i as

$$i_1 = \underset{i \le s_{mp}/\Delta_s}{\arg\min} \left| i\Delta_s - s_{mp} \right|, \tag{3.9}$$

$$i_2 = \underset{i > s_{mp}/\Delta_s}{\arg\min} \left| i\Delta_s - s_{mp} \right|. \tag{3.10}$$

Finally, utilizing the corresponding LF samples $L[m, i_1]$ and $L[m, i_2]$, the intensity value of $\tilde{L}(mh_x, s_{mp})$ is obtained through linear interpolation, i.e.

$$\tilde{L}(mh_x, s_{mp}) = \frac{L[m, i_1](i_2\Delta_s - s_{mp}) + L[m, i_2](s_{mp} - i_1\Delta_s)}{\Delta_s}. \tag{3.11}$$

Compared to the voxel mapping in Eq. (3.6), our solution employs a rigorous signal processing framework, thus avoiding the issues of the original method.

3.1.3 Time-multiplexing ray separated holograms

The hologram (or the complex valued field at the display) is obtained from the resampled LF samples as

$$H(x) = \sum_m \Pi\left(\frac{x - mh_x}{h_x}\right) \sum_p \sqrt{\tilde{L}(mh_x, s_{mp})} \exp\left[jk\left(\tilde{r}_{mp}(x) - \tilde{z}_{mp}\right)\right], \tag{3.12}$$

---

[1] Importantly, if the LF is not densely sampled, it can be computationally reconstructed using sparsification priors [75].

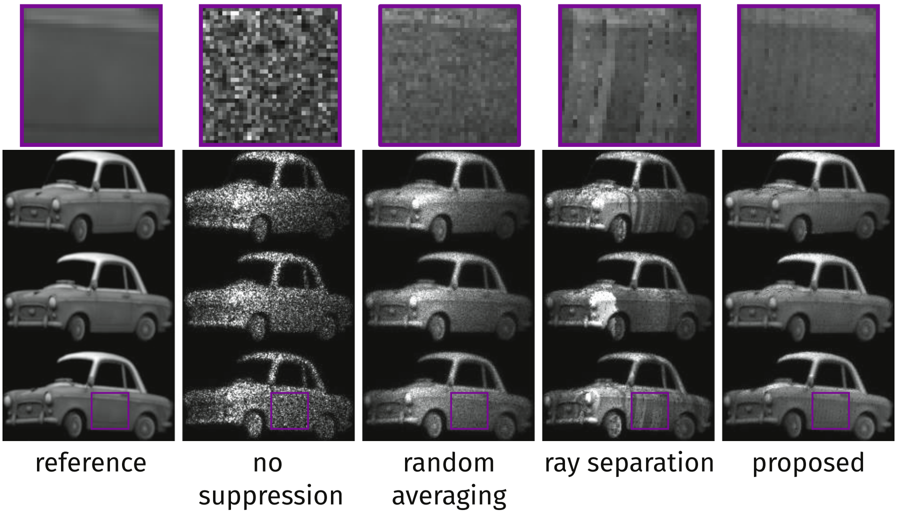


Figure 3.3 Example view simulation results comparing different speckle reduction methods. From left to right: reference image, no speckle suppression, random averaging, ray separation, and proposed method. The top row shows the highlighted regions enlarged for better view of details.

where $\tilde{r}_{mp}(x)$ is the distance between the center of voxel p and the pixels in hogel m. This process generates a hologram of the location-wise quantized version of the recorded scene.

For effective speckle noise reduction, the distance between adjacent scene points recorded on the hologram should be increased as explained in Section 2.3.3. That is, only sparse sets of all quantized points (or the equivalent light rays) are to be included in a single hologram frame, according to the horizontal and vertical coordinates. In practice, this is achieved by including every Nth row and column of the quantization grid into each hologram frame (visualized in Fig. 3.2), which are then displayed sequentially. Choosing the multiplexing factor N is a compromise: larger N results in better ray separation and speckle suppression, at the cost of increasing the number of hologram frames to be N^2. Theoretically $N = 2$ ensures that at least L_x lateral separation is achieved on the retina (for points on same depth plane), though due to contributions from the side lobes of the PSF, increasing N to 4 is highly beneficial. Assuming that the frames are displayed in a fast enough manner, i.e. faster than the temporal resolution of the human vision, the resulting reconstruction combines speckle suppressed frames of partial LFs such that the full scene is perceived with reduced speckle noise.

In Publication II, we demonstrated that the method outperforms the traditional random averaging speckle suppression method [68] when utilizing the same number of frames. Additionally, our LF resampling approach for assigning the quantized LF intensities alleviated the intensity variation issues in the existing ray separation approach [70]. Although small reconstruction errors in the form of dark intensity stripes can still be seen in the results (Fig. 3.3), the proposed approach resulted in significant improvement for suppressing speckle noise in RGB-D LF holograms.

3.2 Analysis of accommodation cues in holographic stereograms

In the context of ray-based CGH methods, in addition to handling speckle noise, there is an important question to consider regarding the level of approximations used for hologram synthesis. Specifically, HSs are an attractive choice due to their fast generation even though the total wave field is modelled as a collection of (unaligned) plane waves. However, as we are interested in using holographic imaging as an accommodation-enabling solution to VAC, it should accurately recreate accommodation cues. To better understand the exact behaviour of accommodation cues in relation to HSs, we developed a simple analysis framework, motivated by similar ones utilized in LF display analysis in Publication III. Aiding our analysis is the fact that HSs can be directly linked between LF displays due to how they model light (i.e. plane wave segments), thus making it easier to identify specific parameters which influence e.g. the spatial and angular resolution of the viewer perceived LF. Furthermore, such analysis specifically for accommodation cues had not been previously done rigorously for HSs, as existing works had covered the topic from a more generic LF sampling point of view [76], [77].

3.2.1 Analysis framework

The analysis relies heavily on accurately quantifying the effective spatial and angular sampling of the LF created by an HS, and especially on linking the relevant hologram parameters to these indicators, as well as to the number of rays within pupil. In this context, the key hologram parameters are the pixel pitch and the hogel size. The hologram pixel pitch Δ_x defines the highest spatial frequency content the hologram

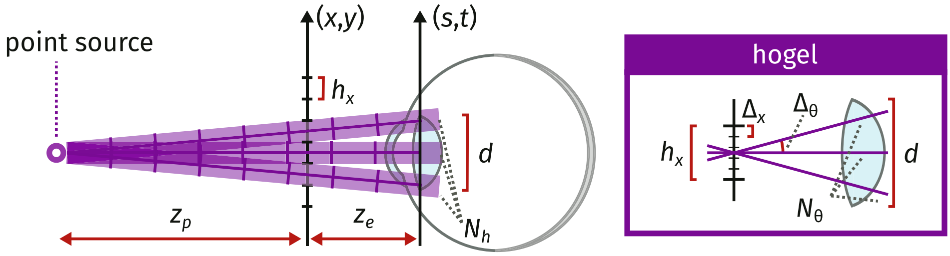


Figure 3.4 The analysis relevant parameters of the HS. The left side illustrates the parameters for determining the number of plane wave segments within the pupil extent from a single point source. The right side illustrates the angular sampling from a single hogel and the resulting number of plane wave segments across the pupil extent.

can record and thus the maximum diffraction angle $\theta_{\max}$, which is defined as

$$\theta_{\max} = \arcsin \frac{\lambda}{2\Delta_x}. \tag{3.13}$$

The other key parameter, hogel size h_x, defines the uniform spatial frequency sampling within a hogel. This corresponds to a non-uniform angular domain sampling according to the grating equation, especially for large angles. Despite this, the hogel size h_x has the largest contribution to the angular sampling Δ_θ, which can be estimated as

$$\Delta_\theta \approx \arcsin\left(\frac{\lambda}{h_x}\right), \tag{3.14}$$

and determines how accurately a hogel can direct the plane waves in the desired direction. Based on the given angular resolution, we define the total number of planar wavefront segments (correspondingly rays) that can be accurately directed within the pupil extent from a hogel as

$$N_\theta = \frac{d}{z_e \tan \Delta_\theta}, \tag{3.15}$$

where d is the pupil diameter and z_e is the distance between the pupil and the hologram. The parameters discussed here are visualized in Fig. 3.4.

The hogel size determines also the perceived spatial resolution at viewing distance, as it represents the smallest horizontal and vertical distance between adjacent beams of light. As such, it is typically chosen based on limitations from human vision. Assuming the human eye as a diffraction limited imaging system, the minimum resolvable distance between two points at (axial) distance z_e is defined by the Rayleigh criterion as [71]

$$\Delta_x^{\mathrm{HVS}} = 1.22\frac{\lambda z_e}{d}. \tag{3.16}$$

For a pupil size d and intended viewing distance z_e, Δ_x^{HVS} is the upper limit for the hogel size which will still ensure that the perceived spatial resolution is maximized. However, such choice of hogel size results in $N_\theta \approx 1$ according to Eq. 3.15. Increasing the hogel size, and thus lowering the perceived spatial resolution, increases N_θ and leads to a similar trade-off between angular and spatial resolution as in LF displays [78].

Another factor to consider in the accommodation analysis is, for a single point source, the number of planar wavefront segments emitted from multiple different hogels and intercepted by the eye. The number of such wavefront segments N_h is defined for a point at depth z_p (w.r.t. the hologram plane) as

$$N_h = \left|\frac{z_p}{z_e - z_p}\right|\frac{d}{h_x}. \tag{3.17}$$

If the pupil size d and viewing distance z_e are fixed, N_h increases as the magnitude of z_p increases, i.e. the point is moved further away from the hologram plane. On the other hand, increasing hogel size reduces N_h contrary to N_θ. Please note that N_h corresponds to the parameter defined as ray density (number of rays incident in the eye pupil) in the context of ray-based LF displays [78].

3.2.2 Simulation results

To round off the theoretical analysis, we employed the following image formation model to experimentally quantify the effects discussed here:

$$I(u, v) = |\mathcal{P}_l\{Q_f(s,t)\mathcal{P}_{z_e}\{H_{\mathrm{HS}}(x,y)\}|^2, \tag{3.18}$$

where Q_f is the lens transmittance function including defocus for focal length f (see Eq. 2.17), l is the sensor-retina distance of the simulated eye and z_e is its distance from the hologram plane. Since the Fresnel kernel is separable, the analysis can be done for a 2D cross section of the full 3D space by considering the input hologram, the lens and the image results as 1D signals. The holograms are generated from a single point source placed at various depths z_p, both behind and in front of the hologram, and a stack of images with varying focal lengths of the eye are generated using the model in Eq. 3.18. The resulting focal stack of images corresponds to PSFs of the full optical system, including the display and the simplified eye model, which quantify the perceived sharpness of the point as a function of accommodation distance. The likeliest depth at which the eye is expected to accommodate at is determined through the maximum value of the modulation transfer function (MTF) focal stack [78]. The depth corresponding to the maximum MTF value is compared against the correct depth of the point source to determine the accommodation error, which essentially specifies the accuracy of the provided accommodation cue. The overall process is visualized in Fig. 3.5.

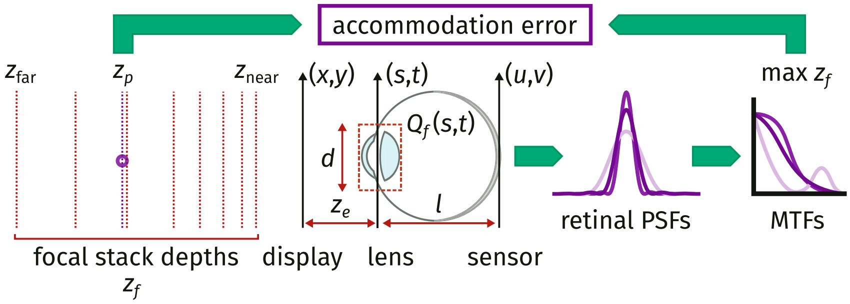


Figure 3.5 The simulation framework for quantifying accommodation error. A point is placed at z_p, the equivalent hologram is calculated and simulated by the illustrated viewer. The retinal PSFs are converted to MTFs, and the depth where the MTF is maximized is found. This is compared against the correct point depth to estimate the accommodation error.

During the simulations, the point is placed at 10 mm intervals within the depth range of -280 mm to 100 mm, where z is defined to be positive in front of the hologram ($z_p > 0$ means the point is between the viewer and the hologram). When illustrating the results, we define the point distance with respect to the hologram plane Δd in diopters as

$$\Delta d = \frac{1}{z_e} - \frac{1}{z_e - z_p}. \tag{3.19}$$

This results in a depth range of -1.6 to 1.6 diopters, where 0 diopters corresponds to the depth of the hologram plane. The simulated eye is set to have diameter $D = 5$ mm and a view distance $z_e = 300$ mm. The focal length of the lens is altered such that the eye is set to focus at different distances within the observed depth range at 2.5 mm intervals, to obtain the focal stack of PSFs, from which the MTFs are evaluated at spatial frequencies of 5, 10, 15 and 20 cycles per degree (cpd) as a function of

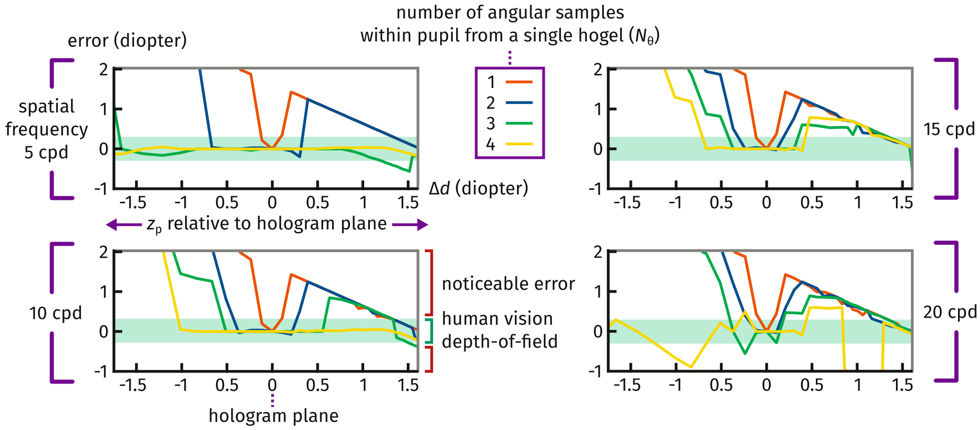


Figure 3.6 The error in accommodation for four different spatial frequencies and four different hogel sizes. The DoF of human vision (±0.3 D) highlighted in light green.

z_p and z_f. The frequencies are chosen according to the limitations of the simulated display and relevance to the HVS contrast sensitivity function [79]. For each z_p an estimate of the likely HVS focus distance $\hat{z}_f$ is obtained as the value of z_f maximizing the MTF. The simulation is replicated for four different hogel sizes, corresponding to four different N_θ values from 1 to 4.

The simulation results, presented in Fig. 3.6, exhibit anticipated behaviour: increasing the hogel size maintains lower accommodation error, and thus preserves correct accommodation cues, further away from the hologram plane. The smallest hogel size, i.e. when $N_\theta = 1$, exceeds the acceptable accommodation shift already at very small Δd values, meaning that the provided accommodation cues are incorrect immediately as scene content is moved away from the hologram plane. Nonetheless, for low spatial frequencies (5 and 10 cpd), the results suggest that even two views within the pupil extent can provide correct accommodation cues, even if only at a limited range of approximately ±0.5 diopters around the hologram. When the hogel size is increased such that $N_\theta = 3$ and $N_\theta = 4$, the accommodation error remains within the acceptable range for further distances away from the hologram: up to ±1.5 diopters at low frequencies, and approximately 2–3 times further when compared to $N_\theta = 2$ at high frequencies. However, this comes at a cost of reduced perceived spatial resolution, effectively increasing the pixel size by 2.5 and 3.3 times, respectively, in comparison to what the (diffraction-limited) HVS could resolve. A key finding of the simulation results was that at this range of hogel sizes, the angular resolution is the dominant factor in determining the accommodation response. Thus, increasing N_h has no similar positive effect on the accommodation response as in the case of ray-based LF displays [78], where the angular resolution of lenses (equivalent to hogels in HSs) is higher than HSs.

3.3 Novel accurate hologram synthesis from light rays

As shown in Publication III, the standard HS suffers from a similar spatio-angular tradeoff as most LF displays, and is incapable of producing accurate accommodation cues for deep scenes without significant compromise on spatial resolution. OLAS alleviates this issue by overlapping the hogels, thus trading computational complexity to provide a visually better alternative. However, while in OLAS depth is utilized to align the plane wave segments, the method is still limited in its ability to produce correctly aligned wave segments from scene points at large depths [51]. The work presented here addresses this issue by utilizing spherical wavefront segments in conjunction with planar ones. Though this further exacerbates the issue of computational complexity, by forfeiting the option of using Fourier transforms to obtain each hogel, the method improves the 3D reconstruction accuracy for deep scenes, while maintaining correct modelling of specular light emission.

3.3.1 Segmented spherical wave fronts from ray and depth information

In OLAS and other HS-based approaches, each light ray in the LF corresponds to a single frequency wave, i.e. a plane wave, where the frequency is determined by the angle of the ray according to Eq. 2.1. If further geometric information is available, e.g. in the form of depth maps, the plane waves from a specific point source can be aligned to better approximate the true spherical wavefront and to avoid wave discontinuities along the edges of the segmented waves, as done e.g. in PASs [47]. The notable advantage of these types of CGH methods is that, since each ray corresponds to a single spatial frequency, a collection of such rays within a single hogel can be efficiently obtained by inverse Fourier transform of the rays.

If computation time is not critical, and the aim is to have as accurate wave model as possible from light rays and depth, we can further model each light ray as a windowed segment of a spherical wave. DSCP is an example of a CGH method utilizing such a wave model [49]. If we opt for this wave model, but also let the hogels overlap as in OLAS, the typical spatio-angular tradeoff of stereograms is avoided, while extending the original model in OLAS. Such a hologram is obtained via the following convolution:

$$H_{\text{spherical}}(x, y) = \iint\limits_{-\infty}^{\infty} w(x - x', y - y') \iint\limits_{-\infty}^{\infty} \sqrt{L(x', y', s, t)}$$
$$\exp[jk(R(x, y, x', y', s, t) - D(x', y', s, t))]\, \mathrm{d}s\, \mathrm{d}t\, \mathrm{d}x'\, \mathrm{d}y', \qquad (3.20)$$

where $R(x, y, x', y', s, t)$ is the Euclidean distance between the hologram coordinates at (x, y) and the origin of the ray (x', y', s, t), i.e. the radius of the spherical wave from the ray origin, and $D(x', y', s, t)$ is the depth map value of the ray (x', y', s, t). In the continuous domain, this corresponds to assigning a frequency bandwidth, rather than a single frequency, for each light ray. Alternatively, it can also be considered as a bandlimited Fresnel zone plate, where the bandwidth is based on the distance and angle between the hologram segment and the light ray origin. Regardless, the wave (segment) within the hogel by itself already converges towards the correct distance, thus providing more accurate depth information than a comparative plane wave setup.

3.3.2 Wave modelling of diffuse and specular surfaces

Diffuse light emission is modelled as a wave having equal amplitudes in all emitted directions. For example, in the case of a self-emitting point source, the amplitude is constant across the entire spherical wave and its equivalent Fresnel zone plate. However, specular light emission and other non-Lambertian reflections are not as straightforward. Minor specularities, such as a glossy surface, are correctly modelled as a spherical wave from its source location (like diffuse light) with amplitude varying across the wave. Model-based holograms, particularly ones from point clouds, cannot typically model specular light correctly as the input data modality does not have such information (i.e. varying light intensity per point) available. This results in incorrect color information during motion parallax, though defocus blur is still correctly modelled due to the optical and hologram-encoded physical distance of the light source matching each other.

However, more severe specular light emission, as depicted in Fig. 3.7, is particularly problematic. In such cases, the true optical distance and physical distance of the

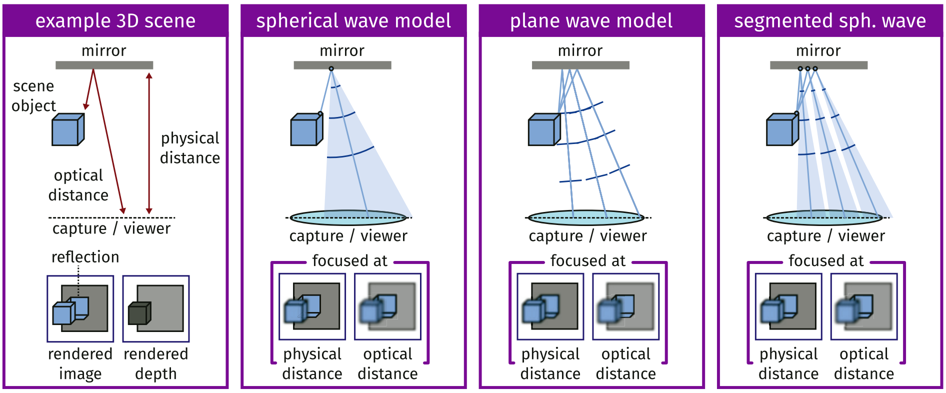


Figure 3.7 Wave models and their behaviour in the case of a mirror reflection. Leftmost figure shows the geometry of the example scene, the other three blocks describe different wave models and how they form images in terms of defocus blur when focused at different distances.

object are at a mismatch. In the Fig. 3.7 example, these distances correspond to the ones between the viewer and the reflected object via the mirror (optical distance) and the viewer and the mirror surface (physical distance). Similar effect is also present in the case of (partially) transparent surfaces. In synthetic data, the depth map information correspond to the latter distance. Thus, spherical wave models based on depth map input physically map the origin point relative to the physical depth rather than the optical depth. This causes incorrect defocus blur in our mirror example, as any reflected content is in-focus on the surface of the mirror rather than when focused at the full optical distance. Contrarily, plane waves encoding the angular information of a LF converge correctly at the optical distance, thus correctly modelling the defocus blur. Segmented spherical waves model the light behaviour as a compromise between the previous two models: each independent wave segment behaves as a spherical wave, while the directional information of the rays guide the individual wave segments as plane waves. This results in conflicting information in terms of the wave origin, thus prohibiting sharply focused image at either the physical or the optical depth.

3.3.3 Planar and spherical wave composite model

To counteract the issue raised in previous section, we propose in Publication V the following solution: for light rays identified to be originating from non-Lambertian objects or surfaces, the holographic data is constructed of plane waves. Everywhere else, we can utilize spherical waves. Thus, the full hologram is obtained as

$$H_{\text{hybrid}}(x,y) = \iint\limits_{-\infty}^{\infty} w(x-x',y-y') \iint\limits_{-\infty}^{\infty} \sqrt{L(x',y',s,t)}$$
$$\left(M(x',y',s,t)\exp\left[j2\pi\left(f_x(s,t)x' + f_y(s,t)y' + \frac{r(x',y',s,t)}{\lambda}\right)\right]\right.$$
$$\left. + (1-M(x',y',s,t))\exp[jk(R(x,y,x',y',s,t) - D(x',y',s,t))]\right) \mathrm{d}s\,\mathrm{d}t\,\mathrm{d}x'\,\mathrm{d}y', \tag{3.21}$$

where M is a binary mask dictating which light ray (or the equivalent point source) from the LF is identified as non-Lambertian:

$$M(x,y,s,t) = \begin{cases} 1 & \text{if } L_{\text{non-Lambertian}}(x,y,s,t) > \tau \\ 0 & \text{otherwise}\,. \end{cases} \tag{3.22}$$

Here we utilize the additional information from the renderer to obtain the relevant render passes, which contain intensities only from non-Lambertian reflections, denoted as $L_{\text{non-Lambertian}}(x,y,s,t)$. If the ray intensity exceeds a threshold (τ), it is considered to be non-Lambertian and treated as such for the purposes of hologram generation. The mask can also be estimated without the additional render passes, e.g. by estimating depth from disparity and comparing it against the depth map values (sufficiently different estimates can be assumed as non-Lambertian sources), or by analyzing the angular frequency content in the LF source and thresholding accordingly (non-Lambertian sources occupy higher frequencies).

The resulting hologram in Eq. 3.21 can be considered as a hybrid between OLAS and DSCP. However, unlike regular DSCP, the hogels are allowed to overlap, thus avoiding the loss of spatial resolution that would otherwise occur. Moreover, the composite model utilizes the benefits of both wave modalities to ensure that the depth is recreated as accurately as possible for Lambertian scene content (via spherical waves) while maintaining correct directional information for non-Lambertian content (via plane waves). Particularly, the spherical wave segments enable more accurate reconstruction of details for deep scenes (i.e. far from the hologram plane), resulting in better perceived image quality. Compared to OLAS, however, adding the more accurate mapping of scene points comes at the cost of computational complexity due to not being able to utilize Fourier transforms to generate the hologram. This increase in computation time is rather significant (approximately 50 times without any additional optimization). However, in our further work presented in Publication V, we aim to utilize the proposed CGH method to generate a data set of reference holograms and train a neural network to perform the hologram synthesis at a highly accelerated rate.

3.3.4 Comparison to other analytical holograms

We compare the proposed analytical CGH method against two important competitor approaches: OLAS [51] as the representative for state-of-the-art ray-based method, and the point-based method (PBM) with Fresnel kernel (similar to the reference holograms in [62]) as an example of a physically accurate spherical wave model (from a single view point). The PBM holograms are generated from a single RGB-D view (i.e. color image and corresponding depth map) without the back-projected occlusion ray culling along each subhologram shown in Fig. 2.6. All holograms are generated at wavelengths 638, 520 and 450 nm for the red, green and blue channels, respectively. The display plane (x, y) is placed at a distance of $z_e = 30$ mm from the eye, and it is assumed to be illuminated by a point light source at the same distance. The hologram pixel pitch is set to $\Delta x = 8$ µm at 512×512 resolution.

We examine the behaviour in the simulated focal stacks using a model based on the basic imaging system discussed in Sec. 2.4.2 with an extension to include chromatic aberrations as follows. In this case, the perceived image $I(u, v)$ focusing at z_f distance is determined as:

$$I(u, v; z_f) = \left|\mathcal{P}_l\{Q_f(s, t)\mathcal{P}_{z_e}\{H_{\mathrm{SLM}}(x, y)\}\}\right|^2, \tag{3.23}$$

where $Q_f(s, t)$ is the pupil function of a thin lens with focal length f and $H_{\mathrm{SLM}}(x, y)$ is the complex field on the display plane magnified by the intended display illumination. The propagation distance z_e corresponds to the distance between the display and the eye, and l to the distance between the simulated pupil and retina. Here we model the pupil function through the defocus term, aberrated chromatically, in the following manner. Assuming a pupil at coordinates (s_0, t_0), the transmittance function $Q_f(s, t)$

is defined as

$$Q_f(s,t) = A_d(s,t)\exp\left(-\frac{j\pi}{\lambda f_{\mathrm{LCA}}}\left[(s_0-s)^2+(t_0-t)^2\right]\right), \tag{3.24}$$

where $A_d(s,t)$ is the circular aperture function (diameter d) and f_{LCA} is the focal length altered by longitudinal chromatic aberration (LCA):

$$f_{\mathrm{LCA}} = \left(\frac{1}{z_{f,\mathrm{LCA}}}+\frac{1}{z_e}\right)^{-1}. \tag{3.25}$$

The intended focus distance z_f at reference wavelength λ_{ref} is shifted by LCA to obtain the aberrated focus distance $z_{f,\mathrm{LCA}}$ as [80]

$$z_{f,\mathrm{LCA}} = \left[\frac{1}{z_f}+633.26\left(\frac{1}{\lambda 10^9-214.10}-\frac{1}{\lambda_{\mathrm{ref}}10^9-214.10}\right)\right]^{-1}. \tag{3.26}$$

To test the relative accuracy of the different models, we consider the following aspects: the accuracy of diffuse objects at far and near depths (relative to the viewer), occlusions boundaries and defocus blur when observing specular reflections from a mirror. The simulated views of the these tests are visualized in Fig. 3.8. The first row of images corresponds to diffuse objects placed at a 0.1 D (yellow) and 5 D (white-dotted) distance from the viewer while focusing the simulated eye at z_f = 0.1 D at λ_{ref} = 520 nm (green channel wavelength). The texture on the far object contains narrow black lines distributed radially along the circle in the middle to provide varying direction horizontal and vertical lines as sharp details, thus demonstrating how well the different methods can recreate details at far distances relative to the hologram. The simulated views show that, while the Fresnel-PBM most accurately recreates the lines, our proposed approach is nearly as accurate and significantly improves over the degraded resolution of OLAS. Although not as profound, a similar effect can be seen on the second row of results on the outlines of the dotted texture, when the simulated eye is set to focus at the near object (z_f = 5 D). This strongly supports our assumption that the spherical wave segments better approximate the true waves as emitted by point sources at a specific distance, thus resulting in more accurately resolvable details.

The third and fourth row results contain occlusions and glossy materials, focusing the eye near the middle of the scene depth range (z_f = 3 D). In these results, our

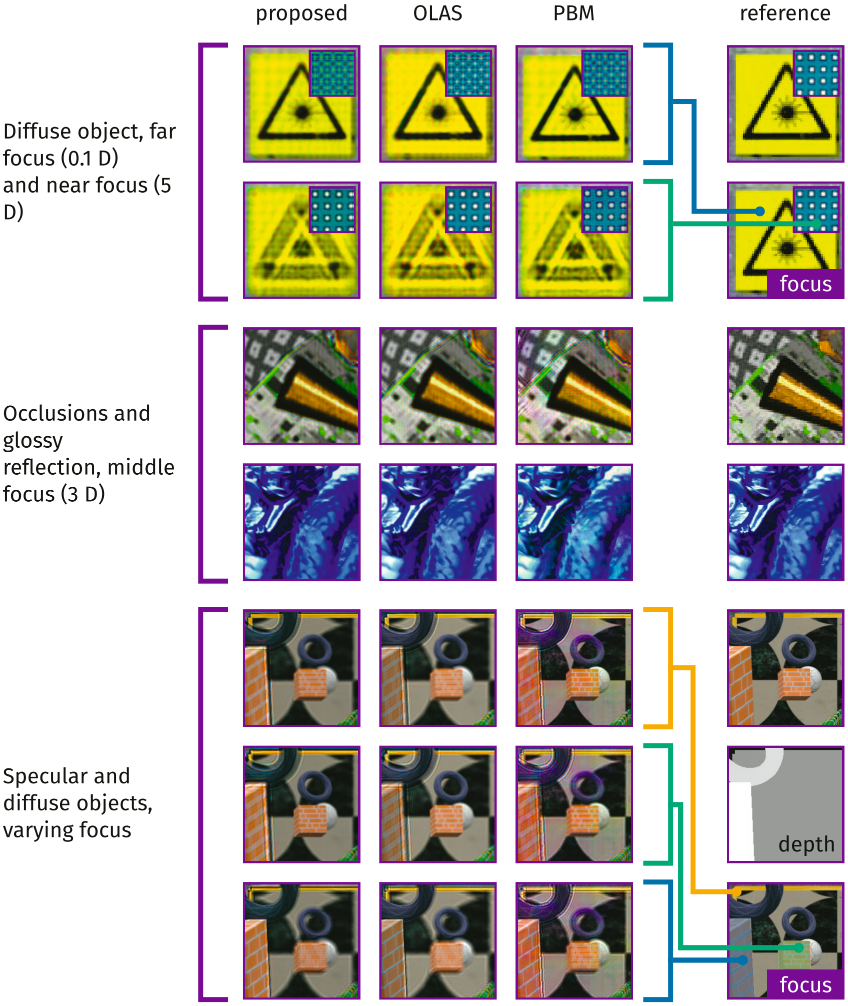


Figure 3.8 Comparison of different focal state images from the proposed analytical method, OLAS and point-based Fresnel. The first two row shows the accuracy of far and near focus of Lambertian objects (yellow object at far focus distance, white-dotted object at near focus distance). The second and third rows show behaviour around occlusion boundaries and glossy reflections. The three final rows demonstrate the defocus blur when focusing the simulated eye on the mirror (top row), the full optical distance to the reflected objects (middle row) and the foreground object (bottom row).

proposed approach and OLAS are nearly identical, while the Fresnel-PBM contains minor distortion around some of the occlusion boundaries. The final test results along the three bottom rows show defocus blur formation of mirror-reflected content. The Fresnel-PBM maps all point seen via the mirror on its surface, as both the mirror frame and all reflections are in-focus when focusing the simulated eye at z_f = 3 D (the physical distance to the mirror). Conversely, the proposed method and OLAS show only the mirror frame in-focus at this focal depth. When the focus is moved further to z_f = 1 D, we can see the reflections in-focus at their (approximate) optical distance. Thus, the proposed approach and OLAS correctly form defocus blur for specular scene sources. As a summary, the simulation results suggest that our proposed approach is capable of matching OLAS in modelling specular light emission, while improving the resolvable detail resolution at large depth ranges to nearly the equivalent of a Fresnel hologram.

4 MACHINE LEARNING FOR COHERENTLY ILLUMINATED NEAR-EYE DISPLAYS

Thus far in the thesis, the focus has been on analytical approaches for obtaining the optimal holographic data. However, the recent resurgence in machine learning has also affected the design space of hologram synthesis, enabling real-time generation of holograms in certain cases, as well as the end-to-end learning-based design of novel near-eye display configurations. To this end, two particular topics are covered in this chapter: a general end-to-end learning framework for coherently illuminated NEDs to tackle VAC through accommodation-invariance (Publication IV), and the acceleration of hologram synthesis via machine learning (Publication V).

4.1 End-to-end learning for accommodation-invariant coherent display

Conventional NEDs recreate 3D content by providing binocular vergence and disparity cues by way of displaying stereoscopic views for the viewer. However, typically they lack the correct monocular accommodation and defocus blur cues, which leads to VAC, and further to visual discomfort for the user [13]. This can be circumvented by either providing the aforementioned depth cues accurately, e.g. by displaying the scene LF, or by decoupling the display from defocus blur entirely and relying on vergence to drive accommodation in a display type known as accommodation-invariant display (Fig. 4.1).

The working principle of accommodation-invariant displays is based on findings from vision science, which have shown that accommodation and vergence are linked to each other [11], [12]. This has also been shown to be effective in practice via subjective experiments of one such display, where none of the 50 participants reported eyestrain or nausea [81]. Furthermore, all participants in the study reported the perceived 3D effect to be realistic and could estimate distances to virtual objects (shown at 1 meter) correctly. In a later subjective study, it was also shown that accommodation can be induced solely by vergence [82], further solidifying the use of accommodation-invariance as a measure to combat VAC.

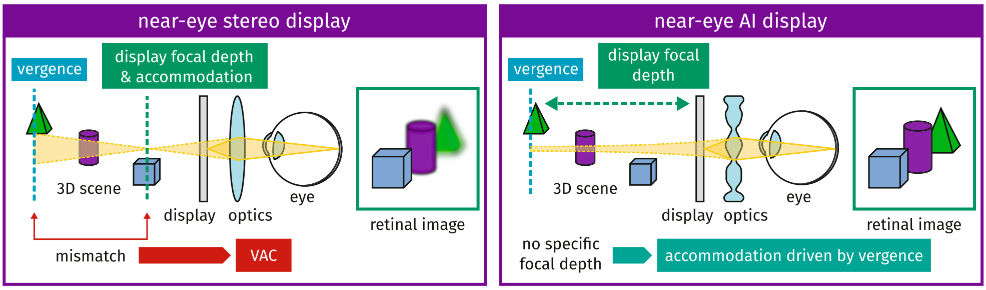


Figure 4.1 Relation between vergence and accommodation in stereo and accommodation-invariant NEDs. In the conventional stereo display (left) the relevant visual cues are mismatched when content is shown away from the display focal depth, whereas the accommodation-invariant display (right) forgoes retinal blur to let vergence drive the accommodation response of the viewer.

The contribution presented here approaches the accommodation-invariant, or extended DoF, display problem through the following problem formulation: given a display setup, where a static diffractive optical element (DOE) is placed in front of the coherently illuminated, amplitude-modulating display panel, what combination of image pre-processing function and DOE minimizes the difference between the perceived retinal image and a perfectly in-focus (sharp) image for a large depth range. This would result in removing defocus blur, thus letting vergence and disparity to drive accommodation. To solve the design problem, we first have to specify how the perceived image is formed through a computational display model, which we discuss in the following section. It should be noted here that, unlike holographic displays, the display configuration does not contain dynamic phase modulation, but only amplitude modulation combined with static phase modulation from the DOE. This reduces the complexity of the display design problem significantly. Moreover, it provides a computationally inexpensive and optically straightforward (and cost effective) method for avoiding VAC, especially when compared to full holographic imaging and accommodation-enabling display design.

4.1.1 Computational display model

The imaging model contains three notable points of interest: the display panel (where the amplitude of the incoming light is modulated), the additional display optics (to-be-optimized DOE) and an aberration-free eye with a planar retina accommodating at a distance of z. Due to the simple eye model, the image can be found on a conjugate plane, or the reference plane (x, y). Assuming monochromatic illumination at wavelength λ, the coherent system PSF $h_{\lambda,z}(x, y)$ on the reference plane is

$$h_{\lambda,z}(x, y) \propto \mathcal{F}\left\{Q(s, t)|_{\left(\frac{x}{\lambda z}, \frac{y}{\lambda z}\right)}\right\}, \tag{4.1}$$

where $Q_{\lambda,z}(s, t)$ is the generalized pupil function and $\mathcal{F}\{\cdot\}$ is the Fourier transform operator. The pupil function is obtained as a combination of the circular aperture function $A(s, t)$, the phase delay of the DOE $\Phi_\lambda(s, t)$ and the refractive lens defocus as

$$Q_{\lambda,z}(s, t) = A(s, t) \exp\left(j\Phi_\lambda(s, t)\right) \exp\left(j\Psi_{\lambda,z}\frac{s^2 + t^2}{r^2}\right), \tag{4.2}$$

where

$$\Psi_{\lambda,z} = \frac{\pi}{\lambda}\left(\frac{1}{z_d} - \frac{1}{z} - \frac{1}{f_\lambda}\right) r_{\text{lens}}^2 \tag{4.3}$$

is the defocus coefficient, r_{lens} is the radius of the circular lens and f_λ is the effective focal length of the refractive lens. The DOE phase delay is related to its height profile $d_{\text{DOE}}(s, t)$ through the wavelength-dependent refractive index n_λ of the DOE material as

$$\Phi_\lambda(s, t) = k\,(n_\lambda - 1)\, d_{\text{DOE}}(s, t). \tag{4.4}$$

The reference plane amplitude in the coherent imaging regime is a convolution between the coherent PSF and the source amplitude from the display panel $A_\lambda^d(x, y) = \sqrt{I_\lambda^d(x, y)}$. Thus, the final reference plane image is obtained as

$$I_{\lambda,z}^p(x, y) = \left|A_\lambda^d(x, y) * h_{\lambda,z}(x, y)\right|^2. \tag{4.5}$$

In practice, the phase of the display illumination is slightly perturbed (in the order of wavelength) in a random fashion, thus modifying Eq. 4.5 by multiplying the display amplitude $A_\lambda^d(x, y)$ by a random phase term $\exp(jk\phi_\lambda^d(x, y))$. As discussed in Sec. 2.3.3 and visualized in Fig. 4.2, this results in speckle noise and lowers the

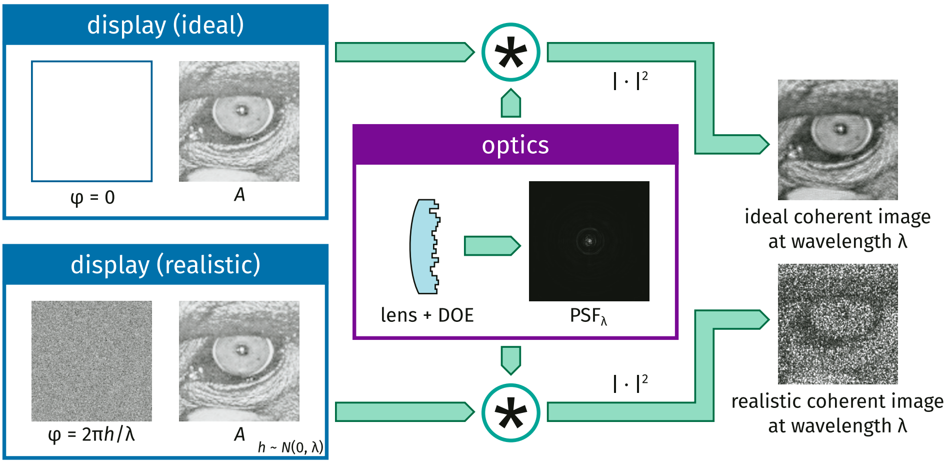


Figure 4.2 The coherent image formation model for an ideal display (equal phase everywhere) and a realistic display (uniformly random phase). When the input complex-valued display field is convolved with the system PSF, the neighbouring points interfere causing high-contrast speckle noise in the realistic display case (lower path).

perceived visual quality. However, realistic light sources are rarely perfectly coherent in the spectral domain, but rather have a finite support which can be modelled as a unimodal Gaussian function [83]. This reduces the severity of the speckle noise issue, even if not completely eliminating it. To account for this, we modify the imaging model when testing the system by discretely sampling the Gaussian spectral density. For each sampled wavelength, the image is formed as in Eq. 4.5 and a weighted sum of the images is taken to obtain the final image:

$$\hat{I}^{p}_{\lambda,z}(x,y) = \sum_{i} w_{\lambda_i} \left| A^{d}_{\lambda}(x,y)\exp\left(jk\phi^{d}_{\lambda_i}(x,y)\right) * h_{\lambda_i,z}(x,y)\right|^2, \tag{4.6}$$

where λ_i are the discretely sampled wavelengths in the spectral density and w_{λ_i} are the corresponding weights, which are dictated by the (Gaussian) spectral density function at the sampled wavelengths. During training, however, the imaging model is assumed to be ideally coherent for simplicity.

