Advanced Algorithmic Approaches for Improving Image Quality in 2D and 3D Holographic Displays

Fan Yang

Department of Engineering

University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

Jesus College

November 2022

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Fan Yang

November 2022

Author Name: Fan Yang

Thesis Title: Advanced Algorithmic Approaches for Improving Image Quality in 2D and 3D Holographic Displays

Abstract

Holography is an advanced three-dimensional (3D) imaging and visualisation technology capable of reconstructing realistic 3D scenes. Despite decades of concentrated effort, holographic 3D displays are still struggling to meet the demands required for a consumer-ready solution. This thesis addresses technical challenges in the practical implementation and focuses specifically on potential image quality improvements based on algorithmic development.

The thesis builds a holographic display system prototype and reconstructs established 3D scenes and real-world scenes using commercially available RGB-D cameras. By closely examining the reconstructed images, experimental reconstruction issues are evaluated. Given that image quality degradation is one of the significant issues in holographic displays, the gradient descent method is introduced to phase-only CGH optimisation. Contemporary image quality metrics (IQMs) considering human visual systems are employed as loss functions to improve the reconstructed image quality. Extensive objective and subjective assessment of experimentally reconstructed images reveal that the perceived quality improves considerably should the appropriate IQM loss be selected. Finally, the gradient descent method is extended to 3D hologram generation. While previous work optimises 3D CGH generation primarily on the in-focus region, this research further combines the method with an incoherent imaging module and a reformulated loss function to improve the defocus effect without sacrificing in-focus image quality. The experimentally acquired result demonstrates its effectiveness in reconstructing realistic 3D images beyond the capabilities of existing 3D hologram generation algorithms.

Acknowledgements

It is likely that I have benefited from the support of the most fantastic PhD advisor in all academia guiding my research. The completion of this dissertation and the genuine enjoyment of my PhD experience would not be possible without the academic supervision, kind encouragement, and unique personality with a good sense of humour provided by my supervisor, Tim Wilkinson, especially during the unexpected global pandemic of Covid-19. I cannot reiterate enough how grateful I am to have been given this opportunity and to have such freedom to learn and discover the things that interest me.

I would like to sincerely show my appreciation to Cambridge Trust and China Scholarship Council for providing financial support for my study at the University of Cambridge. Their generous support with university fees, maintenance funding and extension funding due to Covid allowed me the luxury of fully concentrating on my research.

I would like to thank everyone I have had the fortune of getting to know in the CMMPE group and VividQ. Thank You to all for big or small favours during my study. I am particularly grateful to Ralf, Andrew and Andrzej for countless discussions and numerous academic support efforts. I also want to thank all the friends of mine to share all the happiness and bumpiness together and support me in all aspects.

Finally, I would like to thank my parents and grandparents for their endless support and trust and for always inspiring me to aim for greater heights.

Table of contents

List of Figures

List of Tables

Nomenclature

ADOSPR	Adaptive One Step Phase Retrieval
ADMM	Alternating Direction Method of Multipliers
AIF	All-in-Focus
AR	Augmented Reality
ASM	Angular Spectrum Method
CGH	Computer Generated Hologram
CITL	Camera-in-the-Loop
DPH	Double Phase Hologram
DS	Direct Search
DFT	Discrete Fourier Transform
DISTS	Deep Image Structure and Texture Similarity
DIV2K	DIVerse 2K resolution image dataset
EDSDK	EOS Digital Software Development Kit
FFT	Fast Fourier Transform
FOV	Field of View
FPGA	Field-Programmable Gate-Array
FPS	Frames Per Second

FZP	Fresnel Zone Plate
GD	Gradient Descent
GPU	Graphics Processing Unit
GS	Gerchberg-Saxton
HaarPSI	Haar Wavelet-Based Perceptual Similarity Index
HIO	Hybrid Input-Output
HUD	Head-Up Display
HVS	Human Visual System
ICP	Iterative Closest Point
IQM	Image Quality Metrics
LPIPS	Learned Perceptual Image Patch Similarity
LT	Liu-Taghizadeh
LUT	Look-Up Table
MSE	Mean Squared Error
MS-GMSD	Multi-Scale Gradient Magnitude Similarity Deviation
MS-SSIM	Multi-Scale Structural Similarity
NDH	Natural Defocus Holography
NLC	Nematic Liquid Crystal
NLPD	Normal Laplacian Pyramid Distance
OSPR	One Step Phase Retrieval
PCL	Point Cloud Library
PSF	Point Spread Function
PSNR	Peak Signal-to-Noise Ratio
RDH	Realistic Defocus Holography
RGB-D	Red, Green Blue-Depth
ROI	Region of Interest

SA	Simulated Annealing
SDK	Software Development Kit
SFTF	Spatial Frequency Transfer Function
SLM	Spatial Light Modulator
SL	Structured Light
SRCC	Spearman’s Rank order Correlation Coefficient
SSIM	Structural Similarity
STD	Standard Deviation
TOF	Time-of-Flight
VAC	Vergence and Accommodation Conflict
VIF	Visual Information Fidelity
VSI	Visual Saliency-Induced Index
VR	Virtual Reality
2D	Two Dimensional
2AFC	2-Alternative Forced Choice
3D	Three Dimensional

Chapter 1 Introduction

1.1 Holography: Historical View

In recent years, 3D imaging and visualisation applications have attracted considerable attention in virtual and augmented reality, medical, architecture and human-computer interaction. Among the many promising technologies in this area, holography stands out for its ability to precisely generate arbitrary wavefronts that resemble realistic 3D scenes. Initially invented in 1948, Gabor proposed holography as a solution for wavefront recording in electron microscopy [1]. The term “holography” is derived from the Greek word “holos”, which means to contain all information: it includes the amplitude and phase information of the object light wave. However, the necessity that the object and reference wave be coherent restricted its applicability until the mid-1960s holographic explosion, thanks to the development of laser by Yuri Denisyuk [2]. Leith and Upatnieks recorded the first hologram of a three-dimensional object in 1964 as a by-product of radar research [3,4], sparking the beginning of the holographic technique revolution.

Analog optical holography uses light wave diffraction and interference to record fringe patterns on photosensitive materials. The fringe pattern, also known as a hologram, comprises both the amplitude and phase information of the light wave emanating from the object. When illuminated by a reference light, a hologram can reconstruct the object light faithfully. This optical recording and reconstruction process is purely analogue, inspiring the idea of computational holography that can directly generate holograms on computers by simulating the physical light propagation and interference of the object light [5–7]. The replacement of optical recording with electronic recording leads the computer-generated holograms (CGHs) in compact and digital formats. CGHs can recreate realistic 3D images both digitally and optically.

1.2 Other 3D Display Technologies

Before introducing the fundamental concept of holography, it is useful to briefly recapitulate other typical 3D display technologies. Numerous comprehensive books and reviews on this topic [8–12] comprise a substantial body of research in this fascinating field. The primary goal of this section is to illustrate the general concepts behind popular 3D display technologies rather than detailing specific display designs or implementations.

Stereoscopy served as the very first strategy developed for the challenge of recording and displaying 3D images. *Stereoscopic displays* exploit the binocular human visual system (HVS) by inducing binocular disparity to stimulate depth perception. The display systems require users to wear special eyeglasses so that two slightly different views can be delivered to each eye. The first stereoscope was proposed in 1838 [13]. Later, colour and polarisation were used to encode stereo image pairs and deliver left- and right-eye views [14–16].

Autostereoscopic displays are also based on stereoscopy but without using special glasses. The display systems of this type are realised by different strategies, including using a parallax barrier or a lenticular sheet, attempting to provide a fixed viewing zone for one eye while blocking it from the other [17,18].

Multi-view displays convert 3D images into a range of sub-viewing zones in the horizontal direction, offering multiple stereoscopic perspectives for one or multiple observers. Despite providing a small degree of motion parallax during observer movement, multi-view displays have the drawback of flipping when crossing viewing zones and suffer from limited display resolution due to the spatially multiplexed design [8].

However, a crucial issue that may rule out these stereoscopic displays being the perfect 3D display technology is the vergence and accommodation conflict (VAC) [19]. Vergence refers to the inward or outward rotation of the eyes required to align the images from each eye to create the perception of depth. Accommodation, on the other hand, refers to the ability of the eye to adjust its focus to different distances. This conflict arises because stimulated 3D images are generated on the screen but not at the focusing distance from the viewer. As a result, the HVS receives mismatching cues between the projected stereoscopic images (vergence) and the actual observation distance of the stereoscopic monitor distance (accommodation). This intrinsic visual conflict forces the HVS to physiologically decouple recognition processes of vergence and accommodation, leading to severe visual discomfort or fatigue [8,20–23].

Some true 3D display technologies, such as integral imaging and volumetric display technology, have been proposed concurrently to overcome the VAC conflict. *Integral imaging* can be traced back to 1908 when Lippmann proposed to use a micro-lens array to capture 3D images into a set of 2D elemental images with different perspectives [24]. Based on the reverse principle of light rays, integral imaging aligns the captured elemental images with a lens array to reconstruct 3D images that can be observed from different perspectives within a limited viewing angle. Integral imaging has attracted substantial interest with diverse applications due to its intuitive and straightforward implementation; it provides full motion parallax with incoherent illumination. However, it has a limited depth range and significantly sacrifices spatial resolution in exchange for directional resolution.

Volumetric displays directly generate 3D images in true 3D space by occupying an actual display volume filled with volume pixels or voxels that can be controlled at any desired spatial position. Free-space displays, fixed-volume displays, and swept-volume displays are the three most common forms of volumetric displays. Free-space volumetric displays direct laser beams in free space to activate voxels through visible fluorescence or scattering. As a straightforward extension of 2D display techniques, fixed-volume displays project onto a static volume consisting of stacked depth-separated scene layers. Swept-volume displays rotate or translate emissive displays, or projection surfaces, where fast rotating or translating screens are illuminated by laser beams to create translucent or contour images. Although volumetric 3D displays provide physiological and psychological depth cues, they are still not ideal for practical applications. Since typical volumetric displays require voxels distributed in a volume to generate 3D scenes, the size of the created 3D scene depends on the required volume of such displays, which makes them bulky and limits their applicability. Other disadvantages depend on the specific implementation types, including the scale-up and multi-colour capability, low resolution, crosstalk between voxels, low brightness, and hardware design [11]. Other displays including Super multivew displays may also be a great candidate for 3D imaging and display. The use of other display technologies, such as Super Multiview displays [25], also present a viable option for 3D imaging and display.

1.3 Thesis Motivation

Holography can present a virtual window of a 3D scene by optically generating waves that match the natural light from a real object by wavefront shaping. Compared to traditional 3D display technology, holographic display technology is necessary as it can provide all the depth cues required by the HVS to perceive 3D objects, including both vergence and accommodation. Holographic displays eliminate the need for special glasses or headsets and reduce eye strain and fatigue, making them the most intriguing and elegant technology to visualise 3D images. The use of beam modulation to generate images is another distinctive advantage of holography over other display technologies that rely on the addition of ancillary optical components to block light. This notable distinction yields substantially higher optical efficiency for holographic displays, thereby leading to a superior visual experience characterised by enhanced contrast, vividness, and realism compared to other true 3D display technologies. Moreover, other 3D technologies exhibit resolution limits due to the optical design; a smaller form factor increases the diffraction-limited spot size, making the high-resolution optical design more challenging. The unique virtue of holography is that it inherently requires a smaller display pixel pitch to have a larger field of view (FOV), presenting opportunities for miniaturising the system while increasing the output image size.

The holographic display technique is currently transitioning from laboratory research to commercial markets and from two-dimensional to three-dimensional visualisation. For a holographic display to be suitable for consumers, it must provide users with a comfortable and immersive experience, ensure accessibility, and guarantee reliability. Immersion is related to a wide FOV and high resolution, while comfort is characterised by a compact form factor, large eyebox, visual image quality, low latency, and high colour accuracy [26]. The holographic display should be affordable and accessible to a wide range of consumers, including those with disabilities or specific needs.

Among these requirements, providing a comfortable and immersive viewing experience is the most crucial but has yet to be fully achieved due to significant technical challenges in developing high-quality 3D holographic displays. One major obstacle in the hardware is the limited space-bandwidth product of the core display device, known as the spatial light modulator. This limitation is directly proportional to the product of eyebox size and FOV, leading to a fundamental compromise between the two. Expanding the eyebox, for instance, will lead to a decrease in the FOV and vice versa. Though adding ancillary optics can improve the finite space-bandwidth product, it usually involves sacrificing the form factor of the device [27,28]. Moreover, despite the current 4K resolution of SLMs being adequate for near-eye displays, applications, such as teleconferences that require CGH reconstruction from real-world scenes with significant data compression and transmission, are far from practical. Regarding CGH calculation algorithms, simulating light wave diffraction is a computationally demanding task; synthesizing high-quality CGHs at real-time frame rates that adequately achieve visual comfort for viewers remains impractical. Additionally, representing 3D objects with sufficient detail requires large amounts of data presenting difficulties in storage and processing. Furthermore, the optimisation of CGHs to satisfy the requirements of SLMs and reliably achieve high image fidelity while simultaneously maintaining real-time performance poses a significant challenge in CGH algorithms. Therefore, it remains an unsolved challenge for holography to deliver 3D image data with genuine 3D display capabilities while simultaneously achieving high image quality [26,29].

This thesis aims to address technical challenges in holographic displays, specifically in enhancing the holographic reconstructed image quality. In this regard, the thesis first assesses the feasibility of generating CGHs from real-world and computer-generated scenes to support high-quality holographic display solutions. Such feasibility is assessed by building a holographic display prototype with real-world scene acquisition, addressing practical image degradation issues in holographic reconstruction. This thesis then extensively focuses on developing next-generation 2D and 3D CGH generation algorithms. The goal is to create algorithms that can generate high-quality CGHs in support of a more visually comfortable experience for the user. This research will contribute to a better understanding of the technical challenges and limitations in holographic displays and the development of next-generation algorithms that can enhance the visual quality of holographic images.

1.4 Thesis Organisation

The organisation of the dissertation is as follows. Chapter 2 introduces the fundamental theory of the holographic display technique and reviews current limitations and research progress from the hardware and algorithm perspectives in the field.

Chapter 3 presents classic algorithms for generating CGHs from 2D and 3D images. The 2D CGH algorithms Gerchberg-Saxton and One-Step-Phase-Retrieval are explained in detail. CGH methods of generating 3D scenes based on decomposing primitives are reviewed, including point-based, polygon-based, and layer-based methods.

Chapter 4 attempts to build a primitive holographic display system from the 3D data acquisition to the 3D data reconstruction. In addition, several practical issues and potential improvements to the image reconstruction quality are examined.

Due to the severe degradation of the perceived quality in holographic reconstruction, Chapter 5 focuses on 2D CGH, introducing the gradient descent method to phase-only CGH optimisation. Contemporary image quality metrics are further introduced in the hologram optimisation process as loss functions to improve reconstructed image quality. A comprehensive analysis of the relative performance of IQM losses is presented based on extensive objective quality assessments as well as subjective comparisons informed by more than 10k human judgments.

Chapter 6 introduces using the gradient descent method for 3D hologram generation. The gradient descent method is further extended using an incoherent propagation model to generate target images for image quality improvement in the defocused area. Numerical simulation and optical experiments validate its capability of reproducing naturally defocused reconstructed images.

Chapter 7 is dedicated to providing a concise conclusion and potential future work.

Chapter 2 Fundamentals of Holography

The mathematical foundations of computer-generated holography are presented in this chapter. Although some chapters have different methodology and literature sections, the fundamentals utilised throughout the chapters are presented here. This chapter begins by describing the scalar diffraction theory. The diffraction theory unfolds the wave propagation process, revealing how the amplitudes and phases of the lightwaves from a two-dimensional aperture reach the replay plane and form complex-amplitude distributions. The chapter then introduces computer-generated holography and how to simulate holograms based on the diffraction theory. The last section briefly discusses the major limitations of holography displays from both hardware implementation and algorithmic development perspectives.

2.1 Theory and Simulation

2.1.1 Scalar Diffraction Theory

In holography, the primary concern is with the wave phenomena of light for hologram calculation and reconstruction, including diffraction and interference. Maxwell's equations can analyse the light propagating behaviour as an electromagnetic wave. Although the scalar diffraction theory is extensively elucidated with the derivation of the formulae in the literature *Introduction to Fourier Optics, Holographic Imaging, Optics* and lecture notes of *4B11: Photonic Systems* [5,30–32], it is necessary to review this subject for background concepts and notation consistency.

Maxwell's equations are vector equations describing the spatial and temporal coupling effects between the electric and magnetic fields. When light is propagating through a linear, homogeneous, isotropic, and non-dispersive medium, all components of the electric and magnetic field have identical behaviour and can be summarised by a single *scalar wave equation*:

$$\nabla^2 u(P,t) - \frac{n^2}{c^2}\frac{\partial^2 u(P,t)}{\partial t^2} = 0, \tag{2.1}$$

where $u(P,t)$ represents any scalar field components with position P in space and time t. The medium refractive index is described by n, with c as the speed of light in the vacuum.

The spherical wave and the plane wave are the simplest solutions of the above scalar wave equation. For a simple harmonic oscillation in space with angular frequency ω, the spherical wave at the centre of spherical coordinates (R, θ, ϕ) is represented by:

$$u(R,t) = \frac{U_0}{R} sin(\omega t - kR)\,, \tag{2.2}$$

where k is the wave number given by the wavelength λ: $k = \frac{2\pi}{\lambda}$ and the source strength is given by U_0. The wavefront of a spherical wave will remain itself as a sphere with a larger radius as the spherical wave propagates, with the amplitude decaying as R^{-1}.

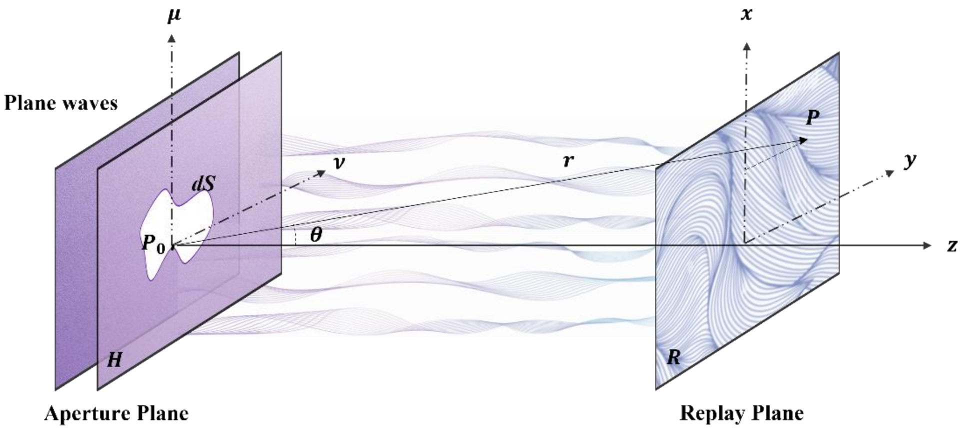

Figure 2.1. Huygens-Fresnel Diffraction geometry in Cartesian coordinates.