4.1.2 Learning approach

For learning the accommodation-invariant response desired for the display, the optical construction is combined with a pre-processing step. That is, the display ampli-

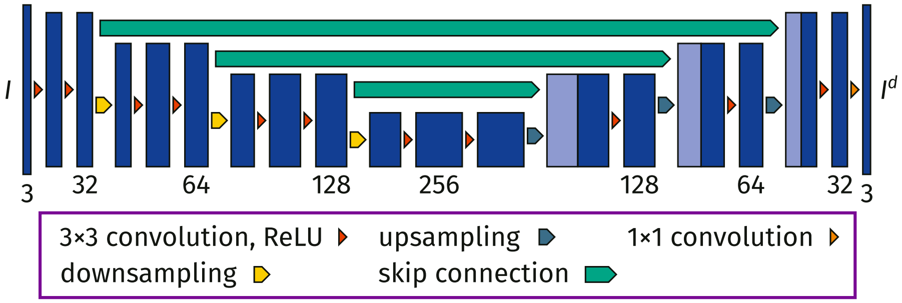


Figure 4.3 Structure of the pre-processing network. Channel sizes indicated by the numbers below each layer (blue rectangles), connections and operations highlighted in different colored arrows.

tude A^d is the output of a function compensating for the optics of the display, which is jointly optimized with the display optics (i.e. the DOE). Utilizing a CNN for this purpose suits the problem well, as the weights of the network can be trained to learn the desired response. Additionally, there is a significant number of different established network structures, which have been shown to suit for various image processing tasks. Here, the U-Net architecture [84] is chosen for this purpose, due to its demonstrated success in various computer vision and image restoration problems including deblurring. It is a multilevel network comprised of a contracting path and an expansive path, which can be considered as encoding and decoding paths for the purposes of this work, along with skip connections at each level. The structure of the network, as depicted in Fig. 4.3, contains the following processing steps. The data in the encoding path is first processed by convolution using 3×3-sized kernels, then by rectified linear unit (ReLU) activation functions, finally followed by a downsampling operation of a factor of two. Inversely, the decoding path begins with upsampling and its output is then concatenated with the corresponding skip connection output of encoding. The resulting data is again convolved and processed by ReLU layers. Finally, the decoding path output is mapped to the original image size using a 1×1 convolution layer.

The final part of the learning approach is the loss function. It drives the learning process by minimizing the difference, or error, between the desired and current output of the system over time. Depending on the desired output and the task at hand, a variety of different image quality metrics can be utilized. Here the metrics are the L_1 distance and the structural similarity index measure (SSIM) [85]. The aggregated loss function attempts to minimize L_1 and maximize the SSIM as a compromise between

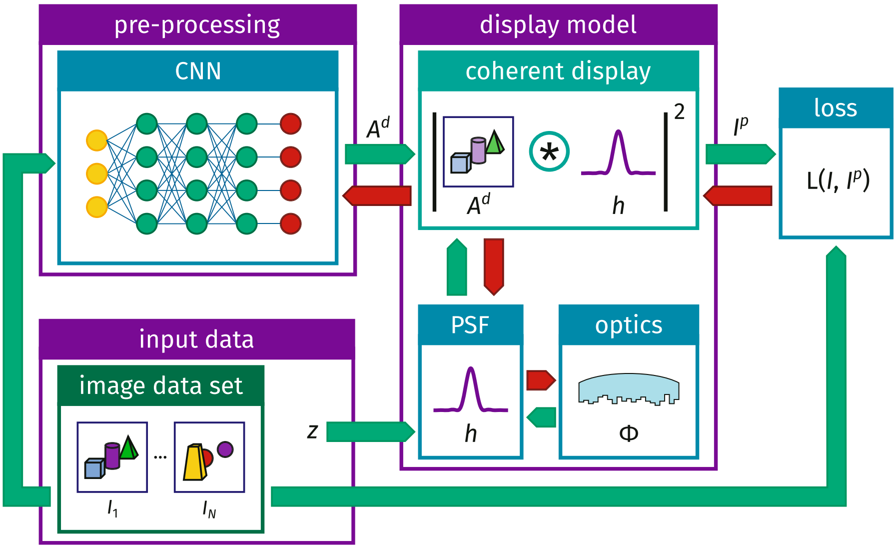


Figure 4.4 The overall structure of the learning based accommodation-invariant NED approach. The input image is fed to the pre-processing network (CNN) which outputs the amplitude on the display. The amplitude then passes through the display model to output the perceived image and compared against the reference sharp input image. Forward passes are indicated by green arrows, backward passes (derivatives) by red arrows.

the texture detail quality and the perceived change in structural information. The overall loss is calculated as

$$\mathcal{L}(I, I^p) = \mathcal{L}_{L_1}(I, I^p) + \mathcal{L}_{\mathrm{SSIM}}(I, I^p) + \alpha\mathcal{R}(I, I^p) + \gamma\mathcal{R}^d(I^d), \tag{4.7}$$

where $\mathcal{L}_{L_1}(I, I^p)$ is the L_1 distance, $\mathcal{L}_{\mathrm{SSIM}}(I, I^p)$ is the SSIM-based loss, $\alpha\mathcal{R}(I, I^p)$ is the network output regularization (equalizes the intensity levels between the network output and the input) and $\gamma\mathcal{R}^d(I^d)$ is the display image regularization (controls the dynamic range of the display image). To maximize SSIM, we define its loss function as

$$\mathcal{L}_{\mathrm{SSIM}}(I, I^p) = 1 - \mathrm{SSIM}(I, I^p). \tag{4.8}$$

During the design stage of the learning approach (Fig. 4.4), i.e. while training, the DOE and the pre-processing CNN are jointly optimized utilizing a large set of (2D) training images. To obtain the desired extended display DoF, the accommodation distance z of the imaging model is randomly changed during training within the target depth range, the perceived image is formed and compared against the sharp input image using the loss metric presented above. Naturally, the extent of the target depth range determines the difficulty of the DoF extension problem, larger extension making it more difficult.

4.1.3 Results

The presented machine learning model was trained in Publication IV, for a target depth range of 0 to 3 D, display resolution corresponding to 20 cpd and 6 µm feature size (spatial sampling) on the DOE. After training, the learnt DOE phase and CNN weights were utilized to evaluate the system in three distinct cases: ideal coherent, partially coherent and partially coherent with speckle suppression. In the first case, the light source is assumed to be perfectly coherent and that no variation is contained within the display phase. The two latter cases were utilized to demonstrate a more realistic scenario, where the display phase is random and the light source has finite support in the spectral domain, both with and without additional speckle reduction. The speckle reduction was assumed to be equivalent to a moving diffuser setup [86], and simulated by averaging the intensities of 25 independent noise realization frames. All simulated views are compared against the reference sharp view using peak signal-to-noise ratio (PSNR) and SSIM.

The zoomed-in reconstructions in Fig. 4.5 highlight both the benefits and the downsides of coherent imaging for an accommodation-invariant NED; the ideal coherent results far exceed any of the incoherent alternatives, however, the partially coherent (realistic) results are degraded by notable speckle noise. It could be argued that the proposed approach produces visually better (sharper) results in some or most of the cases, regardless of the metrics. When the speckle noise is reduced via time-multiplexing, the proposed display can maintain satisfactory reconstruction quality also beyond the intended depth range (0 to 3 D) assumed during training. Furthermore, the proposed design shares the physical construction benefits of the incoherent variant: the additional display optics are thin and static, and DOEs can be effectively fabricated via grayscale [87] or imprint lithography [88] with minimal (< 50 nm) average error in height profile [89].

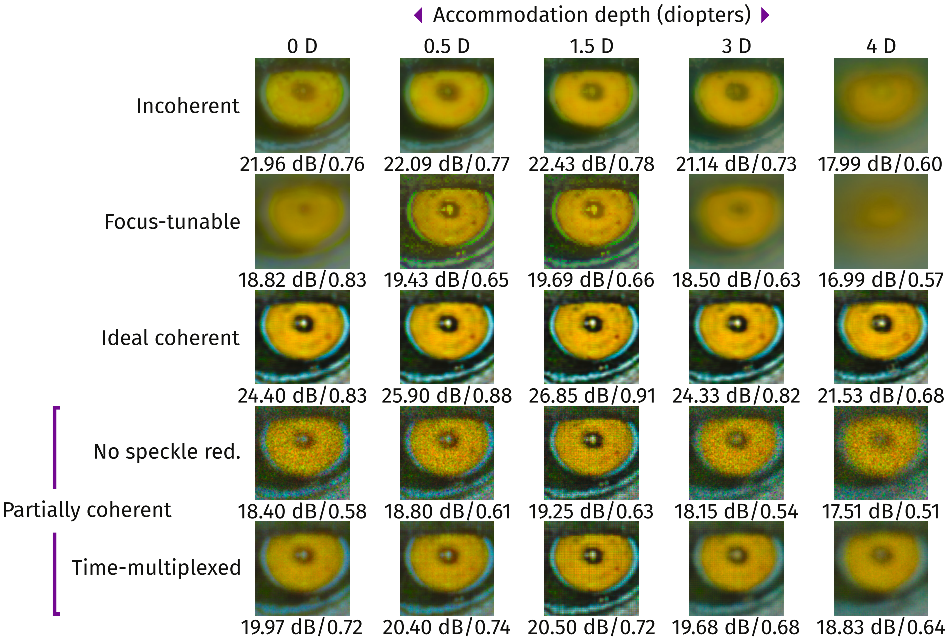


Figure 4.5 Comparing the results of the proposed approach against the state of the art incoherent counterparts. Different accommodation depths shown in columns, and the PSNR and SSIM values given below each image (higher better).

4.2 Accelerating hologram synthesis

Extending the goal from accommodation-invariant to accommodation-enabling displays, holographic imaging provides a desirable alternative. One of the key issues, however, is the computational complexity of synthesizing suitable holograms. To that end, we aimed to obtain real-time hologram synthesis in Publication V, without sacrificing the 3D reconstruction quality. Particularly, we strived to include the benefits of ray and depth -based holography in terms of visual quality while eradicating their main weakness of slow and heavy computation.

We present here our work for accelerating the CGH method introduced in Section 3.3.3 using machine learning. This required developing tools to generate large data sets of random synthetic 3D scenes, to render the appropriate LFs and depth maps, and subsequently, to generate reference holograms using our analytical CGH model. The overall framework follows along the basic guidelines discussed in Section 2.3.2, and shares similarities with the TensorHolo framework [62]. We focus here on expanding upon the novel improvements to random scene generation for generating large synthetic data sets, and examine the full learning framework along with the ablation study to find the best neural network for learning accurate hologram synthesis.

4.2.1 Random scene generation

Our composite CGH model utilizes RGB-D LFs as its input, and thus, for learning the transformation, a large number of synthetic 3D scenes with controllable parameters is needed. Rendering this information using computer graphics is necessitated by the need for accurate depth maps, as well as for easily obtaining the mask information in Eq. 3.22. In [62], a similar approach is taken to obtain RGB-D images; however, we propose several key changes to improve the random scene generation and to better fit the needs of our work.

In terms of the geometrical properties of the 3D scenes, we randomize the following properties. Objects are placed randomly within the field-of-view of the center view and scaled in size such that they occupy approximately 10 to 35 % of the sensor area (and subsequently, the rendered image pixels). The horizontal and vertical position, along with the orientation and the imaged size of the object, are all picked from a uniform random distribution. The geometry of these objects is randomized from the ABC data set [90]. Altogether, the randomization provides a wide variety of object surfaces and occlusion boundaries. As opposed to [62], we propose to place the objects along the z-axis such that they are uniformly distributed in dioptric range rather than metric. Combined with our perceptual loss and its imaging model, this choice ensures that the resulting scenes and the corresponding holograms contain objects equally relative to the amount of retinal defocus blur they create.

For color and material properties, it is crucial to vary the intensities along the three main wavelengths (red, green and blue), as well as along the angular dimension of the LFs (i.e. allocating the angular frequency bandwidth uniformly). To guarantee these properties in the data set and that the objects within the scenes appear as realistic as possible, we opted to use material-texture pairings from ambientCG[1], thus making sure that even when randomizing the materials, they are always paired with a sensible texture. Subsequently, we avoid for example assigning a grass texture with a shiny, metallic material. Furthermore, we included an additional set of highly

---

[1] ambientcg.com, assets licensed under the Creative Commons CC0 1.0 Universal License

non-Lambertian materials (e.g. glass, mirrors) in the pool of potential choices to further enrich the data set and to include light behaviour which requires the use of plane waves in the composite CGH model.

4.2.2 Deep learning framework

Utilizing the presented random scene generation approach, we can efficiently generate a large number of random 3D scenes (and subsequently rendered images) with desirable properties for learning purposes. This is particularly useful for learning-based holography to train a neural network to approximate a specific holographic encoding process. Most analytical hologram synthesis methods are several orders of magnitudes slower than a typical neural network inference rate, thus training a network to approximate one is an attractive option for accelerating hologram synthesis.

As a part of the research work conducted for Publication V, we aimed to accelerate the plane-spherical hybrid analytical method from Sec. 3.3.4 using the learning framework shown in Fig. 4.6. The training data set consists of RGB-D LFs and reference holograms generated using our analytical approach. For finding the most suitable network structure to approximate the holographic encoding, we study three options in the publication: a fully convolutional neural network (FCNN), an FCNN with spectral connections (Fourier-CNN) [91] and U-Net [84]. The first option is shown to successfully learn hologram synthesis in [62], and has a physically explainable link between the number of network layers and the size of hologram patches from each scene point. The second option expands the basic CNN to take advantage of the spectral information within the network by including connected Fourier-transformed layers. The final option, U-Net, is a proven choice for various computer vision and image processing tasks, including the end-to-end accommodation-invariant display learning in Publication IV. The results of the comparative study are presented in the next section.

The final step of the learning framework is how the predicted hologram output and the target hologram are compared against each other to drive the training of the network weights, i.e. the loss. Our approach is similar to [62], combining a data fidelity loss with a perceptual loss: the former determines the similarity between the amplitude and phase of the complex-valued holograms, while the latter compares images obtained from varying propagated scene depths (i.e. a focal stack) [62]. The data fidelity loss is the main driving force in the loss during training, with regular-

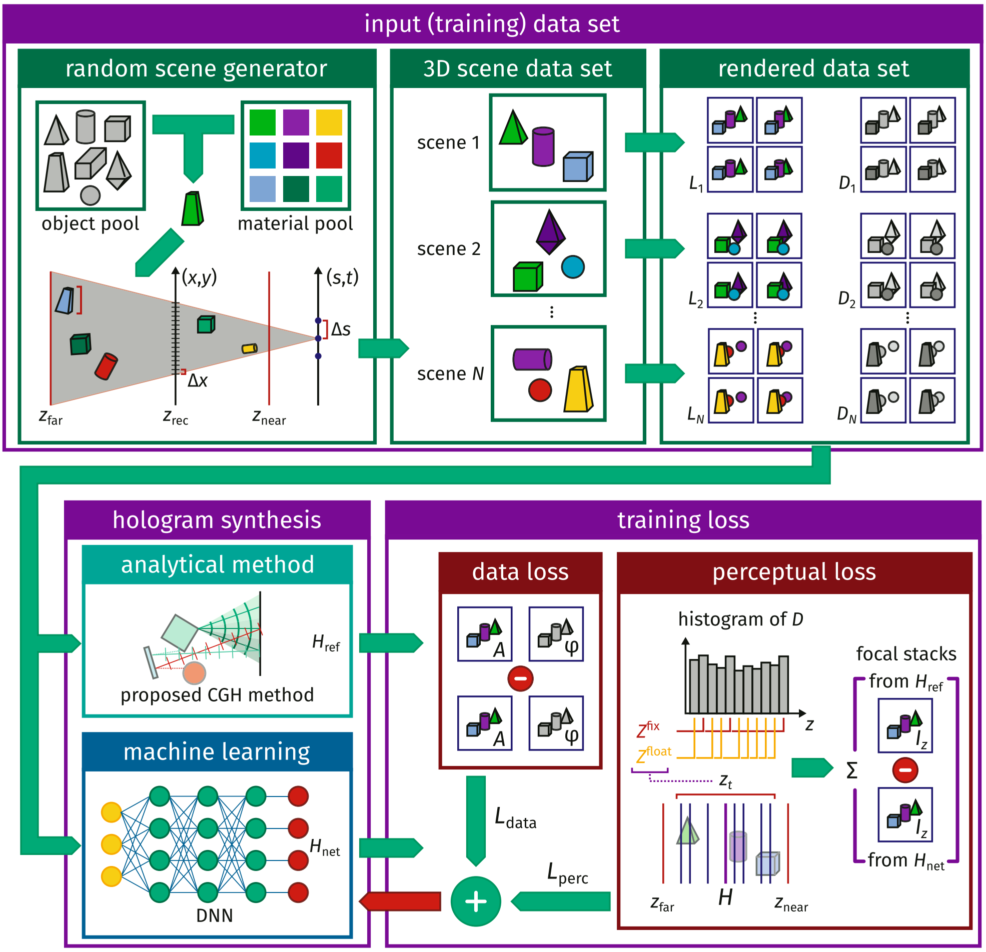


Figure 4.6 The full learning framework for approximating the analytical hologram synthesis method along the top path by a neural network (bottom path).

ization provided by the perceptual loss. The data fidelity loss is determined as a phase-corrected L_2 distance between the reference hologram $H = A\exp(j\phi)$ and the network predicted hologram $\hat{H} = \hat{A}\exp(j\hat{\phi})$, i.e.

$$\mathcal{L}_{\text{data}} = \left\| \hat{A} - A\exp(j[\partial(\hat{\phi},\phi) - \bar{\partial}(\hat{\phi},\phi)]) \right\|_2, \tag{4.9}$$

where $\partial(\hat{\phi},\phi)$ is the signed shortest angular distance between $\hat{\phi}$ and ϕ in polar coordinates

$$\partial(\hat{\phi},\phi) = \operatorname{arctan2}[\sin(\hat{\phi}-\phi), \cos(\hat{\phi}-\phi)], \tag{4.10}$$

and $\bar{\delta}$ denotes the mean of δ. Additionally, the accuracy of the 3D reconstruction is evaluated by determining the defocus blur behaviour in a focal stack. The simplest imaging model for such a focal stack, is to propagate the hologram in the 3D space to the desired focal planes (at depth z_f w.r.t. display plane) and form the intensity images on those planes:

$$I_{z_f}(x,y) = \left|\mathcal{P}_{z_f}\{H(x,y)\}\right|^2, \tag{4.11}$$

where $\mathcal{P}_z$ is a scalar field propagation operation for a distance of z and $H(x,y)$ is the hologram. During each training iteration t, we determine a set of different focal depths $z_t \in \{Z_t^{\text{fix}} \cup Z_t^{\text{float}}\}$ as follows. First, a histogram of the depth map values is determined for T bins. In focal depth set Z_t^{fix}, the k_{fix} highest histogram bins are chosen and uniformly sampled. From the remaining bins, k_{float} are randomly chosen and uniformly sampled to form the set Z_t^{float}. The perceptual loss is then defined as a summation of attention-mask attenuated L_1 distances between the focal stack images of the network predicted hologram $\hat{I}$ and the reference hologram I, and their gradients [62]:

$$\mathcal{L}_{\text{perc}}(t) = \sum_{z_t} \left\| M_{\text{att}}(z_t)(\hat{I}_{z_t} - I_{z_t} + \nabla\hat{I}_{z_t} - \nabla I_{z_t}) \right\|_1, \tag{4.12}$$

where $M_{\text{att}}(z_t)$ is the attention mask and ∇ is the total image gradient operator. The attention mask emphasizes pixels near the in-focus depth z_t within the focal stack images as

$$M_{\text{att}}(z_t) = \exp(\beta[2z_{\text{rec}} - (z_t - D)]), \tag{4.13}$$

where β is a user-selected factor to control the masking effect and D contains the depth map values corresponding to the image pixels.

4.2.3 Ablation study

Publication V concludes with a study to examine the differences between three network models after training and potential redundancies in the input data (RGB-D LF). Thus, we train the to-be-compared options using our custom data set in Publication V of 4000 RGB-D LFs from random 3D scenes with objects placed within a depth range of 0.1 D to 6.67 D. The accuracy of the network predicted holograms

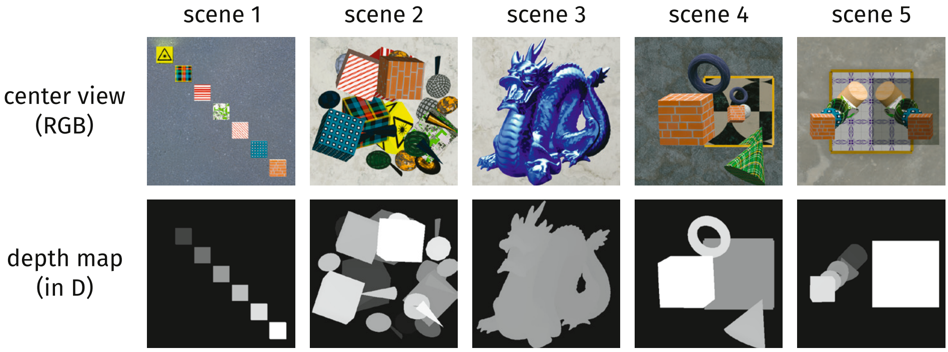


Figure 4.7 Center RGB view and depth map of the generalization test scenes. The depth values are converted to diopters for visualization purposes (near depths white, far depths black).

after training are evaluated in two ways. The first set of metrics is the average PSNR and SSIM values between the reference and predicted hologram amplitudes over the test set, as well as the average focal stack PSNR and SSIM from the reference and predicted holograms with the simple imaging model utilized in the perceptual loss (Eq. 4.11) and the full optical model (Eq. 3.23) utilized to compare the different analytical methods in Sec. 3.3.4. The second evaluation method tests the generalization capabilities of the trained networks by examining scenes not in the training data set and with generally different scene composition (shown in Fig. 4.7). For these, we examine the focal stack PSNR and SSIM averages over focal depths using the full optical model. In the test focal stack simulations, we utilize seven focal depths at each integer diopter depth between 0.1 and 6 D (including the limits). In the network comparison, all models are configured to match each other in inference time. For our reference model complexity we utilize the FCNN with the same setup as in [62], as it was demonstrated to achieve real-time suitable inference rates (up to 60 Hz). After determining the preferred network model, we examine the properties of the input data, specifically testing for the potential redundancies between different modalities. The full input data, i.e. all views of the LF with depth maps, is our reference, against which we compare two models trained with either the LF only (without depth) or the center view of the LF only with its corresponding depth. The first comparison indicates whether the depth map values are redundant and if all required information can be inferred from the angular LF properties. The second comparison highlights the necessity of the monocular LF angular information compared to the center view. As in the model comparison, all three cases are trained

Table 4.1 Average test set metrics from 100 scenes. Comparing different network models with RGB-D LF as input for each one. Values presented as PSNR (dB) / SSIM in each cell for hologram amplitude, simple focal stack and fully modelled focal stack (highest result in each column in bold).

model	amplitude	simple focal stack	focal stack with optics
Fourier-CNN	28.09 / 0.91	26.22 / 0.89	25.72 / 0.91
FCNN	28.89 / 0.93	27.14 / 0.91	26.69 / 0.93
U-Net	**29.90 / 0.95**	**28.23 / 0.94**	**27.71 / 0.95**

Table 4.2 Average metrics from simulated focal stacks (full optics model) of the five custom-designed test scenes. Values presented as PSNR (dB) / SSIM in each cell. Each model uses RGB-D LF as its input.

model	scene 1	scene 2	scene 3	scene 4	scene 5
Fourier-CNN	26.69 / 0.92	25.36 / 0.90	29.39 / 0.96	26.40 / 0.90	23.98 / 0.90
FCNN	**29.55** / 0.94	26.74 / 0.91	31.12 / **0.97**	**28.48** / 0.91	**25.45** / 0.92
U-Net	29.34 / **0.95**	**27.29** / **0.94**	**31.16** / **0.97**	26.06 / **0.92**	24.56 / **0.93**

identically, the only difference being the data modality fed to the network.

The study results for the learnt hologram encoding amplitudes and their equivalent focal stacks are presented in Table 4.1. Over the test data set, the U-Net achieves the highest PSNR and SSIM scores in all three evaluation cases, followed by the FCNN (approximately 2 dB lower PSNRs and 2–3 % lower SSIMs). The Fourier-CNN is third with similar differences in results to the FCNN. The generalization test results in Table 4.2 (with visual example in Fig. 4.8) are closer between the three networks, with the highest scores being shared by the FCNN and U-Net depending on the scene. The FCNN has the highest PSNR averages in most cases, whereas the highest (or joint highest in scene 3) average SSIM is always achieved by the U-Net predicted hologram. Combining the results of both tests, we choose U-Net as the best candidate network to further determine whether there are redundancies in the input data.

The results for the learnt hologram encoding amplitudes and their equivalent focal stacks comparing the input data modalities are presented in Table 4.3. Interestingly, our results seem to indicate that both the LF intensity and depth are needed for the most accurate hologram approximation, and that the angular information present in the LFs is insufficient to infer the ray distances needed for aligning different wave segments. A better result without depth might be achieved by increasing the net-

Table 4.3 Average test set metrics from 100 scenes. Comparing different input data modalities using U-Net as the network model for each option. Values presented as PSNR (dB) / SSIM in each cell for hologram amplitude, simple focal stack and fully modelled focal stack (highest result in each column in bold).

input	amplitude	simple focal stack	focal stack with optics
Center (RGB-D)	25.89 / 0.83	23.94 / 0.79	23.38 / 0.84
LF (RGB)	27.12 / 0.89	25.35 / 0.87	24.71 / 0.90
LF (RGB-D)	**29.90 / 0.95**	**28.23 / 0.94**	**27.71 / 0.95**

Table 4.4 Average metrics from simulated focal stacks (full optics model) of the five custom-designed test scenes. Values presented as PSNR (dB) / SSIM in each cell. Each different input setup uses U-Net as the model.

input	scene 1	scene 2	scene 3	scene 4	scene 5
RGB-D center view	26.44 / 0.88	23.06 / 0.89	30.02 / 0.96	25.51 / 0.86	23.31 / 0.87
RGB LF	20.76 / 0.81	19.43 / 0.77	22.75 / 0.83	20.75 / 0.76	18.17 / 0.76
RGB-D LF	**29.34 / 0.95**	**27.29 / 0.94**	**31.16 / 0.97**	**26.06 / 0.92**	**24.56 / 0.93**

work model complexity, however, this would compromise the targeted fast inference rate. Moreover, it is evident by examining the central view results against the full LF results that the contribution of off-centered views is instrumental to both hologram reconstruction and focal stack accuracy, the latter outperforming the former in PSNR by 6–7 dB. Importantly, while the chosen evaluation setups do not consider off-center viewer positions, the defocus behaviour present in the focal stacks heavily indicate the benefits of angular information. Finally, the generalization results in Table 4.4 support the findings from the test set. Furthermore, visually examining the results of Fig. 4.8 demonstrates how the model utilizing only the central view suffers from the same incorrect defocus blur formation as the point-based Fresnel in our comparison between analytical CGH methods.

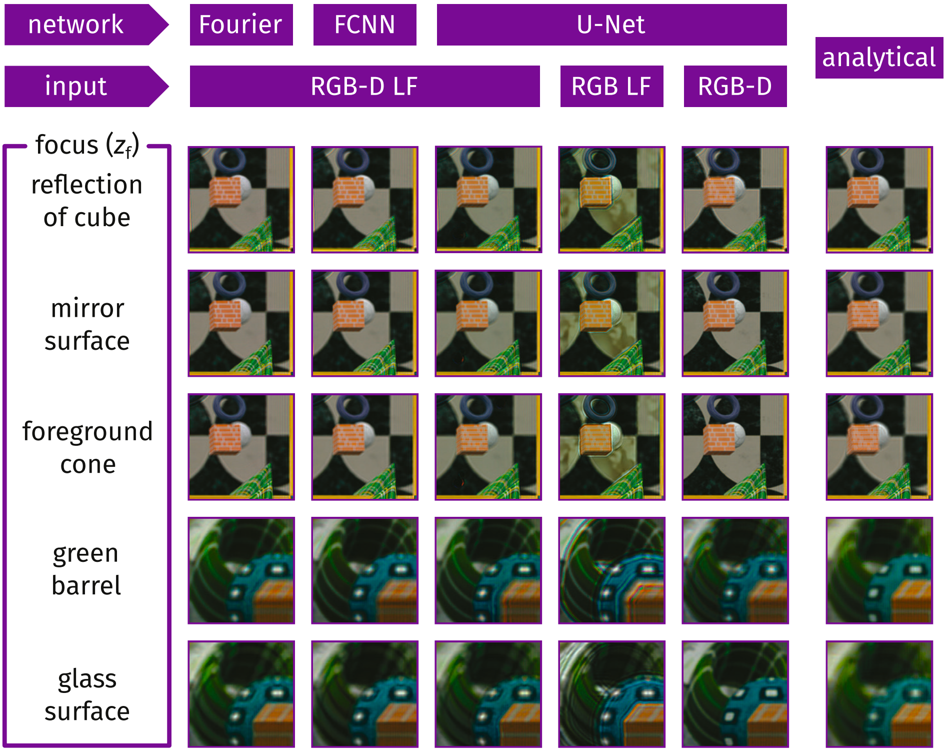


Figure 4.8 Comparison of different focal state images from the predicted holograms after training. Different network models and input modalities placed along columns and different focal states of the simulated eye on along rows. Rightmost column contains the same simulation results from the proposed analytical CGH method. The results on the first three rows are from scene 4, and the bottom two rows are from scene 5.

5 CONCLUSIONS

NEDs are a compelling solution for providing immersive viewing experiences, as they enable the user to freely explore 3D environments in an intuitive manner by simply moving their head. As display technologies and computing power have rapidly advanced in recent times, conventional stereoscopic NEDs have become readily available in the consumer market, especially for entertainment purposes, despite their limited approximation of 3D scene reconstruction via 2D stereo views. However, a notable limiting factor in their comfortable use is VAC, which has been attempted to resolve via accommodation-invariant and accommodation-enabling display designs. For both categories, coherent imaging provides notable benefits: inherently extending DoF compared to incoherent imaging for displays, and empowering holographic imaging to fully recreate the wave field. Nonetheless, there are critical problems in the form of speckle noise, accurate simultaneous wave modelling of specular and diffuse light and computational complexity of holographic encoding.

The research conducted during this thesis work can be broadly categorized into two main categories: the analysis (RQ1) and synthesis (RQ2 and RQ3) of optimal near-eye holograms, and end-to-end learning for coherent and holographic NEDs (RQ4 and RQ5). Specifically, the following outcomes and contributions resulted as a part of this thesis:

- Simple analysis framework for accommodation cues in holographic NEDs, utilized to show how holographic stereograms can provide accurate retinal blur, albeit in a limited depth range near the hologram (Publication III)
- Improvements to analytical hologram synthesis
  - time-multiplexed speckle suppression for 3D holograms from LFs and depth, utilizing ray separation and LF resampling (Publication II)
  - novel 3D hologram synthesis method capable of producing accurate blur and monocular view-dependent effects by combining segmented overlap-

ping plane and spherical wave models (Publication V)

- End-to-end learning framework for coherent imaging displays, successfully applied to accommodation-invariant displays for their DoF extension and holographic displays for accelerating the underlying hologram synthesis (Publications IV & V)

Altogether, the research conducted during this thesis work has resulted in several key improvements for coherent and holographic NEDs that have significant potential future implications. The analysis framework developed in Publication III can be utilized for various hologram synthesis methods and display parameters to examine how accurately they can recreate retinal blur, which is crucial for accommodation-enabling displays to avoid VAC. Furthermore, the findings of the publication in terms of HS link its critical parameters to the degree of retinal blur accuracy and demonstrate the similarity between HS and LF displays. On the other hand, the speckle suppression method in Publication II improves upon existing ray separation approaches by employing a rigorous signal processing algorithm to resample the LF and to better model the light behaviour across the scene. The proposed solution minimizes the perceived unwanted intensity variations that were present in the existing method, thus enabling efficient artifact-free speckle suppression for ray-based holography. Closing the first category of contributions, the analytical hologram synthesis method developed as part of Publication V accurately models both specular and diffuse light emission, thus providing a minimal-compromise method for obtaining high-quality 3D holograms when processing time is not critical.

In the second category of contributions, the end-to-end learning frameworks of Publications IV & V provide two different means for removing VAC. In Publication IV, the results of the coherent (i.e. amplitude-modulating) accommodation-invariant display design demonstrate that the sharp display response region is improved over the incoherent competitors at the cost of introducing speckle noise. Although reducing the speckle noise comes with certain trade-offs, either to the simplicity of the display configuration or to the effective framerate, the overall approach is promising and can lead to new developments in the accommodation-invariant display design space. Finally, Publication V continues with a holographic imaging solution to VAC by training a neural network to approximate the accurate analytical hologram encoding process at its inference rate, thus eradicating the main weakness of the analytical method, namely its computational complexity. The results of Publication V show that the training is successful and that the network-predicted holograms are a close approximation of the analytical ones. As a consequence, the learning approach is full of promise for bringing realistic and accurate real-time hologram encoding closer to practicality.

Beyond the work included in this thesis, which covers the presented research questions and problems thoroughly, there is important and challenging future work. As the proposed methods are evaluated purely in a simulated environment, conducting psychophysical experiments and objective camera-based measurements with coherent and holographic display prototypes will be beneficial to verify the results. Following that, many of the presented approaches must be further extended to account for practical display limitations, including limitations to wave modulation (e.g. phase-only), modulator device limitations (e.g. quantization levels) and distortions caused by display relay optics. Continuing the research work in these directions will have great potential for the development and deployment of the next-generation immersive NEDs.

REFERENCES

[1] D. A. Wismeijer, C. J. Erkelens, R. van Ee, and M. Wexler, "Depth cue combination in spontaneous eye movements", *Journal of Vision*, vol. 10, no. 6, pp. 1–15, 2010. DOI: 10.1167/10.6.25.

[2] I. P. Howard and B. J. Rogers, *Binocular Vision and Stereopsis*. Oxford University Press, 1996, vol. 29.

[3] M. W. Halle, "Autostereoscopic Displays and Computer Graphics", *ACM SIGGRAPH Computer Graphics*, vol. 31, no. 2, pp. 58–62, 1997. DOI: 10.1145/271283.271309.

[4] T.-C. Poon, T. Akin, G. Indebetouw, and T. Kim, "Horizontal-parallax-only electronic holography", *Optics Express*, vol. 13, no. 7, pp. 2427–2432, 2005. DOI: 10.1364/OPEX.13.002427.

[5] J. W. Nadler, M. Nawrot, D. E. Angelaki, and G. C. DeAngelis, "MT Neurons Combine Visual Motion with a Smooth Eye Movement Signal to Code Depth-Sign from Motion Parallax", *Neuron*, vol. 63, no. 4, pp. 523–532, 2009. DOI: 10.1016/j.neuron.2009.07.029.

[6] F. Campbell, "The Depth of Field of the Human Eye", *Optica Acta: International Journal of Optics*, vol. 4, no. 4, pp. 157–164, 1957. DOI: 10.1080/713826091.

[7] S. Marcos, E. Moreno, and R. Navarro, "The depth-of-field of the human eye from objective and subjective measurements", *Vision Research*, vol. 39, no. 12, pp. 2039–2049, 1999. DOI: 10.1016/S0042-6989(98)00317-4.

[8] S. A. Cholewiak, G. D. Love, and M. S. Banks, "Creating correct blur and its effect on accommodation", *Journal of Vision*, vol. 18, no. 9, pp. 1–29, 2018. DOI: 10.1167/18.9.1.

[9] A. B. Watson and A. J. Ahumada, “Blur clarified: A review and synthesis of blur discrimination”, *Journal of Vision*, vol. 11, no. 5, pp. 1–23, 2011. DOI: 10.1167/11.5.10.

[10] R. J. Jacobs, G. Smith, and C. D. Chan, “Effect of defocus on blur thresholds and on thresholds of perceived change in blur: Comparison of source and observer methods.”, *Optometry and vision science*, vol. 66, no. 8, pp. 545–553, 1989. DOI: 10.1097/00006324-198908000-00010.

[11] C. M. Schor, “A dynamic model of cross-coupling between accommodation and convergence: Simulations of step and frequency responses.”, *Optometry and vision science*, vol. 69, no. 4, pp. 258–269, 1992. DOI: 10.1097/00006324-199204000-00002.

[12] G. Hung, “Quantitative analysis of the accommodative convergence to accommodation ratio: linear and nonlinear static models”, *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 4, pp. 306–316, 1997. DOI: 10.1109/10.563300.

[13] D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, “Vergence-accommodation conflicts hinder visual performance and cause visual fatigue”, *Journal of Vision*, vol. 8, no. 3, pp. 1–30, 2008. DOI: 10.1167/8.3.33.

[14] T. Shibata, J. Kim, D. M. Hoffman, and M. S. Banks, “The zone of comfort: Predicting visual discomfort with stereo displays”, *Journal of Vision*, vol. 11, no. 8, pp. 1–21, 2011. DOI: 10.1167/11.8.11.

[15] M. Emoto, T. Niida, and F. Okano, “Repeated vergence adaptation causes the decline of visual functions in watching stereoscopic television”, *Journal of Display Technology*, vol. 1, no. 2, pp. 328–340, 2005. DOI: 10.1109/JDT.2005.858938.

[16] G.-A. Koulieris, B. Bui, M. S. Banks, and G. Drettakis, “Accommodation and Comfort in Head-Mounted Displays”, *ACM Transactions on Graphics*, vol. 36, no. 4, pp. 1–11, 2017. DOI: 10.1145/3072959.3073622.

[17] M. Von Waldkirch, P. Lukowicz, and G. Tröster, “LCD-based coherent wearable projection display for quasi-accommodation-free imaging”, *Optics Communications*, vol. 217, no. 1-6, pp. 133–140, 2003. DOI: 10.1016/S0030-4018(03)01132-5.

[18] R. Konrad, N. Padmanaban, K. Molner, E. A. Cooper, and G. Wetzstein, “Accommodation-Invariant Computational Near-Eye Displays”, *ACM Transactions on Graphics*, vol. 36, no. 4, pp. 1–12, 2017. DOI: 10.1145/3072959.3073594.

[19] U. Akpinar, E. Sahin, and A. Gotchev, “Phase-Coded Computational Imaging For Accommodation-Invariant Near-Eye Displays”, in *2020 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2020, pp. 3159–3163. DOI: 10.1109/ICIP40778.2020.9191236.

[20] M. Levoy and P. Hanrahan, “Light Field Rendering”, in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, ACM, 1996, pp. 31–42. DOI: 10.1145/237170.237199.

[21] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, “The Lumigraph”, in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, ACM, 1996, pp. 43–54. DOI: 10.1145/237170.237200.

[22] C. van Berkel, “Image preparation for 3D LCD”, in *Stereoscopic Displays and Virtual Reality Systems VI*, vol. 3639, SPIE, 1999, pp. 84–91. DOI: 10.1117/12.349368.

[23] K. Matsumoto and T. Honda, “Research of 3D display using anamorphic optics”, in *Stereoscopic Displays and Virtual Reality Systems IV*, vol. 3012, SPIE, 1997, pp. 199–207. DOI: 10.1117/12.274458.

[24] M. Hutley, R. Stevens, and D. Daly, “Microlens arrays”, *Physics World*, vol. 4, no. 7, p. 27, 1991. DOI: 10.1088/2058-7058/4/7/25.

[25] H. E. Ives, “A Camera for Making Parallax Panoramagrams”, *Journal of the Optical Society of America*, vol. 17, no. 6, pp. 435–439, 1928. DOI: 10.1364/JOSA.17.000435.

[26] T. Ueno and Y. Takaki, “Super multi-view near-eye display to solve vergence–accommodation conflict”, *Optics Express*, vol. 26, no. 23, p. 30 703, 2018. DOI: 10.1364/oe.26.030703.

[27] C. Yao, D. Cheng, T. Yang, and Y. Wang, “Design of an optical see-through light-field near-eye display using a discrete lenslet array”, *Optics Express*, vol. 26, no. 14, pp. 18 292–18 301, 2018. DOI: 10.1364/OE.26.018292.

[28] T. Zhan, Y.-H. Lee, and S.-T. Wu, “High-resolution additive light field near-eye display by switchable Pancharatnam–Berry phase lenses”, *Optics Express*, vol. 26, no. 4, pp. 4863–4872, 2018. DOI: 10.1364/OE.26.004863.

[29] J. Zhang, X. Lan, C. Zhang, X. Liu, and F. He, “Switchable near-eye integral imaging display with difunctional metalens array”, *Optik*, vol. 204, p. 163 852, 2020. DOI: 10.1016/j.ijleo.2019.163852.

[30] S. A. Benton and V. M. Bove, *Holographic imaging*, 1st edition. Wiley, 2008.

[31] T. Yoneyama, E. Murakami, Y. Oguro, H. Kubo, K. Yamaguchi, and Y. Sakamoto, “Holographic head-mounted display with correct accommodation and vergence stimuli”, *Optical Engineering*, vol. 57, no. 6, p. 061 619, 2018. DOI: 10.1117/1.OE.57.6.061619.

[32] H. Mizushina, I. Negishi, J. Nakamura, Y. Takaki, H. Ando, and S. Masaki, “Accommodation and Vergence Responses to Electronic Holographic Displays and Super Multiview Holographic Stereograms”, *IEEE Transactions on Industry Applications*, vol. 58, no. 6, pp. 7978–7987, 2022. DOI: 10.1109/TIA.2022.3192739.

[33] J. W. Goodman, *Speckle Phenomena in Optics: Theory and Applications*. Roberts and Company Publishers, 2007.

[34] S. Reichelt, R. Häussler, G. Fütterer, *et al.*, “Full-range, complex spatial light modulator for real-time holography”, *Optics Letters*, vol. 37, no. 11, pp. 1955–1957, 2012. DOI: 10.1364/OL.37.001955.

[35] C. K. Hsueh and A. A. Sawchuk, “Computer-generated double-phase holograms”, *Applied Optics*, vol. 17, no. 24, pp. 3874–3883, 1978. DOI: 10.1364/AO.17.003874.

[36] R. Eschbach, “Comparison of error diffusion methods for computer-generated holograms”, *Applied Optics*, vol. 30, no. 26, pp. 3702–3710, 1991. DOI: 10.1364/AO.30.003702.

[37] T. Leportier, M.-C. Park, and T. Kim, “Numerical Alignment of Spatial Light Modulators for Complex Modulation in Holographic Displays”, *Journal of Display Technology*, vol. 12, no. 10, pp. 1000–1007, 2016. DOI: 10.1109/JDT.2016.2576446.

[38] C. Jang, K. Bang, G. Li, and B. Lee, "Holographic Near-eye Display with Expanded Eye-box", *ACM Transactions on Graphics*, vol. 37, no. 6, pp. 1–14, 2018. DOI: 10.1145/3272127.3275069.

[39] J. An, K. Won, Y. Kim, *et al.*, "Slim-panel holographic video display", *Nature Communications*, vol. 11, 2020. DOI: 10.1038/s41467-020-19298-4.

[40] A. Maimone, A. Georgiou, and J. S. Kollin, "Holographic Near-Eye Displays for Virtual and Augmented Reality", *ACM Transactions on Graphics*, vol. 36, no. 4, pp. 1–16, 2017. DOI: 10.1145/3072959.3073624.

[41] B. R. Brown and A. W. Lohmann, "Complex Spatial Filtering with Binary Masks", *Applied Optics*, vol. 5, no. 6, pp. 967–969, 1966. DOI: 10.1364/AO.5.000967.

[42] C. Slinger, C. Cameron, and M. Stanley, "Computer-generated holography as a generic display technology", *Computer*, vol. 38, no. 8, pp. 46–53, 2005. DOI: 10.1109/MC.2005.260.

[43] J. T. McCrickerd and N. George, "Holographic stereogram from sequential component photographs", *Applied Physics Letters*, vol. 12, no. 1, pp. 10–12, 1968. DOI: 10.1063/1.1651831.

[44] M. W. Halle, S. A. Benton, M. A. Klug, and J. S. Underkoffler, "Ultragram: a generalized holographic stereogram", in *Practical Holography V*, vol. 1461, SPIE, 1991, pp. 142–155. DOI: 10.1117/12.44722.

[45] M. E. Lucente, "Interactive computation of holograms using a look-up table", *Journal of Electronic Imaging*, vol. 2, no. 1, pp. 28–34, 1993. DOI: 10.1117/12.133376.

[46] R. Lange and P. Seitz, "Solid-state time-of-flight range camera", *IEEE Journal of Quantum Electronics*, vol. 37, no. 3, pp. 390–397, 2001. DOI: 10.1109/3.910448.

[47] M. Yamaguchi, H. Hoshino, T. Honda, and N. Ohyama, "Phase-added stereogram: Calculation of hologram using computer graphics technique", in *Practical Holography VII: Imaging and Materials*, vol. 1914, SPIE, 1993, pp. 25–31. DOI: 10.1117/12.155027.

[48] H. Kang, T. Yamaguchi, and H. Yoshikawa, "Accurate phase-added stereogram to improve the coherent stereogram", *Applied Optics*, vol. 47, no. 19, pp. D44–D54, 2008. DOI: 10.1364/AO.47.000D44.

[49] Q. Y. J. Smithwick, J. Barabas, D. E. Smalley, and V. M. Bove Jr., "Interactive holographic stereograms with accommodation cues", in *Practical Holography XXIV: Materials and Applications*, vol. 7619, SPIE, 2010, p. 761 903. DOI: 10.1117/12.840526.

[50] L. Shi, F.-C. Huang, W. Lopes, W. Matusik, and D. Luebke, "Near-Eye Light Field Holographic Rendering with Spherical Waves for Wide Field of View Interactive 3D Computer Graphics", *ACM Transactions on Graphics*, vol. 36, no. 6, pp. 1–17, 2017. DOI: 10.1145/3130800.3130832.

[51] N. Padmanaban, Y. Peng, and G. Wetzstein, "Holographic Near-Eye Displays Based on Overlap-Add Stereograms", *ACM Transactions on Graphics*, no. 6, pp. 1–13, 38 2019. DOI: 10.1145/3355089.3356517.

[52] R. Ziegler, S. Bucheli, L. Ahrenberg, M. Magnor, and M. Gross, "A Bidirectional Light Field - Hologram Transform", *Computer Graphics Forum*, vol. 26, no. 3, pp. 435–446, 2007. DOI: 10.1111/j.1467-8659.2007.01066.x.

[53] J. P. Waters, "Holographic image synthesis utilizing theoretical methods", *Applied Physics Letters*, vol. 9, no. 11, pp. 405–407, 1966. DOI: 10.1063/1.1754630.

[54] D. Leseberg and C. Frère, "Computer-generated holograms of 3-D objects composed of tilted planar segments", *Applied Optics*, vol. 27, no. 14, pp. 3020–3024, 1988. DOI: 10.1364/AO.27.003020.

[55] H. Zhang, Q. Tan, and G. Jin, "Full parallax three-dimensional computer generated hologram with occlusion effect using ray casting technique", *Journal of Physics: Conference Series*, vol. 415, no. 1, p. 012 048, 2013. DOI: 10.1088/1742-6596/415/1/012048.

[56] R. H.-Y. Chen and T. D. Wilkinson, "Computer generated hologram from point cloud using graphics processor", *Applied Optics*, vol. 48, no. 36, pp. 6841–6850, 2009. DOI: 10.1364/AO.48.006841.

[57] S.-C. Kim and E.-S. Kim, "Effective generation of digital holograms of three-dimensional objects using a novel look-up table method", *Applied Optics*, vol. 47, no. 19, pp. D55–D62, 2008. DOI: 10.1364/AO.47.000D55.

[58] Y. Peng, S. Choi, N. Padmanaban, and G. Wetzstein, "Neural Holography with Camera-in-the-Loop Training", *ACM Transactions on Graphics*, vol. 39, no. 6, pp. 1–14, 2020. DOI: 10.1145/3414685.3417802.

[59] R. Horisaki, R. Takagi, and J. Tanida, "Deep-learning-generated holography", *Applied Optics*, vol. 57, no. 14, pp. 3859–3863, 2018. DOI: 10.1364/AO.57.003859.

[60] M. H. Eybposh, N. W. Caira, M. Atisa, P. Chakravarthula, and N. C. Pégard, "DeepCGH: 3D computer-generated holography using deep learning", *Optics Express*, vol. 28, no. 18, pp. 26 636–26 650, 2020. DOI: 10.1364/OE.399624.

[61] S. Choi, M. Gopakumar, Y. Peng, J. Kim, M. O'Toole, and G. Wetzstein, "Time-multiplexed Neural Holography: A flexible framework for holographic near-eye displays with fast heavily-quantized spatial light modulators", in *Proceedings of the ACM SIGGRAPH*, ACM, 2022, pp. 1–9. DOI: 10.1145/3528233.3530734.

[62] L. Shi, B. Li, C. Kim, P. Kellnhofer, and W. Matusik, "Towards real-time photorealistic 3D holography with deep neural networks", *Nature*, vol. 591, pp. 234–239, 2021. DOI: 10.1038/s41586-020-03152-0.

[63] H. Zhang, L. Cao, and G. Jin, "Computer-generated hologram with occlusion effect using layer-based processing", *Applied Optics*, vol. 56, no. 13, F138–F143, 2017. DOI: 10.1364/AO.56.00F138.

[64] L. Shi, B. Li, and W. Matusik, "End-to-end learning of 3D phase-only holograms for holographic display", *Light: Science & Applications*, vol. 11, no. 247, 2022. DOI: 10.1038/s41377-022-00894-6.

[65] T. Shimobaba, T. Kakue, and T. Ito, "Real-time and low speckle holographic projection", in *2015 IEEE 13th International Conference on Industrial Informatics (INDIN)*, IEEE, 2015, pp. 732–741. DOI: 10.1109/INDIN.2015.7281827.

[66] M. Yamaguchi, H. Endoh, T. Honda, and N. Ohyama, “High-quality recording of a full-parallax holographic sterogram with a digital diffuser”, *Optics Letters*, vol. 19, no. 2, pp. 135–137, 1994. DOI: 10.1364/OL.19.000135.

[67] F. Yaraş, H. Kang, and L. Onural, “Real-time phase-only color holographic video display system using LED illumination”, *Applied Optics*, vol. 48, no. 34, H48–H53, 2009. DOI: 10.1364/AO.48.000H48.

[68] J. Amako, H. Miura, and T. Sonehara, “Speckle-noise reduction on kinoform reconstruction using a phase-only spatial light modulator”, *Applied Optics*, vol. 34, no. 17, pp. 3165–3171, 1995. DOI: 10.1364/AO.34.003165.

[69] Y. Takaki and M. Yokouchi, “Speckle-free and grayscale hologram reconstruction using time-multiplexing technique”, *Optics Express*, vol. 19, no. 8, pp. 7567–7579, 2011. DOI: 10.1364/OE.19.007567.

[70] T. Utsugi and M. Yamaguchi, “Speckle-suppression in hologram calculation using ray-sampling plane”, *Optics Express*, vol. 22, no. 14, pp. 17 193–17 206, 2014. DOI: 10.1364/OE.22.017193.

[71] J. W. Goodman, *Introduction to Fourier Optics*, 2nd edition. McGraw-Hill, 1996.

[72] T.-C. Poon and J.-P. Liu, *Introduction to Modern Digital Holography: With Matlab*. Cambridge University Press, 2014.

[73] Z. Lin and H.-Y. Shum, “A Geometric Analysis of Light Field Rendering”, *International Journal of Computer Vision*, vol. 58, no. 2, pp. 121–138, 2004. DOI: 10.1023/B:VISI.0000015916.91741.27.

[74] C. H. J. Howard, “A Test for the Judgment of Distance”, *American journal of ophthalmology*, vol. 2, no. 9, pp. 656–675, 1919. DOI: 10.1016/S0002-9394(19)90180-2.

[75] S. Vagharshakyan, R. Bregovic, and A. Gotchev, “Light Field Reconstruction Using Shearlet Transform”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 1, pp. 133–147, 2018. DOI: 10.1109/TPAMI.2017.2653101.

[76] P. S. Hilaire, “Modulation transfer function and optimum sampling of holographic stereograms”, *Applied Optics*, vol. 33, no. 5, pp. 768–774, 1994. DOI: 10.1364/AO.33.000768.

[77] K. Wakunami and M. Yamaguchi, "Calculation for computer generated hologram using ray-sampling plane", *Optics Express*, vol. 19, no. 10, pp. 9086–9101, 2011. DOI: 10.1364/OE.19.009086.

[78] H. Huang and H. Hua, "Systematic characterization and optimization of 3D light field displays", *Optics Express*, vol. 25, no. 16, pp. 18 508–18 525, 2017. DOI: 10.1364/OE.25.018508.

[79] F. W. Campbell and J. G. Robson, "Application of Fourier analysis to the visibility of gratings", *The Journal of Physiology*, vol. 197, no. 3, pp. 551–566, 1968. DOI: 10.1113/jphysiol.1968.sp008574.

[80] L. N. Thibos, M. Ye, X. Zhang, and A. Bradley, "The chromatic eye: A new reduced-eye model of ocular chromatic aberration in humans", *Applied Optics*, vol. 31, no. 19, pp. 3594–3600, 1992. DOI: 10.1364/AO.31.003594.

[81] P. K. Shrestha, M. J. Pryn, J. Jia, *et al.*, "Accommodation-Free Head Mounted Display with Comfortable 3D Perception and an Enlarged Eye-box", *Research*, vol. 2019, no. 9273723, pp. 1–9, 2019. DOI: 10.34133/2019/9273723.

[82] J. Lyu, C. J. Ng, S. P. Bang, and G. Yoon, "Binocular accommodative response with extended depth of focus under controlled convergences", *Journal of Vision*, vol. 21, no. 8, pp. 1–12, 2021. DOI: 10.1167/jov.21.8.21.

[83] Y. Deng and D. Chu, "Coherence properties of different light sources and their effect on the image sharpness and speckle of holographic displays", *Scientific reports*, vol. 7, no. 1, pp. 1–12, 2017. DOI: 10.1038/s41598-017-06215-x.

[84] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, vol. 9351, Springer, Cham, 2015, pp. 234–241. DOI: 10.1007/978-3-319-24574-4_28.

[85] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity", *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004. DOI: 10.1109/TIP.2003.819861.

[86] M. Akram and X. Chen, "Speckle reduction methods in laser-based picture projectors", *Optical Review*, vol. 23, pp. 108–120, 2016. DOI: 10.1007/s10043-015-0158-6.

[87] J.-S. Sohn, M.-B. Lee, W.-C. Kim, *et al.*, "Design and fabrication of diffractive optical elements by use of gray-scale photolithography", *Applied Optics*, vol. 44, no. 4, pp. 506–511, 2005. DOI: 10.1364/AO.44.000506.

[88] M. Meem, S. Banerji, C. Pies, *et al.*, "Large-area, high-numerical-aperture multi-level diffractive lens via inverse design", *Optica*, vol. 7, no. 3, pp. 252–253, 2020. DOI: 10.1364/OPTICA.388697.

[89] U. Akpinar, E. Sahin, M. Meem, R. Menon, and A. Gotchev, "Learning Wavefront Coding for Extended Depth of Field Imaging", *IEEE Transactions on Image Processing*, vol. 30, pp. 3307–3320, 2021. DOI: 10.1109/TIP.2021.3060166.

[90] S. Koch, A. Matveev, Z. Jiang, *et al.*, "ABC: A Big CAD Model Dataset For Geometric Deep Learning", in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2019, pp. 9593–9603. DOI: 10.1109/CVPR.2019.00983.

[91] R. Suvorov, E. Logacheva, A. Mashikhin, *et al.*, "Resolution-robust large mask inpainting with fourier convolutions", in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, IEEE, 2022, pp. 2149–2159. DOI: 10.1109/WACV51458.2022.00323.

PUBLICATIONS

PUBLICATION
I

Computer-Generated Holograms for 3D Imaging: A Survey

E. Sahin, E. Stoykova, J. Mäkinen, and A. Gotchev

ACM Computing Surveys, vol. 53, no. 2, pp. 1-35
DOI: 10.1145/3378444

Computer-generated holograms for 3D imaging: A survey

ERDEM SAHIN, Tampere University, Finland
ELENA STOYKOVA, Institute of Optical Materials and Technologies, Bulgaria
JANI MÄKINEN, Tampere University, Finland
ATANAS GOTCHEV, Tampere University, Finland

Holography is usually considered as the ultimate way to visually reproduce a three-dimensional scene. Computer-generated holography constitutes an important branch of holography, which enables visualization of artificially-generated scenes as well as real three-dimensional scenes recorded under white light illumination. In this article, we present a comprehensive survey of methods for synthesis of computer-generated holograms classifying them into two broad categories, namely as wavefront-based methods and ray-based methods. We examine their modern implementations in terms of the quality of reconstruction and computational efficiency. As it is an integral part of computer-generated holography, we devote a special section to speckle suppression, which is also discussed under two categories following the classification of underlying computer-generated hologram methods.

CCS Concepts: • **General and reference → Surveys and overviews**; • **Human-centered computing → Visualization**; • **Hardware → Displays and imagers**.

Additional Key Words and Phrases: Computer-generated holograms, 3D displays, 3D imaging

ACM Reference Format:
Erdem Sahin, Elena Stoykova, Jani Mäkinen, and Atanas Gotchev. 2019. Computer-generated holograms for 3D imaging: A survey. *ACM Comput. Surv.* 1, 1, Article 1 (January 2019), 42 pages.

1 INTRODUCTION

The holograms carry intensity, color, depth and directional information of a given three-dimensional (3D) scene and they are able to reconstruct the corresponding true light wavefronts. They provide continuous motion parallax and deliver correct visual cues of binocular disparity, vergence, accommodation (focus) and retinal blur, which are all critical for accurate depth perception. Thus, holographic 3D imaging enables highly realistic visualization and this makes it a very appealing research area.

Advances in digital sensor and display devices as well as computing have introduced new areas into traditional optical (analog) holographic imaging. Significant efforts have been dedicated to development of dynamic holographic displays [Yamaguchi 2016; Yaraş et al. 2010]. Device based limitations still exhibit an important challenge to overcome especially for glasses-free type of visualization due to need for high space-bandwidth product, i.e., correspondingly, wide field of

---

Authors' addresses: Erdem Sahin, Tampere University, Faculty of Information Technology and Communication Sciences, Korkeakoulunkatu 1, Tampere, 33720, Finland, erdem.sahin@tuni.fi; Elena Stoykova, Institute of Optical Materials and Technologies, Bulgarian Academy of Sciences, Acad. Georgi Bonchev Str., Bl.109, Sofia, 1113, Bulgaria, elena.stoykova@gmail.com; Jani Mäkinen, Tampere University, Faculty of Information Technology and Communication Sciences, Korkeakoulunkatu 1, Tampere, 33720, Finland, jani.makinen@tuni.fi; Atanas Gotchev, Tampere University, Faculty of Information Technology and Communication Sciences, Korkeakoulunkatu 1, Tampere, 33720, Finland, atanas.gotchev@tuni.fi.

0360-0300/2019/1-ART1 $15.00
https://doi.org/

view and wide viewing angle. However, this constraint can be significantly relieved in the case of near-eye or head-mounted display scenarios, intended for virtual or augmented type of visualization. This, already makes holographic imaging a viable option for such displays [Maimone et al. 2017]. Furthermore, compared to other advanced 3D display technologies such as ray-based light field displays including super multiview displays [Takaki and Nago 2010], integral imaging [Xiao et al. 2013] and tensor displays [Wetzstein et al. 2012], holographic imaging and displays have the very crucial advantage that it can accurately reproduce deep 3D scenes.

The scene capture or recording methods for holographic displays can be classified into two broad categories, namely digital holography and computer-generated holography [Tsang and Poon 2016]. Digital holography utilizes optically recorded digital holograms. The need for precise and optically stable coherent illumination setup imposes a strong constraint in such holograms, which usually limits the capability of such holograms to static and small-scale scenes. The second method is based on computational synthesis of so-called computer-generated holograms (CGHs). Computer-generated holography seems to have much greater potential due to the capability of hologram synthesis for synthetically generated computer graphics (CG) objects and real scenes recorded under natural white light.

Besides displaying CGHs on dynamic holographic displays, it is also possible to "print" them onto a physical carrier using lithography. The methods such as HS printing [Brotherton-Ratcliffe et al. 2011] or wavefront printing [Kang et al. 2016b] combine computational and analog holography by recording white-light viewable holograms from digital contents. Combined implementation of both concepts is reported in [Matsushima and Sonobe 2018], where a wave field captured for a real 3D object by means of digital holography is incorporated in a CGH. Application areas of CGHs are not limited to holographic displays. They are also utilized in holographic projection [Buckley 2011], or they can serve as diffractive optical elements for various other purposes such as lens, diffraction grating, combiner, phase spatial filter, beam shaper, optical tweezer (photonic manipulation of particles), etc., which make them useful in various advanced scientific and technological fields as well as industry sectors [Cirino et al. 2011].

In the last two decades, much effort in computer-generated holography has been put on i) improvement of object data encoding in the hologram by more accurate rendering techniques for high-quality reconstructions and ii) fast computation by developing more efficient algorithms and utilizing high-performance hardware. This paper presents a survey of the state-of-the-art in synthesis of CGHs with a focus in this time duration where notable advances have been achieved in the field. The structure of the survey is as follows: Section 2 introduces preliminary theoretical background as well as a short historical overview of CGH synthesis. The theoretical background is supplemented in Appendix A with numerical wave propagation algorithms. We divide the existing CGH methods into wavefront-based and ray-based methods, which are presented in Section 3 and Section 4, respectively, with their pros and cons from the point of view of imaging quality and computational efficiency. These two groups consist of several subgroups, which differ by the type of model employed for 3D scene representation as well as wave propagation method adopted for computing the hologram due to given model. Section 5 discusses the speckle suppression techniques for both wavefront and ray-based CGH methods. Comparative summary of the described methods concludes the paper. Please note that in Appendix B we also provide a supplementary discussion on synthesis and numerical reconstruction of different types of CGHs via simulating the process of CGH viewing by the human eye.

2 BACKGROUND

Holography is a two-step process of recording and reconstruction of the wavefront due to a 3D object under coherent illumination. This wavefront, the so-called object field, is described by a complex amplitude $O(x,y) = a_0(x,y)\exp[j\phi_0(x,y)]$ on the hologram plane (x,y) at $z = 0$, where $a_0(x,y)$ and $\phi_0(x,y)$ are the amplitude and the phase of the object field, respectively, and j is the imaginary unit. The amplitude gives the intensity of light, while the phase encodes the depth information. The hologram records interference of the object wave with a mutually coherent reference wave, $R(x,y) = a_R(x,y)\exp[j\phi_R(x,y)]$ with amplitude $a_R(x,y)$ and phase $\phi_R(x,y)$ (see Figure 1(a)), as a two-dimensional (2D) distribution of intensity:

$$I_H(x,y) = |R(x,y) + O(x,y)|^2 = RR^* + OO^* + OR^* + O^*R, \tag{1}$$

where the asterisk denotes a complex conjugate operator. The relevant information is encoded in the last two terms, which are referred to as +1 and -1 diffraction orders, respectively. The sum of the intensities of the reference and the object waves, $RR^* + OO^*$, forms the zero-order term, which is usually much brighter than the first-order terms. In a CGH, this zero-order can be discarded and only the relevant object information can be encoded in the hologram resulting in the so-called bipolar intensity distribution [Lucente 1994] as

$$\tilde{I}_H(x,y) = 2\mathrm{Re}\{O(x,y)R^*(x,y)\}. \tag{2}$$

The object wavefront or its conjugate is reconstructed by multiplication of $I_H(x,y)$ with the replica of $R(x,y)$ or its conjugate. That is, for a unit-amplitude reference wave, $OR^*R = O$ or $O^*RR^* = O^*$. This brings into focus the virtual (orthoscopic) image or the real (pseudoscopic) image of the object. Both reconstructed object wavefronts carry the same information content, so they create twin images. In in-line geometry, when $R(x,y)$ and $O(x,y)$ propagate along the same direction, the twin images overlap focusing on the opposite sides of the in-line hologram. This means that in the plane of the focused reconstruction of a virtual image there is a defocused real image and vice versa. Mathematically, the existence of twin images is due to non-linear encoding of complex information as a 2D array of real numbers.

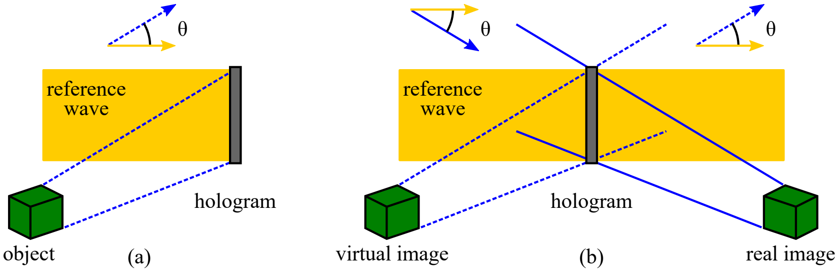


Fig. 1. Holographic recording (a) and holographic reconstruction (b) in off-axis geometry.

The formulation of the holographic principle given by Equation (1) was actually first established with in-line geometry by Dennis Gabor, in 1948. He considered inline illumination of a semi-transparent object with a point light source [Gabor 1948]. Separation of the light beams of the twin images in space was demonstrated later, in 1962, by Leith and Upatnieks [Leith and Upatnieks 1962] via an off-axis geometry, where a spatial carrier frequency was introduced in the hologram through angular separation of the object and reference beams (see Figure 1). Introduction of the carrier frequency shifts the object spectrum from the zero frequency in the spatial frequency domain. This solution decreases the useful space-bandwidth product of the hologram, which is defined as the product of the hologram size and its spatial frequency bandwidth [Claus et al. 2011; Lohmann et al. 1996]. The space-bandwidth product in digital holography characterizes the field of view the hologram can support. At off-axis geometry, only one-fourth of the space-bandwidth product is practically used.

Synthesis of CGHs basically requires acquiring the 3D information of the scene, i.e., the object field $O(x, y)$. Having found the object field, the CGH itself can be calculated by applying Equation (1) following the in-line or off-axis recording geometry. Thus, the CGH synthesis includes three steps: i) digital representation of the 3D scene; ii) digital propagation of the wave field, i.e., object field, from the scene to the hologram plane; iii) digital encoding of the fringe pattern on the hologram plane in accordance with Equation (1). Over the years, implementation of these steps has created different groups of algorithms.

CGH-based holographic 3D imaging is not a trivial task. Crucial factors for the CGH synthesis are image quality and computational burden. High image quality means ability for high spatial resolution, full color and full parallax photorealistic representation and reconstruction in a wide viewing angle of 3D scenes with shading, reflections and occlusions as schematically depicted in Figure 2. Reconstruction of color, shading and reflection depends on illumination direction, surface properties and orientation. Reflection from an opaque surface is given by bidirectional reflectance distribution function which in general may have components describing ambient, diffuse and specular reflection. Occlusion can be mutual occlusion when an object in a 3D scene blocks the light coming from other object or a self-occlusion when some parts of an object shields its other parts. Such properties constitute view-dependent characteristics of CGHs, and, in addition to provided spatial resolution, they are also crucial to ensure high quality realistic reconstructions. In order to support all those, the CGH must possess a very small pixel pitch and extremely large number of pixels. This entails processing of huge amount of data, which makes computational efficiency of vital importance for CGH synthesis, especially for dynamic imaging.

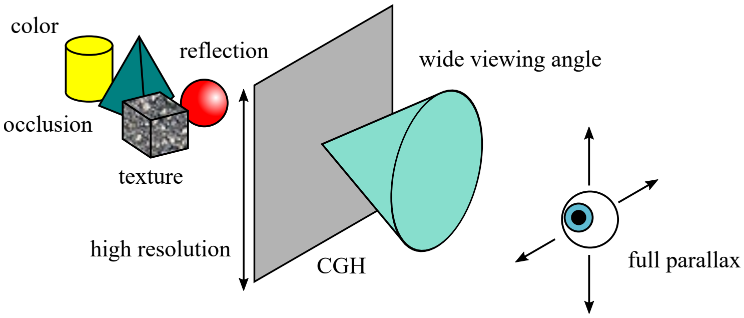


Fig. 2. Requirements for high-quality CGH reconstruction.

Historically, at the beginning of the CGH era, the encoding step dominated the CGH synthesis due to the need of physically fabricating them as optical elements with the required diffractive properties to create the desired visual effect. Design of a CGH as a diffractive optical element followed the development of technology. It started with holograms reproduced by computer plotters as hard copies, which were transferred onto a photographic film in the 1960s, went through photolithographic fabrication in the 1970s and ended with using e-beams in the 1980s. Lohmann's group made the first substantial progress in CGH production technology. Lohmann formulated the idea of encoding a numerically generated hologram in 1956, and various encoding methods such as the "single-sideband" technique and the "detour phase" method [Brown and Lohmann 1966; Lohmann and Paris 1967] emerged in the 1960s to make use of existing computers at that time. In view of inability of the computer plotters for gray-scale drawing, the developed methods produced binary holograms. The complex amplitude $a_0 \exp(j\phi_0)$ was encoded as an aperture with an area proportional to transmission a_0 where desired phase shift ϕ_0 was achieved through displacement with respect to a regular grating grid. Spatial encoding of complex amplitudes formed the sub-class of cell-oriented holograms [Dallas 1980]. Later, to overcome the limitations of binary holograms, development of different gray-scale or phase-only coding schemes proved that the generated amplitude or phase holograms are capable of reconstructing 3D objects. The amplitude hologram inevitably reconstructs the zero-order and the twin images, whereas the phase hologram reconstructs only the object beam. Encoding of only the amplitude or phase information as 2D arrays of real numbers has allowed for production of point-oriented holograms [Dallas 1980]. These types of holograms are especially suitable for the modern pixelated SLMs.

Nowadays, holographic imaging highly benefits from the fast computer controllable SLMs with large throughput and high diffraction efficiency. Liquid crystal based and mirror based devices constitute the two widely used categories of SLMs [Yaraş et al. 2010]. In the former case, the amplitude and phase of the light are modulated based on the polarization and refractive index characteristics of the crystal, respectively. In the latter case, electro-mechanical control of a micro-mirror array produces the desired amplitude or phase modulation. Besides such devices, in an effort of producing horizontal parallax only real-time holographic display, acousto-optic modulators have been also utilized to achieve light modulation through interaction between acoustic waves and coherent light [St-Hilaire et al. 1990]. Despite the recent developments, there are still two critical issues related with SLMs. The first issue is that the conventional pixelated SLMs can modulate only the amplitude or the phase of the incident coherent light. Full complex modulation has been demonstrated via arrangements including multiple amplitude and phase SLMs [Tudela et al. 2002; Zhu and Wang 2014] as well as a single SLM with additional optical filters [Liu et al. 2011; van Putten et al. 2008]. Actually, long before using SLMs, a referenceless hologram was realized by encoding amplitude and phase information of the complex amplitude in the different layers of a color photographic film [Chu et al. 1973]. On the other hand, on the computational side, the Gerchberg-Saxton algorithm has been widely utilized as an effective method to obtain phase-only holograms, where the phase information is optimized through execution of iterative backward and forward Fourier or Fresnel transform with constraints applied in the CGH and image planes [Gerchberg and Saxton 1972; Stoykova et al. 2013]. To improve the computational efficiency and enable real-time operation, recently several non-iterative approaches have been also proposed that directly compute phase-only holograms based on, e.g., random phase noise averaging [Buckley 2011], sampled phase-only hologram [Tsang et al. 2016], and patterned phase-only hologram [Tsang et al. 2017]. The second issue related with SLMs is that CGH encoding using modern SLMs suffers from the bandwidth limitations imposed by pitch and number of the pixels employed in such devices. These two critical parameters are still far from the values required to provide wide field of view and wide viewing angle visualization. The reconstructed objects from the CGHs fed to SLMs are usually small and must be located at a large distance from the SLM due to the small diffraction angle, which is limited by the pixel pitch of the SLM, Δ, as $\theta_d = 2\sin^{-1}[\lambda/(2\Delta)]$. In the case of static imaging, better quality can be achieved thanks to the small pixel pitches provided, e.g., via lithography [Igarashi et al. 2018; Matsushima and Nakahara 2009] or by holographic printing on a silver halide plate [Brotherton-Ratcliffe et al. 2011; Kang et al. 2016b].

Long history of CGH synthesis with contributions from many research groups have resulted in partially inconsistent terminology. Being aware of this fact, we divide the existing CGH design methods in 3D imaging into wavefront-based and ray-based methods. The wavefront-based CGH methods calculate the 3D wave field due to a given object or scene, particularly its 2D distribution on the hologram plane, by simulating the diffraction process. The group of the wavefront-based methods mainly relies on scalar wave diffraction and comprises three main categories, which utilize point cloud, polygon and layered representations of the 3D scene. All such methods utilize 3D positional information of the scene. It is either explicitly available, e.g., in the form of a point cloud, or obtained by some sort of ray tracing for synthetic objects, or extracted from depth estimation in the case of real objects. On the other hand, the ray-based CGH methods form the hologram from incoherently captured 2D images of the 3D scene. That is, they rely on geometric (ray) optics formalism of light and conversion from ray-based representations to wavefront-based holographic information. The ray-based methods comprise two important categories, which are the holographic stereogram (HS) and multiple viewpoint projection (MVP) holography. Depending on the capture setup and encoding scheme, the acquired ray intensities are utilized to generate either a coherent CGH that includes coherent object field information (e.g., MVP), or its incoherent approximation (e.g., HS). The following sections discuss wavefront-based and ray-based CGH methods in detail. Please, refer to Appendix A for a theoretical background on numerical wave propagation algorithms, which are widely utilized in CGH synthesis.

3 WAVEFRONT-BASED CGH METHODS

The wavefront-based methods simulate optical wave propagation to calculate the 3D wave field due to a given 3D scene or object illuminated by a coherent light source. With further implementation of the interference process between the object field and a reference wave, the optical holographic recording process given in Equation (1) is fully simulated. The preparation of the geometrical 3D object information is the very first critical step of CGH generation. The object can be a collection of independent light sources as points or planar segments, or sliced into layers. The dense depth sampling of the scene can be made e.g., via ray casting or ray tracing.

Both the form of the utilized geometric object information and the accuracy of the adopted diffraction model are important factors determining the qualitative characteristics (such as reconstructed image quality, smoothness of parallax, handling of occlusions, etc.) of a given wavefront-based method. Besides accuracy, the computational complexity of the CGH generation is another key factor that has been extensively addressed in the CGH literature. We categorize the wavefront-based methods and discuss the distinctions between various methods under each category based on these factors. In particular, we divide the wavefront-based methods into three main categories: i) Point cloud model, ii) Polygon-based model, iii) Layer-based model.

3.1 Point cloud model

General implementation. One of the most widely used object representations is the point cloud model proposed for the first time by Waters in 1966 [Waters 1966]. According to this model, the 3D object can be represented as a collection of self-emitting point sources of light, which act independently. The wave emitted from the object is formulated as superposition of spherical waves corresponding to such point sources, i.e., the complex amplitude on the hologram plane is found as

$$O(x,y) = \sum_{p=1}^{P} \frac{A_p}{r_p} \exp(jkr_p), \tag{3}$$

where $A_p = a_p \exp(j\phi_p)$ gives the wave field emanated by point source p having real-valued amplitude a_p and phase ϕ_p, $r_p = \sqrt{(x-x_p)^2 + (y-y_p)^2 + z_p^2}$ is the distance between this point source at (x_p, y_p, z_p) and the point $(x, y, 0)$ on the hologram plane located at $z = 0$, P is the total number of point sources and $k = 2\pi/\lambda$ is the wavenumber. It is worth to note that the propagation model given by Equation (3) corresponds to the Rayleigh-Sommerfeld diffraction model without the obliquity factor (see Equation (A.3) in Appendix A). Considering a unit amplitude plane wave $R(x, y)$ normally incident on the hologram plane, the interference pattern due to the object field given by Equation (3) can be written in the form of bipolar intensity distribution as

$$\hat{I}_H(x,y) = \sum_{p=1}^{P} \frac{a_p}{r_p} \cos\left(\frac{2\pi}{\lambda} r_p + \phi_p\right) = \sum_{p=1}^{P} a_p T(x,y,x_p,y_p,z_p). \tag{4}$$

The point cloud model is the most flexible method for 3D object representation. In the reconstruction, it creates highly realistic full-parallax color images of 3D objects with arbitrary shapes at high spatial and angular resolutions. The density of the point sources should satisfy the acuity requirement of the human visual system (HVS). That is, the lateral distance between the neighboring object points subtends an angle of no more than 1/60 degrees at the viewer's eye positioned at the intended viewing depth [Born 1999]. This ensures smooth (continuous) perception of surfaces. The point cloud model is suitable for virtual and real objects; in the latter case, the point cloud is an output of profilometric or tomographic capture [Stoykova et al. 2013]. It also serves well for generation of specific types of holograms, such as image-plane hologram [Yamaguchi and Yoshikawa 2011] and cylindrical hologram [Yamaguchi et al. 2007]. Furthermore, it allows for parallel processing and utilization of field programmable gate arrays (FPGAs) and graphics processing units (GPUs) as well as clusters of these devices [Jackin et al. 2018]. The memory allocated for implementation of the point cloud method contains only the 3D scene data and the output hologram.

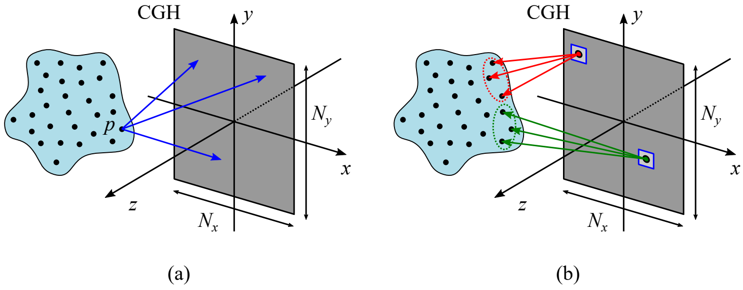


Fig. 3. Point cloud model for CGH generation: (a) object-oriented approach with ray-tracing from a fixed set of object points to the CGH samples; (b) image-oriented approach with ray-casting from the CGH samples and forming different sets of point sources for different parts of the CGH.

Many authors refer to the point cloud method as a ray-oriented, ray-tracing or coherent ray-tracing method [Ichikawa et al. 2013a; Waters 1966; Wei et al. 2016], since all possible rays between a point source and a hologram sample are traced in the hologram calculation. The CGH in the point cloud method consists of overlapping Fresnel zone plates, i.e., $T(x,y,x_p,y_p,z_p)$, $p = 1,2,\ldots,P$. In the basic implementation, the *object-oriented* or *object order approach* propagates the complex amplitudes from a fixed set of point sources representing the 3D scene to the points on the hologram plane (see Figure 3(a)). This straightforward implementation has difficulties in modeling the view-dependent image quality factors such as occlusion and parallax, and view-dependent shading. For instance, in an approximate occlusion culling method implemented in [Chen and Wilkinson 2009], the visibility of each point source from each hologram sampling location is determined based on an approximate visibility formula. Those points that are found to be occluded are then discarded in the CGH calculation. On the other hand, in the *image-oriented* or *image order approach*, illustrated in Figure 3(b), the CG techniques such as ray casting and ray tracing are employed to better handle all such view-dependent factors. In [Zhang et al. 2011], the CG ray-casting technique is used for hidden surface removal. The rays are cast from each sample of the hologram within the diffraction angle determined by the hologram sampling. For each hologram sample, a set of visible points are determined. Such an approach is very effective for expressing the occlusion effect. The ray casting approach introduced in [Ichikawa et al. 2013a] utilizes a coarser viewpoint sampling on the hologram plane. The hologram is divided into a set of elemental hologram segments and the center of each segment is chosen to be a viewpoint to cast rays within the diffraction angle of the hologram. The angular resolution of rays cast from each viewpoint is set to satisfy the requirement of visual acuity. The developed approach is able to express a combination of image properties such as shading, shadowing, multi-reflection and refraction. There are several other methods, e.g., proposed before in [Smithwick et al. 2010] and later in [Zhang et al. 2015], that are very similar to the ray casting introduced in [Ichikawa et al. 2013a]. The common feature of all such methods is that the view-dependent intensity of the scene is sampled on a regular grid on the hologram plane together with depth information for each back projected ray. The coherent wave contribution of each point to each segment is then calculated based on the knowledge of positional and amplitude information. In [Ichikawa et al. 2013b], such an approach builds a CGH for a Fourier transform optical system to enlarge the field of view. Especially for scenes including multiple 3D objects, a multi-plane occlusion processing by means of multiple point cloud sampling planes is proposed in [Jia et al. 2014]. The occlusion culling is achieved based on the orthographic projections along differently orientated sampling planes. After hidden surface removal, the projected images are back-projected to the object domain and then conventional point cloud model is used for CGH synthesis.