Having described the wave propagation in the spherical form, we can introduce diffraction through the two-dimensional aperture, as shown in Figure 2.1. Diffraction problems are among the most challenging encountered in optics. As such, approximate methods are generally used for practical solutions in engineering. Our particular interest is the Huygens-Fresnel principle to model the diffraction phenomena due to its adequacy, intuitive appeal, and simplicity in CGH calculation. The Huygens-Fresnel principle proposes that every point on a wavefront is the source of a spherical emitting point known as a wavelet; the interference of these spherical wavelets forms the secondary wavefront.

For a diffracting aperture H on the (μ, ν) plane illuminated by a monochromatic coherent wave source with wavelength λ and wavenumber k, the principle assumes that any infinitely small differential P_0 at the aperture plane (μ, ν) can be treated as a spherical Huygens wavelet radiating spherical waves. Considering the boundary condition imposed on the solution of the wave equation, the *Huygens-Fresnel diffraction formula* describes the resulting field distribution R at a point P on the plane (x, y) with a distance r away from this source in the $+z$ direction as:

$$R(x,y) = \frac{1}{j\lambda} \iint H(\mu,\nu) \frac{e^{jkr}}{r} cos\theta \, d\mu d\nu, \tag{2.3}$$

where the propagating distance r is geometrically given by:

$$r = \sqrt{(x-\mu)^2 + (y-\nu)^2 + z^2}. \tag{2.4}$$

$H(\mu, \nu)$ is a scalar quantity describing the aperture plane, and $cos\theta$ is the angle between the outward aperture-plane normal vector n̂ and the vector $\hat{r}$ pointing from the wavelet P_0 to the observation point P. Substitute $cos\theta = \frac{z}{r}$, equation 2.4 can be rewritten as:

$$R(x,y) = \frac{z}{j\lambda} \iint H(\mu,\nu) \, \frac{e^{jkr}}{r^2} \, d\mu d\nu \, . \tag{2.5}$$

2.1.2 The Fresnel Diffraction

Depending on the propagating distance, we can classify the diffraction regions into three: the near-field region, the Fresnel region, and the far-field region or the Fraunhofer region, as shown in Figure 2.2. The propagating distance in the near field region is so small that the exact formula must be used for diffraction calculation. If the distance from the aperture plane is relatively large, we can approximate the spherical wavefronts to parabolic wavefronts in the Fresnel region. We can further approximate the wavefront as planar if the propagating distance is sufficiently large.

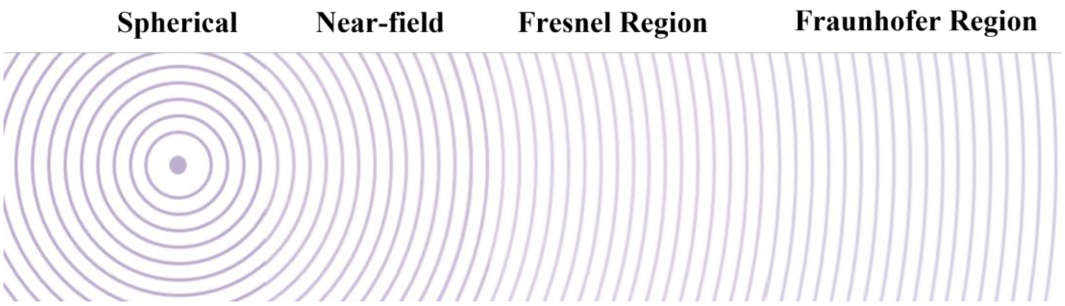

Figure 2.2. A spherical wave is flattened and approximated during propagation.

For Fresnel diffraction, binomial expansion is introduced to approximate the distance r between the aperture and the observation plane. The binomial expansion of a square root is:

$$\sqrt{1+\mathrm{k}} = 1 + \frac{1}{2}\mathrm{k} - \frac{1}{8}\mathrm{k}^2 + \ldots \qquad 2.6$$

The binomial expansion of the distance r discards higher-order terms and factorises z outside:

$$r = z\sqrt{1 + (\frac{x-\mu}{z})^2 + (\frac{y-\nu}{z})^2} \approx z\left[1 + \frac{1}{2}(\frac{x-\mu}{z})^2 + \frac{1}{2}(\frac{y-\nu}{z})^2\right]. \qquad 2.7$$

This approximation assumes the emitting point source P at the aperture plane has a reasonably small propagating angle around the z axis, and the distance r between the observation plane and the aperture is comparatively large. In this case, we apply the binomial approximation equation 2.7 to the r in the exponential term of equation 2.5 and assume $r \approx z$ in the denominator. The Fresnel approximation replaces the spherical wavefront from the Huygens wavelet with a parabolic wavefront. Applying the approximation of the distance r, we can describe the resulting Fresnel Diffraction formula as:

$$R(x,y) = \frac{e^{jkz}}{j\lambda z}\iint H(\mu,\nu)\, e^{\frac{jk}{2z}[(x-\mu)^2+(y-\nu)^2]}\, d\mu d\nu. \qquad 2.8$$

By factorising out the constant term $e^{\frac{jk}{2z}(\mu^2+\nu^2)}$, equation 2.8 is rearranged to:

$$R(x,y) = \frac{e^{jkz}}{j\lambda z} e^{\frac{jk}{2z}(x^2+y^2)} \iint \left\{H(\mu,\nu)e^{\frac{jk}{2z}(\mu^2+\nu^2)}\right\} e^{-\frac{jk}{z}(\mu x+\nu y)}\, d\mu d\nu \qquad 2.9$$

$$= \frac{e^{jkz}}{j\lambda z} e^{\frac{jk}{2z}(x^2+y^2)} \mathcal{F}\left\{H(\mu,\nu)e^{\frac{jk}{2z}(\mu^2+\nu^2)}\right\}.$$

This complex-amplitude field $R(x, y)$ can be expressed as a ***Fourier transform*** of the aperture function $H(\mu, \nu)$ with an additional *quadratic phase term* $e^{\frac{jk}{2z}(\mu^2+\nu^2)}$ scaled by a factor $\frac{e^{jkz}}{j\lambda z} e^{\frac{jk}{2z}(x^2+y^2)}$. Both forms are called ***Fresnel diffraction integral***.

The Fresnel approximation is valid if the near-field distance boundary satisfies the following:

$$z^3 \gg \frac{\pi}{4\lambda}[(x-\mu)^2 + (y-\nu)^2]^2_{max} \qquad 2.10$$

2.1.1 The Fraunhofer Diffraction

The Fresnel diffraction equation 2.9 can be further simplified by imposing a stronger far-field condition:

$$z \gg \frac{k(\mu^2 + \nu^2)_{max}}{2}. \qquad 2.11$$

Then the quadratic wavefront from the Huygens wavelet in the Fresnel diffraction equation 2.9 can be simplified to a planar wavefront by approximating the exponential term of the quadratic phase $e^{\frac{jk}{2z}(\mu^2+\nu^2)}$ to unity over the entire aperture:

$$\begin{aligned} R(x, y) &= \frac{e^{jkz}}{j\lambda z} e^{\frac{jk}{2z}(x^2+y^2)} \iint H(\mu, \nu) e^{-\frac{jk}{z}(\mu x + \nu y)} d\mu d\nu \\ &= \frac{e^{jkz}}{j\lambda z} e^{\frac{jk}{2z}(x^2+y^2)} \mathcal{F}\{H(\mu, \nu)\} \end{aligned} \qquad 2.12$$

The observed field $R(x, y)$ is then a direct Fourier transform of the aperture function $H(x, y)$ with a scaling factor $\frac{e^{jkz}}{j\lambda z} e^{\frac{jk}{2z}(x^2+y^2)}$. Thus, equation 2.12 is called the ***far-field diffraction integral*** or the ***Fraunhofer diffraction integral***.

However, the far-field condition requires an impractical large observation distance z. For a $635nm$ wavelength laser with an aperture width of 2.54cm, the distance z must satisfy $z \gg 1600\text{m}$ [5,32]. A positive focal length lens is therefore imposed in front of the aperture to reduce the required distance, as shown in Figure 2.3, and the far-field diffraction pattern is then displayed at the focal plane of the lens $z = f$. The lens introduces a defocus aberration which appears as a quadratic phase distortion factor in equation 2.12:

$$R(x,y) = \frac{e^{jkf}}{j\lambda f} e^{\frac{jk}{2f}\left(1-\frac{d}{f}\right)(x^2+y^2)} \iint H(\mu,\nu) e^{-\frac{jk}{f}(\mu x+\nu y)} d\mu d\nu. \qquad 2.13$$

To eliminate the phase distortion term to unity in the equation introduced by the additional lens, we can place the positive focal lens one focal distance $d = f$ behind the aperture.

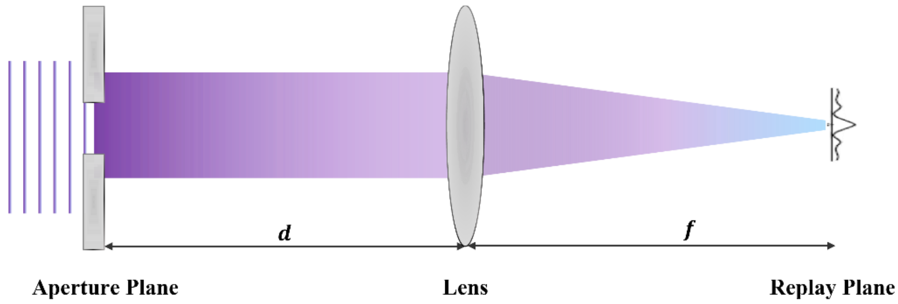

Figure 2.3. Fourier lens placed at distance d behind the aperture.

2.1.2 Discrete Diffraction Functions

Since physical displays are pixelated structures, the continuous aperture function $H(\mu,\nu)$ needs to be sampled, resulting in a discrete aperture function $H_d(\mu,\nu)$. When the discrete aperture function $H_d(\mu,\nu)$ is simulated as a CGH and illuminated by a reconstruction beam, the complex-amplitude light field $R(x,y)$, whether in the Fresnel or Fraunhofer region, is defined as the object field or the replay field. The complex replay field $R(x,y)$ is then related to the discrete aperture function $H_d(\mu,\nu)$ with a ***Discrete Fourier Transform (DFT)*** and other factors dependent on the diffraction region.

The discrete Fresnel diffraction equation is described as follows:

$$R(x,y) = \frac{e^{jkz}}{j\lambda z} e^{\frac{j\pi}{\lambda z}(x^2+y^2)} \sum_{\mu=-\frac{M}{2}+1}^{\frac{M}{2}} \sum_{\nu=-\frac{M}{2}+1}^{\frac{N}{2}} \left\{H_d(\mu,\nu) e^{\frac{j\pi}{\lambda z}(\mu^2+\nu^2)}\right\} e^{-2\pi j\left(\frac{\mu x}{M}+\frac{\nu y}{N}\right)}. \qquad 2.14$$

The discrete Fourier (Fraunhofer) diffraction equation is:

$$R(x,y) = \frac{e^{jkz}}{j\lambda z} e^{\frac{j\pi}{\lambda z}(x^2+y^2)} \sum_{\mu=-\frac{M}{2}+1}^{\frac{M}{2}} \sum_{\nu=-\frac{M}{2}+1}^{\frac{N}{2}} H_d(\mu,\nu)\, e^{-2\pi j\left(\frac{\mu x}{M}+\frac{\nu y}{N}\right)}\ . \qquad 2.15$$

2.1.3 The Angular Spectrum Method

Alternatively, the Angle Spectrum Method (ASM) has been widely applied in CGH calculation as it strictly describes the physical process of light propagation in the near field region and remains a simple form. The relation between a complex function $U(x,y)$ and its angular spectrum $AS_U(f_x, f_y)$ is described by the inverse Fourier Transform:

$$U(x,y) = \iint AS_U(f_x, f_y) exp\left[2\pi j(f_x x + f_y y)\right] df_x df_y. \qquad 2.16$$

It can be regarded as a decomposition of the complex function $U(x,y)$ into plane-wave components with different direction cosines (a, β, γ):

$$a = \lambda f_x, \qquad \beta = \lambda f_y, \qquad \gamma = \sqrt{1-(\lambda f_X)^2-(\lambda f_Y)^2}. \qquad 2.17$$

Then the angular spectrum $AS_H(f_\mu, f_\nu)$ of the hologram field $H(\mu,\nu)$ and the angular spectrum $AS_R(f_x, f_y)$ of the replay field $R(x,y)$ is related by:

$$AS_R(f_x, f_y) = AS_H(f_\mu, f_\nu) \exp\left[j\frac{2\pi}{\lambda} z\sqrt{1-\left(\lambda f_\mu\right)^2-(\lambda f_\nu)^2}\right]. \qquad 2.18$$

This equation describes that for light propagating in free space by a distance z, the spectrum of the replay field is the product of the spectrum of the hologram field multiplied by a phase factor $TF(f_\mu, f_\nu)$, which is called the spatial frequency transfer function (SFTF):

$$TF(f_\mu, f_\nu) = \exp\left[j\frac{2\pi}{\lambda} z\sqrt{1-\left(\lambda f_\mu\right)^2-(\lambda f_\nu)^2}\right]. \qquad 2.19$$

Therefore, the diffraction propagation formula described by the angular spectrum method can be obtained as follows:

$$R(x,y) = \mathcal{F}^{-1}\{\mathcal{F}[H(\mu,\nu)] \cdot TF(f_\mu, f_\nu)\}. \quad 2.20$$

2.2 Computer-Generated Holography

Computer-generated holography can simulate the hologram digitally based on the physical principles of light propagation and interference. The simulated hologram, also known as CGH, can be presented on a physical display and illuminated to reconstruct the object wave.

There are three steps to generate a CGH. The first step is to digitally acquire the sampled complex amplitude light field of a target object. Input images or 3D scenes we wish to reproduce by CGHs are prepared and modelled into a discrete finite set for CGH calculation. The second step is to compute CGHs from the sampled input. During this process, we first need to compute a complex-amplitude CGH by carrying out the diffraction equations introduced above. Due to the physical limitation of current display technologies, we cannot independently control the amplitude and phase of the generated CGH. Therefore, we must choose a suitable encoding method for the calculated complex amplitude CGH in the hologram plane to be displayed on particular physical displays. The last step is to transfer the encoded CGH to the chosen physical display. The replay field reconstruction process of a CGH is no different from that of an optical hologram.

2.3 Major Shortcomings of Modern CGHs

Despite being the ultimate display technology due to its unique capacity to reproduce genuine 3D scenes, holographic 3D display technology is impeded primarily by three challenges: computation of holograms from 3D data, transmission of resulting holograms to the display, and optical reconstruction of 3D data from holograms. Transmission and optical reconstruction are fundamentally constrained by the limitations of the display hardware, while computation is limited by the heavy computational loads required to generate holographic patterns.

2.3.1 Data Transmission Challenge

The transmission of holographic data poses significant challenges due to the large amount of data involved. For instance, a holographic display with a 70cm diagonal screen and an 8μm pixel pitch would have approximately 4×10^9 pixels. To ensure a minimum refresh rate of 60 Hz and at least three channels for RGB colours, each with a grey-level resolution of at least 8 bits, the required data rate for such a display, excluding any encoding or compression algorithm, would be~ 670.55 Gb/sec(4×10^9 $Pixels$ $\times$ 60Hz $\times$ 8 bits $\times$ 3 channel). The limited data rate is further exacerbated when considering the need for a smaller pixel pitch to achieve a larger FOV and a faster switching speed for time-multiplexing methods. The maximum diffraction angle in a holographic display in plane-wave illumination is $\theta_{max} = sin^{-1}(\lambda/2\Delta)$, where λ is the wavelength of the light and Δ is the pixel size of the SLM. For an immersive viewing experience with 120° FOV, the pixel pitch would therefore be approximately on the order of 300 nm, requiring 3×10^{12} pixels of the same physical size. Fast switching speed, which may require refresh rates of up to 10 kHz, further increases the required data rate by orders of 10^3 magnitude.

Due to the high data rate demands, there is a need for compression and codecs to facilitate the transmission of dynamic hologram data. However, no generic compression and codec can be applied to holograms with various content while remaining robust and efficient for a high data rate. Moreover, currently available off-the-shelf SLMs have limited modulation schemes and cannot fully modulate given complex amplitude functions, which further requires a complex-value coding process during data transmission. Although there are SLMs that can modulate phase and amplitude simultaneously, the phase-amplitude coupling effect exists in these SLMs, limiting independent modulation. Therefore, generated complex-amplitude CGHs need to be encoded into either the amplitude-only or the phase-only type based on the properties of SLMs. Since amplitude SLMs limit the transmission of incident light by blocking, phase SLMs are typically used, which delay the phase of incident light instead, resulting in high optical efficiency. Encoding complex-amplitude CGHs to meet the constraints imposed by current phase-only SLMs introduces an additional layer of complexity to the holographic data transmission process.

2.3.2 **Optical Reconstruction Challenge**

A superior holographic display would offer a wide FOV and high-resolution images to enhance immersion while supporting comfortable viewing with a large eyebox and excellent visual quality in a compact form. These features are intrinsically linked to the characteristics of the core display hardware, SLM.

The FOV refers to the range of angles or distance over which an object is visible, while the eyebox refers to the area in front of the display in which the viewer can move their eyes and still maintain a clear and undistorted view of the content. In the case of holographic displays, the FOV of the display system is determined by the maximum diffraction angle of the SLM, which establishes the maximum dimensions of the holographic image. The eyebox is determined by the display area of the SLM. For a holographic display, with the maximum diffraction angle θ, and width w, the FOV under paraxial approximation and eyebox are:

$$FoV \approx 2\frac{f_1}{f_2}\theta, \qquad eyebox = \frac{f_2}{f_1}w, \tag{2.21}$$

where f_1 is the focal length of the Fourier lens, which places the reconstructed image produced by the SLM to its focal plane as discussed in the previous section, and f_2 is an additional lens, together with f_1 that composites an $4f$ system and relays the image to the eye. Therefore, the product of FOV and eyebox, also called étendue [27] along one dimension, is:

$$\mathrm{FOV} \cdot eyebox \approx 2\theta w \approx \frac{\lambda w}{\Delta} = \lambda N_x, \tag{2.22}$$

where N is the pixel number of the SLM in one dimension.