The image-oriented approach is actually a sampling process, where the positional and/or view-dependent intensity information of a 3D object is sampled from different viewpoints as traced on the hologram plane. It is most likely that for each viewpoint, a different set of object points are sampled. As a consequence of that, the coherent wave contribution of a sampled point source is not fully represented on the entire hologram plane. Thus, although image-oriented approaches provide an efficient way of expressing view-dependent quality factors of the CGHs as discussed above, this sampling issue should be noted as their main drawback, which needs to be further addressed.

Acceleration methods. The high computational complexity of the point cloud model is its main drawback. Implementation of Equation (4) requires $\alpha P N_x N_y$ operations, where α is the number of the executed arithmetic operations for calculation of a cosine, square root, additions and multiplications and $N_x \times N_y$ is the size of the CGH [Shimobaba et al. 2009]. Since the representation of solid shapes needs extremely fine sampling, the computation time can be unacceptable in some cases. Therefore, development of accelerated computation has become an integral part of the implementation of this model. The pros and cons of the recently advanced methods for accelerated point-cloud based computation are thoroughly analyzed in [Tsang et al. 2018].

A straightforward way to accelerate hologram computation is to encode the view-dependent properties only in the horizontal direction resulting in so-called horizontal-parallax-only (HPO) holograms. By this way, an order of magnitude acceleration can be achieved [Lucente 1994; Plesniak et al. 2006]. In [Juárez-Pérez et al. 1997] the four-point symmetry of the Fresnel kernel is utilized to accelerate the computation by a factor of 4. In [Matsushima and Takai 2000] recurrence formulas are derived to accelerate computation of the depth-related phase.

An effective acceleration method is to compute the possible contributions of (unit-amplitude) point sources, i.e., separate terms in Equation (4), off-line and store them in a *look-up-table* (LUT). The composed LUT stores precalculated unit amplitude 2D fringe patterns, $T(x, y, x_p, y_p, z_p)$, $p = 1, 2, \ldots, P$ for all possible object points. In addition, the phases ϕ_p are set to zero for all points [Lucente 1993]. Computation of the CGH requires again $P N_x N_y$ loops but with only one addition and one multiplication for each loop. If the memory attributed to store one pixel of a fringe pattern corresponding to an object point is M_Δ, the memory requirement for the LUT storage is $M_\Delta P N_x N_y$, which is in the range of terabytes. Therefore, despite accelerated computation, the computational complexity remains to be high.

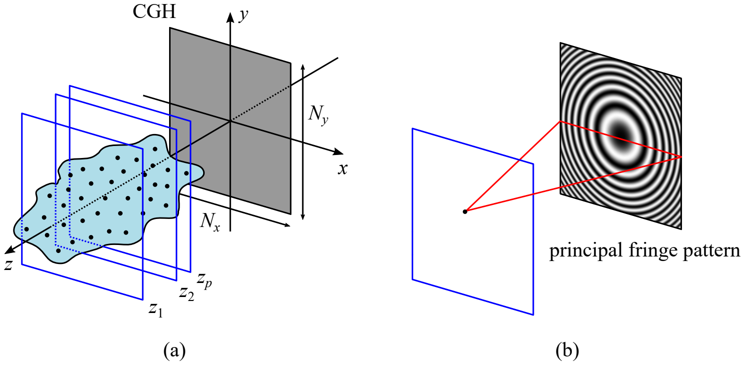


Fig. 4. Schematic representation of the NLUT method (a) and generation of the principal fringe pattern for a given depth (b).

A novel LUT (NLUT) method is proposed in [Kim and Kim 2008] with substantially reduced number of precalculated fringes. The method slices the object by using a set of planes located at different depths along the z-axis (see Figure 4). A plane at a given depth z_p contains a certain number of point light sources, but only the so-called unit amplitude principle 2D fringe pattern, $T(x, y, 0, 0, z_p)$, corresponding to the point source at the center of the plane is precalculated and stored. The fringe patterns of the other light sources on the plane are found by shifting the principle fringe in the hologram plane. The size of the principal fringe pattern is large enough to enable shifts for all possible points. The CGH is obtained by summation of all shifted fringes in all planes, and the memory size is reduced to gigabytes. The NLUT method is based on the shift-invariance of the fringe patterns for points (assigned to be) at the same depth. For holographic video, removing points that do not change in consecutive images and generating CGHs for the residual images is proposed in [Kim et al. 2008b]. Compression efficiency depends on the speed of the objects motion and the method becomes ineffective, if the residual image contains more than 50% of the points in the original images. Higher compression is achieved through evaluation of displacements of objects between two consecutive images along the x-axis and y-axis and then utilization of the estimated motion vectors for motion compensation [Kim et al. 2013]. The drawbacks of this approach are the need for segmentation of the 3D scene for accurate estimation of motion vectors and the decrease of accuracy of this estimation for faster moving objects. Complexity of the task increases for a 3D scene consisting of objects moving at different speeds. For improving the accuracy of motion estimation and compensation, an MPEG-based algorithm is developed in [Dong et al. 2014a] that removes temporal redundancy of the object data. As the data extraction relies not on estimation but on a mathematical model of the 3D scene, the algorithm shows excellent performance when there is more than 50% difference between the points in consecutive images. Because of the shift-invariance property in the lateral direction, the MPEG-based method is not effective for objects with large depth variation. To enable motion compensation in the z-direction, a thin lens property of the Fresnel zone plate is used in [Dong et al. 2014b] to achieve shift-invariance in the z-direction by multiplying two zone plates corresponding to different depths. The result is 3D motion compensation. Compression of the object data for encoding in a set of CGHs by various algorithms is a preprocessing step in NLUT CGH synthesis [Kwon et al. 2016]. For real-time synthesis of a set of CGHs, the composed LUTs need the usage of FPGAs or GPUs [Kwon et al. 2016]. This requires ensuring software and memory compatibility with the GPU structure, which is achieved by implementing an object tracking mask method. Further improvement of the NLUT method by decomposing the 2D principal fringe patterns into a set of one-dimensional (1D) sub-principal fringe patterns is proposed in [Kim et al. 2012]. Generation of a full-color CGH by using tunable NLUT is developed in [Kim et al. 2015].

The split LUT (SLUT) approach proposed in [Pan et al. 2009] builds separate LUTs for modulation factors along the x-axis and y-axis at a given depth. Under the assumption of a small size of the reconstructed image compared to the distance between the scene and the hologram, the contributions $T(x, y, x_p, y_p, z_p)$ of point sources in Equation (4) can be written as a multiplicatively separable function, i.e., as a product of two terms depending on $(x - x_p)$ and $(y - y_p)$, respectively. This allows decreasing the memory usage and computation time in comparison with the ordinary LUT approach. For a line along the y-axis with n point sources, the number of loops for the CGH synthesis decreases from nN_xN_y to $nN_y + N_xN_y$ [Pan et al. 2009]. The method is further evolved by introducing basic 1D modulation light factors based on the Fraunhofer approximation [Jia et al. 2013] and composing the so-called compressed LUT (CLUT). The SLUT and CLUT are used for accelerated computing of a gigabyte hologram by applying several GPUs controlled by a CPU and CUDA architecture with a dynamic parallelism in [Zhang et al. 2016b]. The authors admit that using a GPU is much more effective for processing floating type data. Although the SLUT method demonstrates substantial acceleration for regularly sampled objects such as planar images, it is not that effective for arbitrarily sampled objects and large size holograms. An improved LUT method for the case of arbitrarily sampled 3D objects is developed in [Wei et al. 2016], where a distance-dependent phase factor, $\exp(jkr_p)$, is introduced and the LUTs of precomputed phase values of successive slices of the 3D object are built. An acceleration method, which takes advantage of the concentric redundancy of the Fresnel zone plate created by a point light source on the CGH plane, is proposed in [Lee et al. 2016; Nishitsuji et al. 2012; Su et al. 2016]. The method calculates the distribution of the complex amplitude along a single radial line of a zone plate corresponding to a given depth, where the length of the line is limited based on the maximum diffraction angle. In [Nishitsuji et al. 2012] this is done by recurrent formulas derived from the relation of adjacent phases. Fast CG technique is developed for drawing a discrete circle in order to roll the calculated line and to form a zone plate. The zone plates are stored in a LUT. The stored plate for a given depth is translated correspondingly to positions of other point sources in the same plane. The developed CG algorithm requires random memory access and is prone to errors. This drawback is removed in [Nishitsuji et al. 2015]. In [Su et al. 2016] CG is also used to roll the complex amplitude distribution along a line around the center of the zone plate, whereas in [Lee et al. 2016] a point source CGH is calculated by linear interpolation. The effects of sampling along the radial line and interpolation on the quality of reconstruction are also studied in [Lee et al. 2016]. Numerical simulations prove efficiency of the approach. Further 10 to 20 times compression of the LUT corresponding to the radial symmetry interpolation method is proposed in [Jiao et al. 2017] by using principal component analysis.

In [Yamaguchi and Yoshikawa 2011], acceleration of computation is achieved via *image-plane holography*, where the location of the 3D object is near the hologram and a given point light source on the object surface contributes to a small spatial window on the CGH plane within the diffraction angle supported by the hologram (see Figure 5(a)). The computational complexity is given by $O(\alpha P(R/\Delta)^2)$, where R is the average radius of such windows [Shimobaba and Ito 2017]. The drawback of this method is that the reconstruction, which corresponds to an image-plane hologram, is located in the vicinity of the hologram plane. It is advanced in [Shimobaba et al. 2009] to reconstruct holograms in the Fresnel region by introducing a *wavefront recording plane* (WRP) close to the 3D objects (see Figure 5(b)). The complex amplitude of the wavefront due

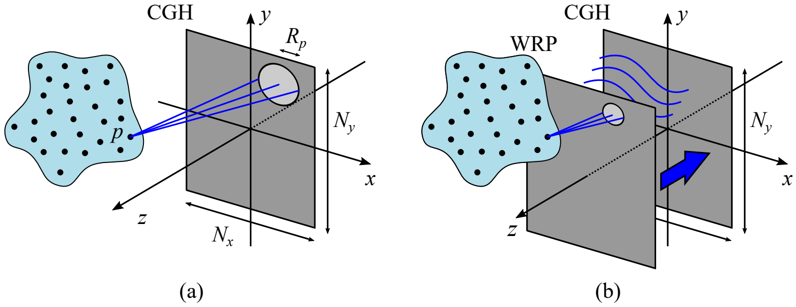


Fig. 5. Computation of an image plane CGH (a) and acceleration of computation by using a wavefront recording plane (b).

to point cloud is calculated at this plane. The wave field due to each point source cover a small area on the WRP (limited by the maximum diffraction angle) and this decreases the number of hologram sampling points taken into consideration. The complex amplitude at the CGH plane is found by Fresnel transform of the complex amplitude on the WRP. For $N_x = N_y$, the computational complexity is evaluated to be $O(\alpha P(R/\Delta)^2) + O(\beta N_x^2 \log N_x)$, where β is the number of arithmetic operations in fast Fourier transform (FFT). For large P the first term dominates. The LUT and WRP methods are combined in [Shimobaba et al. 2010]. The LUT is applied for calculation of the complex amplitude on the WRP plane. Additional acceleration is achieved by using a GPU for propagation from the WRP to the plane of the hologram. Real-time synthesis of a large hologram by GPU implementation of the WRP method is reported in [Weng et al. 2012]. Using multiple WRP planes is proposed in [Phan et al. 2014] for reconstruction of deep scenes. Change of intensity at a single object point affects the whole hologram. Real-time relightening of the reconstructed image without CGH recalculation is proposed in [Tsang et al. 2012] by applying intensity modulation to the WRP. The GPU implementation of this method provides generation of 2K×2K hologram for less than 20 ms. Generation of the WRP directly from a planar intensity image with a Fresnel kernel is proposed in [Tsang and Poon 2015]. The depth related resampling of the WRP encodes the 3D information. The method enables generation of large holograms (2K×2K pixels) of dense objects with rich texture at 100 frames per second. Two WRPs located in front of the objects are used. In [Okada et al. 2014], a discrete set of depths is used with introduction of several WRPs across the point cloud. Backward and forward propagation are utilized. In [Symeonidou et al. 2015], a set of multiple WRPs that slice the point cloud is introduced. This allows choosing the closest WRP for a given point source and to minimize the corresponding spatial support on the WRP for that source. The backward and forward propagation kernels are stored in LUTs for a discrete set of depths. Gaussian interpolation in WRPs blurs the discrete points to simulate propagation of a smooth wavefront surface to the CGH plane. Inverse Gaussian filters mask the occluded points. In [Symeonidou et al. 2018], a Phong illumination model is incorporated in the precomputed LUTs for creation of photorealistic reconstruction. At fixed number of WRPs with predetermined distances between them, the computation time varies depending on the distribution of points in the point cloud. Automatic optimization of the number and locations of WRPs is proposed in [Hasegawa et al. 2017]. A wavelet shrinkage method, which is named as WASABI, is proposed and applied in [Arai et al. 2017; Shimobaba and Ito 2017] to represent the complex amplitudes with a few wavelet coefficients for faster computation [Gilles et al. 2016]. Additional acceleration of computation on GPU with a factor of 30 over conventional approach and improvement of quality of reconstruction compared to WASABI method is reported in [Blinder and Schelkens 2018] by using a sparse basis representation in the short-time Fourier space. Analysis of the WRP implementation with a single plane or multiple planes is made in [Tsang and Poon 2016].

The so-called *phase-added stereogram* (PAS) approach, first proposed in [Yamaguchi et al. 1993], accelerates the computation of a CGH by sampling the directional information through approximation of the spherical wavefronts of the point sources with a set of planar wavefront patches. The first step in PAS computation is partitioning of the hologram into $M \times N$ equal size square segments, i.e., holographic elements (hogels), with $S \times S$ pixels. The contribution of point p from the point cloud to the segment (m, n), $m = 1, 2, \dots, M; n = 1, 2, \dots, N$ with a central point at (x_c^{mn}, y_c^{mn}) is given by the 2D complex sinusoid of $(A_p/r_p)\exp(j\Phi_p^{mn})\exp[j2\pi(u_p^{mn}x' + v_p^{mn}y')]$, where (x', y') are the local spatial coordinates defined with respect to (x_c^{mn}, y_c^{mn}). The spatial frequencies (u_p^{mn}, v_p^{mn}) of this sinusoid are constant across the segment at a given wavelength and are found with respect to the segment central point. The phase of the sinusoid, $\Phi_p^{mn} = \phi_p + kr_p^{mn}$, is a sum of the phase ϕ_p of the wave due to the point p and a phase related to the distance r_p^{mn} between the point p and the central segment point. The distance related phase is added to match the wavefronts of the plane waves diffracted from all segments. The segment size, $S\Delta \times S\Delta$, where Δ is the pixel size at the hologram plane, should be small enough to well approximate the spherical wavefront emitted by the point p with a plane wave across the segment. The fringe pattern across a segment due to all object points is a superposition of 2D complex sinusoids and can be calculated by a single inverse Fourier transform of the spectrum of the segment built by placing the complex amplitudes of the sinusoids to their locations in the spatial frequency domain. Accelerated computation of the hologram requires applying inverse FFT at the expense of quantization in the frequencies of complex sinusoids. Especially for small size segments, the quantization of frequencies can strongly reduce the reconstruction quality due to resulting inaccurate light steering. Over the years, different improvements of the PAS method have been proposed. The first improvement is the compensated phase-added stereogram (CPAS) algorithm [Kang et al. 2007], which improves the steering by phase compensation introduced to all sinusoids. The compensation contains the difference between the spatial frequencies in the continuous and the discrete domains. The CPAS provides better reconstruction in comparison to that of the PAS with FFT practically without increase of the calculation time. To satisfy the controversial requirements set on the segment size, the accurate PAS (APAS) is developed in [Kang et al. 2008a]. The idea is to calculate the FFT in a larger area than the segment and to increase resolution in the sampled spatial frequency domain. The hologram segment is obtained by truncation. The later accurate compensated PAS (ACPAS) combines both improvements by merging phase compensation and directional error reduction into a single step [Kang et al. 2008b] and produces higher quality reconstructions compared to the previous versions. The fast PAS (FPAS) proposed in [Kang et al. 2016a], applies better phase compensation to ACPAS, where the point p contributes to the segment (m, n) by $(A_p/r_p)\exp\left(j\Phi_p^{mn}\right)\exp\left[j2\pi\left(u_p^{mn}x' + v_p^{mn}y'\right)\right]\exp\left\{j2\pi\left[u_p^{mn}\left(x_c^{mn} - x_p\right) + v_p^{mn}\left(y_c^{mn} - y_p\right)\right]\right\}$. Since the FPAS demonstrates finer beam steering than the other PAS approximations, it provides higher peak intensity and larger peak signal-to-noise ratio in calculating the Fresnel zone plates corresponding to point sources. The recent work overlapp-add stereogram (OLAS) [Padmanaban et al. 2019] uses similar phase compensation to PAS, but unlike in all abovementioned methods, the segments are chosen to be overlapping with single pixel shift. By this way depth-independent high resolution reconstructions are achieved, which is advantageous for deep 3D scenes where the optimal hogel size of conventional PAS methods usually depends on depth. The phase compensated stereogram methods combine well with the ray casting approach, and thus, they allow implementation of visibility tests for handling occlusions.

3.2 Polygon-based model

General description. The other popular model in CGH computation is polygonal representation of 3D objects. Representation of object surface with light sources of planar primitives is the basis for polygon-based models [Leseberg and Frère 1988; Tommasi and Bianco 1993]. In the polygon-based 3D model, the non-planar object is given by a mesh of P light sources with a polygonal shape (e.g., triangle). Each polygon represents an aperture that becomes a source of the object wave field and the wave fields from all polygons are superposed at the plane of the hologram; that is why this method is characterized as a field-oriented approach in [Matsushima and Nakahara 2009].

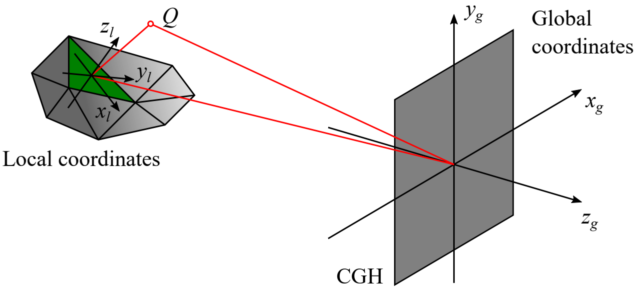


Fig. 6. Schematic representation of the polygon-based CGH computation approach, where the hologram plane is placed at $z_g = 0$ with respect to the global coordinate system and a local coordinate system is utilized on each polygon.

Over the last fifteen years two types of polygon-based CGH algorithms have emerged that are classified as numerical and analytical. In both approaches, a global coordinate system (x_g, y_g, z_g) is introduced with the hologram located in the $(x_g, y_g, z_g = 0)$ plane (see Figure 6). The object field at the hologram plane is found by summing up the wave fields produced by all P polygons:

$$O(x_g, y_g, z_g = 0) = \sum_{p=1}^{P} U^p(x_g, y_g, z_g = 0). \tag{5}$$

In both approaches the polygon fields, $U^p(x_g, y_g, z_g = 0)$, $p = 1, 2, \ldots, P$, are calculated by the angular spectrum method (ASM), see Equation (A.4) in Appendix A, but the angular spectra of the polygons are computed in a different manner. For this purpose, a local coordinate system (x_l, y_l, z_l) is defined on the given polygon with a z_l-axis oriented along its normal and the polygon located at $(x_l, y_l, z_l = 0)$ plane. This system is unique for each polygon. As the polygon plane is tilted with respect to the hologram plane, the formulas describing propagation of a wave field between parallel planes are not directly applicable to the synthesis of a CGH. Rotational transformation is required to relate the plane of a polygon to a plane parallel to the hologram [Matsushima 2008; Matsushima et al. 2003].

Considering an arbitrary point Q as shown Figure 6, the relation between its global and local position vectors $\vec{r}_g$ and $\vec{r}_l$, respectively, can be defined through rotation and translation as [Park 2017]

$$\vec{r}_l = \mathbf{R}\vec{r}_g + \vec{t};\ \vec{r}_l = \begin{bmatrix} x_l \\ y_l \\ z_l \end{bmatrix},\ \vec{r}_g = \begin{bmatrix} x_g \\ y_g \\ z_g \end{bmatrix},\ \vec{t} = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix},\ \mathbf{R} = \begin{bmatrix} \vec{n}_{x_l}^T \\ \vec{n}_{y_l}^T \\ \vec{n}_{z_l}^T \end{bmatrix}, \tag{6}$$

where $\mathbf{R}$ is a 3×3 rotation matrix with rows denoting the representation of unit vectors of the local coordinate in the global system and $\vec{t}$ is a translation vector. Please note that the polygon index p is omitted in the formulas for convenience. The relation between the angular spectra of the wave field on the planes $(x_g, y_g, z_g = 0)$ and $(x_l, y_l, z_l = 0)$, i.e., $G_g(u_g, v_g)$ and $G_l(u_l, v_l)$, respectively, can be obtained as [Park 2017]

$$G_g(u_g, v_g) = \frac{w_l}{w_g} G_l(u_l, v_l) \exp[j2\pi(u_l t_x + v_l t_y + w_l t_z)], \tag{7}$$

where $w_l = \sqrt{(1/\lambda)^2 - u_l^2 - v_l^2}$ and $w_g = \sqrt{(1/\lambda)^2 - u_g^2 - v_g^2}$. The Fourier transforms of the angular spectra of all polygons then give the corresponding object fields $U^p(x_g, y_g, z_g = 0)$, $p = 1, 2, \ldots, P$.

Numerical approach. In the numerical algorithm, a complex function called a property or surface function describes each polygon. The amplitude of this function gives the polygon shape and texture. The phase distribution describes the reflection model, e.g., a random phase can be introduced on each surface function [Matsushima 2005a] to create diffuseness and hence to ensure a wide viewing angle for the polygon. Effective numerical implementation of ASM requires double FFT. The first FFT is executed on a regular sampling grid in the local coordinate system to find the angular spectrum of each surface function. Usage of FFT entails remapping of Fourier coefficients for each polygon when going from the tilted polygon to the global coordinate system. This is due to the rotational transformation, which introduces a shift of the spectrum in the Fourier domain corresponding to the rotated coordinate system and may require interpolation to ensure regular sampling grid in the spatial frequency domain after rotation [Matsushima et al. 2003]. This grid is unique for each polygon, because it is composed according to the geometry of the polygon with respect to the hologram plane. Compensation of the shift in order to place the spectrum into the origin of the rotated system is equivalent to forcing the emitted light to propagate in the direction of z_g-axis. This procedure combined with resampling and bilinear interpolation in the Fourier space is called remapping. Computing the contribution from a single polygon field takes more time than computing a spherical wave from a point source. Nevertheless, the polygonal model offers substantially faster CGH synthesis due to much smaller number of polygons required to represent the object in comparison to the number of point sources in the point cloud model. Therefore, the method is appropriate for synthesis of CGHs of large objects. Remapping is the most computationally extensive step in this FFT-based approach. According to the estimation in [Matsushima 2006], it can take up to 44% of the CPU time. Parallel computation with advanced computing hardware can accelerate the numerical algorithm [Ahrenberg et al. 2006]. Numerical propagation of the wave fields requires storing the complex functions as 2D arrays with the size of the hologram, and hence, restricts the number of pixels in the CGH. Solution to this problem is proposed in [Matsushima and Nakahara 2009] by partitioning the hologram into segments.

Numerical implementation of the polygon-based approach is highly suitable for CGHs providing photorealistic reconstruction. A complex function expressing brightness, surface roughness and illumination conditions characterizes each polygon. Including the object surface properties in the algorithm does not slow down the hologram computation [Lee et al. 2014]. A hidden-surface removal algorithm is demonstrated in [Matsushima 2005b] where a brightness model of the planar surfaces is proposed for introducing desired shading. The idea is based on using a silhouette mask. The method imitates the phenomenon of seeing the light coming from the front surface of the object, $O(x, y)$, and the background field that is not blocked by the object. The latter is found by using a binary mask, $M(x, y)$, which is zero inside the silhouette of the object and unity otherwise. The plane (x, y) coincides with the maximal cross-section of the object for a given viewing direction. Then, for a background wave field, $b(x, y)$, the observed field is $M(x, y)b(x, y) + O(x, y)$. This field is propagated to the hologram plane for the given viewing direction. For multiple objects, the masking procedure is applied to each object and is called object-by-object shielding. It requires low computational costs and is highly appropriate for processing of mutual occlusions. However, this procedure fails in treating self-occlusions when some parts of an object shield its other parts. To avoid occlusion errors in the case of concave surfaces, the silhouette method is applied to polygons themselves and so is transformed into polygon-by-polygon light shielding. The latter is robust, but is very time-consuming due to the requirement of propagating the wave fields given by $M(x, y)b(x, y) + O(x, y)$ for each polygon, where the surface area of polygons are much smaller than the area occupied by the whole wavefront. Acceleration of calculation for the polygon-by-polygon method is proposed in [Matsushima et al. 2014] by developing a switch-back technique that considers each polygon as a silhouette-shaped aperture.

The polygon-based model facilitates modeling of reflectance distributions by encoding them into the phase of the property function. As in CG, the ratio between the specular and diffuse reflection varies for each object material. Different CG reflection models such as the Phong reflection model, the Cook-Torrance reflection model and others are applied. A method based on Blinn and Torrance-Sparrow reflection models in CG is proposed in [Yamaguchi and Sakamoto 2009] to express reflection distributions, which correspond to arbitrary illuminations, and to include background reflections, which occur when metallic or mirror surfaces are present in the object. This study continues in [Yamaguchi et al. 2011b] by analysis of how the surface roughness affects the reflection distributions. An accurate but time-consuming description of reflectance distributions in CGHs is provided in [Ichikawa et al. 2011] by using a finite difference time-domain method. The Phong reflection model is used in [Nishi et al. 2011] with emphasis on fast computation for synthesizing a high-definition CGH. Because the bandwidth of the spectrum for a specular surface is narrower than the spectrum for a diffusely reflecting surface, the surface function of each polygon is modified in accordance with the Phong model to render specular surfaces. The holograms for reflected images are computed and applied as property functions in [Cho et al. 2015]. Methods for reduction of phase mismatches on the boundaries in the polygonal mesh are also proposed [Im et al. 2015; Liu et al. 2010]. To remove angular appearance of the object surface during reconstruction, a smooth shading is proposed in [Yamaguchi et al. 2011a] by expressing a patch as a curved surface. A CGH of size 8K × 4K is calculated in [Yamamoto et al. 2010] by directly applying CG rendering techniques in case of illumination by a point light source and ambient light.

Analytical approach. Analytical computation of the polygons' spectra can strongly accelerate the polygon-based method. In case of a triangular mesh, the angular spectrum of an arbitrary size and orientation triangle is related to the angular spectrum of a unit-amplitude reference triangle through affine transformation [Ahrenberg et al. 2008; Kim et al. 2008a]. The 3D object is a mesh of triangles and the angular spectra of the wave fields coming from them are computed without performing FFTs in the local coordinate systems. All computations are performed analytically in the continuous domain in the global coordinate system. A uniform sampling grid is used only in the hologram plane, where a single FFT is executed to transform the sum of the angular spectra from the triangular mesh to the object wave field at the hologram plane. No remapping is required in the analytical method, which makes it potentially very fast for CGH synthesis. The fully analytical approach with flat amplitude distributions is derived in [Ahrenberg et al. 2008] by using a right triangle as a reference. A carrier plane wave illuminates each triangle and it reconstructs a wave field with uniform amplitude depending on the illumination direction and the normal vector to its surface. Accuracy of formulas derived in [Ahrenberg et al. 2008] is analyzed and improved in [Pan et al. 2013; Zhang et al. 2013, 2018]. Practically concurrently with Ahrenberg *et.al.*, a semi-analytic model of a wave field emitted by a 3D triangular mesh is introduced in [Kim et al. 2008a]. The angular spectrum of an arbitrary triangle is calculated at the direction of the illumination given by a carrier plane wave through representing the triangle as a sum of two right triangles. Surface diffuseness is also included in the model by dividing each triangle in the local coordinate system into a set of $m(m-1)/2$ similar small triangles with different amplitudes and phases of the complex wave field, where m is an integer number. The angular spectrum of each triangle is found as a sum of the spectra of the wave fields emitted by this set of $m(m-1)/2$ smaller triangles, exploiting their similarity for acceleration of computation. A ray-tracing approach is applied for searching hidden triangles to express the occlusion effect. The final CGH is composed as an angular spectrum CGH and the FFT at the hologram plane is avoided by using a Fourier transform optical system for reconstruction. A fully analytical approach is proposed in [Liu et al. 2010], which expresses directly the complex amplitude at the hologram plane through the Fourier spectra of the wave fields emitted by the separate triangles in a triangular mesh model. Under certain approximations, the Fourier spectra relate the hologram plane to a specially devised frequency domain. They are found analytically by using a reference right triangle and the CGH is computed without FFT. A phase adjustment is made to avoid visualizing the edges of the triangular mesh in reconstruction due to the difference in the flat amplitude distributions in the triangles with abutting boundaries. Expressing the light beam emitted by an arbitrary triangle through transformation of a precalculated object beam emitted by a basic triangle is described in [Hosoyachi et al. 2013]. Calculation of the CGH there relies only on transformations in the spatial domain. Recording the basic beam on a plane requires a lot of memory due to the high density of fringes far away from the center of the plane. Its recording on a spherical surface is proposed to decrease the memory [Hosoyachi et al. 2013].

The issue of the expressiveness for the analytical or semi-analytical polygon-based approach is not as straightforward as for its numerical counterpart. First, analytical calculation of the angular spectra entails flat shading of the triangles in the 3D mesh and inevitable visualization of the mesh edges at reconstruction. Solution to this problem is proposed in [Park et al. 2015] by introducing a spatially varying amplitude in each triangle. Three different amplitude values are assigned to the vertices of a given triangle in accordance with the illumination direction and the normal vectors of the vertices that are provided by the used software for modeling the 3D object or through averaging the normal vectors to the plane of the neighboring triangles. The amplitude inside the triangle is found by an interpolation formula. Introduction of texture mapping to the fully analytical method is made in [Lim et al. 2013] by expanding the surface function inside the triangle into a Fourier series. The angular spectrum of the textured triangle is found as a weighted sum of analytically computed angular spectra with Fourier series coefficients as weights. The drawback is the significant amount of increase in computation time due to the requirement of taking many coefficients for better expressiveness. Holographic reconstruction from the textured polygon-based CGH is analyzed in [Lee et al. 2014] for improved semi-analytic approach. This is done by expressing the texture function as a Fourier series and composing a shift-invariant form for the textured angular spectrum at the hologram plane through some mathematical approximation of the formula for the angular spectrum of the unit amplitude triangle. This allows applying convolution and acceleration of computations. The drawback is reconstruction quality degradation for triangles highly inclined with respect to the hologram plane because of approximations [Lee et al. 2014]. The method is tested for the semi-analytic approach developed in [Kim et al. 2008a]. The semi-analytic algorithm for synthesis of the amplitude spectrum CGH for adaptive view direction change is further developed in [Cho et al. 2012]. Encoding of arbitrary angular reflectance distribution for the fully analytical approach is proposed in [Yeom and Park 2016] by relying on the fact that the angular spectra of a triangle in the global and local coordinate systems depend on the carrier wave or illumination direction. This is used to accumulate angular spectra of one and the same unit amplitude triangle corresponding to different carrier waves. Accumulation is based on a narrow diffraction angle around each carrier wave. Each angular spectrum is accumulated at a given diffracted intensity and a phase bias to compose the desired reflectance distribution. It is proved that it is possible to introduce an approximated convolution approach, which avoids unnecessary separate computation of many spectra. In [Ji et al. 2016], a texture mapping method is developed without Fourier series expansion of the surface function, where each triangle is adaptively divided into smaller triangles with uniform amplitudes according to the texture mapping function. Similarity relations facilitate fast calculation of the angular spectra of the small triangles. Occlusion processing in the angular spectrum frequency domain for the fully analytical method is proposed in [Askari et al. 2017]. The occluded parts of the triangles are found by applying convolution in the hologram plane between the global angular spectrum of the rear triangles and the current triangle. The main advantage of the method is the convolution operation performed in the hologram plane. The result is blocking of the light waves from the rear triangles in the tilted planes of these triangles contrary to the silhouette mask approach that stops the light waves from the planes parallel to the hologram. This makes the method free from oblique angle artifacts [Askari et al. 2017].

3.3 Layer-based representation

The CGH synthesis can be highly accelerated by arranging the object data on planes. Such idea is implemented in [Bayraktar and Özcan 2010] by slicing a 3D object with a set of planar layers at equidistant depths parallel to the hologram plane. Each layer contains non-zero data for the visible object part and zeros for the invisible part (see Figure 7). The CGH is a sum of contributions from all layers obtained by using the Fresnel diffraction formula, convolution approach or ASM [Bayraktar and Özcan 2010; Trester 2000; Zhao et al. 2015]. CGH synthesis based on multiple fractional Fourier transforms is also proposed in [Zheng et al. 2009]. To obtain reconstruction quality close to that of the point cloud model, the number of layers should be rather large. Although the layer-based method has less computational complexity and operates with less amount of input data, its basic form can be applied only to diffusive Lambertian surfaces and provides a very narrow viewing zone around the viewing angle normal to the parallel layers.

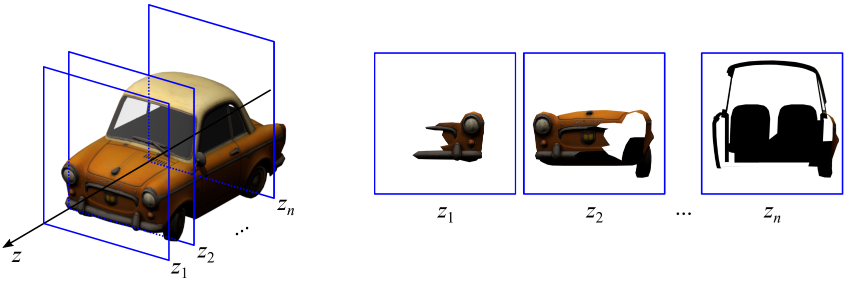


Fig. 7. Schematic representation of a layer-based CGH synthesis by cutting the 3D object with a set of planes parallel to the hologram plane and taking contribution from the visible parts of the object in each plane for a given viewing point.

To overcome the limitations in the basic implementation of the layer-based model, a multi-view layer-based CGH synthesis is proposed in [Chen et al. 2014]. The model is applied to the point cloud of a 3D object and a multiview rendering of the object is performed for fast computation of a full parallax CGH with occlusion and view-dependent shading. Accordingly, the point cloud is sliced with different set of layers for each viewing point. Angular tiling is used to display holograms calculated for the different viewpoints. The backward-forward propagation with a Ping-Pong algorithm [Dorsch et al. 1994] is used to deal with occlusions. In the first version of the developed algorithm, each considered point is projected to the nearest layer. For each layer, the diffraction pattern is calculated, and the patterns obtained for all layers are summed up in the plane of the CGH. Optical reconstruction of points only on the layers creates a depth error for the objects points outside layers. This error is not expected to cause a serious degradation in the perceived reconstructed image, if the distance between consecutive layers is not more than the depth discrimination threshold of the HVS. An improved version of the method is proposed in [Chen and Chu 2015] by projecting the object point to the two closest layers located behind and in front of it with amplitude values determined by depth from defocus fused 3D method. Thus, the depth error is eliminated for observation direction coinciding with alignment direction of the layers, and the computation is four times faster [Chen and Chu 2015]. Composing a CGH from elemental holograms, which are calculated by the layer-based approach, is proposed in [Zhang et al. 2016a]. A separate set of planar layers is built within the viewing frustum of each elemental hologram. The so-called classification method for rapid sorting of points in a point cloud into the layers is introduced in [Su et al. 2016]. The sorting is made with an occlusion mask; the mask being filled with the points for the first layer is updated by removing the occluded points in the next layer and by adding all new points. A layer-based algorithm with rendering for a single viewpoint is developed in [Zhang et al. 2017] for synthesis of a 3D full parallax CGH with occlusion effects. A slab-based orthographic projection is used for slicing the 3D object. These projections are produced also between adjacent layers to create data for occlusion processing by performing silhouette mask culling for each layer. The complex wave field on each layer is propagated to hologram plane via ASM. Angular spectrum layer-based algorithm is used in [Zhao et al. 2016] to generate CGHs for layers at different depths. The CGHs are combined into groups to be exposed on the SLM by time-division multiplexing to increase the space-bandwidth product of reconstruction. In [Gilles et al. 2016], after slicing the 3D scene into layers parallel to the plane of the hologram, light is propagated from one layer to another. Shielding is performed using a point cloud approach and a threshold criterion to determine the number of points. The final CGH is obtained through propagation of the complex field on the layer nearest to the hologram. No visibility test is required for this technique to process occlusions. Fast computation of a CGH from a layer-based based model by using sparse FFT for calculation of diffraction is proposed in [Kim and Ro 2017]. Further development is proposed in [Jia et al. 2018] for the case of layers having a lot of zero-valued pixels due to occlusion effects. The developed two-step algorithm takes into consideration of only non-zero values and applies a sub-sparse 2D FFT calculation through performing two one-dimensional FFTs.

4 RAY-BASED CGH METHODS

Unlike wave-based techniques, the ray-based CGHs do not require knowledge about the geometric description of the scene. Instead, they rely solely on the captured intensity images of the scene under incoherent (white) light illumination. As incoherent holography approaches, the ray-based methods constitute an important category of CGH by enabling generation of holograms of still or dynamic real life scenes without requiring strict coherent illumination conditions and complex optical setups that should be immune to vibrations. Incoherent holography has also inspired some approaches in digital holography, such as optical scanning holography [T.-C. Poon et al. 1996] and Fresnel incoherent correlation holography [Rosen and Brooker 2007]. These methods utilize self-interference based and scanning based structured illumination techniques, respectively, which enable speckle free reconstructions [Liu et al. 2018]. Here we consider two categories of ray-based CGH methods, namely, HS and MVP holography, as incoherent holography techniques that utilize incoherently captured intensity images of the scenes in encoding the corresponding holograms.

4.1 Holographic stereogram

HSs can be either recorded optically by means of interference or calculated numerically. In both cases, the main ingredient is a set of multi-perspective images that define the information on the hologram in hogel-by-hogel basis. Hogels can be interpreted as segments that are usually placed on a uniform rectangular grid and form the complete hologram when tiled. Historically, first optical implementations of the HS technique goes back to late 1960s [McCrickerd and George 1968]. In early 1970s, a set of perspective images calculated by a computer were used to record a HPO hologram by moving a vertical slit mask [King et al. 1970], i.e., the hogel for this HPO arrangement is a vertical segment that vertically covers the entire hologram. Yatagai first proposed computer synthesis of a HS in 1974 [Yatagai 1974], where each hogel was obtained through Fourier transform of a perspective projection of the 3D object. Verification of the concept was done for an object represented as a collection of many small flat surfaces and for a HPO hologram.

The set of multi-perspective images used in HS calculation corresponds to a set of rays that can be parametrized as a light field (LF). The ray optics formalism of the LF represents the light as a collection of rays. That is, at a given time, for a given wavelength, taking only the rays propagating in free space to, e.g., $+z$ direction, any ray can be parameterized by the four-dimensional (4D) LF $L(x, y, u, v)$ using the crossing points of rays on two parallel planes (x, y) and (u, v), respectively [Levoy 2006]. This so-called two-plane parametrization is depicted in Figure 8(a).

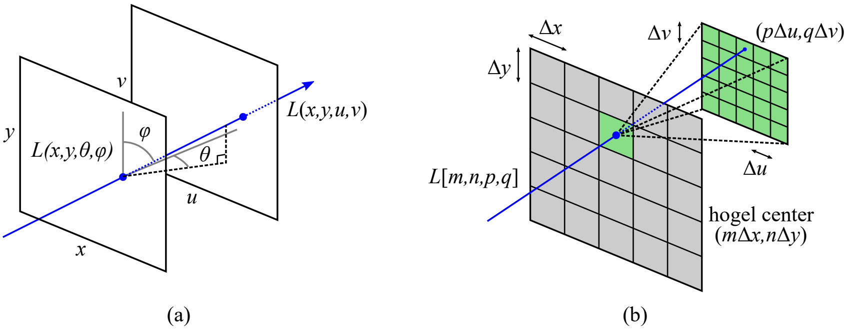


Fig. 8. 4D parametrization of the light field (a) and schematic representation of HS (b).

The two parallel planes can be directly linked to multi-perspective image capture setup by assigning (x, y) as the camera view plane and (u, v) as the image plane of the camera. The LF can be also equivalently represented by a space-angle parametrization $L(x, y, \theta, \phi)$ where θ and ϕ denote the propagation directions of the rays with respect to x and y axes, respectively. Thus, in either case, the (x, y) plane can be treated as the ray-sampling plane (RSP). The capture process involves two recording schemes. The first is recording of perspective projections of the scene from each spatial sampling point on the RSP with a setup consisting of usually a high number of low-resolution cameras. The second is recording of usually a smaller number of images with orthographic projection, i.e., capturing a set of parallel rays each of which corresponds to a different direction in the LF space-angle parametrization.

Let us consider the case illustrated in Figure 8(b), where the RSP coincides with the hologram plane and the HS is divided into a regular grid of hogels of size $\Delta x \times \Delta y$ with abutting boundaries. The corresponding captured discrete LF is denoted as $L[m, n, p, q] = L(m\Delta x, n\Delta y, p\Delta u, q\Delta v)$, where Δu and Δv are the pixel pitches of the captured intensity images. The HS then encodes the intensity and directional information in the complex amplitude of the object field as

$$O_{HS}(x,y)=\sum_{m=1}^{M}\sum_{n=1}^{N}\text{rect}\left(\frac{x-m\Delta x}{\Delta x}\right)\text{rect}\left(\frac{y-n\Delta y}{\Delta y}\right)$$
$$\sum_{p=1}^{P}\sum_{q=1}^{Q}\sqrt{L[m,n,p,q]}\exp[j2\pi(f_x^{mnpq}x+f_y^{mnpq}y)], \tag{8}$$

where $(m\Delta x, n\Delta y)$ are the coordinates of the center of the hogel $[m,n]$ and (f_x^{mnpq}, f_y^{mnpq}) are the spatial frequency components for the ray $[m,n,p,q]$. According to the grating equation, for a reference plane wave incident normally to the plane of the hologram, the spatial frequencies for the ray $[m,n,p,q]$ are related to their angles of incidence with respect to (x,y) plane as

$$f_x^{mnpq}=\frac{\sin\theta_x^{mnpq}}{\lambda},\ f_y^{mnpq}=\frac{\sin\theta_y^{mnpq}}{\lambda} \tag{9}$$

As seen in Equation (8) and Equation (9), the diffractive properties of each hogel are determined by the encoded set of basis sinusoidal fringes with different amplitudes that reconstruct a set of plane waves propagating along a certain view angle in accordance with the spatial frequencies of the fringes [Halle et al. 1991; Lucente 1993, 1994]. The complex amplitude can, thus, be obtained by applying inverse Fourier transform to the captured images, which encodes intensity and directional information in the hogel. Such encoding scheme enables the calculation of CGH via FFT techniques, and thus, it significantly reduces the computation time of the CGH compared to wavefront-based approaches. However, this improvement in computation time comes at the expense of degradation in quality of the reconstructed images in comparison with the wavefront-based CGHs. The HS encoding corresponds to an approximation of the spherical wavefront due to a point source in the scene via a set of discontinuous planar patches, resulting in hogels with single complex exponential components that are unmatched in phase.

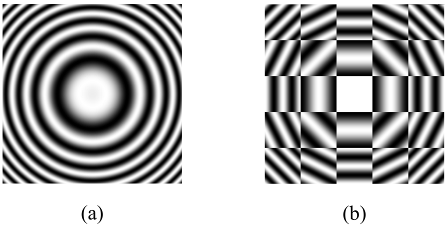

(a) (b)

Fig. 9. Real part of Fresnel diffraction kernel (a) and corresponding HS (b).

Figure 9 illustrates how HS approximates the Fresnel diffraction kernel for a point source. As mentioned in Section 3, PAS methods also approximate the wavefronts by using plane wave patches to take advantage of FFT in CGH calculation. In addition to that, however, the depth information of the scene is also used in those methods to match the phases of planar patches. This results in more accurate wave representation, and thus, higher quality reconstructed images. On the other hand, the HS approach is similar to image-order type of wavefront-based methods in the way of acquisition the scene data by capturing images of it. The CG rendering techniques are, thus, highly appropriate to include hidden surface removal, shading, reflections, texture, glossiness, mutual and self-occlusions in the encoded data for photorealistic reconstruction [Verma and Walia 2010].

The LF sampling setup in Figure 8(b) is impractical for real life scenes. It restricts the scene to be on either side of the hologram plane. The discrete LF $L[m,n,p,q]$ used for HS generation,

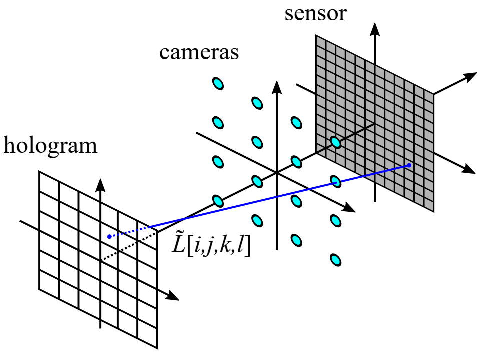


Fig. 10. LF sampling away from hologram plane by a set of cameras.

as in Equation (8), can be also acquired by capturing a set of 2D images at a certain distance from the hologram plane. As shown in Figure 10, the captured discrete LF can be parametrized as $\tilde{L}[i,j,k,l] = \tilde{L}(i\Delta\tilde{x}, j\Delta\tilde{y}, k\Delta\tilde{u}, l\Delta\tilde{v})$ by using another pair of parallel planes $(\tilde{x},\tilde{y})$ and $(\tilde{u},\tilde{v})$ with the former being the actual ray sampling plane on which the cameras are put. Each camera now captures different directional information for a given hogel and the pixels of each camera correspond to hogels, i.e., they sample the spatial information on the hologram plane. As usually there is no straightforward one-to-one correspondence between the samples of $L[m,n,p,q]$ and $\tilde{L}[i,j,k,l]$, it is necessary to resample $\tilde{L}[i,j,k,l]$ to obtain the desired ray samples $L[m,n,p,q]$ for CGH generation. By choosing the appropriate LF sampling parameters for the given scene, i.e., satisfying the prerequisites of the densely sampled LF, any ray is accurately obtained from the samples $\tilde{L}[i,j,k,l]$ via linear interpolation [Lin and Shum 2004]. This limits the disparity between adjacent views during capture such that it is within $[-1,1]$ pixels with respect to hologram plane, thus restricting the distance between adjacent cameras and the boundaries of the captured scene. Such a framework ensures accurate resampling of $L[m,n,p,q]$ from $\tilde{L}[i,j,k,l]$.

A CGH generated as a HS possesses the main drawbacks of ray-based 3D light field displays. The image quality perceived from the reconstructed hologram is mainly determined by the spatio-angular resolution. The spatial sampling gives the perceived image resolution, whereas the angular sampling affects view-dependent image aspects such as motion parallax, specular reflection, occlusions, etc. The characteristics of the modulation transfer function allow analysis of optimum sampling of HS [Hilaire 1994]. The joint spatio-angular resolution of HS is dictated by the hogel size and it is subject to a trade-off due to the uncertainty principle inherent to diffraction. For improved spatial resolution, one needs to decrease the hogel size, which in turn degrades the angular resolution and vice versa. Besides that, to deliver the available angular resolution, the number of pixels within the hogels should be also sufficiently high. In practice, the characteristic of the HVS plays a key role in choosing the sampling parameters [Lucente 1994]. The lateral visual acuity of the HVS, when observing an object at distance d is given by

$$\Delta x^{HVS} = \frac{1.22\lambda d}{D_{eye}}, \tag{10}$$

where D_{eye} is the aperture size of the eye pupil [Goodman 1996]. That is, given an intended observation distance of the HS, the hogel size can be chosen to match with the HVS acuity. On the other hand, the pupil size of the eye puts an upper limit on the perceivable angular resolution as

$$\Delta\theta^{HVS} = 2\tan^{-1}\left(\frac{D_{eye}}{2d}\right), \tag{11}$$

which then can be also utilized to determine angular sampling (or spectral sampling of the fringes).

The HVS together with the depth range and the scene data impose strict requirements on the LF sampling. The LF capture for HS is usually required to be very dense with a high number of images at small parallax between neighboring views. The capture of real scenes needs a scanning camera positioned accurately at the viewpoints [Nikolskij et al. 2012]. Relieving the constraints can ease the capture process for HSs and enable using of multicamera systems instead of scanning camera rigs. The view synthesis and LF reconstruction algorithms are directly applicable in this context. In [Ohsawa et al. 2013], the amount of projections is reduced by using scene depth via shape-from-silhouette technique for a voxel-based model of the 3D object. A depth-corrected LF rendering is used in [Jurik et al. 2012] with a practical real scene capture setup. Reducing the number of captured images in [Hayashi et al. 2011; Kinoshita and Sakamoto 2009] involves ray interpolation by a distance transformation and light wave rotation. In [Sahin et al. 2016], an image based LF reconstruction with a Shearlet transform relieves the LF sampling requirement by a factor of up to 8×8, in the horizontal and vertical dimensions, without significant degradation in the perceived image quality.

Due to the LF sampling and diffraction from finite size hogels, objects far away from the hologram cannot be sharply reconstructed [Hilaire 1994; Wakunami and Yamaguchi 2011]. This makes HSs unsuitable for deep scenes. The concept of virtual light-ray sampling plane proposed in [Wakunami and Yamaguchi 2011; Yamaguchi 2013] reduces the blur from LF sampling. The resolution of reconstructed images for objects far away from the hologram plane is improved by hybrid computation of HS. This is achieved by i) locating a virtual plane close to the object and sampling the light rays from the object at this plane and ii) propagation of the complex amplitude from the virtual plane to the hologram using diffraction theory. The points in the virtual plane encode projection images produced by CG techniques as ray-tracing or image-based rendering. For encoding, each projection image is multiplied by a 2D random phase uniformly distributed in $[0, 2\pi]$ and Fourier transformed to obtain the complex amplitude in the virtual plane. Thus, the virtual plane contains an array of abutting elemental holograms. The complex amplitude in this plane is propagated by Fresnel diffraction to the CGH plane. Several virtual light-ray sampling planes can be introduced to reconstruct objects at different depths. The developed approach is used for multi-occlusion processing in [Wakunami et al. 2013] by locating virtual planes near the objects in the scene. Using the light-ray domain for occlusion processing allows development of a fast and simple algorithm. The proposed method with a virtual plane correctly represents mutual occlusion. In [Plesniak et al. 2006], the so-called reconfigurable image projection CGH is introduced by populating a virtual plane located at some distance from the hologram plane with projection primitives that create the desired distribution of the directional information. The virtual plane is sampled on a regular grid by hogels with abutting boundaries. For each hogel, the basis fringes on the hologram plane are found by Fresnel propagation. The resulting overlapping fringes on the hologram plane reduce artifacts caused by the discontinuity of the fringes at hogel boundaries in the conventional HS computing.

For high spatial and angular resolutions, the HS should consist of very small size of pixels, which can be provided by techniques such as lithography or HS printing. In printing, each hogel image is fed to an amplitude SLM that is focused on a silver-halide plate by a lens. A reference beam illuminates the plate from the opposite side, and a hogel is recorded as a volume reflection hologram. The average grain size of the available silver-halide emulsions can be as little as 10 nm. The HS printers can produce large size, high-quality, color HPO holograms [Brotherton-Ratcliffe et al. 2011]. When illuminated by a point source of white light, the printed HS spatially multiplexes the hogel images reconstructed from different hogels.

4.2 Multiple viewpoint projection holography

Similar to HS, the main motivation of MVP holography is the utilization of a set of 2D intensity images, captured under incoherent white light illumination, in CGH generation [Shaked et al. 2009]. A fundamental difference, however, is that MVP holograms aim at generating smooth (coherent) object wavefronts, in accordance with, e.g., Fourier or Fresnel hologram, which results in a different LF capture setup and problem formulation.

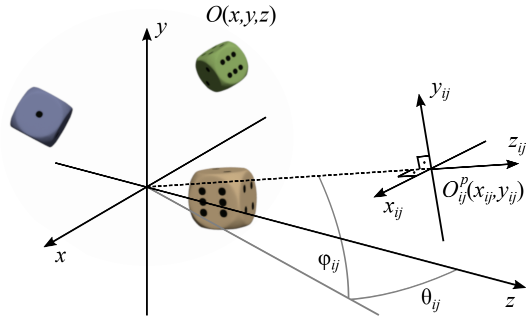


Fig. 11. LF capture geometry for MVP holography.

As illustrated in Figure 11, the LF capture setup in MVP holography involves a set of projection images $O^p_{ij}(x_{ij}, y_{ij})$ recorded from different angles (θ_{ij}, ϕ_{ij}). Those projection images actually correspond to orthogonal projections of the scene intensity function $O(x, y, z)$, which represents the light distribution that is assumed to be isotropically emitted from the surface of the objects in the scene, along the corresponding directions. That is,

$$O^p_{ij}(x_{ij}, y_{ij}) = \int O_{ij}(x_{ij}, y_{ij}, z_{ij}) dz_{ij}, \tag{12}$$

where $O_{ij}(x_{ij}, y_{ij}, z_{ij})$ is the representation of the scene intensity function with respect to the coordinates of the camera used for the projection along (θ_{ij}, ϕ_{ij}) [Sando et al. 2003]. One can rewrite Equation (12) by taking the 2D Fourier transform of both sides as

$$\tilde{O}^p_{ij}(u_{ij}, v_{ij}) = \iiint O_{ij}(x_{ij}, y_{ij}, z_{ij}) \exp\left[-j2\pi(u_{ij}x_{ij} + v_{ij}y_{ij})\right] dx_{ij}dy_{ij}dz_{ij}. \tag{13}$$

Under small-angle approximations, $\cos(\theta_{ij}) \approx 1$ and $\cos(\phi_{ij}) \approx 1$, these coefficients are shown to constitute a Fourier hologram of the scene $O(x, y, z)$ [Abookasis and Rosen 2003; Sando et al. 2003]. When an SLM is placed at the front focal plane of a lens and these complex-valued coefficients are properly coded and combined in a matrix to be fed to the SLM as real and nonnegative transmittance, a normally falling plane wave reconstructs the scene at the back focal plane of the lens.

The method in [Li et al. 2001] forms the complex-valued matrix from the data in the projection images captured at equal angular steps in a horizontal plane by a camera through an imaging lens. The function corresponds to the capture of a LF from a 3D object through a special optical system from two cylindrical lenses. Due to the HPO capability, however, defocused reconstructions are likely to occur normally to the recording axis. A full parallax system is proposed in [Abookasis and Rosen 2003] by acquisition of angular projections in a 2D grid and forming a 2D complex matrix. Each element of the matrix corresponds to a given viewpoint and the matrix is arranged in the order the projection images are captured. The Fourier hologram inherently includes the object field. Thus, having obtained the Fourier coefficients by Equation (13), other types of holograms such as Fresnel can be generated [Abookasis and Rosen 2006; Park et al. 2009; Sando et al. 2003], which enable reconstruction without additional lenses. The full color reconstruction is demonstrated in [Sando et al. 2004] by building three CGHs for the primary colors. Analysis of different magnifications at three wavelengths is also provided. Different variations of the Fresnel hologram for short and long reconstruction distances are presented in [Abookasis and Rosen 2006]. Generation of Fourier and Fresnel holograms uses a slightly different projection geometry, where the image plane is fixed and thus not always normal to projection lines as in the conventional angular projection [Park et al. 2009]. The integral photography implemented in [Shaked et al. 2007] acquires projection images by a lens array. The projection images recorded by a digital camera are then used for a Fourier hologram.

Similar to HS, the main drawback of MVP holography is the need for capture of a large number of images with small angular increments in the projection angles, typically under 1 degree. Simultaneous capture of these images with a lenslet array is questionable because of the low resolution of the capture system [Ichihashi et al. 2012]. The number of projections is reduced in [Katz et al. 2007] and the intermediate projections are calculated by a view synthesis algorithm. The requirement for a distinct anchor point for interpolation of different perspectives entails increasing the number of the projections for textured and smooth scenes. Both anchor points and the number of projections must be chosen individually for each 3D object. Compressive sensing applied in [Brady et al. 2009; Rivenson et al. 2011] uses sparse representation of the input data in the Fourier space. That is, an under-sampled Fourier hologram is calculated from an undersampled set of projections. The Fourier space is non-uniformly sampled in such a way that the density of projections are kept higher in the central low frequency regions than in the high frequency regions. Two equations are built for reconstruction from the under-sampled Fourier hologram based on 2D-2D and 3D-2D reconstruction models. Successful 3D reconstructions are reported with only around 6% of all projections.

5 SPECKLE SUPPRESSION IN CGH

Holograms suffer from an inherent coherent imaging issue known as speckle noise. This issue is also relevant to CGHs and requires careful examination. In order to simulate diffused diffraction of light from the recorded scene and to avoid concentration of light on the CGH, usually random phases are added to the point sources of light (or equivalent primitives). As the image perceived by the HVS can be considered as a sum of point spread functions (PSF) corresponding to the scene points, the random phase added to the points causes random interference patterns, if the PSFs overlap on the retinal surface. This random interference is observed as random speckle patterns with high contrast and spatial frequency that heavily degrade the perceived quality. Thus, eliminating the speckle patterns from the holographic reconstructions is vital in order to achieve satisfactory visual quality.

Due to the different hologram generation in wavefront-based and ray-based CGH techniques, the speckle suppression methods differ in most cases and thus require separate examination for both CGH types. Some solutions, however, can be utilized in both cases as the speckle suppression is achieved through display technique means, i.e., by modifying the holographic display optics, rather than altering the hologram generation. Mostly these methods rely on reducing either the temporal or the spatial coherence of the reconstruction illumination. By using a diffuser, the spatial coherence of the light is decreased [Yamaguchi et al. 1994]. The temporal coherence is reduced by utilizing light-emitting diodes (LEDs) for illumination [Yaraş et al. 2009]. It should be noted, however, that the approaches reducing the coherence of the reconstruction light add blur to the reconstructed image, which can negatively affect the reproduction of deep scenes [Yaraş et al. 2009].

The amount of speckle noise in the reconstructed views is often evaluated as speckle contrast [Goodman 2007]. The contrast is obtained as the ratio of the standard deviation σ and mean of the intensity values $\widehat{I}$ within the region of interest, i.e., $C = \sigma/\widehat{I}$. The region-of-interest should consist of uniform intensity values in view of the signal-dependent nature of the speckle noise. Subsequently, lower speckle contrast signifies better speckle suppression. Values equal to or under 0.05 are typically considered tolerable for sufficient image quality in laser projection displays. For high-end systems, however, achieving a more demanding speckle contrast criterion of 0.01 or less is recommended [Manni and Goodman 2012].

5.1 Speckle suppression methods for wavefront-based CGHs

One of the methods for suppressing speckle noise is to combine several CGH frames with statistically independent speckle patterns, which is known as random averaging [Amako et al. 1995]. The frames are displayed in sequence faster than the temporal response of the HVS, thus combining them in a time-multiplexed manner. As a result, the viewer perceives temporally averaged speckle patterns. The reduction at speckle contrast is proportional to $\sqrt{N}$ where N is the number of hologram frames. Thus, efficient speckle suppression requires a large number of frames, which in turn necessitates a high-speed display device to avoid separate frames to be resolved, as flickering, by the viewer.

As the random interference between the overlapping PSFs on the retina creates speckle noise, it can be suppressed by avoiding or reducing the overlaps. This can be achieved by generating several CGH frames, each including a subset of all the point sources, such that the distance between adjacent points separates the PSFs on the retina [Takaki and Yokouchi 2011]. The frames are then combined in a time-multiplexed manner in order to display them as a single reconstruction frame for the viewer with the entire set of point sources in the scene. The practical implementation of such speckle suppression in a holographic display can be accomplished by including a set of microlenses on the SLM, e.g., by using a digital micromirror device as in [Takaki and Yokouchi 2011]. In further studies, object point separation solution is adopted for lens-less holographic projection with undersampled bitmaps to generate the hologram [Makowski 2013]. To reduce the periodic interference due to the periodic locations of the separated object points, a random pixel separation method is proposed in [Mori et al. 2014]. By separating the points at random locations, the overlapping areas producing periodic interference are dispersed, thus reducing the overall unwanted interference. For polygon-based CGH methods, a time-multiplexing method incorporating angular spectrum interleaving reduces the speckle noise by a single carrier wave for each multiplexed CGH frame, which results in a linear phase distribution for the mesh surfaces, thus avoiding random interference [Ko and Park 2017]. Each frame produces speckle-free reconstructions, though time-multiplexing of multiple CGH frames is required in order to maintain the viewing angle supported by the hologram sampling size.

Similarly to point source separation applied in the object-oriented wavefront-based CGHs mentioned above, in image-oriented wavefront-based CGHs, speckle suppression can be achieved through ray separation. Based on the depth information, the utilized light rays can be separated into subsets of sparse rays, corresponding to sparse points in the scene [Utsugi and Yamaguchi 2014]. Due to finite angular resolution in ray casting, the view-dependent rays, which would correspond to a single point source, are usually mapped to multiple point sources. This issue is handled by quantizing the found point source locations on a uniform grid of voxels, resulting in pseudo-point sources. The time-multiplexed combination of CGHs for sparse subsets of rays results in reconstructions with reduced speckle. The reconstruction artifacts introduced in this approach due to ray quantization errors are addressed in [Mäkinen et al. 2018] by using the concept of densely sampled light field capture that enables accurate light ray resampling.

There also exist speckle suppression methods that do not rely on time-multiplexed reconstruction, but instead manage the speckle patterns without generating additional hologram frames through various means. One of these methods achieves this for the display of phase-only holograms by cyclic sequential shifting of the hologram [Golan and Shoham 2009]. The deterministic shift-averaging method utilizes the properties of phase shifting in order to minimize the interference between different pairs of PSFs within certain squares of interest on the hologram. Based on the iterative Fourier transform algorithm (IFTA), the Fresnel ping-pong algorithm utilizes forward and backward propagation of the field between two different object planes along with amplitude adjustments in order to suppress speckles [Dorsch et al. 1994]. The method is modified in [Makowski et al. 2005] to accommodate phase-only CGHs by using two object planes with the hologram plane and intensity equalization to assure a phase-only CGH, while the amplitude adjustments reduce the speckle noise. An alternative solution for single frame speckle suppression replaces the use of random phase distribution and instead multiplies the object wave with virtual converging light [Shimobaba and Ito 2015]. For phase-only holographic displays, complex modulation is shown to reduce speckle noise without the need for iteration, thus making it suitable for improving image quality in dynamic displays [Qi et al. 2016].

Alternatives to random phase distribution are also proposed to avoid the generation of speckle patterns. In [Wyrowski and Bryngdahl 1989], an iterative Fourier-based algorithm is proposed for finding phase distributions eliminating the formation of speckle patterns. Similarly, by using an iterative optimization approach, object-dependent distributions is calculated [Bräuer et al. 1991]. In both cases, these distributions spread the light over a finite area in the Fourier domain, thus resulting in speckle-free diffraction patterns. Moreover, a deterministic method for object-independent phase distribution acquisition, which is mostly optimal for near constant magnitude objects, is reported. A quasi-band-limited distribution is also shown to be effective at recording Fourier transform holograms with reduces speckle [Yamaguchi et al. 1994].

5.2 Speckle suppression methods for ray-based CGHs

In the case of ray-based CGHs, random averaging is also applicable. In particular, different CGHs can be computed with a different sets of random phases and then they can be reconstructed in a time-multiplexed manner. However, the abovementioned point or light ray separation approaches are not directly applicable due to the requirement of explicit depth data of the scene. HSs can utilize a modified version of the pixel separation method. By separating the hogels on the hologram plane to different CGH frames, a time-multiplexed reconstruction of the hologram results in reduced interference between adjacent hogels [Makowski 2013; Takaki and Yokouchi 2011].

The interference patterns can be suppressed by phase modulation as proposed in [Takaki and Taira 2016]. The method transforms the speckle patterns into regular sinusoidal interference patterns with spatial frequency being higher than the cut-off spatial frequency of HVS. This is achieved by improving the viewing area continuity, i.e., the number of viewpoints is increased by generating virtual views between the originally captured views. As the lowest spatial frequency terms are generated by adjacent views, the analysis in [Takaki and Taira 2016] considers 2×2 views within the extent of the pupil, where the phases are set by modifying the coefficients of the intensity distribution on the retina. By maximizing certain coefficients, the spatial frequency is increased such that the human vision cut-off frequency is exceeded and the speckle patterns become invisible to the human eyes. Alternatively, the spatial bandwidth of the random phase distribution can be limited to reduce speckle noise [Takaki and Ikeda 2013]. Assuming a hologram pixel pitch of Δ and spatial bandwidth of B_p in the parallax images, the spatial bandwidth of the random phase distribution should be limited to $1/\Delta - B_p$ to ensure that the diffraction distribution is as uniform as possible and its deviation is minimized to avoid the generation of speckle noise.

6 CONCLUSIONS

This paper provides a survey of the state-of-the-art in the field of CGHs for 3D imaging. Development of methods for synthesis of such CGHs has shown steady growth over the last few decades due to the rapid progress in computing and optical devices for capture and visualization. In addition to dynamic optical reconstruction of such CGHs via advanced SLMs, it has become also possible to fabricate highly realistic CGHs on holographic plates. Especially for glasses-free visualization, there is still a big room for improvement in terms of optical devices to provide necessary size and bandwidth, i.e., space-bandwidth product. However, in the case of virtual or augmented reality type of visualization via near-eye or head-mounted display technology, the holographic imaging and in particular CGH based visualization is getting more close to “reality” due to significantly relieved requirements on the size and bandwidth of hologram compared to glasses-free display case.

Computation-wise, realistic representation of 3D objects sets extremely high demands in the CGH encoding process. The reason is the need for view-dependent full parallax rendering of information related to depth, color, texture, specular and diffuse reflection, mutual and self-occlusion, etc. The various 3D scene characteristics are supposed to be expressed as amplitude or phase variation within a CGH at different wavelengths. Complexity of rendered visual effects has a direct impact on the efficiency of the CGH computation, i.e., speed and memory usage. Fast generation methods are especially crucial for imaging of dynamic scenes. Thus, achieving better quality of reconstruction and acceleration of computation are the most up-to-date issues in the CGH synthesis.

The contemporary CGHs can be categorized into two broad categories as wavefront-based and ray-based methods depending mainly on the utilized object model and light propagation model. In both cases, the interference of the object and reference waves is computed, thus imitating the optical holographic recording. The wavefront-based methods rely on scalar diffraction in computing the object wave. The 3D object is modeled as a set of light primitives, which scatter light into wavefronts. The object wave is found by superposing such wavefronts. The amount of light primitives dictates the computational complexity. Point cloud model is the most computationally demanding approach, which represents a 3D object as a collection of point light sources. The great potential of this model for photorealistic imaging is the main motivation behind the numerous approaches proposed for acceleration of computation. The solutions make use of similarities in the Fresnel zone plates formed in the hologram by the point sources, which are located in a plane parallel to the hologram or on a line normal to the hologram, for creating look-up-tables. Other solutions exploit the small diffraction angle of the existing SLMs to restrict the contribution of a point source at the CGH plane or divide the hologram into elemental holograms, which allow for replacing the spherical wavefront from a point source by patches of plane waves. Acceleration can be also achieved by modeling the 3D objects as meshes of polygons or plane slices located at different depths parallel to hologram plane. The polygon-based method is more flexible in expressing object features, while the layer-based method is more computationally efficient. The numerical modification of the polygon-based method is well developed for incorporating texture, reflections, occlusions and other visual effects. Using a triangular mesh for 3D object representation offers an option for analytical computation at the expense of constraints set on the light field distribution within each triangle. Many efforts are focused on alleviation of these constraints. The narrow viewing zone is the main drawback of the layer-based approach. Overcoming this requires angular multiplexing of holograms corresponding to different viewing points. Due to assumption of independency between light primitives, parallel computing offers great potential in accelerating the computation of wavefront-based CGHs, which is extensively exploited in the literature. In summary, the current status of the wavefront-based CGHs can be described mainly as active development of better encoding algorithms. In the case of dynamic optical reconstruction, e.g., via SLMs, the pixel pitch limits the bandwidth of CGH. Lithography or wavefront printing based methods enables visualization of static wavefront-based CGHs by means of a very fine resolution holographic plate and thus taking full advantage of highly realistic nature of such CGHs. Besides 3D imaging, another interesting application area of wavefront-based CGHs is holographic projection where CGHs are synthesized for 2D intensity images relying on Fresnel diffraction model. Holographic projection allows for lensless magnification.

The ray-based methods for CGH synthesis rely on acquiring spatial and angular intensity distribution of the propagating light rays, i.e., light field, due to given 3D object or scene. Encoding of this distribution can be based on the 2D images that are incoherently captured from different viewing points, which is a great advantage compared to wavefront-based methods in real life capture scenarios. Such multi-perspective images provide the light field as a collection of rays. CGH encoding follow two main schemes. The HS approach divides the hologram into holographic elements, which are called as hogels, and it encodes a set of fringes in each hogel, including a plane wave patch corresponding to each light field ray, based on the spatially and directionally varying intensity information in the captured light field. The HS is a significantly faster approach compared to wavefront-based CGHs, however it suffers from the spatio-angular resolution trade-off, which is dictated by the hogel size. The plane wave patches that are encoded incoherently in hogels are unmatched in phase along the hologram and this directly degrades the image quality for deep scenes compared to coherent case. An effective solution to this problem is achieved by the MVP approach, which makes it possible to reconstruct smooth wavefronts from a set of orthographic images. Other than dynamic optical reconstruction, the HS printing technology enables producing scalable static HSs. In summary, incoherent capture and fast computation are two important factors that make the ray-based methods attractive in creating dynamic or large size static holograms of real 3D scenes.

Speckle noise can significantly degrade reconstructed images in both ray-based and wave-based CGHs, which makes speckle suppression an important problem. Speckle suppression methods that are based on time-multiplexed computation (and reconstruction) of spatially separated object points or light rays demonstrate more effective implementations compared to the naive random (phase) averaging approach. However, there is still room for improvement in this problem.

ACKNOWLEDGMENTS

The paper is partially supported by the "Competitive Funding to Strengthen University Research Profiles", Academy of Finland, decision number 292477; and the National Science Fund of Bulgaria, project DH-08/13, "Holographic imaging, beam shaping and speckle metrology with computer generated holograms". Jani Mäkinen would like to acknowledge the support of the graduate school funding of Tampere University.

REFERENCES

David Abookasis and Joseph Rosen. 2003. Computer-generated holograms of three-dimensional objects synthesized from their multiple angular viewpoints. *J. Opt. Soc. Am. A* 20, 8 (Aug 2003), 1537–1545.

David Abookasis and Joseph Rosen. 2006. Three types of computer-generated hologram synthesized from multiple angular viewpoints of a three-dimensional scene. *Applied Optics* 45, 25 (Sep 2006), 6533–6538.

Lukas Ahrenberg, Philip Benzie, Marcus Magnor, and John Watson. 2006. Computer generated holography using parallel commodity graphics hardware. *Optics Express* 14, 17 (Aug 2006), 7636–7641.

Lukas Ahrenberg, Philip Benzie, Marcus Magnor, and John Watson. 2008. Computer generated holograms from three dimensional meshes using an analytic light transport model. *Applied Optics* 47, 10 (Apr 2008), 1567–1574.

Jun Amako, Hirotsuna Miura, and Tomio Sonehara. 1995. Speckle-noise reduction on kinoform reconstruction using a phase-only spatial light modulator. *Applied Optics* 34, 17 (Jun 1995), 3165–3171.

Daisuke Arai, Tomoyoshi Shimobaba, Takashi Nishitsuji, Takashi Kakue, Nobuyuki Masuda, and Tomoyoshi Ito. 2017. An accelerated hologram calculation using the wavefront recording plane method and wavelet transform. *Optics*

Communications 393 (2017), 107 – 112.

Mehdi Askari, Seong-Bok Kim, Kwang-Soo Shin, Seok-Bum Ko, Sang-Hoo Kim, Dae-Youl Park, Yeon-Gyeong Ju, and Jae-Hyeung Park. 2017. Occlusion handling using angular spectrum convolution in fully analytical mesh based computer generated hologram. *Optics Express* 25, 21 (Oct 2017), 25867–25878.

Muharrem Bayraktar and Meriç Özcan. 2010. Method to calculate the far field of three-dimensional objects for computer-generated holography. *Applied Optics* 49, 24 (Aug 2010), 4647–4654.

David Blinder and Peter Schelkens. 2018. Accelerated computer generated holography using sparse bases in the STFT domain. *Optics Express* 26, 2 (Jan 2018), 1461–1473.

Max Born. 1999. *Principles of Optics. Electromagnetic theory of propagation, interference and diffraction of light* (7. ed.). Cambridge University Press, Cambridge.

David J. Brady, Kerkil Choi, Daniel L. Marks, Ryoichi Horisaki, and Sehoon Lim. 2009. Compressive Holography. *Optics Express* 17, 15 (Jul 2009), 13040–13049.

Ralf Bräuer, Frank Wyrowski, and Olof Bryngdahl. 1991. Diffusers in digital holography. *J. Opt. Soc. Am. A* 8, 3 (Mar 1991), 572–578.

David C. Brotherton-Ratcliffe, Stanislovas J. Zacharovas, Ramunas J. Bakanas, Julius Pileckas, Andrej Nikolskij, and Jevgenij Kuchin. 2011. Digital holographic printing using pulsed RGB lasers. *Optical Engineering* 50 (2011), 50(9) 091307.

B. R. Brown and A. W. Lohmann. 1966. Complex Spatial Filtering with Binary Masks. *Applied Optics* 5, 6 (Jun 1966), 967–969.

Edward Buckley. 2011. Holographic Laser Projection. *J. Display Technol.* 7, 3 (Mar 2011), 135–140.

Jhen-Si Chen and Daping Chu. 2015. Improved layer-based method for rapid hologram generation and real-time interactive holographic display applications. *Optics Express* 23, 14 (Jul 2015), 18143–18155.

Jhen-Si Chen, Daping Chu, and Quinn Smithwick. 2014. Rapid hologram generation utilizing layer-based approach and graphic rendering for realistic three-dimensional image reconstruction by angular tiling. *Journal of Electronic Imaging* 23, 2 (2014), 023016–023016.

Rick H.-Y. Chen and Timothy D. Wilkinson. 2009. Computer generated hologram from point cloud using graphics processor. *Applied Optics* 48, 36 (Dec 2009), 6841–6850.

Jaebum Cho, Joonku Hahn, and Hwi Kim. 2012. Fast reconfiguration algorithm of computer generated holograms for adaptive view direction change in holographic three-dimensional display. *Optics Express* 20, 27 (Dec 2012), 28282–28291.

Jaebum Cho, Hwi Kim, Jiwoon Yeom, Gang Li, and Byoungho Lee. 2015. Rendering of reflective surfaces in polygon based computer generated holograms, In Imaging and Applied Optics 2015. *Imaging and Applied Optics 2015*, JT5A.21.

D. C. Chu, J. R. Fienup, and J. W. Goodman. 1973. Multiemulsion On-Axis Computer Generated Hologram. *Applied Optics* 12, 7 (Jul 1973), 1386–1388.

Giuseppe A. Cirino, Patrick Verdonck, Ronaldo D. Mansano, Jr. José C. Pizolato, Daniel B. Mazulquim, and Luiz G. Neto. 2011. Digital Holography: Computer-Generated Holograms and Diffractive Optics in Scalar Diffraction Domain.

Daniel Claus, Daciana Iliescu, and Peter Bryanston-Cross. 2011. Quantitative space-bandwidth product analysis in digital holography. *Applied Optics* 50, 34 (Dec 2011), H116–H127.

WJ Dallas. 1980. Computer-generated holograms. *The computer in optical research* 41 (1980), 291–366.

Xiao-Bin Dong, Seung-Cheol Kim, and Eun-Soo Kim. 2014a. MPEG-based novel look-up table for rapid generation of video holograms of fast-moving three-dimensional objects. *Optics Express* 22, 7 (Apr 2014), 8047–8067.

Xiao-Bin Dong, Seung-Cheol Kim, and Eun-Soo Kim. 2014b. Three-directional motion compensation-based novel-look-up-table for video hologram generation of three-dimensional objects freely maneuvering in space. *Optics Express* 22, 14 (Jul 2014), 16925–16944.

Rainer G. Dorsch, Adolf W. Lohmann, and Stefan Sinzinger. 1994. Fresnel ping-pong algorithm for two-plane computer-generated hologram display. *Applied Optics* 33, 5 (Feb 1994), 869–875.

Dennis Gabor. 1948. A New Microscopic Principle. *Nature* 161, 4098 (1948), 777–778.

R. W. Gerchberg and W. O. Saxton. 1972. Practical algorithm for determination of phase from image and diffraction plane pictures. *OPTIK* 35, 2 (1972), 237–&.

Antonin Gilles, Patrick Gioia, Rémi Cozot, and Luce Morin. 2016. Hybrid approach for fast occlusion processing in computer-generated hologram calculation. *Applied Optics* 55, 20 (Jul 2016), 5459–5470.

Lior Golan and Shy Shoham. 2009. Speckle elimination using shift-averaging in high-rate holographic projection. *Optics Express* 17, 3 (Feb 2009), 1330–1339.

Joseph W. Goodman. 1996. *Introduction to Fourier optics* (2nd., international ed.). McGraw-Hill, New York.

Joseph W. Goodman. 2007. *Speckle Phenomena in Optics: Theory and Applications.* Roberts and Company Publishers.

Michael W. Halle, Stephen A. Benton, Michael A. Klug, and John S. Underkoffler. 1991. Ultragram: a generalized holographic stereogram. In *Practical Holography V*, Vol. 1461. International Society for Optics and Photonics, 142–156.

Naotaka Hasegawa, Tomoyoshi Shimobaba, Takashi Kakue, and Tomoyoshi Ito. 2017. Acceleration of hologram generation by optimizing the arrangement of wavefront recording planes. *Applied Optics* 56, 1 (Jan 2017), A97–A103.

Noriyuki Hayashi, Yuji Sakamoto, and Yasuhiro Honda. 2011. Improvement of camera arrangement in computer-generated holograms synthesized from multi-view images. In *Proceedings of SPIE - The International Society for Optical Engineering*, Vol. 7957.

Pierre St Hilaire. 1994. Modulation transfer function and optimum sampling of holographic stereograms. *Applied Optics* 33, 5 (Feb 1994), 768–774.

Kohei Hosoyachi, Kazuhiro Yamaguchi, Tsubasa Ichikawa, and Yuji Sakamoto. 2013. Precalculation method using spherical basic object light for computer-generated hologram. *Applied Optics* 52, 1 (Jan 2013), A33–A44.

Yasuyuki Ichihashi, Ryutaro Oi, Takanori Senoh, Kenji Yamamoto, and Taiichiro Kurita. 2012. Real-time capture and reconstruction system with multiple GPUs for a 3D live scene by a generation from 4K IP images to 8K holograms. *Optics Express* 20, 19 (Sep 2012), 21645–21655.

Tsubasa Ichikawa, Yuji Sakamoto, Agus Subagyo, and Kazuhisa Sueoka. 2011. Calculation method of reflectance distributions for computer-generated holograms using the finite-difference time-domain method. *Applied Optics* 50, 34 (Dec 2011), H211–H219.

Tsubasa Ichikawa, Kazuhiro Yamaguchi, and Yuji Sakamoto. 2013a. Realistic expression for full-parallax computer-generated holograms with the ray-tracing method. *Applied Optics* 52, 1 (Jan 2013), A201–A209.

Tsubasa Ichikawa, Takuo Yoneyama, and Yuji Sakamoto. 2013b. CGH calculation with the ray tracing method for the Fourier transform optical system. *Optics Express* 21, 26 (Dec 2013), 32019–32031.