Therefore, the finite and limited number of pixels in SLMs introduces a trade-off between the FOV and the eyebox size. For example, high-end commercial SLMs with 4K resolution can achieve a horizontal FOV of 90° at 532nm but with an eyebox size of only 1.4 mm, which means that even a slight angular deviation of the eye can result in the image disappearing. One solution to this issue is to increase the pixel count of the SLM, which also enhances the perceived resolution by expanding the space-bandwidth products. However, to achieve a 15 mm eyebox with a 90° FOV that is comparable to human vision, an SLM with approximately

4×10^4 pixels in the horizontal axis would be required, which exceeds current technological capabilities.

Moreover, fill rate, refresh rate and other characteristics of current SLMs are also directly related to the visual quality in the replay field. Further optimisation of the optical structure of existing holographic display systems is also required to realise a compact design, especially for near-eye display applications. It is necessary to miniaturise the size of the hardware while simultaneously achieving high reconstructed image quality.

2.3.3 Computational Complexity of Generating CGHs

The physical model of light propagation has been analytically well defined as mathematical equations, but practical computation is far from trivial. As previously mentioned, the vast quantity of data to be processed significantly increases the computational load. CGH generation algorithms are heavily involved with performing Fourier Transforms, as demonstrated by the Fresnel and Fraunhofer diffraction equations. However, Fourier transforming such a large amount of data with a high frame rate is still challenging. Even when implementing the Fast Fourier Transform (FFT) [33] to perform Fourier Transforms, the process is computationally expensive and is scaled with the number of pixels on SLMs. Generating CGHs from 3D images further increases computational complexity due to the numerous primitives required for calculation.

Moreover, since no commercially available SLMs can arbitrarily modulate the incident light in amplitude and phase simultaneously by far, hologram-generating algorithms are necessary to generate compromise solutions that meet the constraints of SLMs. Thus, the CGH generation process is changed from analytical computation to an optimisation process that can be represented in the following way: *find an optimal hologram from a subset of complex number spaces imposed by the selected type of SLMs, such that the hologram can maximise the image quality in the replay field.*

Another challenge is the realistic rendering of scenes with all depth cues for the 3D CGH generation process. While diffractive wave propagation and interference in holography can directly address the majority of depth cues, the degree of realism remains limited compared to the cutting-edge techniques employed in computer graphics. Several depth cues, including occlusion, shading, defocus blur, and parallax, are often not adequately implemented in CGH calculations, thereby restricting the depiction of virtual 3D objects to simple geometries.

Most CGH generation algorithms are iterative or require additional operations for optimisation, which further increases the computational complexity and impedes real-time generation. As a result, researchers have explored hardware and algorithmic accelerations to alleviate the computational burden of generating CGHs [33,34]. Parallel operations with large processing core counts on modern Graphics Processing Units (GPUs) accelerate the CGH computational process, allowing real-time hologram generation. Field programmable gate arrays (FPGAs) are highly-configurable and hardware-programmable integrated circuits that facilitate flexible logical hardware designs. Recent work from the HORN-8 group [35] demonstrated that FPGA chips could be integrated into large scale and are more dedicated to real-time CGH generation. One of the iconic algorithmic accelerations is the One-Step Phase Retrieval algorithm from the work of Cable and Buckley [36], resolving the computational problem by using time-multiplexing to avoid the iterative optimisation process.

Chapter 3 Hologram Generation Algorithms

3.1 Introduction

The preceding chapter introduced the fundamentals of holography. In this chapter, we will look into hologram generation algorithms. We will start with classic 2D algorithms that apply the diffraction theory to generate CGHs from target images. Furthermore, using the diffraction theory, we can numerically calculate the complex-amplitude optical field in the hologram plane from a 3D scene. The calculated field is a 2D CGH containing the 3D information of the scene, which can be numerical or optical reconstructed later.

3.2 2D CGH Generation Algorithms

Encoding an ideal CGH to a phase-only CGH can be done with either iterative or non-iterative approaches. Classic Iterative CGH algorithms include Gerchberg-Saxton (GS) [37], and hybrid input-output (HIO) [38,39] methods are based on Fourier Transform to find the optimal hologram. Direct search (DS) [40] and simulated annealing (SA) [41] algorithms are examples of iterative pixel-by-pixel approaches. Non-iterative CGH algorithms such as double phase and error diffusion methods can directly encode complex-amplitude diffraction fields into phase-only holograms to overcome the phase-only restriction imposed by these SLMs. Another option is the one-step phase retrieval algorithm (OSPR) [36], which relies on fast-switching SLMs to time-multiplex a set of holograms. Parallel to their application in computer vision, trained deep learning approaches have also emerged as non-iterative solutions in CGH optimisation [42–45]. Researchers have also explored other phase-retrieval methods using first-order nonlinear optimisation, alternative direction methods for phase retrieval [46,47], and non-convex optimisation [48]. The following section mainly introduces the iterative GS algorithm and the non-iterative OSPR Algorithm.

The GS algorithm was first proposed in 1972 to recover electron microscopy and astronomy phase distributions by iteratively alternating between the two fields to satisfy restrictions. The OSPR algorithm is a temporal averaging method, displaying multiple phase-only holograms within a short time interval to statistically average out errors in the replay field.

Algorithm 1: Gerchberg Saxton algorithm

Input: Target image $\boldsymbol{I(u,v)}$, Number of iterations N

Output: Hologram $\boldsymbol{H_N^Q}$ $\boldsymbol{(x,y)}$,

1 Generate a complex-amplitude target object field with a uniform random phase:

$$\boldsymbol{T_1(u,v) = \sqrt{I(u,v)} \cdot e^{j\varphi(u,v)}}$$

For $\boldsymbol{k = 1{:}N}$ do

2 Inverse Fourier Transform the complex-amplitude distribution in the object plane:

$$H_{\mathrm{k}}(x,y) = \mathcal{F}^{-1}\{T_{\mathrm{k}}(u,v)\}$$

3 Extract and quantise the complex-amplitude hologram:

$$H_k^Q\ (x,y) = Quantization\{H_k(x,y)\}$$

4 The hologram is propagated to the object plane by Fourier Transform:

$$R_k(u,v) = \mathcal{F}\{H_k^Q\ (x,y)\} = A_k(u,v)e^{j\varphi_{R_k}(u,v)}$$

5 Apply the target amplitude to form the new target of the complex replay field:

$$T_{k+1}(u,v) = \left|\sqrt{I(u,v)}\right|e^{j\varphi_{R_k}(u,v)}$$

end

3.2.1 Gerchberg-Saxton Algorithm

Algorithm Description

The GS algorithm iterates between the hologram and replay fields and applies constraints to optimise an initial random phase hologram. In CGH generating application, the constraint in the hologram plane is the quantised phase level of each pixel value (and amplitude equals one for planar illumination), and the constraint in the image plane is the equality between the replay field intensity and the target intensity. It can be shown that GS minimises the mean squared error (MSE) between the two intensities for each iteration and quickly converges to a local minimum.

Algorithm Variants

The GS algorithm is generalised as an error-reduction algorithm by Fienup [38] and further recognised as an alternating projection algorithm by Bauschke [49]. The initial random phase can be replaced by a phase vortex with a finite circular aperture [50] to remedy phase errors introduced by SLM and other optical components. The Liu-Taghizadeh (LT) and Double-constraint GS algorithms modify amplitude or phase constraints in the replay field to reduce speckle noise [51–55]. The Ping Pong algorithm utilises another intermediate replay plane, resulting in a lower speckle noise [56,57]. The over-compensation algorithm [58] and the up-scaling algorithm [59,60] have similarities in applying weights to the changes during each iteration. The fractional Fourier transform [61,62] and the gyrator transform [63–65] have been explored in replacing Fourier Transform in the GS algorithm.

3.2.2 One-Step Phase-Retrieval Algorithm

Algorithm Description

The algorithm generates a set of independent and identically distributed (i.i.d.) phase-only holograms k by a direct Fourier Transform from the same target image. The generated holograms are displayed within a very short interval so that the HVS responds to the set of replay fields instead of a single one. The perceived image is thus the statistically average intensities of the replay fields from the set of phase-only holograms:

$$I_{Recon}(u,v) = \frac{1}{k}\sum_{i=1}^{k}\left|\mathcal{F}\{H_i^Q(x,y)\}\right|^2 \qquad 3.1$$

The use of OSPR holograms and the temporal averaging method optimises the noise variance instead of the total noise energy mean in the replay field. The display of the set of holograms within a very short time interval can sum the replay images incoherently in the eye, which statistically decreases the noise variance and substantially improves the perceived image quality.

Algorithm 2: OSPR algorithm

Input: target image $\boldsymbol{T(u,v)}$

Output: Hologram: $\boldsymbol{k}$ set of phase holograms $\boldsymbol{H_i^Q(x,y), i = 1, 2, \ldots k}$

1 Convert an input target image into the amplitude function:

$$\boldsymbol{T(u,v) = \sqrt{I(u,v)}}$$

for $\boldsymbol{i = 1 : k}$ do

2 *Add a random phase to the amplitude of the target images*

$$T'(u,v) = T(u,v) \cdot e^{j\varphi_i(u,v)}$$

3 *Perform a Fourier Transform to get the hologram*

$$H_i(x,y) = \mathcal{F}^{-1}\{T'(u,v)\}$$

4 *Quantize the complex hologram to the multilevel phase-only states:*

$$H_i^Q(x,y) = Quantisation\{H_i(x,y)\}$$

end

Algorithm Variants

The variant of OSPR, the Adaptive one-step phase-retrieval algorithm (ADOSPR), rectifies the noise mean by utilising an adaptive parameter α. As a result, the ADOSPR algorithm significantly improves the image quality in the replay field than the standard OSPR algorithm [66].

$$T_{i+1}(u,v) = \begin{cases} \sqrt{(i+1)I(u,v) - \dfrac{I_{Recon}(u,v)}{\alpha^2}} & if\ (i+1)I(u,v) > \dfrac{I_{Recon}(u,v)}{\alpha^2} \\ 0 & otherwise \end{cases} \qquad 3.2$$

Other variants combine OSPR with the GS or the LT algorithm to provide better individual OSPR frames but may result in higher computational complexity.

3.3 3D CGH Generation Algorithms

The preceding section introduced the algorithms applied to 2D holography. However, the merit of holography is primarily reflected in its ability to reconstruct a 3D scene effectively. Many algorithms have been proposed to decompose a 3D scene into different primitives for CGH calculation, including point cloud, polygon and layer for wave propagation [67,68]. Each primitive is then propagated and accumulated in the hologram field to synthesis a CGH for the entire 3D scene.

3.3.1 Point-based Method

For the point-based method, a 3D scene is decomposed into a sum of object points [69,70]. These indexed object points can be regarded as self-illuminating point sources, emitting spherical waves in the object field. The corresponding hologram is the interference pattern of all these spherical waves propagating to the hologram field. The point-based method is straightforward in expressing 3D scene features with points, and the resulting reconstruction has high quality. However, numerical propagating spherical waves from enormous object points to the hologram plane requires a tremendous computation load, one of the most unsolved issues for point-based hologram calculation. Especially when the 3D scene is densely sampled, the required computational load in the hologram formation makes it a challenging task in real-time applications.

If we assume the coordinates of object points are (x, y, z) and the coordinates of the optical field in the hologram plane are (α, β, γ), then the distance r between the object point and the point on the hologram plane is given by:

$$r = \sqrt{(\alpha - x)^2 + (\beta - y)^2 + (z - \gamma)^2} \tag{3.3}$$

The optical spherical wave propagating from the point source to the hologram plane is given by the spatial impulse response of propagation [71]:

$$H_i(\alpha, \beta) = \frac{A_i}{r_i} exp\left(-j\frac{2\pi}{\lambda} r_i\right) \tag{3.4}$$

Note that a constant complex-amplitude factor representing the initial phase and the amplitude in front of the exponential term is neglected. The exponential function $F(\alpha,\beta,\gamma)$ is sometimes referred to as the Fresnel zone plate (FZP):

$$F(\alpha,\beta,\gamma) = \exp\left(-j\frac{2\pi}{\lambda}r\right) \qquad 3.5$$

For a 3D scene with a total P object points, and each point with amplitude A_m and distance r_m from the hologram plane, the optical field on the hologram plane can be computed by summing the FZP of each object point. Then the hologram for the entire 3D scene can be expressed as:

$$H(\alpha,\beta) = \sum_{m=0}^{P}\frac{A_m}{r_m}F_m = \sum_{p=0}^{P}\frac{A_m}{r_m}exp\left(-\frac{j2\pi}{\lambda}r_m\right) \qquad 3.6$$

It can be seen from equation 3.6 that the majority of the computational load is involved in calculating the FZP of the individual object point. Therefore, the look-up table (LUT) method has been proposed to optimise the point-based approach by explicitly trading physical computer memory for computational speed [69]. The LUT method pre-computes the FZPs for all possible object points and stores the FZPs in the computer memory as a look-up table. However, the LUT method requires enormous computer memory to store the pre-computed FZP, which is still a challenging issue even for a modern computer. GPU acceleration has also been brought up for solving this intensive computation [72–74]. Other modified methods, namely, novel-LUT(N-LUT) [75,76], N-LUT with run-length coding [77,78], LUT with non-uniform sampling [79,80], compressed-LUT(C-LUT) [81,82], split-LUT(S-LUT) [83], have been proposed to reduce the computational complexity and mathematical operations.

More recently, Tensor holography suggests modelling Fresnel diffraction and occlusion using a convolutional residual network [44]. Since Fresnel diffraction propagation is the convolution of a wave field with many FZPs, tensor holography approximates FZPs with a set of spatially invariant kernels.

3.3.2 Polygon-based Method

The polygon-based method decomposes the 3D scene into a triangular mesh, a polygon mesh typically used in computer graphics. It comprises a set of triangular facets represented by the coordinates of their vertices. The complex amplitude light distribution in the hologram plane is calculated from each triangular facet object and added to form the final hologram of the 3D scene. Recent reviews and progress on polygon-based computer-generated holography have been reported in [67,84,85].

Leseberg first proposed the polygon-based CGH method in 1988 [86], and its core concept, the angular spectra of rotated planes, was analysed by [87,88]. Matsushima [89] introduced the FFT to calculate holograms from solid shapes. This FFT-based approach calculates the local angular spectrum of the individual triangular mesh by the FFT in a local tilted plane of each triangle. The local angular spectrum is then resampled and interpolated to obtain the global angular spectrum [90–93]. The fully-analytic technique was proposed to find the global angular spectrum directly from the analytic formula of the local angular spectrum without performing the FFT of individual triangles to the local plane [94–96]. Obtaining the local angular spectrum by performing either FFT or analytic formula from individual triangles, both methods consider the rotation and translation of the angular spectrum from the local plane to the global plane of each triangle and propagate the global angular spectrum to the hologram plane.

In the fully analytic polygon-based CGH, individual triangles of a 3D scene are modelled as triangular apertures illuminated by plane carrier waves. The complex amplitude light distribution from each triangular facet object is calculated and added together in the hologram plane to form the final hologram of the 3D scene. The global angular spectrums of the individual triangle are calculated and added to the hologram plane using the analytic formula of the reference triangle in the local plane. Finally, the aggregated global angular spectrum is then Fourier transformed in the hologram plane to give the complex wave field. The procedure is shown in Figure 3.1.

Due to the use of the analytic formula on the individual triangle, the reconstruction has a uniform amplitude and phase only over the individual triangular area, resulting in a flat shading and a noticeable mesh structure of the 3D object. This could be addressed by material functions which act as diffusers varying the amplitude of triangles to achieve a smooth structure [97]. In the fully analytic method, we assume a homogeneous texture pattern of the 3D model. Further research applies texture properties to the triangular mesh [98].

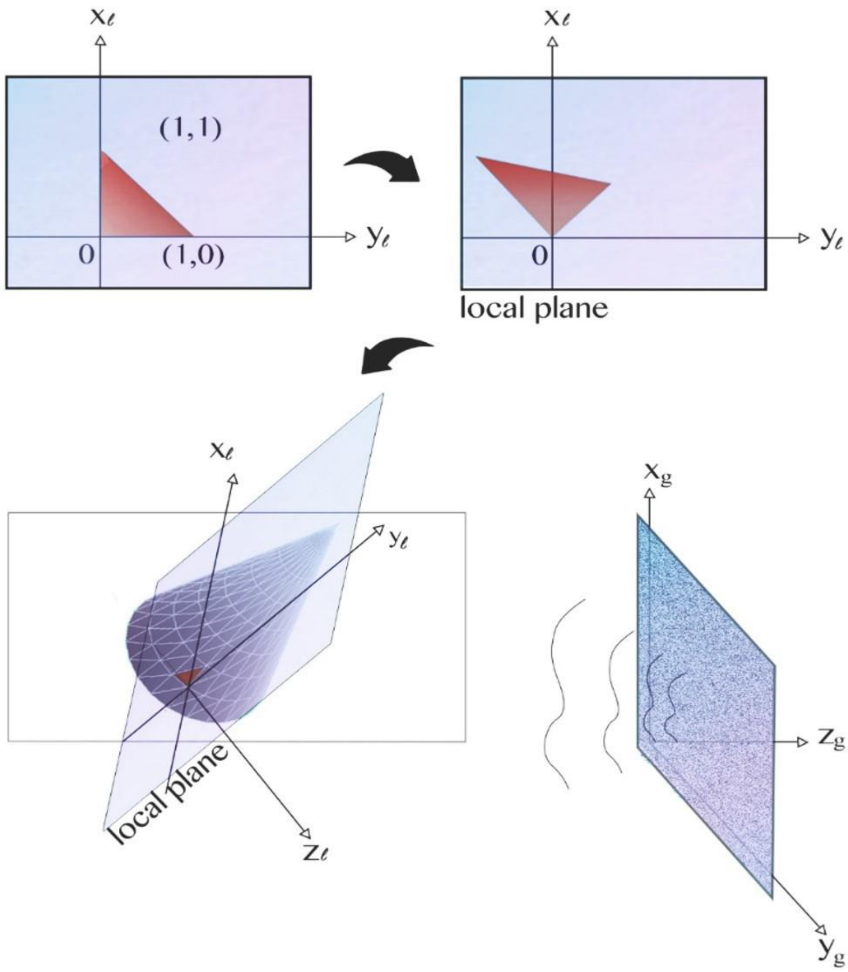

Figure 3.1. Triangular mesh method. The global angular spectrum of a global triangle can be calculated from the local angular spectrum of a local reference triangle. The hologram is calculated by propagating global angular spectrums from all global triangles.