Shunsuke Igarashi, Tomoya Nakamura, Kyoji Matsushima, and Masahiro Yamaguchi. 2018. Efficient tiled calculation of over-10-gigapixel holograms using ray-wavefront conversion. *Optics Express* 26, 8 (Apr 2018), 10773–10786.

Dajeong Im, Jaebum Cho, Joonku Hahn, Byoungho Lee, and Hwi Kim. 2015. Accelerated synthesis algorithm of polygon computer-generated holograms. *Optics Express* 23, 3 (Feb 2015), 2863–2871.

Boaz Jessie Jackin, Shinpei Watanabe, Kanemitsu Ootsu, Takeshi Ohkawa, Takashi Yokota, Yoshio Hayasaki, Toyohiko Yatagai, and Takanobu Baba. 2018. Decomposition method for fast computation of gigapixel-sized Fresnel holograms on a graphics processing unit cluster. *Applied Optics* 57, 12 (Apr 2018), 3134–3145.

Yeong-Min Ji, Hanju-Yeom, and Jae-Hyeung Park. 2016. Efficient texture mapping by adaptive mesh division in mesh-based computer generated hologram. *Optics Express* 24, 24 (Nov 2016), 28154–28169.

Jia Jia, Juan Liu, Guofan Jin, and Yongtian Wang. 2014. Fast and effective occlusion culling for 3D holographic displays by inverse orthographic projection with low angular sampling. *Applied Optics* 53, 27 (Sep 2014), 6287–6293.

Jia Jia, Jhen Si, and Daping Chu. 2018. Fast two-step layer-based method for computer generated hologram using sub-sparse 2D fast Fourier transform. *Optics Express* 26, 13 (Jun 2018), 17487–17497.

Jia Jia, Yongtian Wang, Juan Liu, Xin Li, Yijie Pan, Zhumei Sun, Bin Zhang, Qing Zhao, and Wei Jiang. 2013. Reducing the memory usage for effectivecomputer-generated hologram calculation using compressed look-up table in full-color holographic display. *Applied Optics* 52, 7 (Mar 2013), 1404–1412.

Shuming Jiao, Zhaoyong Zhuang, and Wenbin Zou. 2017. Fast computer generated hologram calculation with a mini look-up table incorporated with radial symmetric interpolation. *Optics Express* 25, 1 (Jan 2017), 112–123.

J. L. Juárez-Pérez, A. Olivares-Pérez, and L. R. Berriel-Valdos. 1997. Nonredundant calculations for creating digital Fresnel holograms. *Applied Optics* 36, 29 (Oct 1997), 7437–7443.

Joel Jurik, Thomas Burnett, Michael Klug, and Paul Debevec. 2012. Geometry-corrected light field rendering for creating a holographic stereogram. In *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. 9–13.

Hoonjong Kang, Tomohiko Fujii, Takeshi Yamaguchi, and Hiroshi Yoshikawa. 2007. Compensated phase-added stereogram for real-time holographic display. *Optical Engineering* 46 (2007), 095802 – 0958011.

Hoonjong Kang, Elena Stoykova, Youngmin Kim, Sunghee Hong, Joosup Park, and Jisoo Hong. 2016b. Color Holographic Wavefront Printing Technique for Realistic Representation. *IEEE Transactions on Industrial Informatics* 12, 4 (2016), 1590–1598.

Hoonjong Kang, Elena Stoykova, and Hiroshi Yoshikawa. 2016a. Fast phase-added stereogram algorithm for generation of photorealistic 3D content. *Applied Optics* 55, 3 (Jan 2016), A135–A143.

Hoonjong Kang, Takeshi Yamaguchi, and Hiroshi Yoshikawa. 2008a. Accurate phase-added stereogram to improve the coherent stereogram. *Applied Optics* 47, 19 (Jul 2008), D44–D54.

Hoonjong Kang, Takeshi Yamaguchi, Hiroshi Yoshikawa, Seung-Cheol Kim, and Eun-Soo Kim. 2008b. Acceleration method of computing a compensated phase-added stereogram on a graphic processing unit. *Applied Optics* 47, 31 (Nov 2008), 5784–5789.

Barak Katz, Natan T. Shaked, and Joseph Rosen. 2007. Synthesizing computer generated holograms with reduced number of perspective projections. *Optics Express* 15, 20 (Oct 2007), 13250–13255.

Hwi Kim, Joonku Hahn, and Byoungho Lee. 2008a. Mathematical modeling of triangle-mesh-modeled three-dimensional surface objects for digital holography. *Applied Optics* 47, 19 (Jul 2008), D117–D127.

Hak Gu Kim and Yong Man Ro. 2017. Ultrafast layer based computer-generated hologram calculation with sparse template holographic fringe pattern for 3-D object. *Optics Express* 25, 24 (Nov 2017), 30418–30427.

Myung K. Kim. 2011. *Digital Holographic Microscopy : Principles, Techniques, and Applications.* Springer New York, New York.

Seung-Cheol Kim, Xiao-Bin Dong, and Eun-Soo Kim. 2015. Accelerated one-step generation of full-color holographic videos using a color-tunable novel-look-up-table method for holographic three-dimensional television broadcasting. *Scientific Reports* 5, 1 (2015), 14056.

Seung-Cheol Kim, Xiao-Bin Dong, Min-Woo Kwon, and Eun-Soo Kim. 2013. Fast generation of video holograms of three-dimensional moving objects using a motion compensation-based novel look-up table. *Optics Express* 21, 9 (May 2013), 11568–11584.

Seung-Cheol Kim and Eun-Soo Kim. 2008. Effective generation of digital holograms of three-dimensional objects using a novel look-up table method. *Applied Optics* 47, 19 (Jul 2008), D55–D62.

Seung-Cheol Kim, Jae-Man Kim, and Eun-Soo Kim. 2012. Effective memory reduction of the novel look-up table with one-dimensional sub-principle fringe patterns in computer-generated holograms. *Optics Express* 20, 11 (May 2012), 12021–12034.

Seung-Cheol Kim, Jung-Hoon Yoon, and Eun-Soo Kim. 2008b. Fast generation of three-dimensional video holograms by combined use of data compression and lookup table techniques. *Applied Optics* 47, 32 (Nov 2008), 5986–5995.

M. C. King, A. M. Noll, and D. H. Berry. 1970. A New Approach to Computer-Generated Holography. *Applied Optics* 9, 2 (Feb 1970), 471–475.

M. Kinoshita and Y. Sakamoto. 2009. Computer-generated holograms at an arbitrary viewpoint synthesized from multi-view images. *Proc. SPIE* 7233, 72330Z.

Seok-Beom Ko and Jae-Hyeung Park. 2017. Speckle reduction using angular spectrum interleaving for triangular mesh based computer generated hologram. *Optics Express* 25, 24 (Nov 2017), 29788–29797.

Min-Woo Kwon, Seung-Cheol Kim, and Eun-Soo Kim. 2016. Three-directional motion-compensation mask-based novel look-up table on graphics processing units for video-rate generation of digital holographic videos of three-dimensional scenes. *Applied Optics* 55, 3 (Jan 2016), A22–A31.

Douglas Lanman, Ramesh Raskar, and Gabriel Taubin. 2008. Modeling and Synthesis of Aperture Effects in Cameras. In *Computational Aesthetics in Graphics, Visualization, and Imaging*. The Eurographics Association, 81–88.

Seok Lee, Hyunsung Chang, Hocheon Wey, and Dongkyung Nam. 2016. Sampling and error analysis of radial symmetric interpolation for fast hologram generation. *Applied Optics* 55, 3 (Jan 2016), A104–A110.

Wooyoung Lee, Dajeong Im, Jeongyeup Paek, Joonku Hahn, and Hwi Kim. 2014. Semi-analytic texturing algorithm for polygon computer-generated holograms. *Optics Express* 22, 25 (Dec 2014), 31180–31191.

Emmett N. Leith and Juris Upatnieks. 1962. Reconstructed Wavefronts and Communication Theory. *J. Opt. Soc. Am.* 52, 10 (Oct 1962), 1123–1130.

Detlef Leseberg and Christian Frère. 1988. Computer-generated holograms of 3-D objects composed of tilted planar segments. *Applied Optics* 27, 14 (Jul 1988), 3020–3024.

Mark Levoy. 2006. Light Fields and Computational Imaging. *Computer* 39, 8 (2006), 46–55.

Youzhi Li, David Abookasis, and Joseph Rosen. 2001. Computer-generated holograms of three-dimensional realistic objects recorded without wave interference. *Applied Optics* 40, 17 (Jun 2001), 2864–2870.

Hong-Gi Lim, Na-Young Jo, and Jae-Hyeung Park. 2013. Hologram synthesis with fast texture update of triangular meshes, In Digital Holography and Three-Dimensional Imaging. *Digital Holography and Three-Dimensional Imaging*, DW2A.8.

Zhouchen Lin and Heung-Yeung Shum. 2004. A Geometric Analysis of Light Field Rendering. *International Journal of Computer Vision* 58, 2 (2004), 121–138.

Jung-Ping Liu, Wang-Yu Hsieh, Ting-Chung Poon, and Peter Tsang. 2011. Complex Fresnel hologram display using a single SLM. *Applied Optics* 50, 34 (Dec 2011), H128–H135.

Jung-Ping Liu, Tatsuki Tahara, Yoshio Hayasaki, and Ting-Chung Poon. 2018. Incoherent Digital Holography: A Review. *Applied Sciences* 8 (01 2018), 143.

Yuan-Zhi Liu, Jian-Wen Dong, Yi-Ying Pu, Bing-Chu Chen, He-Xiang He, and He-Zhou Wang. 2010. High-speed full analytical holographic computations for true-life scenes. *Optics Express* 18, 4 (Feb 2010), 3345–3351.

Adolf W. Lohmann, Rainer G. Dorsch, David Mendlovic, Zeev Zalevsky, and Carlos Ferreira. 1996. Space–bandwidth product of optical signals and systems. *J. Opt. Soc. Am. A* 13, 3 (Mar 1996), 470–473.

A. W. Lohmann and D. P. Paris. 1967. Binary Fraunhofer Holograms, Generated by Computer. *Applied Optics* 6, 10 (Oct 1967), 1739–1748.

Mark E. Lucente. 1993. Interactive computation of holograms using a look-up table. *Journal of Electronic Imaging* 2, 1 (1993), 28–34.

Mark E. Lucente. 1994. *Diffraction-specific Fringe Computation for Electro-holography.* Ph.D. Dissertation. Cambridge, MA, USA.

Andrew Maimone, Andreas Georgiou, and Joel S. Kollin. 2017. Holographic Near-eye Displays for Virtual and Augmented Reality. *ACM Trans. Graph.* 36, 4, Article 85 (July 2017), 16 pages.

Jani Mäkinen, Erdem Sahin, and Atanas Gotchev. 2018. Speckle reduction method for image-based coherent stereogram generation. *Optics Express* 26, 5 (Mar 2018), 5381–5394.

Michał Makowski. 2013. Minimized speckle noise in lens-less holographic projection by pixel separation. *Optics Express* 21, 24 (Dec 2013), 29205–29216.

Michał Makowski, Maciej Sypek, Andrzej Kolodziejczyk, and Grzegorz Mikuła. 2005. Three-plane phase-only computer hologram generated with iterative Fresnel algorithm. *Optical Engineering* 44, 12 (2005), 125805.

Jeffrey G. Manni and Joseph W. Goodman. 2012. Versatile method for achieving 1% speckle contrast in large-venue laser projection displays using a stationary multimode optical fiber. *Optics Express* 20, 10 (May 2012), 11288–11315.

Kyoji Matsushima. 2005a. Computer-generated holograms for three-dimensional surface objects with shade and texture. *Applied Optics* 44, 22 (Aug 2005), 4607–4614.

Kyoji Matsushima. 2005b. Exact hidden-surface removal in digitally synthetic full-parallax holograms, In Practical Holography XIX: Materials and Applications. *Proc. SPIE* 5742, 25–32.

Kyoji Matsushima. 2006. Performance of the polygon-source method for creating computer-generated holograms of surface objects. In *Proceedings of ICO Topical Meeting on Optoinformatics/Information Photonics.* 99–100.

Kyoji Matsushima. 2008. Formulation of the rotational transformation of wave fields and their application to digital holography. *Applied Optics* 47, 19 (Jul 2008), D110–D116.

Kyoji Matsushima and Sumio Nakahara. 2009. Extremely high-definition full-parallax computer-generated hologram created by the polygon-based method. *Applied Optics* 48, 34 (Dec 2009), H54–H63.

Kyoji Matsushima, Masaki Nakamura, and Sumio Nakahara. 2014. Silhouette method for hidden surface removal in computer holography and its acceleration using the switch-back technique. *Optics Express* 22, 20 (Oct 2014), 24450–24465.

Kyoji Matsushima, Hagen Schimmel, and Frank Wyrowski. 2003. Fast calculation method for optical diffraction on tilted planes by use of the angular spectrum of plane waves. *J. Opt. Soc. Am. A* 20, 9 (Sep 2003), 1755–1762.

Kyoji Matsushima and Noriaki Sonobe. 2018. Full-color digitized holography for large-scale holographic 3D imaging of physical and nonphysical objects. *Applied Optics* 57, 1 (Jan 2018), A150–A156.

Kyoji Matsushima and Masahiro Takai. 2000. Recurrence formulas for fast creation of synthetic three-dimensional holograms. *Applied Optics* 39, 35 (Dec 2000), 6587–6594.

J. T. McCrickerd and Nicholas George. 1968. Holographic stereogram from sequential component photographs. *Applied Physics Letters* 12, 1 (1968), 10–12.

Yutaka Mori, Takahiko Fukuoka, and Takanori Nomura. 2014. Speckle reduction in holographic projection by random pixel separation with time multiplexing. *Applied Optics* 53, 35 (Dec 2014), 8182–8188.

Andrej Nikolskij, David C. Brotherton-Ratcliffe, Ramunas Bakanas, Stanislovas Zacharovas, and Julius Pileckas. 2012. Image capture system for a digital holographic printer.

Hirohito Nishi, Kyoji Matsushima, and Sumio Nakahara. 2011. Rendering of specular surfaces in polygon-based computer-generated holograms. *Applied Optics* 50, 34 (Dec 2011), H245–H252.

Takashi Nishitsuji, Tomoyoshi Shimobaba, Takashi Kakue, and Tomoyoshi Ito. 2015. Fast calculation of computer-generated hologram using run-length encoding based recurrence relation. *Optics Express* 23, 8 (Apr 2015), 9852–9857.

Takashi Nishitsuji, Tomoyoshi Shimobaba, Takashi Kakue, Nobuyuki Masuda, and Tomoyoshi Ito. 2012. Fast calculation of computer-generated hologram using the circular symmetry of zone plates. *Optics Express* 20, 25 (Dec 2012), 27496–27502.

Yusuke Ohsawa, Kazuhiro Yamaguchi, Tsubasa Ichikawa, and Yuji Sakamoto. 2013. Computer-generated holograms using multiview images captured by a small number of sparsely arranged cameras. *Applied Optics* 52, 1 (Jan 2013), A167–A176.

Naohisa Okada, Tomoyoshi Shimobaba, Yasuyuki Ichihashi, Ryutaro Oi, Kenji Yamamoto, Takashi Kakue, and Tomoyoshi Ito. 2014. Fast calculation of a computer-generated hologram for RGB and depth images using a Wavefront recording plane method. *Photonics Letters of Poland* 6, 3 (2014), 90–92.

N. Padmanaban, Y. Peng, and G. Wetzstein. 2019. Holographic Near-Eye Displays Based on Overlap-Add Stereograms. *ACM Trans. Graph. (SIGGRAPH Asia)* 6 (2019). Issue 38.

Yijie Pan, Yongtian Wang, Juan Liu, Xin Li, and Jia Jia. 2013. Fast polygon-based method for calculating computer-generated holograms in three-dimensional display. *Applied Optics* 52, 1 (Jan 2013), A290–A299.

Yuechao Pan, Xuewu xu, Sanjeev Solanki, Xinan Liang, Ridwan Bin Adrian Tanjung, Chiwei Tan, and Tow-Chong Chong. 2009. Fast CGH computation using S-LUT on GPU. *Optics Express* 17 (10 2009), 18543–18555.

Jae-Hyeung Park. 2017. Recent progress in computer-generated holography for three-dimensional scenes. *Journal of Information Display* 18, 1 (2017), 1–12.

Jae-Hyeung Park, Min-Su Kim, Ganbat Baasantseren, and Nam Kim. 2009. Fresnel and Fourier hologram generation using orthographic projection images. *Optics Express* 17, 8 (Apr 2009), 6320–6334.

Jae-Hyeung Park, Seong-Bok Kim, Han-Ju Yeom, Hee-Jae Kim, HuiJun Zhang, BoNi Li, Yeong-Min Ji, Sang-Hoo Kim, and Seok-Bum Ko. 2015. Continuous shading and its fast update in fully analytic triangular-mesh-based computer generated hologram. *Optics Express* 23, 26 (Dec 2015), 33893–33901.

Anh-Hoang Phan, Md A. Alam, Seok-Hee Jeon, Jeong-Hyeon Lee, and Nam Kim. 2014. Fast hologram generation of long-depth object using multiple wavefront recording planes. *Proc. SPIE* 9006, 900612.

Wendy J. Plesniak, Michael W. Halle, V. Michael Bove, James Barabas, and Ravikanth Pappu. 2006. Reconfigurable image projection holograms. *Optical Engineering* 45, 11 (2006), 115801.

Ting-Chung Poon and Jung-Ping Liu. 2014. *Introduction to Modern Digital Holography with MATLAB.*

Yijun Qi, Chenliang Chang, and Jun Xia. 2016. Speckleless holographic display by complex modulation based on double-phase method. *Optics Express* 24, 26 (Dec 2016), 30368–30378.

Yair Rivenson, Adrian Stern, and Joseph Rosen. 2011. Compressive multiple view projection incoherent holography. *Optics Express* 19, 7 (Mar 2011), 6109–6118.

Joseph Rosen and Gary Brooker. 2007. Digital spatially incoherent Fresnel holography. *Optics Letters* 32, 8 (Apr 2007), 912–914.

Erdem Sahin, Suren Vagharshakyan, Jani Mäkinen, Robert Bregovic, and Atanas Gotchev. 2016. Shearlet-domain light field reconstruction for holographic stereogram generation. In *2016 IEEE International Conference on Image Processing (ICIP).* 1479–1483.

Yusuke Sando, Masahide Itoh, and Toyohiko Yatagai. 2003. Holographic three-dimensional display synthesized from three-dimensional Fourier spectra of real existing objects. *Optics Letters* 28, 24 (Dec 2003), 2518–2520.

Yusuke Sando, Masahide Itoh, and Toyohiko Yatagai. 2004. Color computer-generated holograms from projection images. *Optics Express* 12, 11 (May 2004), 2487–2493.

Natan T. Shaked, Barak Katz, and Joseph Rosen. 2009. Review of three-dimensional holographic imaging by multiple-viewpoint-projection based methods. *Applied Optics* 48, 34 (Dec 2009), H120–H136.

Natan T. Shaked, Joseph Rosen, and Adrian Stern. 2007. Integral holography: white-light single-shot hologram acquisition. *Optics Express* 15, 9 (Apr 2007), 5754–5760.

Tomoyoshi Shimobaba and Tomoyoshi Ito. 2015. Random phase-free computer-generated hologram. *Optics Express* 23, 7 (Apr 2015), 9549–9554.

Tomoyoshi Shimobaba and Tomoyoshi Ito. 2017. Fast generation of computer-generated holograms using wavelet shrinkage. *Optics Express* 25, 1 (Jan 2017), 77–87.

Tomoyoshi Shimobaba, Tomoyoshi Ito, Nobuyuki Masuda, Yasuyuki Ichihashi, and Naoki Takada. 2010. Fast calculation of computer-generated-hologram on AMD HD5000 series GPU and OpenCL. *Optics Express* 18, 10 (May 2010), 9955–9960.

Tomoyoshi Shimobaba, Nobuyuki Masuda, and Tomoyoshi Ito. 2009. Simple and fast calculation algorithm for computer-generated hologram with wavefront recording plane. *Optics Letters* 34, 20 (Oct 2009), 3133–3135.

Quinn Y. J. Smithwick, James Barabas, Daniel E. Smalley, and V. Michael Bove. 2010. Interactive holographic stereograms with accommodation cues. *Proc.SPIE* 7619, 761903–7619013.

Pierre St-Hilaire, Stephen A. Benton, Mark E. Lucente, Mary Lou Jepsen, J. Kollin, Hiroshi Yoshikawa, and John S. Underkoffler. 1990. Electronic display system for computational holography. In *Practical Holography IV*, Stephen A. Benton (Ed.), Vol. 1212. International Society for Optics and Photonics, SPIE, 174 – 182.

Elena Stoykova, Fahri Yaraş, Ali Özgür Yontem, Hoonjong Kang, Levent Onural, Philippe Hamel, Yves Delacrétaz, Isabelle Bergoënd, Cristian Arfire, and Christian Depeursinge. 2013. Optical reconstruction of transparent objects with phase-only SLMs. *Optics Express* 21, 23 (Nov 2013), 28246–28257.

Ping Su, Wenbo Cao, Jianshe Ma, Bingchao Cheng, Xianting Liang, Liangcai Cao, and Guofan Jin. 2016. Fast Computer-Generated Hologram Generation Method for Three-Dimensional Point Cloud Model. *Journal of Display Technology* 12, 12 (2016), 1688–1694.

Athanasia Symeonidou, David Blinder, Adrian Munteanu, and Peter Schelkens. 2015. Computer-generated holograms by multiple wavefront recording plane method with occlusion culling. *Optics Express* 23, 17 (Aug 2015), 22149–22161.

Athanasia Symeonidou, David Blinder, and Peter Schelkens. 2018. Colour computer-generated holography for point clouds utilizing the Phong illumination model. *Optics Express* 26, 8 (Apr 2018), 10282–10298.

T.-C. Poon, M. H. Wu, K. Shinoda, and Y. Suzuki. 1996. Optical scanning holography. *Proc. IEEE* 84, 5 (1996), 753–764.

Yasuhiro Takaki and Kyohei Ikeda. 2013. Simplified calculation method for computer-generated holographic stereograms from multi-view images. *Optics Express* 21, 8 (Apr 2013), 9652–9663.

Yasuhiro Takaki and Nichiyo Nago. 2010. Multi-projection of lenticular displays to construct a 256-view super multi-view display. *Optics Express* 18, 9 (Apr 2010), 8824–8835.

Yasuhiro Takaki and Kengo Taira. 2016. Speckle regularization and miniaturization of computer-generated holographic stereograms. *Optics Express* 24, 6 (Mar 2016), 6328–6340.

Yasuhiro Takaki and Masahito Yokouchi. 2011. Speckle-free and grayscale hologram reconstruction using time-multiplexing technique. *Optics Express* 19, 8 (2011), 7567–7579.

Tullio Tommasi and Bruno Bianco. 1993. Computer-generated holograms of tilted planes by a spatial frequency approach. *J. Opt. Soc. Am. A* 10, 2 (Feb 1993), 299–305.

Seymour Trester. 2000. Computer-simulated Fresnel holography. *European Journal of Physics* 21, 4 (2000), 317–331.

P. W. M. Tsang, K. W. K. Cheung, and T.-C. Poon. 2012. Real-time relighting of digital holograms based on wavefront recording plane method. *Optics Express* 20, 6 (Mar 2012), 5962–5967.

P. W. M. Tsang, Y. T. Chow, and T.-C. Poon. 2016. Generation of complementary sampled phase-only holograms. *Opt. Express* 24, 20 (Oct 2016), 23390–23395.

P. W. M. Tsang, Y. T. Chow, and T.-C. Poon. 2017. Generation of patterned-phase-only holograms (PPOHs). *Opt. Express* 25, 8 (Apr 2017), 9088–9093.

P. W. M. Tsang and T.-C. Poon. 2015. Fast generation of digital holograms based on warping of the wavefront recording plane. *Optics Express* 23, 6 (Mar 2015), 7667–7673.

P. W. M. Tsang and T.-C. Poon. 2016. Review on the state-of-the-art technologies for acquisition and display of digital holograms. *IEEE Trans. Industrial Informatics* 12, 3 (June 2016), 886– 901.

P. W. M. Tsang, T.-C. Poon, and Y. M. Wu. 2018. Review of fast methods for point-based computer-generated holography [Invited]. *Photonics Research* 6, 9 (Sep 2018), 837–846.

Raul Tudela, Ignasi Labastida, Estela Martin-Badosa, Santiago Vallmitjana, Ignacio Juvells, and Artur Carnicer. 2002. A simple method for displaying Fresnel holograms on liquid crystal panels. *Optics Communications* 214, 1 (2002), 107–114.

Takeru Utsugi and Masahiro Yamaguchi. 2014. Speckle-suppression in hologram calculation using ray-sampling plane. *Optics Express* 22, 14 (Jul 2014), 17193–17206.

E. G. van Putten, I. M. Vellekoop, and A. P. Mosk. 2008. Spatial amplitude and phase modulation using commercial twisted nematic LCDs. *Appl. Opt.* 47, 12 (Apr 2008), 2076–2081.

Vishal Verma and Ekta Walia. 2010. 3D Rendering - Techniques and challenges. *International Journal of Engineering and Technology* 2, 2 (2010), 29–33.

Koki Wakunami and Masahiro Yamaguchi. 2011. Calculation for computer generated hologram using ray-sampling plane. *Optics Express* 19, 10 (May 2011), 9086–9101.

Koki Wakunami, Hiroaki Yamashita, and Masahiro Yamaguchi. 2013. Occlusion culling for computer generated hologram based on ray-wavefront conversion. *Optics Express* 21, 19 (Sep 2013), 21811–21822.

James P. Waters. 1966. Holographic image synthesis utilizing theoretical methods. *Applied Physics Letters* 9, 11 (1966), 405–407.

Hui Wei, Guanghong Gong, and Ni Li. 2016. Improved look-up table method of computer-generated holograms. *Applied Optics* 55, 32 (Nov 2016), 9255–9264.

Jiantong Weng, Tomoyoshi Shimobaba, Naohisa Okada, Hirotaka Nakayama, Minoru Oikawa, Nobuyuki Masuda, and Tomoyoshi Ito. 2012. Generation of real-time large computer generated hologram using wavefront recording method. *Optics Express* 20, 4 (Feb 2012), 4018–4023.

Gordon Wetzstein, Douglas Lanman, Matthew Hirsch, and Ramesh Raskar. 2012. Tensor Displays: Compressive Light Field Synthesis Using Multilayer Displays with Directional Backlighting. *ACM Trans. Graph.* 31, 4, Article 80 (July 2012), 11 pages.

Frank Wyrowski and Olof Bryngdahl. 1989. Speckle-free reconstruction in digital holography. *J. Opt. Soc. Am. A* 6, 8 (Aug 1989), 1171–1174.

Xiao Xiao, Bahram Javidi, Manuel Martinez-Corral, and Adrian Stern. 2013. Advances in three-dimensional integral imaging: sensing, display, and applications [Invited]. *Applied Optics* 52, 4 (Feb 2013), 546–560.

Kazuhiro Yamaguchi, Tsubasa Ichikawa, and Yuji Sakamoto. 2011a. Calculation method for CGH considering smooth shading with polygon models. *Proc. SPIE* 7957, 795706.

Kazuhiro Yamaguchi, Tsubasa Ichikawa, and Yuji Sakamoto. 2011b. Calculation method for computer-generated holograms considering various reflectance distributions based on microfacets with various surface roughnesses. *Applied Optics* 50, 34 (Dec 2011), H195–H202.

Kazuhiro Yamaguchi and Yuji Sakamoto. 2009. Computer generated hologram with characteristics of reflection: reflectance distributions and reflected images. *Applied Optics* 48, 34 (Dec 2009), H203–H211.

Masahiro Yamaguchi. 2013. Image-based computational holography for deep 3D scene display. *Proc. SPIE* 9042, 90420A.

Masahiro Yamaguchi. 2016. Light-field and holographic three-dimensional displays [Invited]. *J. Opt. Soc. Am. A* 33, 12 (Dec 2016), 2348–2364.

Masahiro Yamaguchi, Hideaki Endoh, Toshio Honda, and Nagaaki Ohyama. 1994. High-quality recording of a full-parallax holographic stereogram with a digital diffuser. *Optics Letters* 19, 2 (Jan 1994), 135–137.

Masahiro Yamaguchi, Hideshi Hoshino, Toshio Honda, and Nagaaki Ohyama. 1993. Phase-added stereogram: calculation of hologram using computer graphics technique. *Proc. SPIE* 1914, 25–31.

Takeshi Yamaguchi, Tomohiko Fujii, and Hiroshi Yoshikawa. 2007. Computer-generated cylindrical hologram.

Takeshi Yamaguchi and Hiroshi Yoshikawa. 2011. Computer-generated image hologram. *Chinese Optics Letters* 9, 12 (2011), 20–23.

Kenji Yamamoto, Takanori Senoh, Ryutaro Oi, and Taiichiro Kurita. 2010. 8K4K-size computer generated hologram for 3-D visual system using rendering technology. In *2010 4th International Universal Communication Symposium*. 193–196.

Fahri Yaraş, Hoonjong Kang, and Levent Onural. 2009. Real-time phase-only color holographic video display system using LED illumination. *Applied Optics* 48, 34 (Dec 2009), H48–H53.

Fahri Yaraş, Hoonjong Kang, and Levent Onural. 2010. State of the Art in Holographic Displays: A Survey. *Journal of Display Technology* 6, 10 (2010), 443–454.

Toyohiko Yatagai. 1974. Three-dimensional displays using computer-generated holograms. *Optics Communications* 12, 1 (1974), 43–45.

Han-Ju Yeom and Jae-Hyeung Park. 2016. Calculation of reflectance distribution using angular spectrum convolution in mesh-based computer generated hologram. *Optics Express* 24, 17 (Aug 2016), 19801–19813.

Fucai Zhang, Ichirou Yamaguchi, and L. P. Yaroslavsky. 2004. Algorithm for reconstruction of digital holograms with adjustable magnification. *Optics Letters* 29, 14 (Jul 2004), 1668–1670.

Hao Zhang, Liangcai Cao, and Guofan Jin. 2017. Computer-generated hologram with occlusion effect using layer-based processing. *Applied Optics* 56, 13 (May 2017), F138–F143.

Hao Zhang, Liangcai Cao, Song Zong, and Guofan Jin. 2016a. Computer-generated hologram calculation using layered stereogram based inverse Fresnel diffraction, In Imaging and Applied Optics 2016. *Imaging and Applied Optics 2016*, DW5I.9.

Hao Zhang, Neil Collings, Jing Chen, Bill A. Crossland, Daping Chu, and Jinghui Xie. 2011. Full parallax three-dimensional display with occlusion effect using computer generated hologram. *Optical Engineering* 50, 7 (2011), 074003.

Hao Zhang, Yan Zhao, Liangcai Cao, and Guofan Jin. 2015. Fully computed holographic stereogram based algorithm for computer-generated holograms with accurate depth cues. *Optics Express* 23, 4 (Feb 2015), 3901–3913.

Yingxi Zhang, Juan Liu, Xin Li, and Yongtian Wang. 2016b. Fast processing method to generate gigabyte computer generated holography for three-dimensional dynamic holographic display. *Chinese Optics Letters* 14, 3 (Mar 2016), 030901.

Yaping Zhang, Jianqiang Zhang, Wei Chen, Jialing Zhang, Peng Wang, and Wei Xu. 2013. Research on three-dimensional computer-generated holographic algorithm based on conformal geometry theory. *Optics Communications* 309 (2013), 196–200.

Ya-Ping Zhang, Fan Wang, Ting-Chung Poon, Shuang Fan, and Wei Xu. 2018. Fast generation of full analytical polygon-based computer-generated holograms. *Optics Express* 26, 15 (Jul 2018), 19206–19224.

Yan Zhao, Liangcai Cao, Hao Zhang, Dezhao Kong, and Guofan Jin. 2015. Accurate calculation of computer-generated holograms using angular-spectrum layer-oriented method. *Optics Express* 23, 20 (Oct 2015), 25440–25449.

Yan Zhao, Liangcai Cao, Hao Zhang, Wei Tan, Shenghan Wu, Zheng Wang, Qiang Yang, and Guofan Jin. 2016. Time-division multiplexing holographic display using angular-spectrum layer-oriented method (Invited Paper). *Chinese Optics Letters* 14, 1 (2016), 10005–10009.

Huadong Zheng, Yingjie Yu, Tao Wang, and Linmao Dai. 2009. High-quality three-dimensional holographic display with use of multiple fractional Fourier transform. *Chinese Optics Letters* 7, 12 (Dec 2009), 1151–1154.

Long Zhu and Jian Wang. 2014. Arbitrary manipulation of spatial amplitude and phase using phase-only spatial light modulators. *Sci. Rep.* 4, 7441 (2014).

PUBLICATION
II

Speckle reduction method for image-based coherent stereogram generation

J. Mäkinen, E. Sahin, and A. Gotchev

Optics Express, vol. 26, no. 5, pp. 5381-5394
DOI: 10.1364/OE.26.005381

Research Article
Vol. 26, No. 5 | 5 Mar 2018 | OPTICS EXPRESS 5381
Optics EXPRESS

Speckle reduction method for image-based coherent stereogram generation

JANI MÄKINEN,* **ERDEM SAHIN, AND ATANAS GOTCHEV**

Laboratory of Signal Processing, Tampere University of Technology, Korkeakoulunkatu 10, 33720 Tampere, Finland
**jani.makinen@tut.fi*

Abstract: We propose a speckle noise reduction method for generation of coherent holographic stereograms. The method employs densely sampled light field (DSLF) of the scene together with depth information acquired for each ray in the captured DSLF. Speckle reduction is achieved based on the ray separation technique where the scene is first described as a superposition of sparse sets of point sources corresponding to separated sets of rays and then the holographic reconstructions corresponding to these sparse sets of point sources are added incoherently (intensity-wise) to obtain the final reconstruction. The proposed method handles the light propagation between the sparse scene points and hologram elements accurately by utilizing ray resampling based on the notion of DSLF. As a result, as demonstrated via numerical simulations, significant speckle suppression is achieved at no cost of sampling related reconstruction artifacts.

OCIS codes: (030.6140) Speckle; (090.1760) Computer holography; (090.2870) Holographic display.

References and links

1. D. Gabor, "A new microscopic principle," Nature **161**(4098), 777–778 (1948).
2. J. W. Goodman, *Speckle Phenomena in Optics: Theory and Applications* (Roberts and Company Publishers, 2007).
3. Q. Y. J. Smithwick, J. Barabas, D. E. Smalley, and V. M. Bove, Jr., "Interactive holographic stereograms with accommodation cues," Proc. SPIE **7619**, 761903 (2010).
4. P. W. McOwan, W. J. Hossack, and R. E. Burge, "Three-dimensional stereoscopic display using ray traced computer generated holograms," Opt. Commun. **82**(1–2), 6–11 (1991).
5. J. Amako, H. Miura, and T. Sonehara, "Speckle-noise reduction on kinoform reconstruction using a phase-only spatial light modulator," Appl. Opt. **34**(17), 3165–3171 (1995).
6. P. Memmolo, V. Bianco, M. Paturzo, B. Javidi, P. Netti, and P. Ferraro, "Encoding multiple holograms for speckle-noise reduction in optical display," Opt. Express **22**(21), 25768–25775 (2014).
7. V. Bianco, P. Memmolo, M. Paturzo, A. Finizio, B. Javidi, and P. Ferraro, "Quasi noise-free digital holography," Light: Science & Applications **5**(9), e16142 (2016).
8. L. Rong, W. Xiao, F. Pan, S. Liu, and R. Li, "Speckle noise reduction in digital holography by use of multiple polarization holograms," Chin. Opt. Lett. **8**(7), 653–655 (2010).
9. F. Yaraş, H. Kang, and L. Onural, "Real-time phase-only color holographic video display system using LED illumination," Appl. Opt. **48**(34), H48–H53 (2009).
10. M. Yamaguchi, H. Endoh, T. Honda, and N. Ohyama, "High-quality recording of a full-parallax holographic stereogram with a digital diffuser," Opt. Lett. **19**(2), 135–137 (1994).
11. Y. Takaki and M. Yokouchi, "Speckle-free and grayscale hologram reconstruction using time-multiplexing technique," Opt. Express **19**(8), 7567–7579 (2011).
12. Y. Takaki and K. Taira, "Speckle regularization and miniaturization of computer-generated holographic stereograms," Opt. Express **24**(6), 6328–6340 (2016).
13. T. Kurihara and Y. Takaki, "Speckle-free, shaded 3d images produced by computer-generated holography," Opt. Express **21**(4), 4044–4054 (2013).
14. T. Utsugi and M. Yamaguchi, "Speckle-suppression in hologram calculation using ray-sampling plane," Opt. Express **22**(14), 17193–17206 (2014).
15. M. Levoy and P. Hanrahan, "Light field rendering," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, (ACM, 1996), pp. 31–42.
16. M. Yamaguchi, H. Hoshino, T. Honda, and N. Ohyama, "Phase-added stereogram: calculation of hologram using computer graphics technique," Proc. SPIE **1914**, 25–31 (1993).
17. H. Kang, T. Yamaguchi, and H. Yoshikawa, "Accurate phase-added stereogram to improve the coherent stereogram," Appl. Opt. **47**(19), D44–D54 (2008).

#313969 https://doi.org/10.1364/OE.26.005381
Journal © 2018 Received 21 Nov 2017; revised 3 Feb 2018; accepted 3 Feb 2018; published 22 Feb 2018

18. M. Lucente, "Diffraction-specific fringe computation for electro-holography," Ph.D. dissertation (Massachusetts Institute of Technology 1994).
19. Z. Lin and H.-Y. Shum, "A geometric analysis of light field rendering," International Journal of Computer Vision **58**(2), 121–138 (2004).
20. C. H. J. Howard, "A Test for the Judgment of Distance," American Journal of Ophthalmology **2**, 656–675 (1919).
21. J. W. Goodman, *Introduction to Fourier Optics* 2nd ed. (McGraw-Hill, 1996).
22. Blender Foundation, `http://www.blender.org`.

---

1. Introduction

As a three-dimensional (3D) display method, holography [1] is often considered as the ultimate way to visually replicate a 3D scene, i.e. including all of the relevant visual cues necessary for proper 3D perception such as continuous motion parallax, correct spatial relations between objects, and accommodation. Due to utilized coherent imaging techniques, however, holographic reconstructions suffer from the speckle noise issue. The speckle noise patterns are random in nature with high contrast and frequency, thus heavily degrade the visual quality of the reconstructed images. In traditional optical holography, coherent light illumination creates random phase distributions on the (rough) scene surface. This results in random interference of scene points on the hologram plane, which is observed as speckle noise in the reconstructed images [2].

On the other hand, computer-generated holography provides a way to obtain the hologram numerically by simulating the physical wave propagation phenomenon during recording. A hologram obtained in this manner is commonly referred to as a computer-generated hologram (CGH). Stereograms constitute an important and widely used category of CGHs especially due to their ease of application for real life scenes. In particular, the necessary content can be captured by conventional cameras as a set of multiview images. The stereograms can be categorized into incoherent and coherent types, depending on the data they utilize. Incoherent stereograms are purely image-based, that is, they are generated from a set of multiperspective images. On the other hand, coherent stereograms require information about object location. This can be either in the form of a 3D model (e.g. point cloud) or multiperspective images coupled with depth information (i.e. depth maps). The availability of additional 3D information in coherent stereograms brings critical improvements in different aspects of reconstruction quality compared to incoherent ones, such as delivering correct accommodation cues [3]. Although 3D object models are often very precise and therefore beneficial in CGH applications, utilization of multiperspective images and depth information possesses some important advantages. For example, when recording a real scene the explicit object information is not available. Furthermore, utilizing such scene representation benefits from advanced rendering techniques in computer graphics enabling reduction of computational burden [4]. Occlusions are also intrinsically handled by image and depth based holograms as the perspective views record them correctly on the hologram.

It is a common practice in CGHs to employ random phase distributions (e.g. they are assigned to a set of point sources) so as to simulate diffused diffraction of light from the object. This creates similar speckle noise patterns on the reconstructed images as in optical holograms. One widely used method for speckle suppression in electro-holography is the so-called averaging method, where several CGH frames with statistically independent speckle patterns are averaged (over intensity) through multiple recordings by time-multiplexed reconstruction [5–8]. Although this speckle averaging approach comes at the expense of computational cost or more complicated optics, unlike the methods that suppress speckle by reducing the temporal [9] or spatial [10] coherence of the light, it does not suffer from loss of resolution. Nevertheless, the speckle reduction performance of the speckle averaging method is limited in its efficiency. That is, for N holograms utilized in the averaging, the speckle contrast is reduced by a factor of $1/\sqrt{N}$ [5]. Alternative solutions proposed for incoherent stereograms include spatially separating the fringe patterns in each single hologram of multiple recordings, which are to be averaged again by time-multiplexed reconstruction [11], and phase distribution manipulation [12]. Such methods do not suffer from the abovementioned theoretical limitation inherent to speckle averaging methods, and thus are able reduce the speckle noise more effectively.

When examining speckle suppression for coherent stereograms, on the other hand, it is important to consider the object information utilized in the hologram generation. Speckle suppression in model-based coherent stereograms has been achieved, for instance, through superposition of CGH frames obtained from sparse sets of point sources [11,13]. Similar solution regarding sparse light rays for coherent stereograms utilizing multiperspective images and depth information has been also proposed [14]. These methods also reduce the speckle noise more effectively than the speckle averaging methods [14].

The speckle noise reduction method that we propose in this paper for coherent stereograms also relies on multiperspective images and depth data. The proposed method is mainly based on the ray separation solution previously presented in [14]. That is, the speckle patterns are suppressed through generating several CGH frames for different sets of sparse (quantized) scene points, corresponding to sparse sets of rays, and combining them in a time-multiplexed manner. Although this approach suppresses the speckle noise effectively, it does not accurately solve the light propagation problem between the quantized scene points and the hologram elements (hogels). This degrades the accuracy of the subsequent reconstructions. The method that we propose in this paper approaches the problem through signal processing means by utilizing the notion of densely sampled light field (DSLF). In particular, the light propagation between the quantized scene points and the hogels is accurately defined via accurate ray resampling from DSLF.

Below in section 2, we start with describing generation of coherent stereogram from captured light field and depth maps. In section 3, we first discuss the basic principle behind the ray separation method and then in section 4 we present the proposed approach. Finally, in section 5, numerical simulations are presented where the speckle suppression performance of the proposed method is compared with the existing techniques.

2. Discrete LF capture and coherent stereograms

2.1. Coherent stereogram generation from discrete LF and depth information

Considering geometrical optics and rays as the fundamental light carrier, a defined space can be interpreted as a collection of light rays. LF describes the intensities of light rays traveling from different points in space to different directions. Assuming monochromatic illumination and static scenes with a transparent medium for the light to travel in, the LF can be represented as a 4D radiance function [15]. This function can be parametrized in different ways, including point and direction, points on two (parallel) planes or point pairs on a 3D surface. To simplify the analysis and visualizations, let us consider a 2D cross-section of the 3D space. Thus, utilizing the two-plane parametrization, a continuous LF function $L_1(x, s)$ is defined between planes denoted by x and s.

One can calculate the CGH of a scene based on the captured LF together with depth information of each ray. Let us define the hologram on the x plane as denoted in Fig. 1. In practice, the continuous LF is rarely available and capturing the LF requires usually discretization on the two parametrization planes (e.g. when capturing as multiperspective images and depth). That is, a discrete set of samples is taken from the continuous LF by sampling the planes x and s at intervals of Δ_x and Δ_s, respectively, resulting in the discrete LF $L_1(m\Delta_x, i\Delta_s) = L_1[m, i]$, where $m = 1, 2, \ldots, M$ and $i = 1, 2, \ldots, N$. Utilizing such information for the hologram generation results in segmentation on the hologram plane due to discretization on the x plane. Thus, the discretization on the x plane determines perceived spatial resolution when viewing the hologram. On the other hand, the sampling on the s plane corresponds to view-dependent quality aspects of the hologram such as parallax, occlusion.

There are several segmented coherent holographic representations such as phase-added

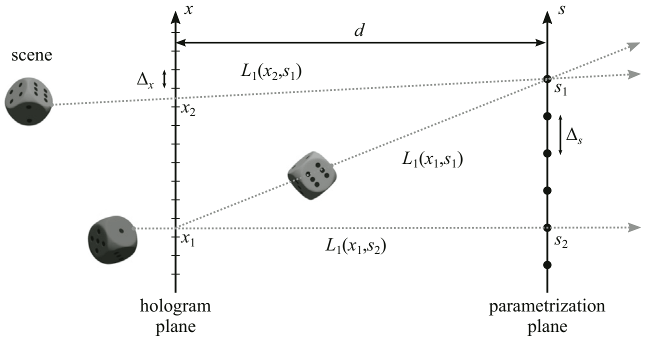


Fig. 1. Discrete light field and hologram definitions according to two-plane parametrization.

stereogram (PAS) [16] and diffraction specific coherent panoramagram (DSCP) [3]. Similarly to the incoherent alternative holographic stereograms (HSs), these CGHs also segment the hologram into elements containing several hologram pixels. In HS and PAS these segments are called holographic elements and they produce a piecewise planar approximation of the wavefield. The DSCP segments, on the other hand, are divided into so-called wavefront elements (wafels), which have controllable curvatures to produce sharp points even for deep scenes. DSCP, therefore, provides a more accurate representation compared to other coherent segmented CGH representations, such as PAS and accurate PAS [17]. Thus, here in this paper we consider the DSCP as an accurate coherent stereogram representation. The wavefield segments of DSCP approximate the field as segments of varying radius spherical waves, the radius depending on the distance between the hologram and the emission location of each light ray. Based on the discrete LF $L_1[m,i]$ together with depth information giving a corresponding point source location (x_{mi}, z_{mi}) for each ray, the object wavefield for the DSCP is defined as [3]

$$O_{DSCP}(x) = \sum_m \operatorname{rect}\left(\frac{x - m\Delta_x}{\Delta_x}\right) \sum_i \frac{\sqrt{L_1[m,i]}}{r_{mi}} \exp\left[\frac{j2\pi}{\lambda}\left(\sqrt{(x - x_{mi})^2 + z_{mi}^2} - z_{mi}\right)\right], \quad (1)$$

where λ is the wavelength of the monochromatic light; x_{mi} is the x-coordinate, z_{mi} is the z-coordinate and r_{mi} is the Euclidean distance between the point source (corresponding to ray $[m,i]$) and the hologram segment indexed by m.

The hologram generation imposes strict restrictions on the discrete LF. The properties of the human visual system (HVS), e.g. its resolution limitations, usually play critical role in determining the sampling parameters Δ_x and Δ_s [18]. Nevertheless, within the scope of this paper, the LF sampling requirement is derived to ensure accurate ray resampling. Because, as will be discussed in section 3, in the ray separation method the point sources corresponding to captured rays are to be quantized on to different grids than the original (capture) grid. Hence, in order to accurately calculate the intensity distribution for the new set of quantized rays, accurate resampling from captured LF $L_1[m,i]$ is required. The following section discusses LF capture and sampling fulfilling such requirements.

2.2. DSLF capture as multiperspective images

The discrete LF required for hologram generation can be captured with a multiperspective camera setup, placing cameras at the parametrization plane s (i.e. camera plane) separated by the LF sampling distance Δ_s from each other. A third plane containing the camera sensors is introduced behind the camera apertures (see Fig. 2), thus capturing the discrete LF between the camera and sensor planes according to two-plane parametrization. The camera and sensor planes are denoted by s and u respectively and the distance between the two planes by l. The continuous LF between the two planes is denoted by $L_2(s, u)$. This LF can be obtained from the LF $L_1(x, s)$ defined between the hologram and camera (or viewer) plane, as the radiance along each light ray remains constant (due to transparent medium assumption).

When capturing multiperspective images, two different camera models are often considered: regular and recentering. Depending on the chosen camera model, the capture process has varying requirements. In the regular camera model, high resolution images with wide field-of-view are required to maintain the entire scene within the image boundaries across the entire range of perspective views. On the other hand, the recentering camera model shifts the sensor behind camera center of projection, as seen in Fig. 2, thus wasting less area on the images. Additionally, the correct LF rays are directly captured, assuming correct setup parameters. Although its practical implementations can be problematic, here we utilize the recentering camera model due to its lower resolution requirement and direct LF correspondence.

Capturing the discrete LF with a multiperspective imaging setup results in the discrete LF $L_2[i, k]$ being stored in the images. However, as the hologram generation process requires the LF in relation to the hologram plane, i.e. the LF $L_1[m, i]$, a solution for obtaining this LF accurately from the captured one is needed. The notion of DSLF provides a structured framework for this purpose. Considering the recentering camera model for LF capture, the DSLF defines the LF sampling such that the continuous function can be retrieved from the discrete LF (with sufficient accuracy) using bilinear interpolation under the condition that the disparity between adjacent camera images is kept to be between $[-1, 1]$ pixels during capture [19].

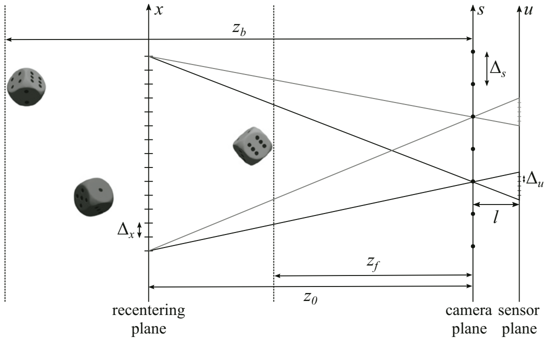


Fig. 2. Capture setup and parameters for DSLF. A set of cameras capture the scene at intervals of Δ_s, sampling the (recentered) image plane at Δ_x. The area between z_b and z_f fulfills the DSLF requirement.

Let us define the scene and capture parameters in accordance with the Fig. 2. A set of cameras sampled at distance Δ_s capture multiperspective views according to the recentering camera model. The recentering plane is placed at distance z_0 from the camera plane, and the camera parameters are chosen such that the recentering plane is sampled at distance Δ_x. The captured scene is limited from the back and front by z_b and z_f, respectively, with respect to the camera plane. Let us assume that $z_f \leq z_0 \leq z_b$ holds, i.e. that the scene boundaries z_b and z_f are behind and in front of the recentering plane, respectively. The maximum camera spacing to achieve dense LF sampling is then defined as the minimum of the camera spacing values corresponding to -1 and 1 pixel disparity:

$$\Delta_s = \min\left\{\frac{\Delta_x z_b}{z_b - z_0}, \frac{\Delta_x z_f}{z_0 - z_f}\right\}. \tag{2}$$

If either of the scene limiting inequalities does not hold, the camera spacing for the corresponding depth limit results in a negative value and can thus be omitted. The camera spacing is maximized when both front and back disparity values are equal [19], i.e.

$$z_{opt} = \frac{2 z_b z_f}{z_b + z_f}. \tag{3}$$

Alternatively, the camera sampling distance Δ_s can be first chosen and then the scene can be limited to meet the DSLF requirement as

$$z_b = \frac{\Delta_s z_0}{\Delta_s - \Delta_x}, \tag{4}$$

$$z_f = \frac{\Delta_s z_0}{\Delta_s + \Delta_x}. \tag{5}$$

By placing the recentering plane at the hologram plane during DSLF capture such that each pixel corresponds to a hologram segment, the DSLF capture can be connected to the discrete LF required for CGH generation.

3. Speckle suppression by ray separation

The image perceived by the HVS can be modeled as a sum of point spread functions (PSF) corresponding to point sources on the object. Those point sources having overlapping PSFs on the retina interfere with each other under coherent illumination. The random phase distribution on the object points results in random interference pattern, which is observed as speckle noise. The significant part of the speckle occurs due to interaction between the main lobes of PSFs. For a diffused surface, the average size of the speckle on the retina can be estimated for laterally separated point sources based on the lateral PSF main lobe width as [2, 14]

$$L_x \approx 2.44\frac{\lambda l}{T}, \tag{6}$$

where T is the human eye lens diameter and l is the distance between the lens and the retina.

The approach of describing the object as a superposition of sets of sparse object points and then calculating and reconstructing the corresponding holograms for these sets separately (i.e. superposing them incoherently) suppresses the speckle noise [11, 14]. This is due to having less overlap between the PSFs of the object points in each sparse set. In order for the speckle suppression to be effective, the separation between adjacent points in each multiplexed set needs to be larger than speckle size. Furthermore, the PSF side lobes also contribute to the interference causing speckles, thus further increase in separation is beneficial.

Since our method is mainly based on the ray separation (correspondingly sparse object points) technique introduced in [14], let us first briefly discuss this technique to better address the problems of the existing approach. Although there are no explicit point sources in image-based scene representations, one can associate point sources (ray emission points) with the light rays utilizing the available depth information as demonstrated in Fig. 3(a). The emission point (x_{mi}, z_{mi}) for the light ray corresponding to the pixel m of captured image at view i is obtained as

$$x_{mi} = m\Delta_x + \frac{z_{mi}(i\Delta_s - m\Delta_x)}{d}, \tag{7}$$

$$z_{mi} = D[m, i], \tag{8}$$

where $D[m, i]$ is the corresponding depth value and d is the distance between the camera and hologram planes. In order to achieve efficient speckle suppression, the ray emission points need to be quantized on a predefined scene grid. If the grid is defined such that the quantization step is less than the Rayleigh resolution limit, no spatial resolution is lost in terms of HVS capabilities. Because of similar reasons, it is also a good and common practice to choose the hologram segment size according to the Rayleigh resolution limit. Consequently, the lateral quantization step can be chosen as [14]

$$\Delta_{\tilde{x}} \leq 1.22\frac{\lambda d}{T}. \tag{9}$$

The quantization grid can, thus, be placed such that the horizontal spacing is equal to the hologram segment size and aligned with the segment center points, i.e. $\Delta_{\tilde{x}} = \Delta_x$. The axial quantization step, on the other hand, can be determined based on the depth acuity of the HVS. The depth acuity of the HVS depends on several factors and, therefore, it has been mostly studied experimentally. However, the stereoacuity can be used to obtain a rough estimate in most of the scenarios: at depth z, the depth difference $\delta_z(z)$ just detectable by the HVS can be estimated as [20]

$$\delta_z(z) = \frac{z^2\delta_\gamma}{cB}, \tag{10}$$

where δ_γ is the angular measure for the stereoacuity of the HVS, which is typically around 0.5 arcmin, B is the baseline between the two eyes of the human, which is typically around 6.5 cm, and $c = 3437.75$ is a constant defining conversion from radian to arcmin. The quantization step in depth can, thus, be chosen as $\Delta_{\tilde{z}} = \delta_z(d)$.

Having formed the quantization grid, the quantized point coordinates on the grid can be obtained by finding the nearest grid points to the emission coordinates as [14]

$$(\tilde{x}_{mi}, \tilde{z}_{mi}) = \underset{(\tilde{x},\tilde{z})\in S_q}{\arg\min}\left\{\sqrt{(\tilde{x} - x_{mi})^2 + (\tilde{z} - z_{mi})^2}\right\}, \tag{11}$$

where S_q is the entire set of points on the quantization grid. As shown in Fig. 3(b), this divides the scene space into equal size quantization volumes (voxels) surrounding each quantization point.

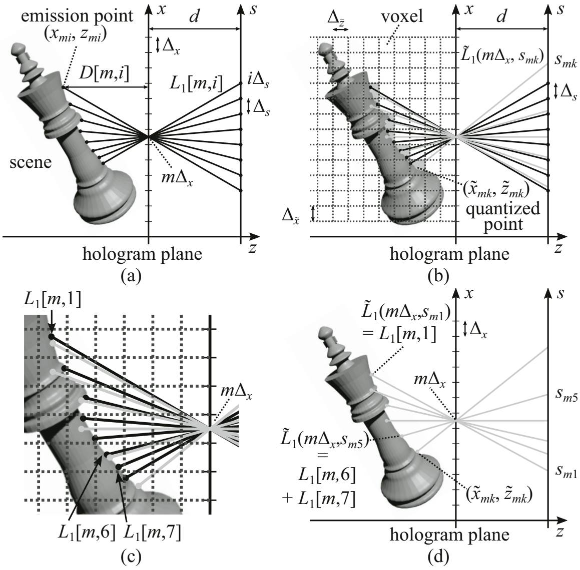
After obtaining the set of voxels representing quantized scene points, it is critical to accurately define the light propagation between these voxels and the hologram segments. Let us denote the LF between such voxels and the hologram segment m as $\tilde{L}_1(m\Delta_x, s_{mk})$, where the corresponding rays intersect the camera plane at s_{mk}, $k = 1, 2, \ldots, K$; $K \leq N$ and N represents the total number angular samples per hogel. Please note that k is used for voxel indexing. As proposed in [14], one can define these new LF samples using the captured discrete LF $L_1[m, i]$ as

$$\tilde{L}_1(m\Delta_x, s_{mk}) = \sum_{i\in V^{mk}} L_1[m, i], \tag{12}$$

where V^{mk} denotes the set of indices i for which the emission points corresponding to captured rays for hogel m are inside the voxel k. These sets of indices can be found as

$$V^{mk} = \frac{1}{\Delta_s}\left[m\Delta_x - \frac{d}{Z^{mk}}(X^{mk} - m\Delta_x)\right], \tag{13}$$

where (X^{mk}, Z^{mk}) is the set of original ray emission coordinates that are quantized on to the point $(\tilde{x}_{mk}, \tilde{z}_{mk})$ representing the voxel k for hogel m. As demonstrated in Figs. 3(c) and 3(d), this summation based mapping procedure can create originally nonexistent intensity variations along the angular dimension, i.e. on the s plane. For example, the value of $\tilde{L}_1(m\Delta_x, s_{m5})$ is the sum of



(a) (b) (c) (d)

Fig. 3. General structure of the ray separation based speckle suppression method. (a) Assigning emission coordinates to the captured light rays, (b) quantizing each emission point to a voxel (highlighted in red) center point, (c) detailed look at the original and quantized ray distributions, (d) assigning quantized ray intensities based on summation of original ray intensities within the voxels.

captured light rays corresponding to $i = 6$ and $i = 7$, whereas $\tilde{L}_1(m\Delta_x, s_{m1})$ is obtained from the single ray corresponding to $i = 1$. Thus, the new LF values vary in relation to the number of rays within the voxel. This creates undesired intensity variations along the quantized scene recorded on the hologram, which are observed as dark and light regions along the scene surfaces in the perceived images, as will be demonstrated in section 5.

4. Proposed speckle suppression method

The speckle reduction method we propose addresses the issues related to light propagation between the quantized voxels and hologram segments highlighted in the previous section. The problem is how to accurately obtain the desired unknown samples $\tilde{L}_1(m\Delta_x, s_{mk})$ from the known data samples $L_1[m, i]$. The accurate signal processing oriented solution for the problem is, thus, to reconstruct the continuous function (i.e. continuous LF between $m\Delta_x$ and the camera plane s, $L_1(m\Delta_x, s)$) and resample that function at the sample positions of $s = s_{mk}$. Such a solution is made available by using DSLF capture and then obtaining $\tilde{L}_1(m\Delta_x, s_{mk})$, $k = 1, 2, \ldots, K$ from $L_1[m, i]$ via linear interpolation.

The resampling procedure is performed for each hologram segment as follows. For the light ray emitted from the quantized point $(\tilde{x}_{mk}, \tilde{z}_{mk})$ corresponding to voxel k for hogel m, first the corresponding intersection point on the camera plane s_{mk} is obtained as

$$s_{mk} = m\Delta_x - \frac{d(\tilde{x}_{mk} - m\Delta_x)}{\tilde{z}_{mk}}. \tag{14}$$

The surrounding ray indices i_1 and i_2 are then acquired from the two nearest camera plane coordinates of the captured set of rays i as

$$i_1 = \underset{i \leq s_{mk}/\Delta_s}{\arg\min} \, |i\Delta_s - s_{mk}|, \tag{15}$$

$$i_2 = \underset{i > s_{mk}/\Delta_s}{\arg\min} \, |i\Delta_s - s_{mk}|. \tag{16}$$

Finally, utilizing the corresponding LF samples $L_1[m, i_1]$ and $L_1[m, i_2]$, the intensity value of $\tilde{L}_1(m\Delta_x, s_{mk})$ is obtained through linear interpolation, i.e.

$$\tilde{L}_1(m\Delta_x, s_{mk}) = \frac{L_1[m, i_1](i_2\Delta_s - s_{mk}) + L_1[m, i_2](s_{mk} - i_1\Delta_s)}{\Delta_s}. \tag{17}$$

When compared to the voxel mapping in Eq. (12), the major difference is that our solution provides a structured signal processing framework enabling accurate calculation of light propagation between the quantized voxels and hogels.

The DSCP object field, similarly to Eq. (1), is obtained from the resampled LF samples as

$$O(x) = \sum_m \mathrm{rect}\left(\frac{x - m\Delta_x}{\Delta_x}\right) \sum_k \frac{\sqrt{\tilde{L}_1(m\Delta_x, s_{mk})}}{\tilde{r}_{mk}} \exp\left[\frac{j2\pi}{\lambda}\left(\sqrt{(x - \tilde{x}_{mk})^2 + \tilde{z}_{mk}^2} - \tilde{z}_{mk}\right)\right], \tag{18}$$

where $\tilde{r}_{mk}$ is the distance between the centers of voxel k and hogel m. This process generates a hologram of the quantized version (in terms of point source locations) of the recorded scene. In order to achieve effective speckle noise reduction, the distance between adjacent scene points recorded on the hologram should be increased as explained in section 3, i.e. from the complete set of quantized points (corresponding light rays), only sparse sets are included in separate frames. The separation is done in horizontal and vertical direction by including every Nth row and column of the quantization grid. Each hologram frame generated from such set of light rays is displayed (or propagated) separately, in sequence. The end result is a combination of the speckle suppressed frames containing parts of the scene corresponding to the sparse set of light rays. Combining these frames results in a speckle-reduced reconstruction of the entire scene as a collection of the quantized point emitted light rays.

5. Experiments

The validity of the proposed method is evaluated by computational simulations including comparisons to random averaging and ray separation method [14]. For the proposed method, the DSCP object wavefields are generated using the rays sampled away from the hologram plane (at z_0) in the form of multiperspective images recentered with respect to the hologram plane. These images satisfy the DSLF criterion given by Eq. (2) for the given scene, hogel size Δ_x and camera sampling distance Δ_s. For the random averaging and ray separation methods, the rays are sampled by conventional pinhole cameras placed at the hogel centers. The angular sampling rate is chosen in accordance with the DSLF capture setup to be $2\tan^{-1}(\Delta_s/(2z_0))$. The images and depth maps are acquired by the 3D-modeling software Blender [22]. Only the green color channel is utilized in hologram calculations with the corresponding wavelength $\lambda = 534$ nm.

In order to evaluate the different speckle suppression methods more reliably, only the complex object wave is utilized, i.e. the reconstruction noise that would otherwise be introduced by the conjugate object wave is avoided. The HVS viewing process is simulated by obtaining the perceived image $I(u, v)$ by the viewer via the Fresnel diffraction model as [21]

$$I(u, v) = |\mathcal{F}_l\{T(s, t)\mathcal{F}_{z_{eye}}\{O_{DSCP}(x, y)\}\}|^2, \tag{19}$$

where $T(s,t)$ is the lens transfer function of the human eye and $\mathcal{F}_z\{\cdot\}$ is the Fresnel propagation operation by distance z. The eye is considered as a camera with a circular aperture and a thin lens placed at a distance z_{eye} from the hologram plane. The distance between the lens and sensor (i.e. pupil and retina) l is fixed at 25 mm. The eye is focused at distance d_f by choosing the focal length f as

$$f = \left(\frac{1}{d_f} + \frac{1}{l}\right)^{-1}. \tag{20}$$

The speckle suppression capabilities of three different methods (random averaging, ray separation [14], and proposed accurate ray separation) are first evaluated by comparison of speckle contrasts. The speckle contrast C is defined as [2]

$$C = \frac{\sigma}{\tilde{I}}, \tag{21}$$

where σ is the standard deviation and $\tilde{I}$ is the mean intensity of the reconstructed image. That is a better speckle suppression method results in lower speckle contrast. Three scenes are utilized for this purpose, each consisting of a planar monochromatic object placed 6 mm, 9 mm or 12 mm behind the hologram plane. The hologram parameters as well as the LF capture parameters for the proposed method are given in Table 1. The rays used by the random averaging and ray separation methods are sampled on the hologram plane with the angular sampling step of 0.29°. The multiplexing factor for the ray separation and proposed methods is chosen to be 4×4 and the random averaging method is applied with 16 frames, i.e. in all cases 16 different reconstructions are superposed intensity-wise to obtain the final reconstructed image. The lateral quantization step is chosen to be hogel size, i.e. $\Delta_{\tilde{x}} = \Delta_x = 64$ μm. The simulated human eye is set to be 300 mm away from the hologram plane and the pupil size is set as 3 mm. The eye is focused on the object surface in each case.

Table 1. Parameters of the CGH and LF capture for the car scene.

Hologram		LF capture	
Parameter	**Value**	**Parameter**	**Value**
Segment size (Δ_x)	64 μm	Capture distance (d)	300 mm
Number of pixels	8192×8192	Camera sampling (Δ_s)	1.5 mm
CGH size	16.4×16.4 mm	Number of views	35×35
Pixel pitch	2 μm	Camera-sensor distance (l)	36.6 mm
Wavelength (λ)	534 nm	Sensor sampling (Δ_u)	7.8 μm

The simulated reconstructed images in Figs. 4(a)–(d) together with the speckle contrasts presented in Table 2 demonstrate the speckle suppression performance of each method. The ray separation and its proposed accurate version suppress speckles more effectively than the random averaging method. However, the ray separation method produces varying results depending on the depth of the planar object (as seen in top and middle row of Fig. 4(c)), which is an undesirable feature in the case of more realistic scenes with 3D objects. On the other hand, the proposed method successfully suppress speckle in all simulated depth cases without introducing periodic intensity variations on the object surface. The speckle contrast values agree with the visual analysis.

In order to properly evaluate the speckle suppression capabilities in a more realistic scenario, another experiment utilizing a scene containing a 3D object is performed. The scene and LF capture setup for the second experiment is shown in Fig. 5. The LF capture setup for the proposed method as well as the ray sampling parameters for the random averaging and ray separation methods are the same as in the previous experiment. From the captured data, four different

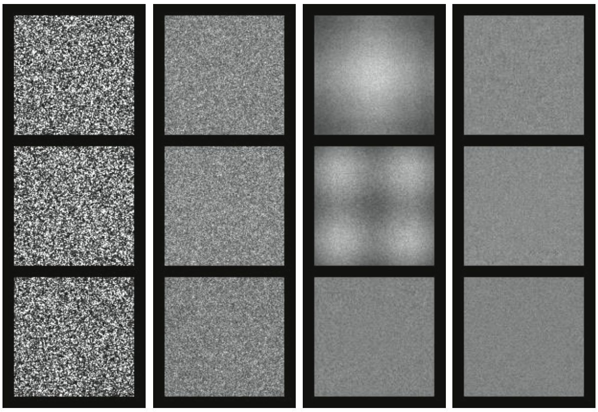

(a) (b) (c) (d)

Fig. 4. The reconstructed images via the viewing simulation for different speckle suppression methods and object distances. The plane distance from the hologram plane is 6 mm (top row), 9 mm (middle row) or 12 mm (bottom row). The speckle suppression methods used: (a) without speckle reduction, (b) random averaging, (c) ray separation [14] and (d) proposed method.

Table 2. Speckle contrasts of each different scene and speckle suppression method.

Plane distance	Without	Random averaging	Ray separation	Proposed
6 mm	0.649	0.163	0.198	0.054
9 mm	0.646	0.164	0.142	0.049
12 mm	0.652	0.164	0.054	0.045

variants of the DSCP object field are generated, i.e. one without speckle suppression, one utilizing random averaging, one with the ray separation method [14] and one with the proposed method. For the ray separation as well as the proposed method, the scene is quantized with quantization steps $\Delta_{\tilde{x}} = \Delta_x = 64\ \mu$m and $\Delta_{\tilde{z}} = 0.20$ mm in accordance with Eq. (9) and Eq. (10), respectively. The final hologram reconstructions are again obtained by intensity-wise summation of 16 reconstructions for random averaging and 4×4 sparse object reconstructions for the ray separation and proposed methods. The HVS viewing process is simulated for each hologram from three different view positions: $(-15, -15)$ mm, $(0, 0)$ mm and $(15, 15)$ mm. The aperture diameter T of the eye is chosen again as 3 mm and it is focused at the hologram plane, i.e. $d_f = 300$ mm.

The simulation results, along with the reference views, are shown in Fig. 6. The reference view $I_{ref}(u, v)$ simulates the aperture effects of the human eye and is generated as a superposition of elementary apertures, i.e. as a sum of several pinhole images within the extent of the lens. Each of these views is then compared against the corresponding reference view and the peak signal-to-noise ratio (PSNR) is evaluated as the visual image quality criterion. The results are presented in Table 3. Please note that as the dynamic ranges of reconstructed images by different methods are different, the PSNRs are calculated after mean normalization with respect to the reference image.

The simulation results show that both the ray separation and its proposed accurate version suppress the speckle noise effectively. However, the basic ray separation method images suffer

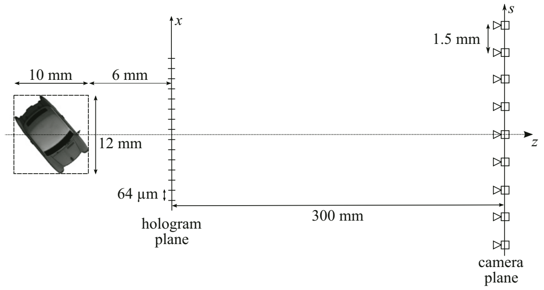


Fig. 5. Scene and LF capture setup for the second experiment utilizing the Pony car model. (The 3D model Pony Cartoon by Slava Zhuravlev is licensed under CC BY 4.0.)

from undesirable intensity patterns on the object surfaces due to its simplistic voxel mapping solution. Though the speckle patterns are adequately suppressed locally, the depth-dependent patterns on the reconstructed views of the scene degrade the overall visual quality. This can be clearly seen in the zoomed-in images shown in Fig. 6. On the other hand, the proposed solution alleviates these issues due to its more accurate and robust mapping approach. The analysis of the perceived visual quality is supported by the corresponding PSNR values shown in Table 3 as the proposed method achieves the best results in all three views.

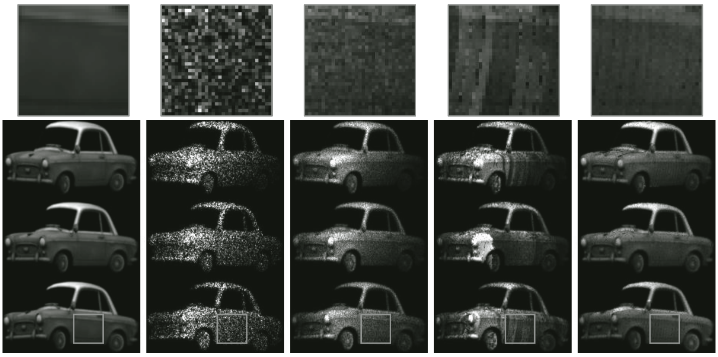


Fig. 6. Reconstructed images obtained via the viewing simulation. The viewer position is $(-15, -15)$ mm in the top row, $(0, 0)$ mm in the middle row and $(15, 15)$ mm in the bottom row. From left to right: reference image, no speckle suppression, random averaging, ray separation [14], and proposed accurate ray separation.

Table 3. PSNRs (dB) for the view images corresponding to different methods.

View	Without	Random averaging	Ray separation	Proposed
$(-15, -15)$ mm	13.93	18.66	17.68	27.74
$(0, 0)$ mm	14.86	19.47	16.32	27.59
$(15, 15)$ mm	15.36	20.19	19.13	27.91

Furthermore, as the proposed method preserves the spectral content of the underlying incoherent image that is to be perceived by the viewer (as in the case of random averaging), no resolution is lost during the multiplexing procedure [14]. This is further demonstrated in Fig. 7 comparing detailed regions of the reference image with the reconstructed image obtained by using the proposed method.

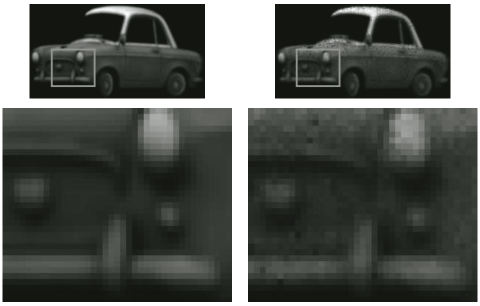


Fig. 7. Resolution comparison of reference image (left) and the reconstructed image for the proposed method (right), magnifying the detail on the front of the car.

The quantization of emission coordinates to the voxel grid can cause occlusion related issues in certain areas. The problem is mostly present in areas where the emission points cross quantization step in depth. In such cases, several emission points can be quantized to voxels that have same lateral position but are at different depths. As some of these voxels are actually occluded, including all of them in the hologram calculations causes errors in the reconstructed images, and therefore needs to be taken into account. As utilized in our method, the simplistic approach of including only the front-most of such voxels provides a reasonable solution for this problem. Small amounts of error, nonetheless, remains in the reconstructed images in the form of dark stripes at locations corresponding to quantized depth transitions. This can be seen in the right-most zoomed-in image shown in Fig. 6.

6. Conclusion

We have presented an improved speckle suppression method for coherent stereograms that is mainly based on the light ray separation technique previously proposed in [14]. The scene points corresponding to captured rays are first quantized on to a 3D uniform grid resulting a voxel-based representation. Then the scene is described as a superposition of sparse sets of voxels that are obtained by undersampling the grid in the lateral directions. The holographic reconstruction is performed by incoherent (intensity-wise) superposition of several reconstructions corresponding to such sparse sets of voxels.

It has been demonstrated with the numerical simulations that the speckle suppression is successfully achieved for 3D scenes. The speckle suppression capability has been shown to be significantly better than the random averaging approach that also uses time-multiplexed reconstruction. The accurate ray resampling enabled with the notion of DSLF provides an accurate tool for calculating the light propagation between the quantized sparse scene points, i.e. voxels, and the hogels. By this way, speckle suppression is achieved without introducing sampling related artifacts in the reconstructed images. This is the main improvement over the ray separation method proposed in [14].

The current implementation utilizes a simplistic approach in dealing with the occluded voxels. In particular, if there are multiple voxels with same lateral positions but different depths, all voxels but the front-most one are ignored. This approach has been shown to provide a reasonable solution. However, there still remain artifacts in the reconstructed images in the form of dark stripes at locations corresponding to quantized depth transitions of the object. A more sophisticated treatment of this occlusion issue can, thus, further improve the reconstruction quality.

PUBLICATION
III

Analysis of accommodation cues in holographic stereograms

J. Mäkinen, E. Sahin, and A. Gotchev

In *2018 – 3DTV-Conference*, IEEE, 2018, pp. 1-4
DOI: 10.1109/3DTV.2018.8478586

ANALYSIS OF ACCOMMODATION CUES IN HOLOGRAPHIC STEREOGRAMS

Jani Mäkinen, Erdem Sahin, Atanas Gotchev

Laboratory of Signal Processing, Tampere University of Technology, Tampere, Finland

ABSTRACT

The simplicity of the holographic stereogram (HS) makes it an attractive option in comparison to the more complex coherent computer generated hologram (CGH) methods. The cost of its simplicity is that the HS cannot accurately reconstruct deep scenes due to the lack of correct accommodation cues. The exact nature of the accommodation cues present in HSs, however, has not been investigated. In this paper, we provide analysis of the relation between the hologram sampling properties and the perceived accommodation response. The HS can be considered as a generator of a discrete light field (LF) and can thus be examined by considering the light ray oriented nature of the hologram diffracted light. We further support the analysis by employing a numerical reconstruction tool simulating the viewing process of the human eye. The simulation results demonstrate that HSs can provide accommodation cues depending on the choice of hologram segmentation size. It is further demonstrated that the accommodation response can be enhanced at the expense of loss in perceived spatial resolution.

Index Terms — Holographic stereogram, light field, accommodation

1. INTRODUCTION

In stereoscopic 3D displays, the conflict of visual cues due to their inability to deliver the accommodation (focus) cues results in the so-called vergence-accommodation conflict. Due to this problem, several alternative 3D display techniques have been proposed to resolve this issue, including light field [1, 2] and holographic displays [3, 4]. These methods can provide a wide variety of the visual cues required for 3D perception, including motion parallax, occlusions and accommodation cues. In holographic display methods the coherent accurate representations, such as the phase-added stereogram [5], rely on intrinsic information about the object shape and location, either through a model description (e.g. point cloud or polygon mesh) or depth information (e.g. depth maps). The accurate location allows such methods to provide accurately the focus cues even for scenes with pronounced depth (i.e. deep scenes) [6]. However, the requirement for object information limits the usability of these methods often to synthetic scenes or to a cumbersome capture procedure and equipment for a real scene. Incoherent image-based CGHs provide a simplified solution to this issue at the cost of 3D reconstruction accuracy. In particular, the HS [7] is a widely utilized model due to its use of only images and fast computation. However, it is widely acknowledged that the HS cannot provide correct accommodation cues in the case of deep scenes [8, 9, 10]. The exact properties of the accommodation cues provided by HSs based on the parameters of the hologram, in addition to the range for which the correct accommodation response is achieved, has not been provided yet.

The objective of this paper is to study the accommodation properties of HSs. The characterization of such properties has been analyzed in [11] for 3D ray-based LF displays, which rely on ray optics formalism of light. Here we carry out a similar analysis to examine the effects of hologram parameters on the accommodation response of HSs. The theoretical analysis is supplemented with simulations of a human viewer perceived images, obtained through a wavefield propagation-based pipeline taking into account the properties of the human visual system (HVS). The results of this paper can be utilized as guidelines for determining the parameters of HSs and the recorded scene in accommodation critical applications.

2. HOLOGRAPHIC STEREOGRAM

The HS represents the content recorded on the hologram, i.e. the complex wavefield of the 3D scene, as a collection of planar wave segments, which together form a piecewise approximation of the complete wavefield. Thus, the HS object field (in 1D for simplicity) is defined along the hologram plane x as

$$O_{\mathrm{HS}}(x) = \sum_{m=1}^{M} \mathrm{rect}\left(\frac{x - m\Delta x}{\Delta x}\right) \times \sum_{p=1}^{P} \sqrt{L[m,p]}\exp\left(j2\pi f_x^{mp} x\right), \tag{1}$$

where rect is the rectangular function, Δx is the holographic element (hogel) size, M is the total number of hogels, P is the total number of LF samples within a hogel, f_x^{mp} is the spatial frequency value and $L[m,p]$ is the intensity of a discrete LF sample at $m\Delta x$ and $p\Delta u$ (see Fig. 1). The planar wavefronts are emitted to different directions from the hologram plane based on their spatial frequencies f_x, which can be interpreted as diffraction angles θ_x along the x-axis, based on the grating equation

$$\sin\theta_x = \lambda f_x, \tag{2}$$

where λ is the wavelength of the monochromatic light. That is, by interpreting the set of plane wave segments as a set of light rays according to their propagation direction, the HS can be considered to be information-wise equivalent to a discrete LF, where the hologram plane is sampled at the hogel centers and a collection of light rays is emitted in directions specified by the hologram parameters. Each hogel contains several hologram pixels, thus segmenting the hologram. The hologram generation can be accelerated by realizing the inner sum in Eq. 1 as an inverse discrete Fourier transform. Thus, fast Fourier transform (FFT) based implementation is very commonly utilized and should, therefore, be considered in the analysis. Although in the FFT-based implementation the spatial frequency grid is fixed, which introduces quantization error, the following analysis is valid for both cases.