The basic polygon-based method does not appropriately incorporate the occlusion effect, which is a crucial factor in enhancing the depth of information. The lack of global hidden surface removal could result in light from occluded surfaces. Several methods have been proposed, including spatial masking techniques and angular spectrum convolution, to realise the occlusion effect [99–101]. Finally, speed enhancement is achieved by GPU parallel computing and pre-calculated base triangles in [102,103].

3.3.3 Layer-based Method

Layer-based Method Overview

Compared with the point-based and the polygon-based methods, the layer-based method calculates CGHs rapidly with sufficient quality. In the layer-based method, the calculating primitive is the layer. 3D models are sliced into parallel layers that are orthogonal to the hologram plane, and each of the layers is numerically propagated and accumulated on the hologram plane by the FFT with the Fresnel diffraction form [104–106], angular spectrum method [107,108] and other methods [109,110]. The layer-based method has fewer primitives than others, sufficiently improving computational speed. However, the limited number of layers may lead to a discontinuous and layered appearance in the replay field, and the supported viewing angle is limited to several degrees from the axis normal to the hologram plane.

Layer-based Method Principle

If we assume the intensity distribution of an object slice at the plane $z = z_i$ is $I(\mu, \nu; z = z_i)$, then the complex optical field of this slice is given by:

$$U(\mu, \nu;\ z = z_i) \ = \ \sqrt{I(\mu, \nu;\ z \ = \ z_i)}\, e^{j\varphi(\mu,\nu)}, \tag{3.7}$$

where we use a random phase distribution $\varphi(\mu, \nu)$ from 0 to 2π. The random phase distribution smooths out the power spectrum of each object layer and works as a diffuser. Since the human eye works as an intensity detector and is insensitive to the phase distribution, the random phase distribution presented in the complex amplitude of the object slice does not affect the image intensity.

The hologram $H_z(x, y)$ of a layer is obtained by numerically propagating this layer $U(\mu, \nu, z\ )$ to the hologram plane using the Fresnel transform:

$$\begin{aligned} H_z(x, y) &= \frac{e^{jkz}}{j\lambda z}\, e^{\frac{jk}{2z}(\mu^2+\nu^2)} \iint \left\{U(\mu, \nu, z) e^{\frac{jk}{2z}(\mu^2+\nu^2)}\right\} e^{\frac{-jk}{z}(\mu x+\nu y)} d\mu d\nu \\ &= \frac{e^{jkz}}{j\lambda z}\, e^{\frac{jk}{2z}(\mu^2+\nu^2)} \times \mathcal{F}\left\{U(\mu, \nu, z) e^{\frac{jk}{2z}(\mu^2+\nu^2)}\right\}. \end{aligned} \tag{3.8}$$

We loop over all layers to calculate their corresponding complex-amplitude fields and stack them to produce the final hologram W_H:

$$W_H = \sum_{i=1}^{T} H_{z_i} = \sum_{i=1}^{T} \frac{e^{jkz_i}}{j\lambda z_i} e^{\frac{jk}{2z_i}(\mu^2+\nu^2)} \times \mathcal{F}\left\{U(\mu,\nu; z = z_i) e^{\frac{jk}{2z_i}(\mu^2+\nu^2)}\right\}, \quad 3.9$$

where T is the total number of layers. The quadratic phase introduced in the Fourier transform can be regarded as a lens attached to the layer with a specific focal length. With layers having different focal depths, we can propagate the hologram W_H at multiple depths with their corresponding lenses. Since the quadratic phase of each layer can be pre-calculated, the primary computational load of the layer-based method is to perform the FFT of each layer. The computational load is heavily dependent on the number of layers T.

Implementation and Result

The proposed method is verified for demonstrating purpose using a target 3D space station model with 255 layers. The basic idea is to calculate the Fresnel hologram at each depth and accumulate them in the hologram plane. In Figure 3.2, (a) depicts the space station model in 3D space with 255 slices, and (b) depicts the object image.

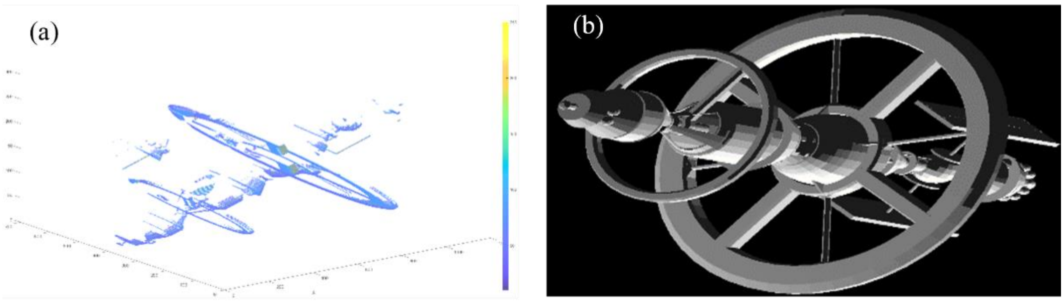

Figure 3.2. Layered images of the 3D space model (a) and its original picture (b).

The pixel pitch is $8\mu m$, and the wavelength is 532nm. The first layer of the space station model is $z\_near = 0.8m$ away from the hologram plane, and the distance between each layer is $delta\_layers = 0.001m$. The quadratic phase is computed and added to the Fourier transform of each layer, respectively. The generated complex field in the hologram plane from each layer is then stacked up as a superposed hologram at a resolution of 1280x1280. The amplitude and phase of the superposed hologram are shown in Figure 3.3.

Two reconstructed fields at layers 5 and 225 of the space station 3D model are calculated and shown in Figure 3.4. The in-focus layers are ideally reconstructed without quantisation from the complex-amplitude superposed hologram, with the same intensity distribution as the corresponding input target layers. The out-of-focus layers appear blurry with speckle. According to the Huygens-Fresnel wavelet principle, the speckle effect is due to the constructive or destructive interference of defocused layers at different planes. Each pixel of the defocused layer emits Huygens wavelets, resulting in either a saturated pixel intensity (constructive interference) shown as a bright spot or a deemed pixel intensity (destructive interference) in the reconstructed field. This speckle effect is not due to quantisation, and the intensity normalisation factor for individual layers could potentially be employed to reduce the speckle effect in the simulated reconstruction.

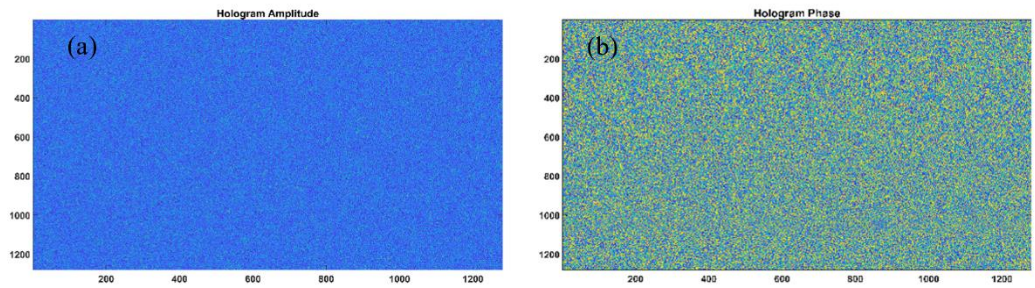

Figure 3.3. The Superposed hologram with the amplitude distribution (a) and the phase distribution (b) from the 3D space station model.

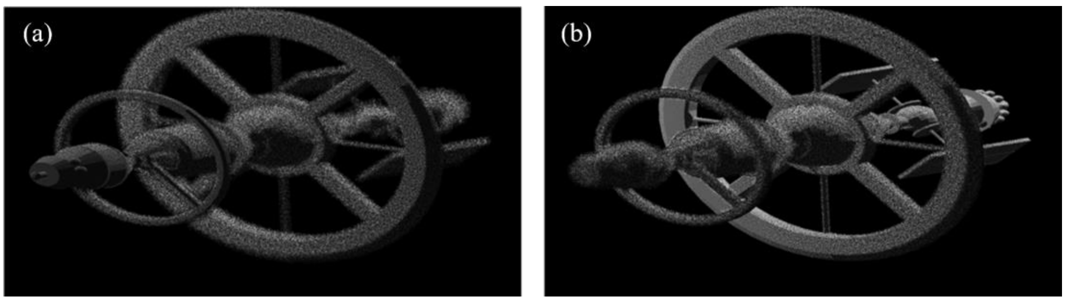

Figure 3.4. The reconstructed images are at layer 5 (a) and layer 225 (b) from the superposed hologram.

Discussion

Although the layer-based method provides fast calculation, the supported viewing angle is limited to several degrees from the axis normal to the hologram plane. A larger viewing angle would result in apparent gaps between reconstructed image layers. Furthermore, due to the limited viewing angle and independence of image layers, the occlusion and shading effects are hard to be implemented in practice. The angular tiling approach divides a 3D scene into several groups of sub-views to provide occlusion and shading effects for the layer-based algorithms, and each sub-view corresponds to a narrow viewing direction [105,106]. Instead of calculating the hologram only from the orthogonal direction, sub-holograms can be calculated by rotating the 3D model or dividing the 3D image within narrow viewing angles. The sub-holograms are spatially multiplexed to form the whole hologram that can reconstruct the 3D scene with multi-viewpoints. The layer-based method with the occlusion effect has also been investigated in [111] by implementing silhouette mask culling.

Practically, for SLMs to display the hologram, the layered images go through several iterations to optimise their phase-only CGH by the GS algorithm. The adaptive addictive iterative Fourier transform algorithm has been reported in [112] to support real-time applications. Computational and image processing techniques, including the depth fused 3D (DF3D) method and the fraction method, have been proposed in [113]. The DF3D method assigns each point in the 3D scenes to its two closest layers instead of one layer in the standard layer-based method, and the DF3D principle decides the magnitudes. This method effectively reduces the depth error during the depth assignment process and could require fewer layers to support the same depth resolution. Sub-sparse two-dimensional Fast Fourier transform (SS-2DFFT) algorithm is proposed to reduce the calculation for sparse image layers [114]. The algorithm is further accelerated by performing two 1D FFT without calculating zero-value columns and rows.

Chapter 4 Holographic Display System

This chapter is primarily based on "Holographic Rendering of a Real-World Scene Captured with a Low-cost RGB-D Camera," F. Yang, Y. Wang, R. Mouthaan, and T. D. Wilkinson, published in Imaging and Applied Optics Congress, The Optical Society (Optica Publishing Group, 2020), paper HF4D.3.

4.1 Introduction

As the previous chapter introduced 2D and 3D algorithms for computer-generated holography, this chapter will demonstrate the outline of building a holographic display. The holographic display has long been considered the pinnacle of display technology, offering accurate depth and focus cues for digital content to resolve problems such as VAC caused by traditional stereoscopic displays. By retaining depth information, holographic displays enable comfortable and natural images.

This chapter proposes a holographic system consisting of a 3D data acquisition module, a data processing module, a CGH generating module, and an optical holographic display module. 3D data can be generated by gaming engines or acquired by depth cameras, potentially allowing the creation of dynamic and interactive holograms required for augmented reality (AR) applications. The resulting data is then processed by applying depth-colour alignment, image normalisation and resampling for CGH generation. The previously introduced algorithms can calculate and reconstruct CGHs of the 3D data. CGHs are then transmitted to the optical holographic display module, allowing optical 3D reconstructions. Several artefacts and potential approaches in implementation are further demonstrated to improve the reconstruction quality.

4.2 3D Data Acquisition

A holographic display system requires 3D data to calculate CGHs. 3D data can be broadly divided into two groups: artificial 3D data generated from computer graphics and real-world 3D data captured by depth cameras. Most of the research on holographic displays utilises artificial 3D data as they are easily manipulated and mathematically well-defined by graphic descriptions [115,116]. Created from gaming engines and numerous design software, artificial 3D data consists of object 3D models, object properties and other rendering information for the scene.

On the other hand, in some practical applications, including telecommunications and video conferences, real-world 3D data is preferred and has become an active research topic over recent years. Kim et al. presented a point source-based CGH method using stereoscopic images of a real 3D object captured by a Wasol 3D camera system. The depth image was calculated by stereo image matching [116]. Lee et al. obtained a digital hologram of a real 3D object from a prototype sensor with a time-multiplexed method for colour and depth image acquisition to generate a digital hologram of a real 3D object using the point-based approach. However, pixel saturation limits the depth resolution to below 1m [117]. Yamaguchi et al. presented a scanning system with a vertical camera array using the ray-sampling plane method and applied it to holography [118]. Li et al. extracted 360-degree depth images of a real object from a Kinect depth camera and merged depth images into a single 3D model [119]. Yanagihara et al. demonstrated a real-time, 3D holographic display using Kinect v2 to reconstruct a 3D video at ~14 frames per second [120]. Compared with artificial 3D scenes, real-object-based methods have accurate spatial imaging and are simple to afford efficient natural visualisation without rendering techniques.

Several existing 3D object datasets recorded from various off-shelf cameras will be introduced later, providing pixel-perfect ground truth for quality evaluation and depth quality compensation methods for the holographic display acquisition system. Despite RGB-D images or point clouds being acquired from various sources, a 3D holographic display acquisition module can convert various forms to the same form for CGH calculation.

4.2.1 3D Data from Unity

3D models can be represented as RGB-D images. An RGB-D image is a single view of a 3D model with colour information typically stored in the RGB image and depth information in the depth image. Each pixel in a depth image corresponds to a distance between the image plane and its matching object in an RGB image. In Unity, the ray casting technique provides scene rendering options scene and can record RGBD images. A ray leaves the camera through a grid of pixels in a specific direction and travels until it hits the closest object. RGB intensities of each pixel from the contributions of light in all directions with its texture properties are calculated at each surface intersection location. The depth information is created by adding a depth shader to the camera to calculate the depth of each pixel rendered based on the input of the camera depth texture. RGB and depth images are saved separately as 2D images shown in Figure 4.1.

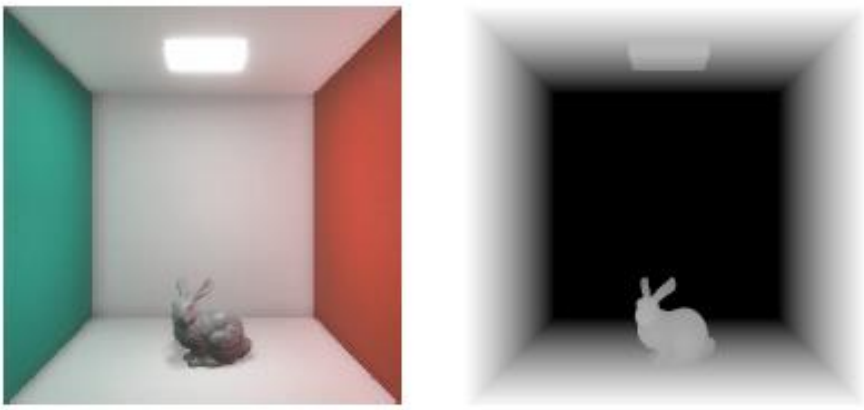

Figure 4.1. RGB-D images from the Unity game engine.

3D data acquisition can also be achieved using 3D models in Unity as point clouds, as shown in Figure 4.2. A point cloud defines XYZ coordinates, which precisely record the geometric shapes of 3D models in space. Models are typically represented as the boundary of objects, not solid volumetric objects. Each point of 3D models is associated with an RGB intensity value. The RGB value of each point is dependent on the rendering options in the scene.

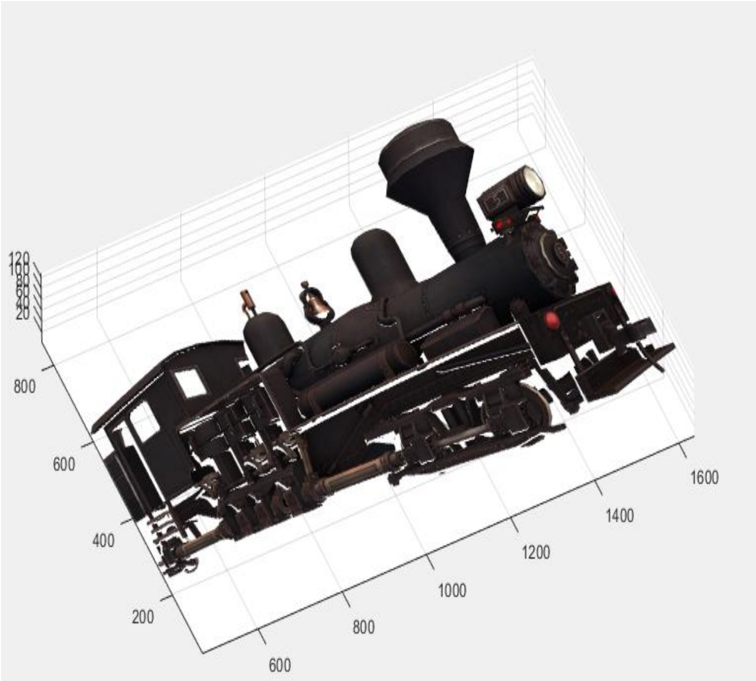

Figure 4.2. A point cloud sample from the Unity game engine.

4.2.2 3D Data Acquisition Using Depth Sensors

Depth cameras can also achieve 3D data acquisition. The advance of 3D imaging enables inexpensive depth sensors in consumer products (e.g. Kinect, Zed Camera, RealSense). These depth cameras operate based on Structured Light (SL), Time-of-Flight (TOF) or stereo triangulation to measure the depth information in a 3D scene. The acquired depth information and colour information from these cameras are then stored as RGB-D images or point clouds.

An Intel RealSense D435i camera and a Zed camera are used for the 3D image acquisition system. Both cameras provide Software Development Kit (SDK) with well-documented examples to acquire 3D images represented by point clouds. A quantitative comparison is performed between two cameras to decide which is more applicable to the CGH generation.

Quantitative Depth Quality Comparison

Various measures could be used to characterise the depth quality acquired by depth cameras. The measurement should be based on a standard test environment compatible with different depth cameras. A typical example of an acceptable test is a white, flat wall with an 85% reflectivity, which has been utilised as the traditional test target for TOF and SL. Intel RealSense Camera introduces the depth testing methodology for depth quality evaluation [121].

The methodology introduces primary depth quality metrics, including Z-accuracy, fill rate, spatial RMS noise and temporal noise (or uniformity) computed within the desired region of interest (ROI) from a camera. These metrics are independent and can be further combined for advanced depth quality metrics to evaluate the depth-related performance of cameras.

Z-accuracy describes the disparity between the measured depth values and ground truth values. It is determined by the median value of the signed difference between the depth images measured and the ground truth. The fill rate measures the point density and valid pixels (with non-zero depth values) of the image. It is the percentile of pixels with a valid depth value relative to the total pixels within the ROI. The spatial noise evaluates spatial uniformity by measuring the intrinsic variation in depth values. The standard deviation (STD) of the measured depth values is utilised to evaluate the variation in the spatial noise within the ROI. Temporal noise evaluates the temporal uniformity of the variation in depth values over time, defined by STD in depth values within a sequential frame.

Two cameras are set to high definition (HD) resolution to compare the imaging performance with 30 frames per second (FPS). The ROI is 40% at the centre of the image. The testing distance ranges from 0.5 to 1.5 metres to capture a depth image from the test target. All other configurations are based on their default configurations.

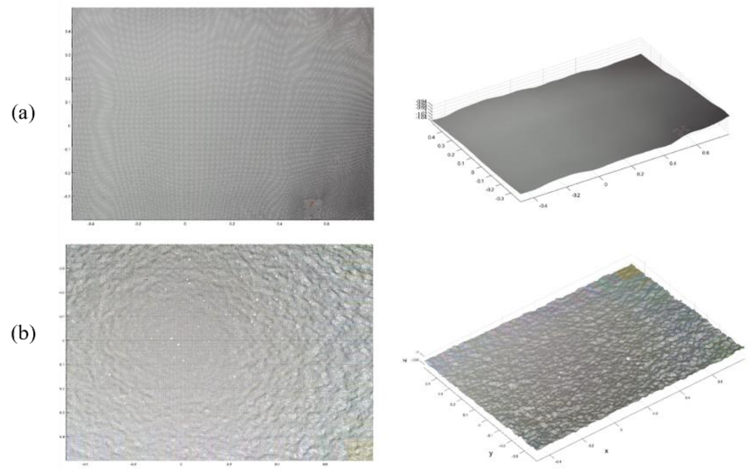

Figure 4.3. Captured point clouds of the test target from Zed (a) and Realsense (b) cameras at 1.0m.

Table 4.1 shows the Z-accuracy, fill rate, means $\mu1$, $\mu2$ and STD values $\sigma1$, $\sigma2$ relative to distance and time respectively, for both cameras. We calculate the mean and STD values within five frames to measure the temporal noise. The point clouds in Figure 4.3 are acquired at a distance of 1m from the test white wall using the same viewer.

Both cameras have a Z-accuracy of less than 1% over their ROI, providing accurate depth data at a distance of around 1m. As the fill rate shows, the RealSense camera provides less valid depth data than the Zed camera. Due to the significant disparity between the colour sensor and the depth sensor, the Zed camera is incapable of producing depth pictures for distances less than 0.65m. For the spatial noise, The STD values $\sigma 2$ of the depth error at each distance reveal that the point cloud from zed is more uniform and smoother with less amplitude variation than the point cloud from the RealSense camera. The RealSense camera has less noise within the given time interval than the Zed camera for the temporal noise. Additionally, the zed camera sometimes has an unstable issue.

More quantitative and systematic studies of the capability of Zed and Realsense cameras can be found in [122–125]. It should be noted that the overall point cloud quality generated from both devices is the combined performances of the depth-sensing hardware with their optimised point cloud generating software algorithms.

Table 4.1. Depth quality comparison for the ZED and the Realsense Camera.

Device	Distance	Z-accuracy	Fill rate	μ1 [m]	σ1[m]	μ2[m]	σ2[m]
RealSense	0.5m	0.20%	99.68%	0.4938	0.1321	0.4933	0.0462
Zed	0.5m	N/A	N/A	N/A	N/A	N/A	N/A
RealSense	0.75m	0.30%	99.68%	0.7529	0.0109	0.7532	0.0081
Zed	0.75m	0.10%	100.00%	0.7573	0.0105	0.7704	0.0131
RealSense	1m	-0.50%	99.99%	0.9919	0.0420	0.9920	0.0045
Zed	1m	0.12%	100.00%	0.9973	0.0148	0.9836	0.0273
RealSense	1.25m	0.16%	99.70%	1.2496	0.0151	1.2495	0.0070
Zed	1.25m	0.44%	100.00%	1.2518	0.0028	1.2488	0.0201
RealSense	1.5m	0.01%	99.95%	1.4991	0.0110	1.4989	0.0273
Zed	1.5m	0.08%	100.00%	1.5039	0.0064	1.5031	0.1317

Qualitative Depth Quality Comparison

Using a flat wall to compare depth quality is relatively simple. Complex sceneries, on the other hand, are challenging to quantify. The performance of both cameras with a complex scene is further evaluated to identify essential factors for depth cameras qualitatively. A test scene in Figure 4.4 is captured 70 cm away from occluded objects.

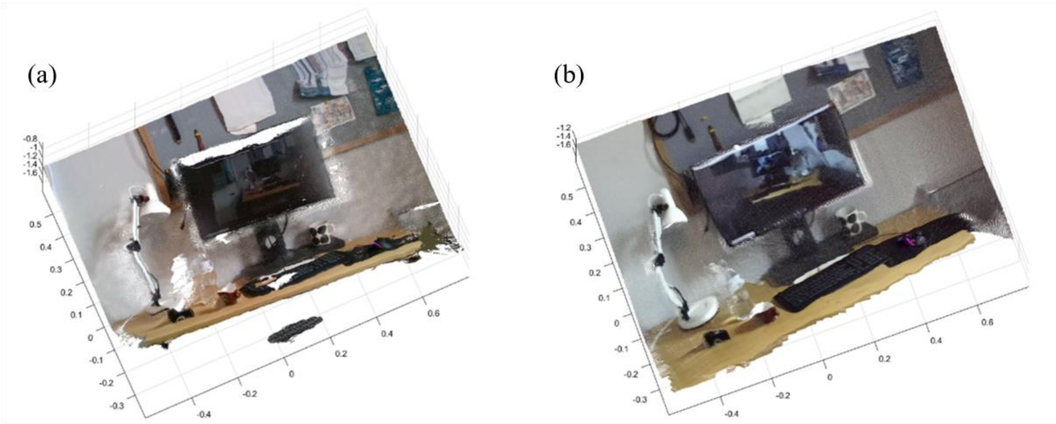

Figure 4.4. Point clouds of a complex scene from the Zed (a) and the RealSense (b) Cameras.

Occlusion Effect. The phenomenon of colour and depth images coming from different positions that are not entirely aligned could result in the occlusion issue, as shown in Figure 4.5 (a). This phenomenon arises due to the physical position disparity between colour and depth sensors. The position disparity could result in invalid data, showing holes around the object. This is most likely to occur along edges. Due to the zed camera having a significant disparity between the RGB and the depth sensors, it does not support a proper depth and colour image alignment for a short distance.

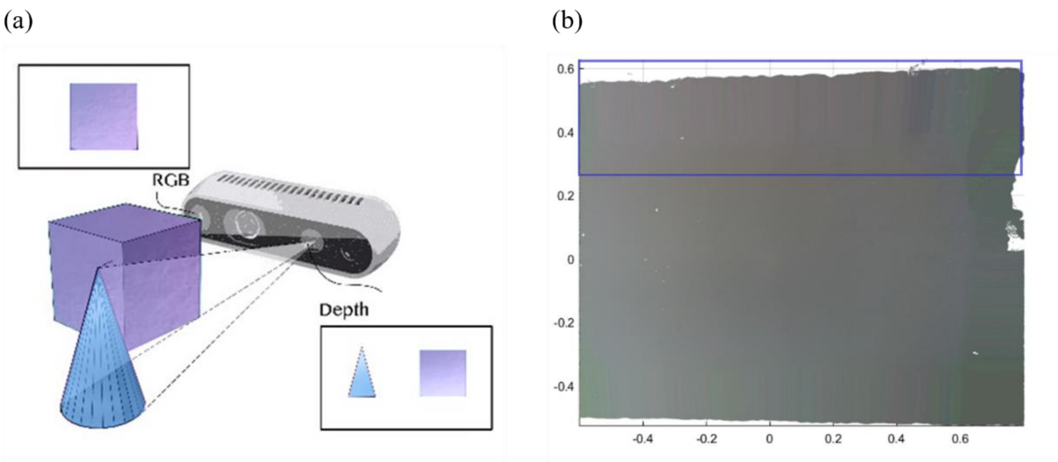

Figure 4.5. (a) Occlusion due to sensor position disparity. (b) Smeared texture due to the sensor FOV disparity. The RGB image boundaries are stretched to match the resolution of the depth image.

Smeared Texture. The available colour boundaries of the RealSense camera are stretched to provide texture data to the depth coordinates according to the OpenGL rendering configuration shown in Figure 4.5 (b). The texture smearing is a consequence of the depth and colour sensors having a different field of view, which is particularly noticeable with the RealSense camera.

The colour sensor has a smaller FOV than the depth sensor. Therefore, the area covered by the depth sensors does not have corresponding pixels from the colour sensors. In contrast, the depth and the RGB camera FOV are more consistent for the Zed camera.

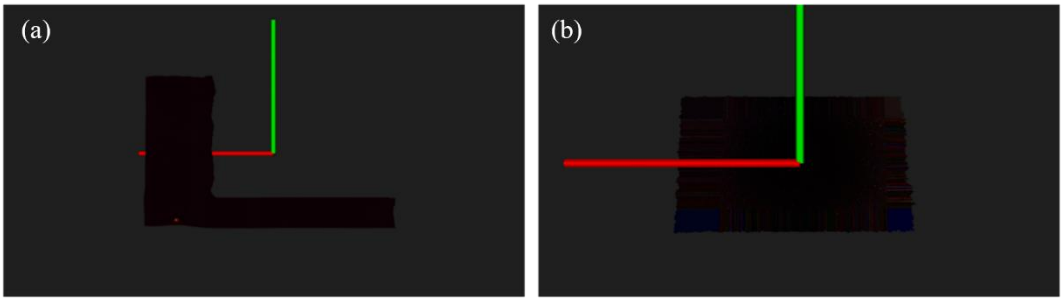

Figure 4.6. (a)The ZED camera uses triangulation to create a depth image, capturing only part of the point cloud under dark lighting conditions. (b) The RealSense camera uses ambient light and an IR pattern and can work in low-light conditions.

Lighting and Materials. Since depth cameras have different mechanisms to collect light, it is recommended to test depth cameras in different light conditions: dark lighting, natural sunlight, and home lighting. Light can come from the depth camera itself or the ambient light in the scene. Figure 4.6 depicts the acquired point clouds from both cameras in the dark lighting condition. The ZED camera uses triangulation (re-projection) from the geometric model to create a depth image, resulting in only a part of the point cloud being captured. The RealSense camera uses ambient light and a projected infrared (IR) pattern. IR-based depth-sensing methods such as TOF and SL can function under dark lighting conditions. However, the quality of the depth data acquired by IR-based sensors can be degraded for materials that do not reflect IR. Included in these materials are leather, black clothes and reflective surfaces. Continuous reflective surfaces in 3D space can be fragmented into pieces.

4.2.3 Synthesis Point Cloud Datasets

The quality of real-world 3D data is compromised by measurement noise, missing depth observations, and position disparity of sensors. Several post-processed RGB-D datasets obtained from depth cameras have been proposed to meet specific needs that may not be found in the original, real data [126–133]. As shown in Figure 4.7, these datasets enable us to set a baseline with real-world 3D data acquisition for evaluating and testing hologram-generating algorithms. Moreover, these datasets also provide measurement error compensation techniques, which can be applied to the depth camera acquisition module.

The NYU Dataset v2 is introduced in ECCV 2012, using Kinect v1 to collect real-world 3D data [133]. Missing values in the raw depth images result from the disparity between the infrared emitter and RGB sensor due to the triangulation-based nature of the Kinect. The missing values are filled by the colourisation scheme of Levin et al. [134]. In applications like perceptual organisation, amodal completion and semantic segmentation, the missing depth observations and in-depth error observation could lead to a nonlinear noise and fragmented common surfaces. The lack of time synchronisation between colour and depth channels could result in misalignment in the dataset. The same issues have been encountered by the SUN RGB-D dataset proposed in CVPR 2015 for collecting real-world 3D data from various depth cameras, including Kinect and RealSense [129]. The depth quality is degraded mainly by the measurement noise, view angle to the regularly reflective surface, and occlusion boundary. Issues are resolved by using nearby frames which contain 3D rotation and translation information to align and warp the depth image. Recently, Stanford presented a 3D point cloud dataset of large-scale indoor areas [135]. The proposed dataset using a Matterport Camera collects 70,496 regular RGB images with depth and their corresponding 695,878,620 points, semantic annotations, and camera metadata. It is possible to generate holograms utilising this large-scale point cloud dataset.

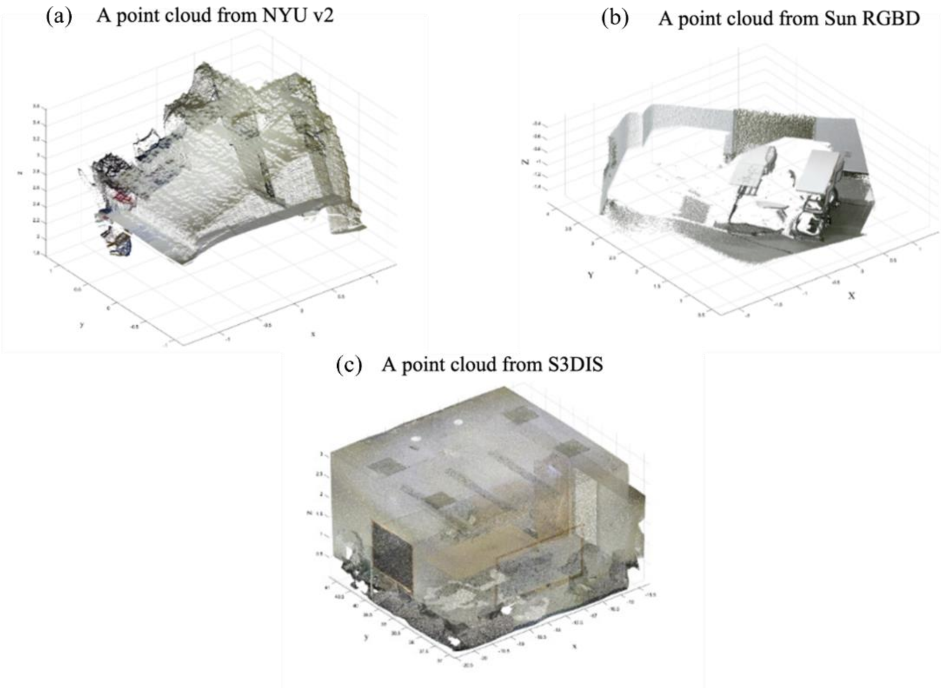

Figure 4.7. Synthesis point cloud datasets: (a) NYUv2 [132], (b) SUN RGBD [129], (c) S3DIS [135].

4.3 Data Processing Module

The last section introduced the 3D data acquisition module. A depth camera is used to simultaneously acquire depth and colour data of real scenes to extract colour point clouds or RGB-D images for CGH generation. However, the acquired point clouds or RGB-D images from depth sensors often suffer from missing values. Completion methods will be introduced to achieve better 3D data quality. The layer-based method for CGH generation considers a 3D scene composed of multiple layered images. Therefore, 3D data are resampled into layered 2D images with occluded values deleted. These layered 2D images are used for CGH generation.

4.3.1 Point Cloud Processing for CGH Generation

Point Cloud Completion

The point clouds produced by depth cameras usually contain substantial depth information and have exact coordinates and colour information. Each point of the real 3D object is represented by a six-component vector $v = (X\ ,Y\ ,Z, R, G, B)$. To achieve a better point cloud, we can use the Iterative Closest Point (ICP) algorithm to register point clouds from nearby frames [136,137]. ICP algorithm transforms the reference point cloud $Q \triangleq \{\boldsymbol{q_i}\}_{i=1}^{N_q}$ to the target point cloud $P \triangleq \{\boldsymbol{p_i}\}_{i=1}^{N_p}$ set and revises the transformation iteratively to optimal align the matched pairs.

Essentially, for kth iteration, the algorithm first establishes a correspondence between the two-point sets with the last estimated rotation matrix R_{k-1} and the translation matrix T_{k-1} using the MSE metric to minimise the point-to-point matching error and optimise the alignment of each source point to its last-founded match. For N_p points, we find the correspondence $C_k(i)$ which gives the best estimate of the closest point in the reference point set to the given point in the target point set:

$$C_k(i) = \underset{j\in\{1,2,\ldots N_q\}}{\operatorname{argmin}} \left(\frac{1}{2}\left\|(R_{k-1}\boldsymbol{p_i} + T_{k-1}) - \boldsymbol{q_j}\right\|^2\right), for\ i = 1,2,3 \ldots N_p \qquad 4.1$$

Second, we can compute new transformation matrices R^*, T^* between the last transformed point set $T \triangleq \{R_{k-1}\boldsymbol{p_i} + T_{k-1}\}_{i=1}^{N_p} = \{t_i\}_{i=1}^{N_p}$ and the corresponding closest point set $\{\boldsymbol{q}_{C_k(i)}\}_{i=1}^{N_p}$ from the target set P by minimising the distance:

$$(R^*, T^*) = \underset{R^T R = I_m,\, \det(R)=1}{\operatorname{argmin}} \left(\sum_{i=1}^{N_p} \left\| (Rt_i + T) - \boldsymbol{q}_{C_k(i)} \right\|^2 \right) \tag{4.2}$$

Then R_k and T_k are updated according to the new transformation matrices. Using the reference point cloud model, we can register other point clouds from nearby frames by using the ICP algorithm to synthesise an accurate point cloud $P^* \triangleq \{\boldsymbol{p_i^*}\}_{i=1}^{N_{p^*}}$.

Depth Layers from Point Clouds

Having collected a reasonably accurate point cloud P^* from the depth camera, we then normalise and rasterise the point cloud into depth layers with RGB channels according to the z coordinate information of each point. Each point initially has real-world coordinates and is then transformed to match the resolution of the hologram. The z coordinate of all points is resampled and indexed into depth grids based on the trade-off between the high fidelity and the computational load. All points on each depth grid have the same z coordinate or the same depth. This step effectively transforms the point cloud into layered 2D images, where the pixels of each image represent the points of each depth layer and contains RGB values. The resulting layered 2D images can be expressed as $L \triangleq \{I_i\}_{i=1}^{T}$, where I_i is an individual RGB image, and T is the number of layers. The layered 2D images L are then divided into sublayers for RGB channels, where the pixel of each image only contains a single value for each RGB channel. The layered 2D images for each channel are represented by $L_m \triangleq \{I_{i,m}\}_{i=1}^{T}$Where m is the colour channel.