The hologram pixel pitch Δ_ξ defines the largest spatial frequency content the hologram is able to produce and thus the max-

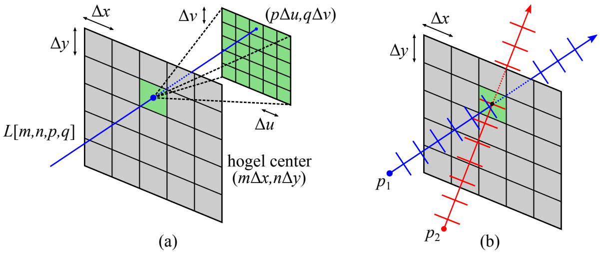


Figure 1: The HS is parametrized by discrete locations on two planes: the hologram plane (x, y) and a parametrization plane (u, v) (a). Plane wave decomposition of the HS (b).

imum diffraction angle θ_{max}, which is defined as

$$\theta_{max} = \arcsin \frac{\lambda}{2\Delta_{\xi}}. \tag{3}$$

The spatial frequency sampling within a hogel is uniform and is dictated by the size of the hogel Δx. According to Eq. 2, this corresponds to a non-uniform sampling step in the angular domain, especially for large angles. However, the main contribution to the angular sampling Δ_θ is due to the hogel size Δx and can be, therefore, estimated as

$$\Delta_\theta \approx \arcsin \left(\frac{\lambda}{\Delta x} \right). \tag{4}$$

The angular sampling is vital in terms of the accommodation response of the human eye viewing the hologram, as it determines how accurately a hogel can direct the plane waves in the desired direction. Based on the given angular resolution, in one dimension, we define the total number of planar wavefront segments (correspondingly rays) that can be accurately directed inside the eye pupil from a given hogel as

$$N_\theta = \frac{D}{z_{eye} \tan \Delta_\theta}, \tag{5}$$

where D is the pupil diameter and z_{eye} is the distance between the pupil and the hologram. On the other hand the hogel size determines the perceived spatial resolution, which is usually chosen according to the properties of the HVS. When the human eye is assumed to be a diffraction limited imaging system, the minimum resolvable distance between two points at distance z_{eye} is defined by the Rayleigh criterion as [12]

$$\Delta_x^{HVS} = 1.22 \frac{\lambda z_{eye}}{D}. \tag{6}$$

That is, for a given eye pupil size D and intended viewing distance z_{eye}, Δ_x^{HVS} is the upper limit for the hogel size which will still ensure that the perceived spatial resolution is maximized. Please note that such choice of hogel size results in $N_\theta \approx 1$ according to Eq. 5. Thus, due to the trade-off between the angular and spatial resolutions, one should sacrifice from the spatial resolution to improve the angular resolution. As will be shown in Sec. 3, this is indeed necessary to enhance the accommodation response and thus maintain correct focus cues.

Another factor to consider in the accommodation analysis is the number of planar wavefront segments emitted from multiple different hogels, for a given 3D point, and intercepted by the eye. The number of such wavefront segments N_h is defined for point at depth z_p away from the hologram plane as

$$N_h = \left| \frac{z_p}{z_{eye} - z_p} \right| \frac{D}{\Delta x}. \tag{7}$$

If the pupil size D and viewing distance z_{eye} are fixed, N_h increases as the point is moved further away from the hologram plane. Increasing hogel size, on the other hand, reduces N_h contrary to N_θ. Please note that N_h corresponds to the parameter defined as the ray density (number of rays incident in the eye pupil) in the context of ray-based LF displays [11].

3. RESULTS AND DISCUSSION

The theoretical analysis is validated by generating HSs with varying parameters and simulating the images perceived by a human eye viewing the hologram at a certain location, focusing at different depths of the scene. These view simulations are implemented by utilizing numerical Fresnel propagation [13] and simulating the HVS as a sensor (retina) behind a thin circular lens (pupil). The eye can be set to focus at different depths by altering the focal length of the lens, thus modifying its transmittance function $T(s, t)$. The image intensity values $I(u, v)$ are retrieved from the wavefield at the sensor plane as

$$I(u, v) = |\mathcal{F}_l\{T(s, t)\mathcal{F}_d\{O_{\text{HS}}(x, y)\}|^2, \tag{8}$$

where $\mathcal{F}_z$ is the Fresnel propagation operation by distance z, l is the sensor-retina distance of the simulated eye and d is its distance from the hologram plane.

We simplify the simulations to a 1D case, which is a valid assumption since the Fresnel kernel is separable. A single point source of light is placed in the recorded scene and the corresponding (1D) HS object field is generated. The process is repeated for several different depths of the point (w.r.t. the hologram plane). The simulated eye is set to focus at different depths to obtain a set of perceived point spread functions (PSFs). These can be utilized to evaluate the perceived sharpness of the point as a function of focus distance. Furthermore, by studying the modulation transfer function (MTF), which is defined as the magnitude of the Fourier transform of the PSF, at various spatial frequency ranges, the scene depth at which the eye is most likely going to focus can be determined for each different point depth. Namely, the focus distance which maximizes the MTF in the given range is estimated as the distance where the human eye would accommodate [11]. Please note that in this paper we do not consider the gradient of the MTF while assessing the accommodation distance, though a narrower MTF peak can be expected to provide a stronger accommodation trigger than a wide and shallow peak. The estimate is then compared to the correct depth of the point source and the shift in accommodation is evaluated, which can be compared to the depth-of-field of the HVS to determine the correctness of the provided accommodation cue.

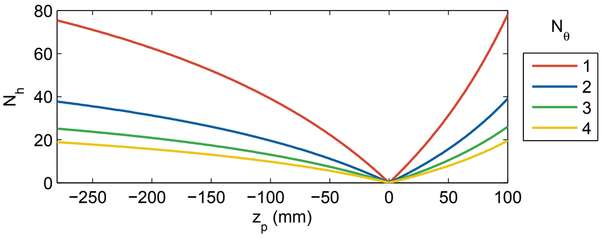


Figure 2: Values of N_h as a function of point distance from the hologram plane for four different N_θ values.

Let us utilize the following experimental setup. A point is placed at a distance z_p away from hologram plane, from 100 mm in front to 280 mm behind the hologram at 10 mm intervals. The

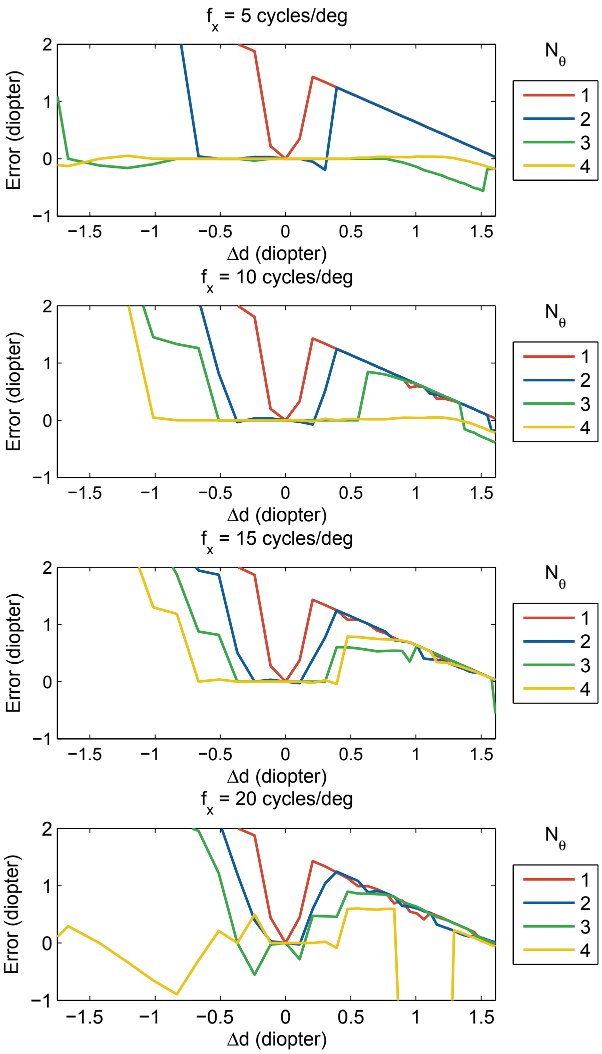


Figure 3: Accommodation shift for regular HSs. Four different spatial frequency values and four different hogel sizes (corresponding to four different N_θ values).

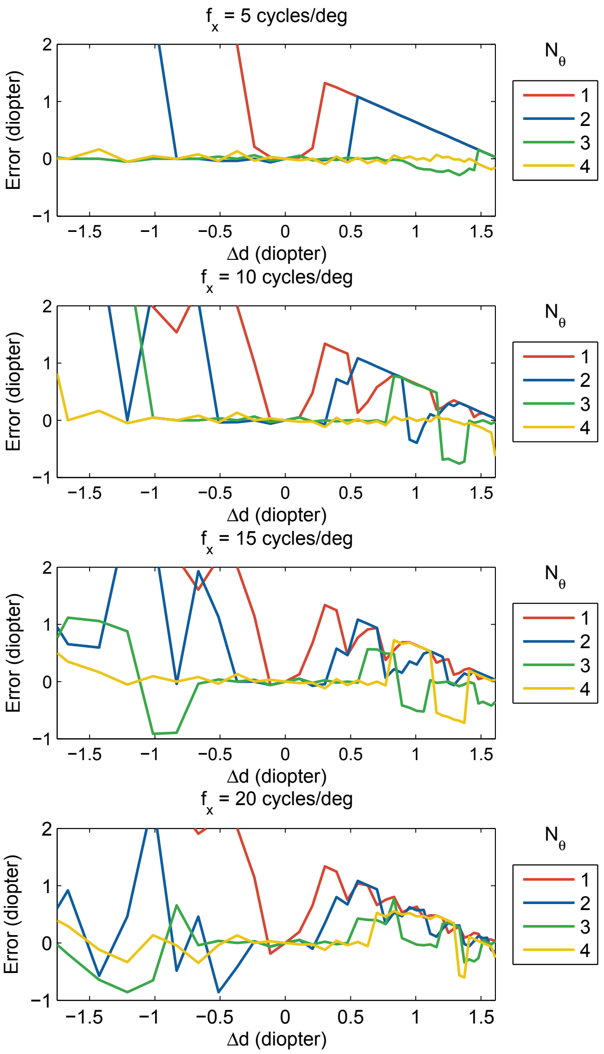


Figure 4: Accommodation shift for FFT HSs. Four different spatial frequency values and four different hogel sizes (corresponding to four different N_θ values).

z-axis is defined to be positive in front of the hologram plane (i.e. between the viewer and the hologram). The point distance with respect to the hologram plane Δd is determined (in diopters) as

$$\Delta d = \frac{1}{z_{eye}} - \frac{1}{z_{eye} - z_p}. \qquad (9)$$

Thus, the range of Δd values for the chosen point distances is approximately between -1.6 and 1.6 diopters, where 0 diopters corresponds to the depth of the hologram plane. The simulated eye (D = 5 mm) is placed at distance z_{eye} = 300 mm and is focused at a distance z_f at an oversampled grid with respect to the point locations (by a factor of 4). For each z_p and z_f, the perceived PSF is obtained through the simulation tool. We evaluate the MTF at spatial frequencies 5, 10, 15 and 20 cycles per degree (cpd), as all of these values are reproducible by the chosen HS configurations and are also relevant in terms of the HVS contrast sensitivity [14]. For a given spatial frequency above, we actually evaluate the corresponding response by integrating the set of MTF values in the range of [-2.5, 2.5] cpd. As a result, the MTF values are assessed as a function of z_p and z_f for a discrete set of spatial frequencies. The value of z_f maximizing the MTF is evaluated, and thus, for each z_p an estimate of the likely HVS focus distance $\hat{z}_f$ is obtained. The simulation is replicated for different hogel sizes, corresponding to four different N_θ values from 1 to 4. For the specified point depth range [-280, 100] mm with respect to hologram plane, the corresponding ranges of N_h values for these four cases of N_θ are given in Fig. 2.

The experimental results for regular HSs are presented in Fig. 3. As can be seen from the accommodation shift values, increasing the hogel size preserves the correct accommodation cues further away from the hologram plane. On the other hand, the smallest hogel size (corresponding to slightly denser sampling than Δ_x^{HVS}) has large accommodation shift already at very small depth values, indicating the lack of correct accommodation cues immediately as the objects are moved away from the hologram plane. Furthermore, the general trend regarding the relation between hogel size and accommodation shift is present also in FFT-based HS, as seen in Fig. 4. If the shift in accommodation exceeds a threshold of 0.3 diopters, the accommodation cue is outside the HVS depth-of-field and can be considered to be incorrect [15]. The results for the low spatial frequencies (5 and 10 cpd) suggest that even two views within the pupil extent can provide correct accommodation cues, though at a very limited range (approximately between -0.6 . . . 0.4

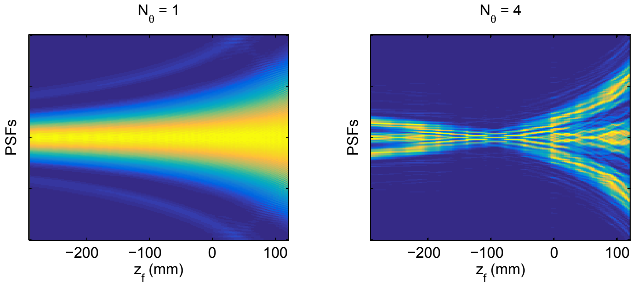


Figure 5: Comparison of perceived PSFs given two different hogel sizes corresponding to N_θ = 1 and N_θ = 4 at different values of z_f for a point at z_p = -100 mm.

diopters away from hologram plane). Significant improvement in terms of accommodation is gained with hogel sizes corresponding to N_θ = 3 and N_θ = 4, which are able to maintain correct cues within (almost) the complete evaluated range of -1.6 . . . 1.6 diopters away from hologram plane. It should be noted, though, that in these cases the perceived spatial resolution is reduced by a factor of 0.4 or 0.3 relative to the Rayleigh diffraction limit, effectively increasing the pixel size by 2.5 and 3.3 times in comparison to what the HVS could resolve. On the other hand, simulation results also indicate that at this range of hogel sizes, the angular resolution is the dominant factor in determining the accommodation response. Thus, increasing N_h has no similar positive effect on the accommodation response as in the case of ray-based LF displays [11], where the angular resolution of lenses (which correspond to hogels in HSs) is quite high compared to HSs.

The noise in the two larger spatial frequency results makes them less conclusive, yet a similar ordering between hogel sizes and general trend as in the lower frequency results can be seen, albeit the accommodation shift exceeds the acceptable threshold already at lower point depth values. Interestingly, in the case of N_θ = 1 and N_θ = 2, the accommodation shifts start to drop for point depths after around 0.4 diopters. Fig. 5 further illustrates different PSFs (vertical slices) for different focused depths given in the horizontal axis. In this figure it can be seen more clearly that for low angular resolution, sharper PSFs are observed at further (more negative) focal depths, which results in failure of producing correct accommodation cues. Further analysis of this effect is required. It is, however, outside the scope of this paper and as such left as future work. For higher N_θ, on the other hand, the behaviour of PSFs justifies the availability of focus cues.

4. CONCLUSIONS

In this paper we have presented analysis of the accuracy of accommodation cues in HSs. Specifically, we have examined the effects of hogel size by utilizing the ray-based interpretation of the hologram and characterizing the number of wavefront segments (corresponding to light rays) incident the viewer's pupil extent. The numerical simulations employing Fresnel propagation to evaluate the perceived accommodation shift have verified the analysis assumptions, thus indicating the presence of correct accommodation cues at a limited range based on the chosen hogel size. Critically, increasing the number of wavefront segments seen from a single hogel by the eye improves such cues. Furthermore, when compared to the minimum resolvable distance between two points of the human eye, a trade-off between accommodation and perceived spatial resolution is seen. That is, if the perceived spatial resolution is maximized by choosing the hogel size according to the Rayleigh criterion, the depth range for correct accommodation cues is extremely limited, and extending this range requires sacrificing spatial resolution. Importantly, we have demonstrated that correct accommodation cues can be provided by HSs. Moreover, HS displays that can provide a limited accommodation response to the viewer can be designed, thus permitting their use in applications where accurate 3D perception particularly regarding accommodation is required, without needlessly increasing the complexity of the model by including model information.

5. REFERENCES

[1] Y. Takaki, "High-density directional display for generating natural three-dimensional images," *Proceedings of the IEEE*, vol. 94, no. 3, pp. 654–663, 2006.

[2] D. Teng, Z. Pang, Y. Zhang, D. Wu, J. Wang, L. Liu, and B. Wang, "Improved spatiotemporal-multiplexing super-multiview display based on planar aligned OLED microdisplays," *Opt. Express*, vol. 23, no. 17, pp. 21549–21564, Aug 2015.

[3] F. Yaraş, H. Kang, and L. Onural, "Real-time phase-only color holographic video display system using LED illumination," *Applied Optics*, vol. 48, no. 34, pp. H48–H53, 2009.

[4] H. Zhang, N. Collings, J. Chen, B. A. Crossland, D. Chu, and J. Xie, "Full parallax three-dimensional display with occlusion effect using computer generated hologram," *Optical Engineering*, vol. 50, no. 7, pp. 074003–074003–5, 2011.

[5] M. Yamaguchi, H. Hoshino, T. Honda, and N. Ohyama, "Phase-added stereogram: calculation of hologram using computer graphics technique," in *Proc. SPIE*, 1993, vol. 1914, pp. 25–31.

[6] Q. Y. J. Smithwick, J. Barabas, D. E. Smalley, and V. M. Bove, Jr., "Interactive holographic stereograms with accommodation cues," in *Proc. SPIE*, 2010, vol. 7619, p. 761903.

[7] J. T. McCrickerd and N. George, "Holographic stereogram from sequential component photographs," *Applied Physics Letters*, vol. 12, no. 1, pp. 10–12, 1968.

[8] M. W. Halle, "Holographic stereograms as discrete imaging systems," in *Proc. SPIE*, 1994, vol. 2176, pp. 73–84.

[9] P. St. Hilaire, "Modulation transfer function and optimum sampling of holographic stereograms," *Applied Optics*, vol. 33, no. 5, pp. 768–774, Feb. 1994.

[10] K. Wakunami and M. Yamaguchi, "Calculation for computer generated hologram using ray-sampling plane," *Opt. Express*, vol. 19, no. 10, pp. 9086–9101, May 2011.

[11] H. Huang and H. Hua, "Systematic characterization and optimization of 3D light field displays," *Opt. Express*, vol. 25, no. 16, pp. 18508–18525, Aug. 2017.

[12] J. W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, 2nd edition, 1996.

[13] D. P. Kelly, "Numerical calculation of the Fresnel transform," *Journal of the Optical Society of America A*, vol. 31, no. 4, pp. 755–764, Apr. 2014.

[14] F. W. Campbell and J. G. Robson, "Application of Fourier analysis to the visibility of gratings," *The Journal of Physiology*, vol. 197, no. 3, pp. 551–566, 1968.

[15] S. Marcos, E. Moreno, and R. Navarro, "The depth-of-field of the human eye from objective and subjective measurements," *Vision Research*, vol. 39, no. 12, pp. 2039 – 2049, 1999.

PUBLICATION
IV

Computational coherent imaging for accommodation-invariant near-eye displays

J. Mäkinen, E. Sahin, U. Akpinar, and A. Gotchev

In *2021 IEEE International Conference on Image Processing*, IEEE, 2021, pp. 3433-3437
DOI: 10.1109/ICIP42928.2021.9506773

COMPUTATIONAL COHERENT IMAGING FOR ACCOMMODATION-INVARIANT NEAR-EYE DISPLAYS

Jani Mäkinen, Erdem Sahin, Ugur Akpinar, and Atanas Gotchev

Faculty of Information Technology and Communication Sciences, Tampere University,
Korkeakoulunkatu 1, Tampere, Finland

ABSTRACT

We present a computational accommodation-invariant near-eye display, which relies on imaging with coherent light and utilizes static optics together with convolutional neural network-based preprocessing. The network and the display optics are co-optimized to obtain a depth-invariant display point spread function, and thus relieve the conflict between accommodation and ocular vergence cues that typically exists in conventional near-eye displays. We demonstrate through simulations that the computational near-eye display designed based on the proposed approach can deliver sharp images within a depth range of 3 diopters for an effective aperture (eyepiece) size of 10 mm. Thus, it provides a competitive alternative to the existing accommodation-invariant displays.

***Index Terms*—** Computational near-eye displays, Optics, Neural networks, Coherent imaging

1. INTRODUCTION

In the pursuit of realistic recreation of 3D content in near-eye displays (NED), depth cues are of vital importance. Conventional NEDs provide the binocular vergence and disparity cues by displaying a separate stereoscopic view for each eye. However, in such displays the monocular accommodation and retinal defocus blur cues are typically missing. As a result, such displays suffer from the well-known phenomenon called vergence-accommodation conflict (VAC), which creates visual discomfort and fatigue [1].

Solving the VAC can be approached by either providing the depth cues accurately (i.e., accommodation enabling displays) or by making the display accommodation-invariant (AI). The former can be achieved, for instance, through varifocal [2], multifocal plane [3] or light field (LF) display [4, 5], all of which usually work in the incoherent imaging regime. In the case of coherent imaging, holographic displays are capable of producing accommodation enabling content with correct defocus blur cues, due to their accurate recreation of the complete wavefield [6, 7]. On the other hand, AI displays provide depth-independent defocus blur cue, and thus the accommodation is mainly (cross-)driven by the disparity signal [8]. By this way, the VAC is avoided as both the vergence and accommodation are consistently driven by the same signal. In a Maxwellian-view display [9], such accommodation response can be achieved by focusing the light rays (or image) onto the entrance pupil of the eye, thus projecting the image to the retina through an aperture effectively smaller than the eye pupil size. The display point spread function (PSF) at the retina can be also designed to be depth-invariant by using varifocal optics, where the display content is projected rapidly at different depths at a rate faster than the human temporal resolution [10]. Here the effective pupil size is equal to the full pupil size, thus avoiding the limitations in Maxwellian-view displays regarding, e.g., light throughput and eyebox.

Alternatively, an attractive approach to AI displays is to employ end-to-end learning to co-design the optics of the display and the preprocessing algorithm, e.g., where the former design (optimization) component is chosen to be a static diffractive optical element (DOE), working in tandem with a refractive lens, and the latter one is implemented via a convolutional neural network (CNN) [11, 12]. Here we propose a computational AI NED based on similar end-to-end learning based design framework, considering the promising results reported in those works despite the simplicity of the optical setup. The main distinction of our approach is that it relies on coherent imaging, employing coherent light sources, which can provide crucial advantages in designing AI displays compared to its incoherent counterparts [11, 12]. Importantly, coherent imaging can provide wider depth of field (DoF) compared to incoherent imaging assuming same imaging optics [13]. Furthermore, coherent illumination ensures operation on distinct wavelengths, whereas incoherently illuminated displays operate in broadband spectrum. As typically three channel (red, green, blue) images are utilized in the modeling and design of display systems, in the former case the wavelength-dependent behaviours of the display system components can be more accurately incorporated in the display design process. Though the coherent illumination for the display imposes further constraints and issues to be addressed, e.g, speckle noise, the inherently larger DoF is demonstrated to ease the end-to-end optimization task.

The paper is partially supported by the Academy of Finland research project “Modeling and Visualization of Perceivable Light Fields”, decision number 325530.

2. COMPUTATIONAL NEAR-EYE DISPLAY

Our display design comprises of a coherently illuminated display panel and the display optics, which are a combination of a refractive lens and a DOE (to provide desired phase modulation), as depicted in Fig. 1. The aim of an AI NED is to produce as sharp images as possible within a certain accommodation depth range, thus the goal is to find the optimal set of display optics to achieve such a response. We can extend the display design by incorporating a computational element in the form of a preprocessing algorithm. The algorithm further assists in achieving an AI display response by compensating for the effect of the display optics. This is analogous to the post-processing deblurring step of similar computational camera designs [14]. To find the optimal combination of display optics (i.e. DOE) and preprocessing algorithm (CNN), we propose utilizing a learning based approach; the design (optimization) procedure of the computational NED system is illustrated in Fig. 2. During the design procedure, the optics and the preprocessing CNN are jointly optimized utilizing a set of training images, such that the display response is as sharp as possible across the desired depth range. Such depth range is covered by randomly changing the accommodation distance z during learning.

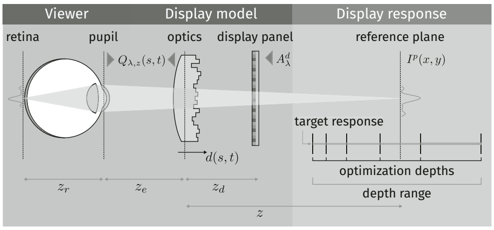


Fig. 1. Display design and imaging model.

As illustrated in Fig. 2, the training starts with feeding the sharp input image I to the CNN, which pre-processes it and outputs the display amplitude A^d. We then simulate, based on the current state of the optics, the image I^p as it would be perceived by a viewer given the distance of accommodation. The perceived image is compared against the ground truth sharp image using the loss function $\mathcal{L}\{I, I^p\}$. The loss function drives the optimization procedure based on its gradient, optimizing the entire display system. The preprocessing network in our proposed method is based on the well-known U-net architecture [15], similar to our previous work [11].

2.1. Computational display model

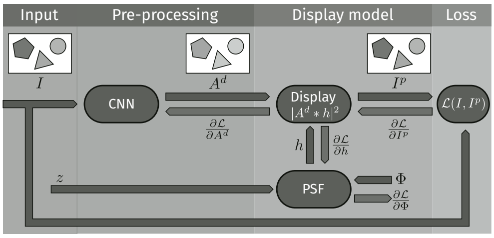


Fig. 2. The proposed learning based computational NED system design approach.

We simulate the viewing process of the proposed display design, as depicted in Fig. 1, by assuming an aberration-free eye accommodating at distance z. Assuming a planar retina, the image can be found on a conjugate plane, denoted here as the reference plane (x,y). Assuming monochromatic illumination at wavelength λ, the coherent system PSF $h_{\lambda,z}(x,y)$ on the reference plane is

$$h_{\lambda,z}(x,y) \propto \mathcal{F}\left\{ Q(s,t)|_{\left(\frac{x}{\lambda z}, \frac{y}{\lambda z}\right)} \right\}, \tag{1}$$

where $Q_{\lambda,z}(s,t)$ is the generalized pupil function and $\mathcal{F}\{\cdot\}$ is the Fourier transform operator. The pupil function is obtained as a combination of the (circular) aperture function $A(s,t)$, the phase delay of the DOE $\Phi_\lambda(s,t)$ and the refractive lens defocus as

$$Q_{\lambda,z}(s,t) = A(s,t) \exp\left(j\Phi_\lambda(s,t)\right) \exp\left(j\Psi_{\lambda,z}\frac{s^2+t^2}{r^2}\right), \tag{2}$$

where

$$\Psi_{\lambda,z} = \frac{\pi}{\lambda}\left(\frac{1}{z_d} - \frac{1}{z} - \frac{1}{f_\lambda}\right) r^2 \tag{3}$$

is the defocus coefficient, r is the radius of the circular lens and f_λ is the effective focal length of the refractive lens. The DOE phase delay is related to its height $d(s,t)$ through the wavelength-dependent refractive index n_λ of the DOE material as

$$\Phi_\lambda(s,t) = k\left(n_\lambda - 1\right) d(s,t). \tag{4}$$

In the coherent image formation model, the reference plane amplitude is a convolution between the coherent PSF and the source amplitude, i.e., the amplitude from the display panel $A^d_\lambda(x,y) = \sqrt{I^d_\lambda(x,y)}$. Thus, the reference plane image is obtained as

$$I^p_{\lambda,z}(x,y) = \left|A^d_\lambda(x,y) * h_{\lambda,z}(x,y)\right|^2. \tag{5}$$

The coherent image formation poses an issue not present in incoherent imaging, as the display amplitude $A^d_\lambda(x,y)$ contains additional phase values, which are randomly distributed in nature across the display pixels due to randomness in the heights of the pixels at the scale of wavelength [16]. When such values are convolved with the coherent PSF, the neighbouring points (or pixels on the display) interfere such that the resulting intensity pattern contains high frequency and contrast noise known as speckle noise. Though it significantly degrades the image quality, it can be diminished by, for instance, reducing the temporal coherence of the light source [17]. During training, however, we assume the light to be perfectly coherent and have equal phase across the display, thus discarding the speckle effects, for simplicity.

We optimize, in addition to the preprocessing network, the phase delay of the DOE during the training phase. The optimization is done for a nominal wavelength λ_0, meaning that the delay $\Phi_{\lambda_0}(s,t)$ is optimized, from which the delay for another wavelength λ is calculated as

$$\Phi_\lambda(s,t) = \Phi_{\lambda_0}(s,t)\frac{\lambda_0\,(n_\lambda - 1)}{\lambda\,(n_{\lambda_0} - 1)}. \tag{6}$$

The focal length of the underlying refractive lens f_λ is also wavelength-dependent and can be related to the nominal wavelength focal length f_{λ_0}:

$$f_\lambda = f_{\lambda_0}\frac{n_{\lambda_0} - 1}{n_\lambda - 1}. \tag{7}$$

2.2. Loss function

The loss function drives the learning process by minimizing the difference, or error, between the desired and current output of the system over time. Here we employ a combination of two metrics: the L_1 distance and the structural similarity index measure (SSIM). The aggregated loss function attempts to minimize L_1 and maximize the SSIM as a compromise between the texture detail quality and the perceived change in structural information. The overall loss is calculated as

$$\mathcal{L}(I, I^p) = \mathcal{L}_{L_1}(I, I^p) + \mathcal{L}_{\mathrm{S}}(I, I^p) + \alpha\mathcal{R}(I, I^p) + \gamma\mathcal{R}^d(I^d), \tag{8}$$

where $\mathcal{L}_{L_1}(I, I^p)$ is the L_1 distance, $\mathcal{L}_{\mathrm{S}}(I, I^p)$ is the SSIM-based loss, $\alpha\mathcal{R}(I, I^p)$ is the network output regularization and $\gamma\mathcal{R}^d(I^d)$ is the display image regularization. In order to maximize the SSIM, we define its loss function as

$$\mathcal{L}_{\mathrm{S}}(I, I^p) = 1 - \mathrm{SSIM}(I, I^p). \tag{9}$$

The regularization terms in Equation (8) serve for two purposes: the first one equalizes the intensity levels between the network output and the input, whereas the latter term applies to the dynamic range of the display image. Here the dark channel prior [18] is utilized as $\mathcal{R}(I, I^p)$, weighted by α. As we aim to obtain display images within the range $[0, 1]$, the cut regularization term punishes (i.e., increases the overall loss value) intensity values outside this range. Specifically, the function $\mathcal{R}^d(I^d)$ indicates whether a pixel is within the range or not, which is then weighted by γ to achieve the desired strictness of the regularization.

3. RESULTS

The display design is optimized for color images, where the corresponding wavelengths are chosen as $\lambda_R = 675$ nm for red, $\lambda_G = 540$ nm for green, and $\lambda_B = 450$ nm for blue. The green channel is utilized as the nominal wavelength during training. We assume physically feasible materials for the optics: fused silica for the refractive lens and S1813 photoresist film for the DOE. We set the aperture radius r as 5 mm and the distance between the optics and the display z_d as 28.6 mm. The underlying refractive lens is assumed to have an effective focal length of 30 mm at 587.6 nm wavelength, a central thickness of 2.90 mm and is set up to focus at 1.5 D at the nominal wavelength. We aim to learn AI PSF response within the depth range of 0–3 D, thus limiting the defocus coefficient in Equation (3) by $|\Psi| < 327.3$. As per the discussion in [19], this imposes a minimum display sampling rate; the spectrum of the optimized DOE operates in the same band as the defocus term, therefore band-limiting the PSF spectrum. The corresponding minimum sampling derived from the limited defocus coefficient is 6 μm. The display pixel pitch is set as 12.48 μm, corresponding to 20 cycles per degree in spatial resolution. During training, the accommodation depth is set randomly within the target depth range at 0.5 D discrete steps.

The optimized coherent display properties are presented in Fig. 3. Notably, the PSFs across the target depth range illustrate the relative sharpness when accommodating at different depths. Here the PSFs remain relatively sharp in the middle of the depth range and begin to spread out when approaching 0 and 3 D depths. Though not optimal in terms of an AI response, the proposed display setup is still capable of delivering a large DoF when combined with the optimized preprocessing network, as will be demonstrated in the simulations. Also note that in the coherent imaging regime, the final reconstructed images include not only the linear superposition of such PSFs but also their interference patterns (cross-terms).

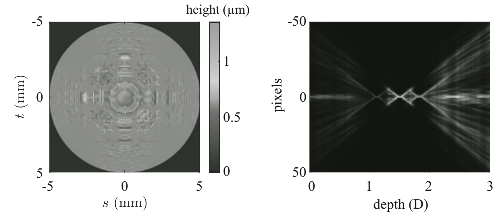


Fig. 3. The optimized DOE height map of the proposed design (left), and the 1D cross-sections of PSF intensities across the target depth range (right).

In the case of coherent light assumption, the resulting reconstructed images are naturally expected to suffer from severe speckle noise. However, in reality the available light sources, even the most coherent ones, have typically finite temporal coherence length, i.e., they are actually partially coherent. In such case, the noise contrast can be observed to be reduced compared to the ideal perfectly coherent case, as the perceived images are effectively integrated (intensity-wise) in spectrum. To take advantage of this and also to better comply with practical scenarios, in our simulations we consider partially coherent illumination. Based on the method in [16], for each color channel, we take the weighted sum of reconstructed images corresponding to different wavelengths (within the spectrum of the simulated light source) and independently applied random surface height profiles for the display phase components. The random height profiles are assumed to follow a Gaussian distribution, where the standard deviation is equal to the wavelength. We further assume that the partially coherent light source has a Gaussian profile spectral response with a standard deviation of 13 nm, which corresponds to a superluminescent LED [17].

To compare our simulation results against the state of the art approaches in computational AI NEDs, we consider three notable cases: an ideal coherent display, where the phase components of all display pixels are assumed to be zero and thus the reconstructed images have no speckle noise, a partially coherently illuminated display exhibiting speckle noise, and a time-multiplexed partially coherent display, which averages 25 frames of independent noise realizations intensity-wise to reduce the speckle noise. Such time-multiplexed reconstruction can be implemented in practice, for instance, by introducing a moving diffuser in the illumination path [20]. In addition to these, we evaluate the results of an equivalent optimized display using incoherent illumination (similar to [11]), and the AI display proposed by Konrad et. al. [10]. The latter display sweeps through the desired accommodation range using a focus tunable lens, thus resulting in the perceived PSF being modeled as the average of the individual PSFs corresponding to different focal depths. Based on the averaged PSF, the method applies a Wiener deconvolution-based preprocessing step. We simulate this approach with three different focal depths of 0.5 D, 1.5 D, and 2.5 D as a fair compromise between the spatial resolution and extended DoF range.

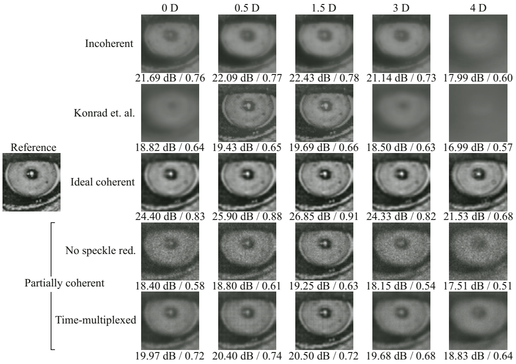


Fig. 4. Zoomed-in simulation results for the Baboon test image at five different accommodation depths, and the corresponding PSNR/SSIM values of the full image.

The zoomed-in simulated reconstructions are presented in Fig. 4 for the well-known Baboon test image, where we compare the reconstructed images with the sharp ground truth image via PSNR and SSIM. In both metrics, the ideal coherent setup achieves the highest scores, followed by the incoherent setup and then the focus tunable lens AI display. The results of the partially coherent display setups are as expected: without any additional speckle reduction methods, the overall scores are among the lowest. However, by introducing even a simplistic speckle reduction approach, the result scores improve noticeably and are relatively competitive with the two rivaling approaches. One can argue that the proposed approach produces visually better (sharper) results in some or most of the cases, despite the PSNR and/or SSIM metrics suggesting the opposite. Although, we keep the more elaborate discussion on perceptual metrics out of scope of this paper, the presented results in Fig. 4 addresses the importance of that topic, especially when comparing (reconstructed) images for displays with different characteristics of introduced distortions. Importantly, when time-multiplexing based speckle reduction method is employed, the proposed display can keep the reconstruction quality also beyond the intended depth range of 0 to 3 D that is assumed during training. Besides these inspiring results, the reconstruction quality of images observed in the ideal coherent case indicate the potential of our approach: if the speckle noise can be further eliminated, via computational or optical means, the resulting display can outperform the state-of-the-art methods.

4. CONCLUSIONS

In this work we present an AI NED that employs computational coherent imaging to address the VAC. By using jointly learned (co-optimized) preprocessing network and hybrid refractive-diffractive optics, the presented display achieves sharp display response within the intended (training) depth range of 0–3 D, which can be further kept beyond the training depth range (e.g., up to 4D) without significant loss in the reconstruction quality. The comparative analysis shows the benefits, and compromises, of replacing an incoherent imaging setup with a coherent one: the reconstructed images are noticeably sharper, however, at the cost of introducing speckle noise. We demonstrate that the reconstruction results can be significantly improved by incorporating speckle reduction methods. Additionally, as the proposed design uses static optical elements, our design provides a simpler construction in comparison, e.g., to varifocal optics based solutions. In the future work, we target including the modeling of partial coherence in the training stage, which can potentially provide better optimized results based on the coherence length of the display illumination source. Furthermore, we plan to incorporate more sophisticated learning-based speckle reduction methods, which is to be also optimized in the initial training stage.

5. REFERENCES

[1] D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, "Vergence–accommodation conflicts hinder visual performance and cause visual fatigue," *Journal of Vision*, vol. 8, no. 3, pp. 33–33, 03 2008.

[2] N. Padmanaban, R. Konrad, T. Stramer, E. A. Cooper, and G. Wetzstein, "Optimizing virtual reality for all users through gaze-contingent and adaptive focus displays," *Proceedings of the National Academy of Sciences*, vol. 114, no. 9, pp. 2183–2188, 2017.

[3] K. Akeley, S. J. Watt, A. R. Girshick, and M. S. Banks, "A stereo display prototype with multiple focal distances," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 804–813, Aug. 2004.

[4] T. Ueno and Y. Takaki, "Super multi-view near-eye display to solve vergence–accommodation conflict," *Optics Express*, vol. 26, no. 23, pp. 30703, 2018.

[5] H. Hua and B. Javidi, "A 3D integral imaging optical see-through head-mounted display," *Opt. Express*, vol. 22, no. 11, pp. 13484–13491, Jun 2014.

[6] T. Yoneyama, E. Murakami, Y. Oguro, H. Kubo, K. Yamaguchi, and Y. Sakamoto, "Holographic head-mounted display with correct accommodation and vergence stimuli," *Optical Engineering*, vol. 57, no. 6, pp. 1 – 9 – 9, 2018.

[7] A. Maimone, A. Georgiou, and J. S. Kollin, "Holographic near-eye displays for virtual and augmented reality," *ACM Trans. Graph.*, vol. 36, no. 4, July 2017.

[8] C. Schor, "A dynamic model of cross-coupling between accommodation and convergence: simulations of step and frequency responses.," *Optometry and vision science : official publication of the American Academy of Optometry*, vol. 69 4, pp. 258–69, 1992.

[9] M. Von Waldkirch, P. Lukowicz, and G. Tröster, "LCD-based coherent wearable projection display for quasi-accommodation-free imaging," *Optics Communications*, vol. 217, no. 1-6, pp. 133–140, 2003.

[10] R. Konrad, N. Padmanaban, K. Molner, E. A. Cooper, and G. Wetzstein, "Accommodation-invariant computational near-eye displays," *ACM Trans. Graph.*, vol. 36, no. 4, July 2017.

[11] U. Akpinar, E. Sahin, and A. Gotchev, "Phase-coded computational imaging for accommodation-invariant near-eye displays," in *2020 IEEE International Conference on Image Processing (ICIP)*, 2020, pp. 3159–3163.

[12] U. Akpinar, E. Sahin, and A. Gotchev, "Computational multifocal near-eye display with hybrid refractive-diffractive optics," in *2020 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, 2020, pp. 1–6.

[13] J. W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, 1996.

[14] U. Akpinar, E. Sahin, M. Meem, R. Menon, and A. Gotchev, "Learning wavefront coding for extended depth of field imaging," *IEEE Transactions on Image Processing*, vol. 30, pp. 3307–3320, 2021.

[15] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015, vol. 9351 of *LNCS*, pp. 234–241.

[16] D. G. Voelz, K. A. Bush, and P. S. Idell, "Illumination coherence effects in laser-speckle imaging: modeling and experimental demonstration," *Applied optics (2004)*, vol. 36, no. 8, pp. 1781–1788, 1997.

[17] Y. Deng and D. Chu, "Coherence properties of different light sources and their effect on the image sharpness and speckle of holographic displays," *Scientific reports*, vol. 7, no. 1, pp. 1–12, 2017.

[18] 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, no. 12, pp. 2341–2353, 2011.

[19] U. Akpinar, E. Sahin, and A. Gotchev, "Learning Optimal Phase-Coded Aperture for Depth of Field Extension," *Proceedings - International Conference on Image Processing, ICIP*, vol. 2019-September, pp. 4315–4319, 2019.

[20] M. Akram and X. Chen, "Speckle reduction methods in laser-based picture projectors," *Optical Review*, vol. 23, 12 2015.

PUBLICATION
V

Computer-generated holography using hybrid planar-spherical wave primitives

J. Mäkinen, U. Gudelek, L. Varjo, E. Sahin, and A. Gotchev

ACM Transactions on Graphics, submitted