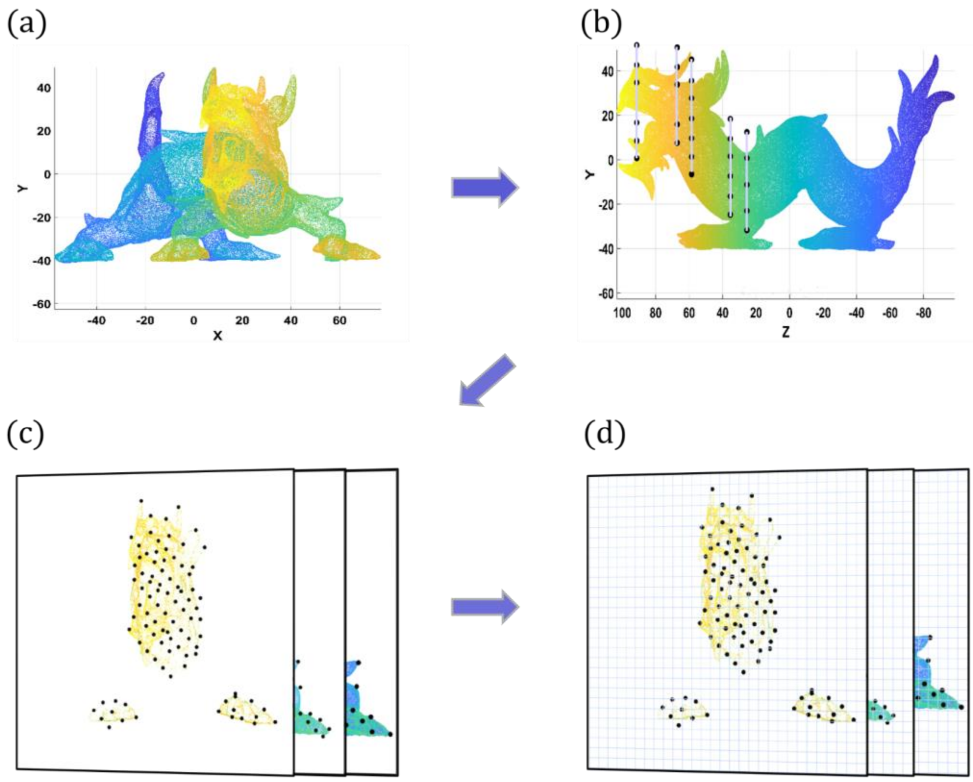

Figure 4.8. The point cloud data processing module.

The process can be seen in Figure 4.8. We normalise the collected data in Figure 4.8(a) and slice the point cloud into layers by the z coordinate with RGB channels shown in Figure 4.8(b). The sliced layers are shown in Figure 4.8(c). We resample all the layers and rasterise them into grids. The repeated points are overwritten during this process. The resultant layers are shown in Figure 4.8(d) for a single RGB channel.

A zed 3D depth camera captures the experimental point cloud displaying a complex 3D scene of a room layout. The captured point cloud is first selected within the ROI to filter out the smeared edges and unwanted areas spatially. The ICP algorithm registers the point cloud with another one from the nearby frame to improve registration accuracy. The Lidar and Point Cloud Processing Toolbox in Matlab implements both the ROI selection and ICP algorithm. The Point Cloud Library (PCL) provides point cloud registration functions for C/C++ applications. The resulting point cloud is shown in Figure 4.9.

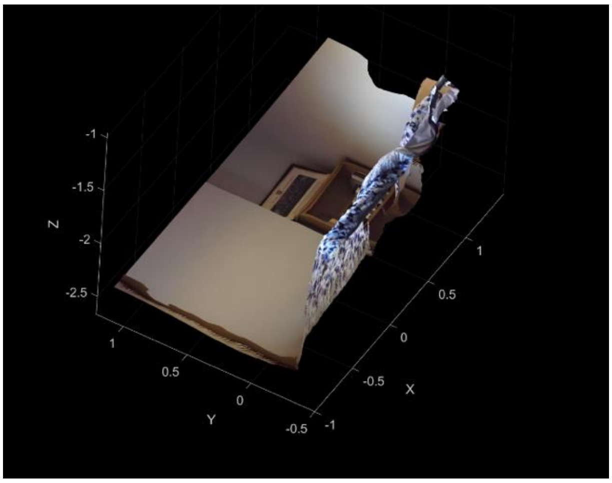

Figure 4.9. Registered Point cloud from the Zed camera.

The point cloud collected has spatial information from the real-world scene. The point cloud is normalised and rasterised into depth layers in 1080×1080 with red, green, and blue (RGB) channels to match the resolution of the simulated hologram. Thus, the point cloud is stretched to 1080×1080 in the horizontal plane. The number of layers depends on the trade-off between the computational load and the reconstructed image quality. For demonstration purposes, we slice the point cloud into $T = 30$ depth layers. Although the limited number of layers could result in a discrete reconstruction, each depth plane contained more points and gave a better in-focus and defocus impression. The point cloud is first shifted to the centre and normalised to match the resolution of the hologram. The z coordinates of all points are rounded and indexed into 30 depth layers. The resulting point cloud is shown in Figure 4.10.

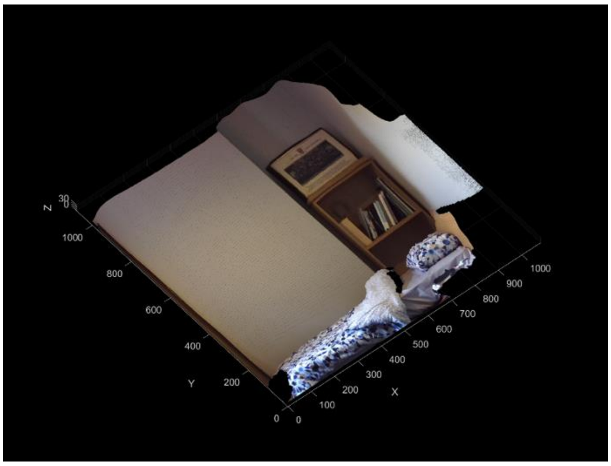

Figure 4.10. The point cloud is stretched to a 1080*1080*30 grid.

The round function assigns each point to the nearest layer index and deletes repeated depth values, reducing matching errors and enhancing the quality of the reconstruction. Therefore, the point cloud contains unique positive integers with fewer matching errors during the layer index assignment. The resulting layered 2D images $L = \{I_i\}_{i=1}^{T}$ is then divided into individual RGB channel layered images $L_m = \{I_{i,m}\}_{i=1}^{T}$ by iterating all the images and storing the single R, G, B values correspondingly.

4.3.2 RGB-D Image Processing for CGH generation

Raw depth images captured with RGB-D cameras could have invalid pixels. The position disparity between the RGB and the depth sensors could result in missing data, especially when scenes are close to RGB-D cameras. The missing values of depth images acquired from the cameras are in-painted using the colourisation scheme [134]. The raw calibrated RGB and depth images are shown in Figure 4.11(a), (b) and the in-painted image is shown in Figure 4.11(c). Both depth images shown are normalised in 8 bits. The colourisation method converts RGB images to grayscale and then applies raw depth images as the weighting factor to colourise the grayscale images. Therefore, seeing from the other way, the raw depth images are in-painted according to the ground truth grayscale images. An essential assumption from the algorithm is that if the brightness of two adjacent pixels is similar, their colour values should also be similar. This assumption is valid for most cases of the depth inpainting process since the missing values most likely occur at the edges of objects due to occlusion, where adjacent pixels typically have different brightness. The in-painted depth images have measurement errors during data acquisition and depth estimation errors during depth completion. In-painted RGB-D images can be easily converted into individual RGB depth layers by rounding depth values to the nearest layer indices.

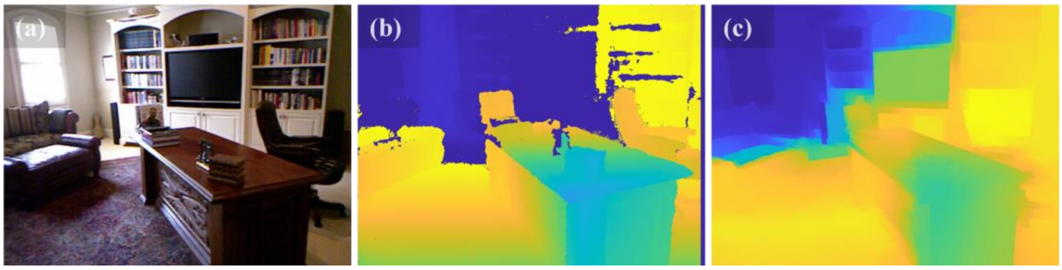

Figure 4.11. Depth image in-painting using the colourisation scheme. (a) Raw captured RGB image. (b) Raw captured depth image. (c) In-painted depth image.

4.4 CGH Generating Module

The layer-based method decomposes a 3D model into layered 2D images to calculate CGHs on each RGB channel, as shown in Figure 4.12. Each 2D image intensity $I_{i,m}(x, y; z = z_i)$ can formulate a complex amplitude object field $U_{i,m}(x, y; z = z_i)$ with a random phase distribution $\varphi_{i,m}(x, y)$, where subscript i denotes the index of its corresponding depth layer. k_m is its corresponding wavenumber with m as the RGB channel. Then the complex amplitude field $H_{i,m}(\mu, \nu)$ in the hologram plane per channel at a fixed depth $z = z_i$ can be calculated from equation 3.8 by:

$$H_{i,m}(\mu, \nu) = \frac{e^{jz_i k_m}}{j\lambda z_i} e^{\frac{jk_m}{2z_i}(\mu^2+\nu^2)} \times \mathcal{F}^{-1}\left\{U(x, y; z = z_i)e^{-\frac{jk_m}{2z_i}(x^2+y^2)}\right\}. \qquad 4.3$$

The Fourier Transform with the quadratic phase factor numerically propagates the object field to the hologram plane for each layer per channel. Quadratic phase factors for different RGB channels can be precalculated to reduce the computational load of the layer-based method. The full hologram W_m per channel is then calculated by stacking the sub-holograms $H_{i,m}$ of each layer, which can be expressed as:

$$W_m = \sum_{i=1}^{T} H_{i,m} \qquad 4.4$$

To optical display the resulting hologram on SLMs, we need further encode the complex-amplitude hologram W_m into phase-only holograms. The required phase level is dependent on the specification of the SLM. For a 256-level phase-only SLM, we can perform the basic GS algorithm on each sub-hologram from layers. Then the phase-only sub-hologram can be expressed as $H_{i,m}^{p} = gs\{ H_{i,m}\}$ and the full phase-only hologram is $W_m^p = \sum_{i=1}^{T} H_{i,m}^{p}$.

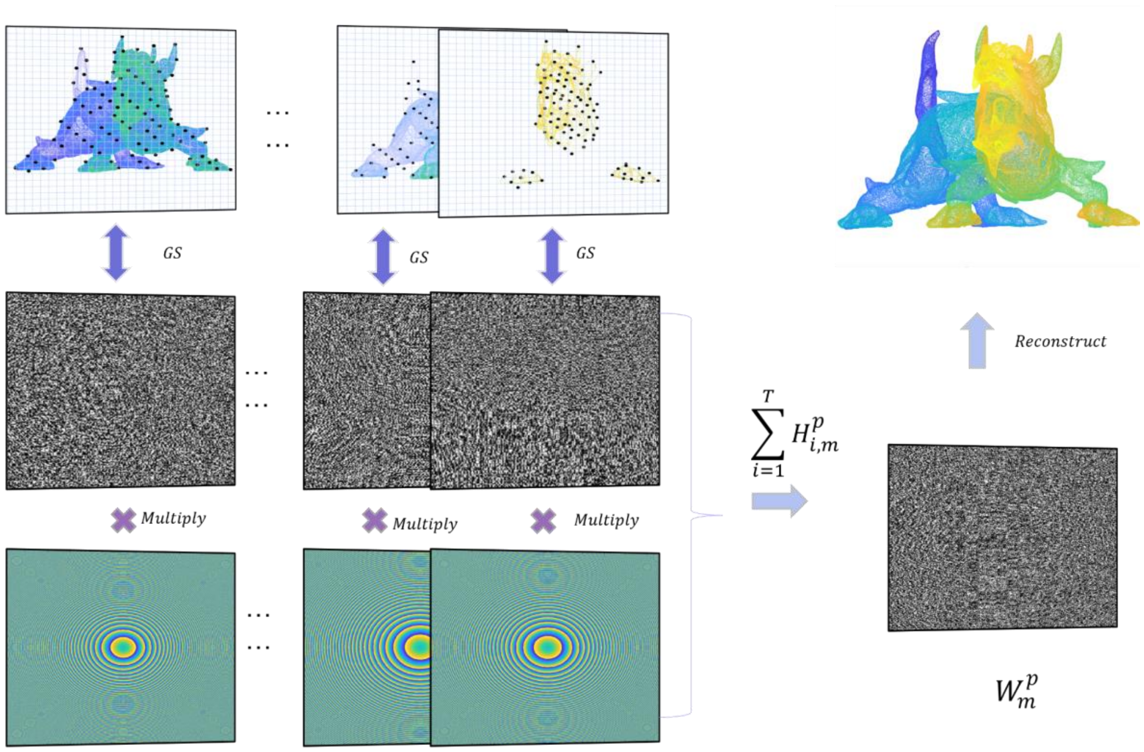

Figure 4.12. Hologram generation from each layer per channel.

In summary, the procedure for computing the full-colour CGH of a real 3D object can be summarised as follows:

1. Acquire and process an accurate point cloud or an RGB-D image from the depth camera.
2. Obtain layered 2D images for each R, G, and B channel.
3. Generate CGHs based on the layered 2D images for each RGB channel using the layer-based method.
4. Combine the RGB CGHs.
5. Perform the GS algorithm and write the CGHs for the specific SLM display.

4.4.1 Numerical Simulation from Real Captures

We generate CGHs based on the layered 2D images for each RGB channel L_m by performing the layer-based method. The pixel pitch is 8μm; wavelengths are 630nm, 532nm, and 465nm; the distance from the hologram plane to the first 2D image $I_{1,m}$ $z\_near = 0.8m$. The distance between each layer is set to $delta_layers = 5$mm, which can be calculated by the range of the real-world z coordinates from the acquired point cloud.

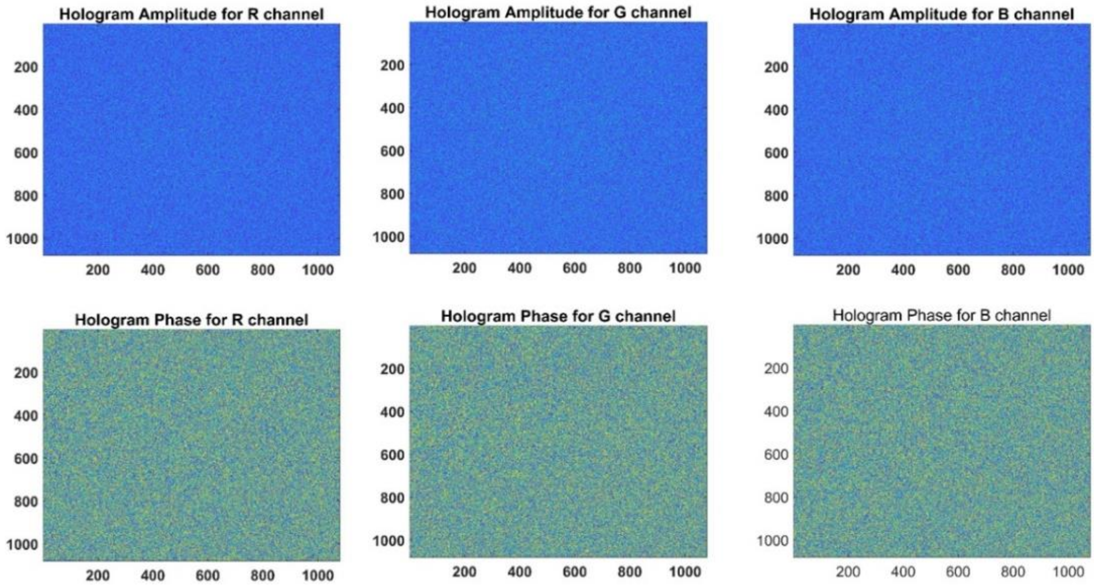

Figure 4.13. Amplitudes and phases of generated R (left), G(middle) and B(right) holograms.

We perform FFT on the optical field $U_{i,m}$ for each 2D image $I_{i,m}$ and add individual quadratic phases, respectively, for each RGB channel based on the layer-based method. The generated sub-hologram $H_{i,m}$ from each layer per channel is then stacked to form the superposed hologram W_m at a resolution of 1080x1080. W_m is a 2D complex-amplitude field in the hologram plane which can reconstruct the 3D layered images in the object plane. The amplitude and phase of the hologram for each colour channel W_R, W_G, W_B are shown in Figure 4.13.

Two reconstruction fields of the 3D model at layers 3 and 24 are simulated in Figure 4.14. Perfect reconstruction can be achieved on the in-focus layers. The defocusing effect can be clearly shown in the RGB reconstructed images. However, the defocus effect is rather noisy due to the multiplane crosstalk. The direct superposition of sub-holograms does not consider the correlation of each complex-amplitude layer, resulting in reconstruction at different depths negatively interfering with each other.

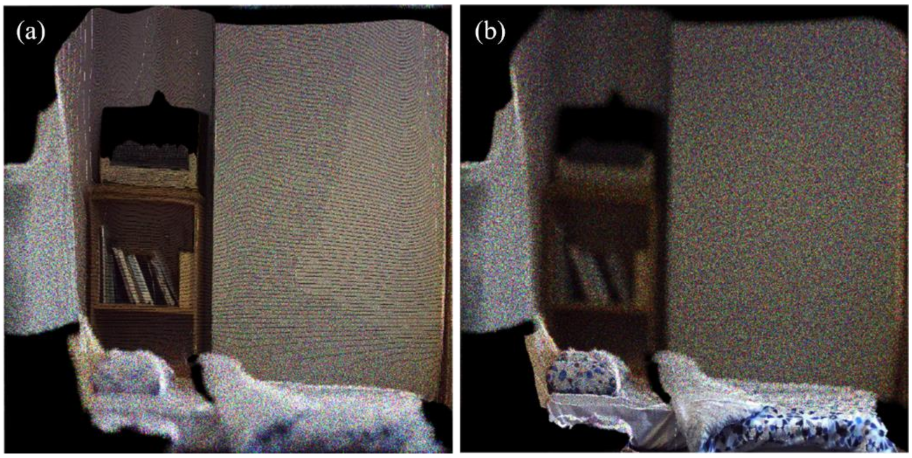

Figure 4.14. Reconstructed Images in the RGB channel at layers 4(a) and 24(b).

4.5 Optical System Design

This section describes the construction process of an experimental holographic display system used in this work to reconstruct holograms optically. The holographic display system is a prototype demonstrating proof-of-concept and theoretical test conclusion. The optical design can be further developed to compensate for the aberration associated with the optical system. The schematic diagram outlining the layout of the 3D holographic display is presented in Figure 4.15.

The optical design consists of a 532nm collimated laser source (A) mounted by a rotation positioner. A half waveplate (B) rotates the polarisation state of the laser beam from the laser. The beam is expended by the lens (C) and spatially filtered by a circular aperture (D) and a slit aperture (E). The expanded beam then passes through a collimating lens (F) and is linearly polarised by a polariser (G). The collimated beam illuminates the SLM (H), travelling through a beam splitter cube (I). The reflected beam from the SLM then passes through a Fourier lens (J) to produce a reconstructed image at its focal distance. The reconstructed image is expanded by another lens (K) and recorded at the required size on a CCD camera (L).

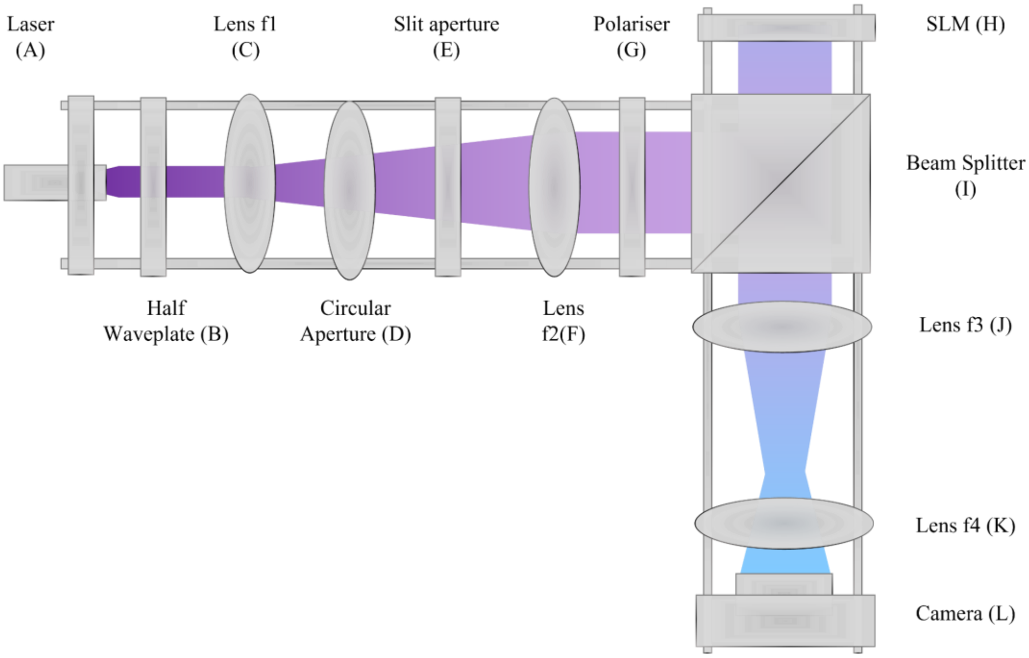

Figure 4.15. The 3D holographic display schematic diagram.

The following steps determine the designing parameters of optical elements in the holographic display system:

1. SLM parameters (H)

The primary element in the holographic display system is the SLM (H). A smaller dimension would reduce the required diameter of the collimating beams incident on the SLM, and a smaller pixel pitch would increase the scale of the replay field. Our available SLM is a reflective phase-only Holoeye LETO SLM, with specifications in Table 4.2.

2. Laser parameters (A)

The available laser source is a Thorlabs 532nm, 0.9mW collimated laser source (A) emitting a Ø3.5mm Gaussian beam. We can calculate the required magnification of the beam dimension to illuminate the selected SLM. Since the collimated beam emitting from the laser is a Gaussian profile, only the approximately flat-top (or top-hat) region of the Gaussian beam is used to illuminate the SLM near-uniformly. Thus, the magnification is determined by the diameter of the flat-top region of the Gaussian beam and the diagonal of the SLM.

Table 4.2. Holoeye LETO SLM Micro-display Features.

Display Type:	Reflective LCOS (Phase Only)
Resolution:	1920 x 1080
Pixel Pitch:	6.4 µm
Fill Factor:	93 %
Active Area / Diagonal	12.5 x 7.1 mm (0.55″ Diagonal)
Addressing	8 Bit (256 Grey Levels)
Signal Formats	HDMI – HDTV Resolution
Input Frame Rate	60 Hz / 180 Hz

The diameter of the Gaussian beam from the laser source is $\emptyset 3.5\text{mm}$ in $1/\mathrm{e}^2$ width, describing the radial distribution of the Gaussian beam at 13.5% of its maximum. The equation describing the radial intensity profile of the Gaussian beam $I(r)$ is given by

$$I(r) = I_0 \exp\left(-\frac{2r^2}{w^2}\right), \tag{4.5}$$

where I_0 is the maximum intensity value, w is half of the beam width, and r is the radial distance from the centre axis of the beam. We empirically choose the radial distance at $r_d = 0.2w$, so that the intensity is 92.31% of its maximum, which provides a nearly-uniform beam in this region. Thus, the ratio between the diameter of the nearly flat-top region to the beam diameter is 0.2 and the nearly-uniform collimated laser source diameter D_{Laser} used for SLM illumination is $\emptyset 0.7\text{mm}$. The magnification is given by the ratio between D_{Laser} and the diagonal of the SLM D_{SLM}. From Table 4.2, D_{SLM} is $0.55''$, corresponding to a diameter of 13.97mm. Therefore, if the SLM is illuminated by the flattop of the Gaussian beam, the magnification required is $\mathrm{M}_{SLM/Laser} = D_{\text{SLM}}/D_{Laser} \approx 20$.

3. Lens parameters (C) and (F)

We use an expanding beam system as required by the magnification to enlarge the collimating beam from the laser. The collimated input beam is expanded from the first lens (C) and then collimated by the second lens (F). The magnification of the Galilean beam expander and the required optical track length L are described by:

$$M_{f_2/f_1} = \frac{f_2}{f_1}, \qquad L = f_1 + f_2. \tag{4.6}$$

Thus, to design a relatively compact optical system, the beam expander with a short focal distance f_1 lens and a large focal distance f_2 lens is preferable to meet the magnification without exceeding the system length requirements. The available lenses are the first lens (C) $f_1 = 13mm$ and the second lens (F) $f_2 = 250mm$, resulting $M_{f_2/f_1} \approx 20$ and $L = 263mm$. A circular aperture (D) and a slit aperture (E) are placed between the lenses to spatially filter out the beam. To physically check whether the beam passing through the beam expander is collimated, a practical method is to record the beam profiles at a few centimetres and a longer distance (~5m) away. The beam profiles should be roughly the same. The collimated beam is shown in Figure 4.16.

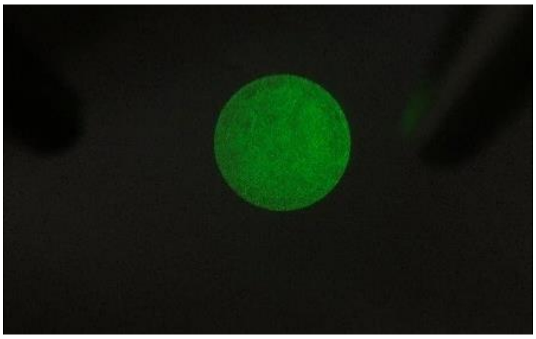

Figure 4.16. The experimental expanded collimated beam from the laser.

4. Linear polariser (G)

Before illuminating the SLM through a beam splitter cube (I), the expanded collimated beam is linearly polarised to align the effective polarisation state of the SLM with the half waveplate. The chosen SLM modulates the incident laser beam by nematic liquid crystals (NLC), which are inherently polarisation sensitive.

5. Fourier lens (J) parameters

The SLM modulates the collimated beam and reflects through the beam splitter (I). The Fourier lens (J) is placed a focal distance away from the SLM, performing the Fourier transform on the modulated beam. The reconstructed image is then shown at its focal distance. This lens (J) has a focal length $f_3 = 150mm$, chosen according to the propagating parameter in CGH generation.

Figure 4.17. Experimental holographic projector system designed for demonstration.

6. Lens (K) parameters

The lens (K) is an objective lens to constitute a 4F system with the CCD camera (L) since the camera is already attached to a lens. The focal length of the lens (K) varies according to the focal distance of the CCD camera. This 4F system magnifies the replay field on the CCD camera.

The constructed holographic projector system designed is shown in Figure 4.17. A test hologram is generated from the 3D space station model shown in Figure 4.18 based on the layer-based method. Sub-holograms H_{zi} generated from each layer are optimised into 256-level phase-only H_{zi}^{Q} using the GS algorithm for 10 iterations. The simulated and optical replay fields are shown in Figure 4.18.

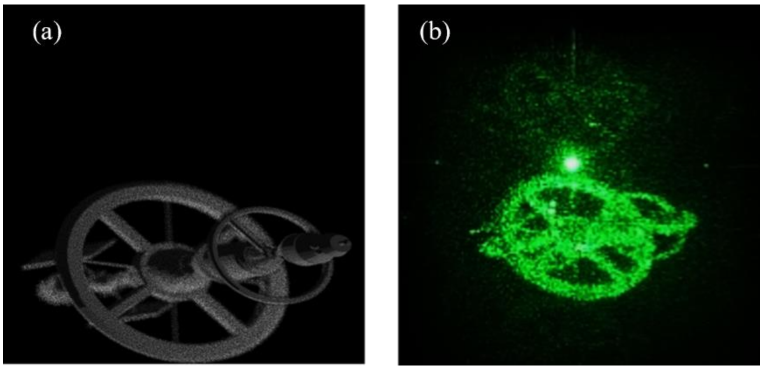

Figure 4.18. The simulated (a) and optical (b) reconstructed images of the 3D space station model.

4.6 Demonstrations

In this section, we demonstrate practical issues of the holographic display system for real-world scene acquisition and display. Holographic displays often show poor image quality due to severe real-world deviations in data acquisition and reconstruction compared to the ideal scenario. Therefore, we demonstrate the holographic reconstruction quality of a real-world scene captured with a low-cost RGB-D camera to identify failure modes. We further perform experimental holographic reconstruction to demonstrate practical artefacts and possible improvements for the reconstruction quality.

4.6.1 Real-World Scene vs Computer-Generated Scene

We compare the holographic reconstruction using a Unity-rendered computer-generated (CG) scene to demonstrate practical issues with a real-world (RW) scene.

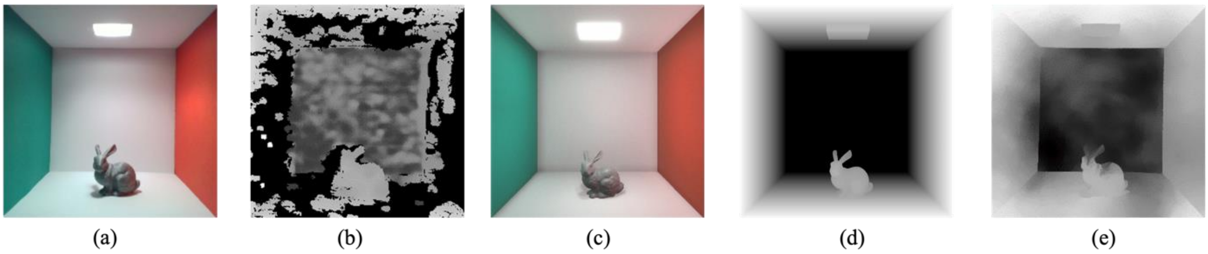

Figure 4.19. RGB and Depth images from RealSense camera (a) (b) and from Unity (c) (d). Filled depth image (e) using the colourisation scheme.

3D data creation and acquisition. A Cornell Box with a 3D printed Stanford bunny placed inside is built as a physical model for the RGB-D sensor acquisition process. The Cornell Box consists of a light source from a laboratory lamp in the centre of the white ceiling, diffusive green and red walls on the left and right sides, and white walls on the back and floor. A Stanford bunny is printed with PLA materials is placed in the box. The same scene is then rendered in Unity with the ray casting technique. The rendering options of the scene are determined by matching the visual properties in Unity to the photograph of the RW scene.

RGB-D images of the RW scene are acquired with an Intel RealSense D435i camera. Figure 4.19(a), (b) shows the raw colour and aligned depth images. The missing values of the depth image acquired from the camera were filled, as shown in Figure 4.19(e), using the colourisation scheme [134]. We directly capture RGB images of the CG scene once rendered, and the depth information is recorded by adding a depth shader to the viewing camera for calculating the depth of each pixel based on the camera depth texture. Captured RGB-D images of the CG scene are shown in Figure 4.19(c), (d).

CGH generation from RGB-D images. The aligned depth and colour images of both scenes were normalised and resampled into 30 layered 2D images with RGB channels based on the trade-off between the high fidelity and the computational load. Occlusion culling was performed by deleting occluded values during the resampling process. We generated CGHs using the layer-based method for each RGB channel from layered images. The generated holograms for each colour channel were stacked to form a superposed 720x720 hologram for each colour channel.

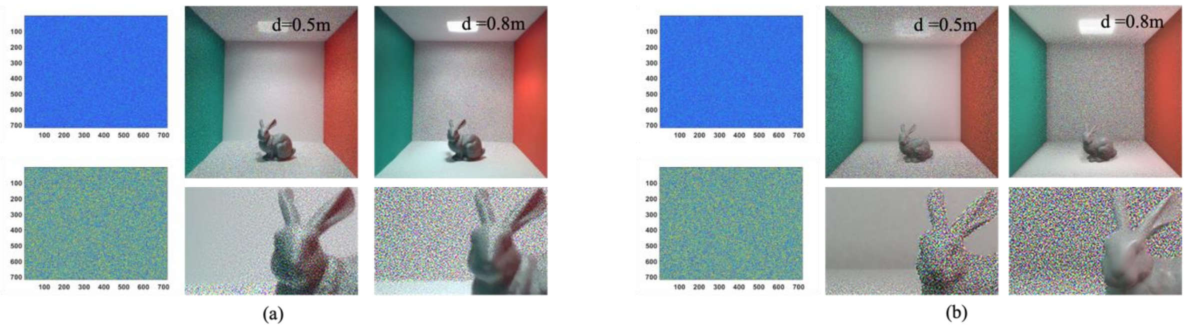

Figure 4.20. Single R channel holograms with amplitude, phase, and numerical reconstructions of the real scene (a) and the CG scene (b) on different focal distances.

Results and Discussion. The amplitude and the phase of holograms for a single colour R channel are shown in Figure 4.20. The full reconstructed fields at two different depth planes for the RW and CG scenes are also shown in Figure 4.20, alongside zoomed-in versions.

The general depth information is conserved, and the real scene can be successfully reconstructed at different depths. Comparing the reconstructed real scene in Figure 4.20(a) with the CG scene in Figure 4.20(b), reconstructed images are visually similar, with reconstruction errors at the boundaries of the bunny. For the same depth plane, the defocus boundaries for the RW scene do not entirely coincide with the boundaries of the object and are generally diffuse instead of sharp. This is due to the less accurate and noisy depth image at both the acquisition completion stages, especially at occlusion boundaries, which contain discontinuities in depth and much of the information of object structures. Both measurement errors in data acquisition and depth estimation errors in depth completion. Errors are also observed near the complex features of the bunny.

4.6.2 Implementation Issues on Experimental Reconstruction

We collect colour and depth images in Unity and generate CGH to demonstrate practical issues in experimental holographic reconstruction. The collected colour and depth images of car models are shown in Figure 4.21. The depth image is normalised into 255 depth levels. The simulated and optical replay fields are reconstructed in the Fresnel region, shown in Figure 4.22, with highlights on the in-focus parts of reconstructions.

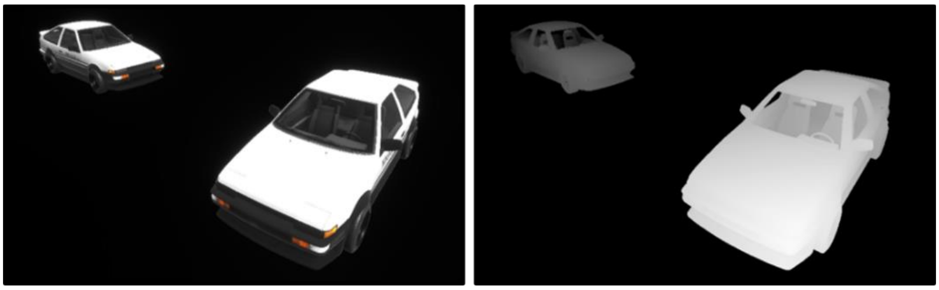

Figure 4.21. RGB-D images of car models from the Unity game engine.

Zero-order. In the optical reconstruction experiment, the illuminated light cannot be fully diffracted due to the limited diffraction efficiency of SLMs, resulting in a zero-order region at the origin of optically reconstructed images. The zero-order issue can be resolved by inserting a polariser after the SLM to filter out undiffracted light or using apertures for spatial filtering.

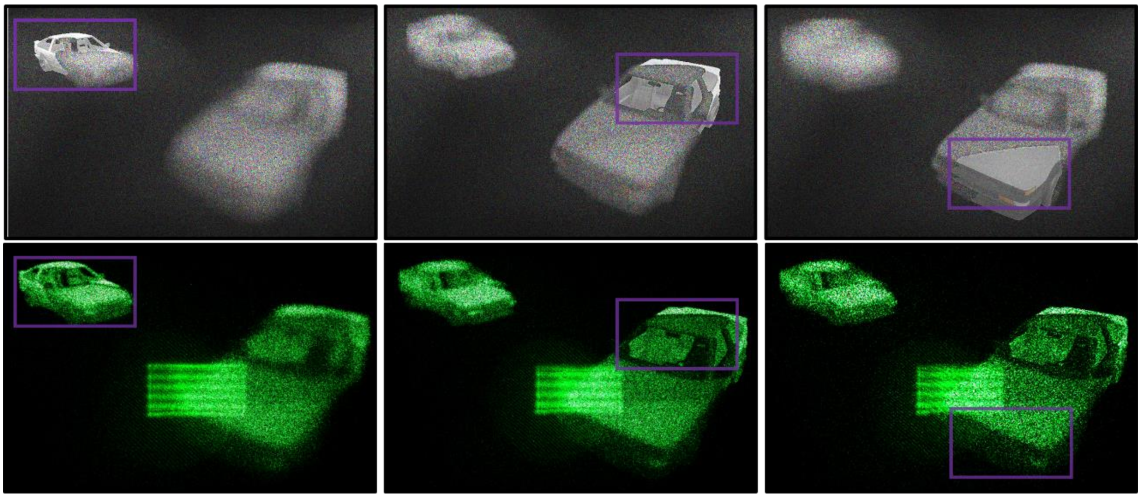

Figure 4.22. Simulated (first row) and optical (second row) replay fields of the car models.

Image quality degradation. Speckle-like artefacts can be observed in the simulated reconstruction. The GS algorithm immediately stagnates at a local minimum, having limited ability to optimise the image quality of reconstructions. Imperfections due to the laser profile, SLM flatness and modulation properties, and aberration in other optical elements reduce the quality of optically reconstructed replay fields. Speckle noise is presented in reconstructed images when the observer plane is a diffuse surface illuminated by coherent, collimated lasers. Time multiplexing methods such as OSPR can effectively suppress the speckle noise. However, generating multiple CGHs for 3D scenarios with a high refresh rate of the spatial light modulator to display CGHs in real time is challenging. Alternatively, active feedback optical aberration correction techniques can be introduced to suppress speckle noise and other optical aberrations.

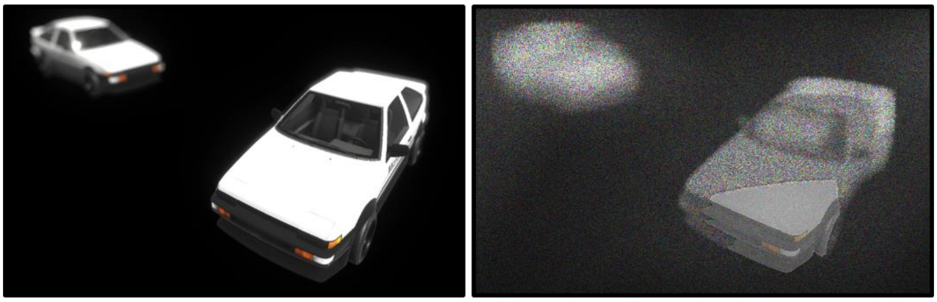

Figure 4.23. The visual difference in defocus blur under Unity and CGH reconstructions.

Image Defocus. Figure 4.23 demonstrates the visual difference in defocus blur under typical Unity and holographic reconstructions. Intuitively, the Unity construction has a significantly smooth spatial variation at the defocus areas. The image defocus generated by CGHs relies on coherent light propagation, having different defocus behaviour than typical incoherent light propagation rendered in Unity. Further discussion on the difference in defocus behaviours will be covered in Chapter 6.

Gamma correction. Every display has an inherent property known as the *gamma value* γ, which describes the transfer function between input and output pixel energy. As a result, gamma correction is often applied during an image pre-processing stage to display systems so that, visually, the display exhibits compatible gamma characteristics. If such gamma correction is not applied, there will be a contrast mismatch between displays displaying the same image. The gamma correction curve is simply the inverse of the gamma response curve. A CGH of a grey-scale ramp is typically calculated to determine the gamma correction function for holographic displays [66]. The gamma response curve is obtained by curve fitting the intensity values of the ramp area of the captured reconstructed image.

Tone Mapping. The layer-based method decomposes an image into piecewise smooth base layers, resulting in intensity variations between layers. Especially for layers with high intensity with less signal area, the reconstructed image intensity seems brighter than other layers. Furthermore, the removal of amplitude information in the hologram plane will perturb the energy distribution of the sliced target images in the replay field, increasing the dynamic range of the reconstruction. Certain regions of the reconstructed image with low pixel intensity are invisible from those with high pixel intensity. Through tone mapping, the intensities of layered images can be better distributed, allowing areas of lower local contrast to gain a higher contrast.

4.7 Conclusion

This chapter presents the design and operation of a primitive holographic display system consisting of a 3D data acquisition module, a data processing module, a CGH calculation and encoding module, and an optical holographic display module. The 3D data acquisition is performed from two depth cameras, and their qualities are compared for CGH calculation. For demonstration purposes, an experimental holographic display is implemented for displaying calculated CGHs. This chapter has achieved an end-to-end chain from the 3D data acquisition to the 3D data reconstruction through holographic displays. However, the system prototype has several practical issues, from data acquisition to optical reconstruction. Moreover, the algorithm used to generate CGHs is the standard layer-based method with the GS algorithm based on CPU calculation for a 256-level phase-only SLM. The following chapters will address more advanced algorithms exploiting high parallel GPUs for computational acceleration to improve reconstructed image quality.

Chapter 5 CGH Optimisation with Image Quality Metrics

This chapter is primarily based on "Perceptually motivated loss functions for computer generated holographic displays". Yang, F., Kadis, A., Mouthaan, R. et al. Sci Rep 12, 7709 (2022).

5.1 Introduction

Understanding and improving the perceived quality of reconstructed images is key to developing hologram generation algorithms. The last chapter demonstrated a holographic display system using the layer-based method with the GS algorithm to optimise CGHs. However, severe image artefacts occurred in the reconstruction process. The GS algorithm relies on iterative applying constraints in both the spatial and Fourier domains with Fourier transforms and is prone to local minimum solutions due to its strong enforcement of constraints. While the double-phase method offers a straightforward and efficient approach for encoding complex amplitude CGHs into phase-only CGHs, the use of the down-sampling operation introduces significant noise encoded in the reconstructed image, thereby limiting its effectiveness.

Recent advancement in complex-amplitude differentiation has enabled the gradient descent method to be applied to phase-only CGH optimisation [29,43,44,138–142]. Unlike the GS algorithm or the double-phase method, the gradient descent method allows for explicit optimisation of CGHs based on an objective function, which provides better control and flexibility over the optimisation process. This is particularly advantageous for improving the quality of reconstructed images in 3D CGHs, where reconstruction is more challenging due to the more significant number of constraints and the requirement for high-resolution encoding.

By optimising the objective function directly, the gradient descent method can avoid issues with local minimum solutions and convergence, resulting in a more robust and globally optimal solution.

This chapter focuses on improving the perceived quality of 2D CGHs, introducing the gradient descent method to phase-only CGH optimisation, an iterative first-order optimisation algorithm typically used in machine learning and deep learning. We validate the proposed method in simulation and use an experimental holographic display prototype to demonstrate the improved optical reconstruction quality. Furthermore, using the gradient descent method, we introduce different image quality metrics (IQMs) as losses for CGH optimisation. We present a comprehensive analysis employing extensive objective and subjective assessment of experimentally reconstructed images to reveal the relative performance of IQM losses for hologram optimisation. This extensive analysis provides guidance for finding a specific perceptually-motivated loss function for CGH generation.

5.2 Gradient Descent Method for CGH Optimisation

The gradient descent method is an iterative optimisation algorithm typically applied to deep learning models to optimise the model weights so that the loss function is as small as possible. For deep learning model optimisation, the loss function can be regarded as a parametric function with the model weights as parameters. The optimal weights of the model can be obtained by calculating the gradient of the loss function with respect to the weights and then optimising along the direction of the gradient for a given dataset. The standard loss function for image-related applications is the mean squared error, quantifying the per-pixel error between the reconstructed and target images.

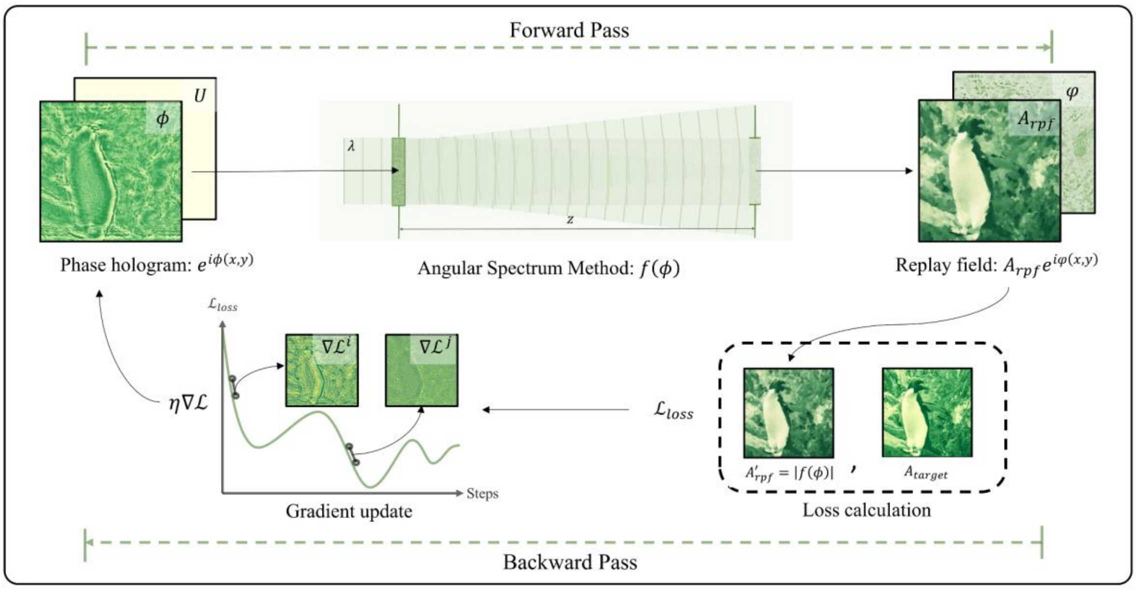

Figure 5.1. CGH optimisation model based on the gradient descent method.

However, the propagation models are given for CGH optimisation based on the gradient descent method, and we need to obtain the optimal input (phase CGHs). In this case, for a selected propagation model, the gradient of the loss function is calculated with respect to the input, and the input is optimised in the gradient direction. Thus, there are three essential components in the gradient descent method for CGH optimisation. First, we need to establish the propagation model to determine the direction of the gradient calculation, and second, we need to calculate the gradient of the loss function. Finally, we use the gradient to optimise the initial phase hologram.

We can treat the gradient descent method for CGH optimisation as a forward-backward optimisation process to minimise a given loss function. In the forward pass, the selected wave propagation model propagates a phase hologram to the replay plane to produce a reconstructed image, which is used to calculate the loss by comparing it to the target image. In the backward pass, the model traverses backwards from the output, collecting the derivatives of the loss function with respect to the phase hologram and updating the hologram to minimise the loss. The model iteratively goes through the forward pass and backward pass to obtain the optimised phase hologram. This process is illustrated in Figure 5.1.

5.2.1 Forward Pass

In the forward pass, we select the angular spectrum method [5,107] with a planar illuminating wave as the diffraction propagation model as introduced in section 2.1.3:

$$f(\phi) = \mathcal{F}^{-1}\left\{\mathcal{F}\{e^{i\phi(x,y)}\} \times exp\left[j2\pi z\sqrt{\frac{1}{\lambda^2} - f_x^2 - f_y^2}\right]\right\}. \quad 5.1$$

Here, $\phi(x, y)$ is the phase hologram that has been quantised so that it can be displayed on a binary or 8-bit SLM, λ is the wavelength, f_x , f_y are spatial frequencies, and z is the propagating distance between the hologram plane and the replay field plane. $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the Fourier transform and the inverse Fourier transform, respectively. The resulting field $f(\phi)$ is a complex replay field whose amplitude is related to the reconstructed image intensity by $I(\mu, \nu) = |f(\phi)|^2$.

To evaluate the perceived image quality, the amplitude of the replay field A_{rpf} is compared with the target amplitude A_{target} using a loss function $\mathcal{L}$. Though intensity-based objective functions can also be utilised for image quality evaluation, amplitude-based objective functions have been found to yield better algorithmic performance and are preferable in hologram optimisation [143,144]. Therefore, the CGH optimisation algorithm aims to find the optimal quantised phase hologram $\hat{\phi}$ that minimises the loss function $\mathcal{L}$ describing the visual quality, calculated from the reconstructed image amplitude $|f(\phi)|$ and the intended target image amplitude A_{target}:

$$\hat{\phi} = \underset{\phi}{argmin}\, \mathcal{L}\big(s \cdot |f(\phi)|, A_{target}\big), \quad 5.2$$

where s is a scaling factor for normalisation. The MSE for a m by n sampling points is commonly used as the loss function, computed by averaging the squared amplitude differences of reconstructed and target image pixels:

$$\mathcal{L}_{MSE} = \frac{1}{mn}\sum_{m,n}\big[|f(\phi)| - A_{target}\big]^2. \quad 5.3$$

5.2.2 Backward Pass

Having selected the model, in the backward pass, we calculate the gradient $\partial\mathcal{L}/\partial\phi^{k-1}$ of the loss function with respect to the current estimate of the phase hologram ϕ^{k-1} to update the next estimate phase ϕ^{k}.

Since the propagation model is composed of multiple complex amplitude functions, we calculate the gradient by the chain rule, which involves the calculation of complex derivatives:

$$\frac{\partial\mathcal{L}}{\partial\phi^{k-1}} = \frac{\partial\mathcal{L}}{\partial A'_{rpf}} \cdot \frac{\partial A'_{rpf}}{\partial f} \cdot \frac{\partial f}{\partial\phi^{k-1}}, f\colon \mathbb{C} \to \mathbb{C}. \tag{5.4}$$

Therefore, in the forward pass, we apply equations 5.1 and *5.3* to calculate the loss function, and in the backward pass, we start by calculating the derivatives of the loss function with respect to the last result and traverse backwards to recursively calculate the derivatives with respect to the one before the last result.

However, in complex analysis, the holomorphic requirement for functions to be complex-differentiable is very strict. Wirtinger calculus relaxes this requirement and allows approximate complex derivatives of nonholomorphic functions to be more easily calculated using a conjugate coordinate system [140,145–147]. Recently, Wirtinger calculus has been implemented in automatic differentiation packages in machine learning libraries such as TensorFlow and PyTorch. These automatic differentiation packages keep a record of all the data and operations done in the forward pass in a direct acyclic graph and automatically compute gradients using the chain rule.

5.2.3 Optimiser Based on the Gradient

The reason for calculating the gradient of the loss function with respect to the current estimate of the phase hologram is to use its gradient to update the next estimate. The gradient descent method updates the next estimate of the phase hologram ϕ^{k} in the opposite direction of the gradient of the loss function $\nabla\mathcal{L}(\phi^{(k-1)})$ with respect to $\phi^{(k-1)}$. In other words, if we follow the direction of the gradient of the loss function downhill, the value of the loss function will decrease until we reach a valley. For a learning rate η, the next estimated phase hologram $\phi^{(k)}$ is given by:

$$\phi^{(k)} = \phi^{(k-1)} - \eta\nabla\mathcal{L}\left(s \cdot \left|f\left(\phi^{(k-1)}\right)\right|, A_{target}\right). \quad 5.5$$

The learning rate η is a tuning hyperparameter that determines the step size at each iteration. The optimisation speed is relatively slow for a small learning rate (10^3). However, as the step size is small, the optimisation is relatively smooth and stable to reach a (local) minimum. If the learning rate is too big, the optimisation may converge too quickly to a suboptimal result or diverge completely. Figure 5.2 demonstrates the optimisation behaviour of the gradient descent method for different learning rates. The changing trend of the overall parameters is to make the loss function keep getting smaller. A smaller learning rate can guarantee that the method gradually approaches the minimum, and a bigger learning rate can result in oscillating before converging.

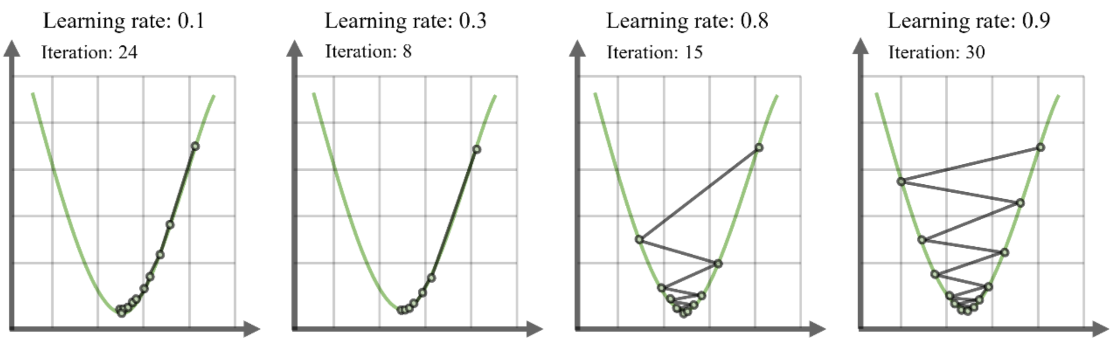

Figure 5.2. Optimisation behaviour of the gradient descent method for different learning rates.

Several update strategies, such as Adagrad [148] and Adaptive Moment Estimation (Adam) [149], propose adaptive learning rates to improve accuracy and convergence speed.

5.2.4 Hologram Generation based on DIV2K Dataset

Before we dive into the implementation details of the gradient method for CGH optimisation, we first introduce the dataset as target images to generate computer-generated holograms. The DIVerse 2K resolution image dataset (DIV2K) was introduced in 2017 for benchmarking computer vision topics, including single-image image superresolution [150,151]. It contains 1000 RGB low-resolution images with corresponding high resolution and diverse image contents. For hologram generation, we select 100 images from the DIV2K dataset. The images are preprocessed using the same tools such that they are all monochrome and have 1,920 × 1080 resolution, as shown in Figure 5.3.

Figure 5.3. Monochrome target image amplitudes from the DIV2K dataset

Based on 100 high-resolution images in the DIV2K dataset, we first compute the phase-only holograms by the gradient descent method to validate the method in numerical simulation. For the convenience of parameter settings, we use argparse in python to process CGH and hyperparameters. CGH parameters include wavelength at $532nm$ in green, propagation distance at $15cm$, SLM pixel pitch at $6.4\mu m$, and the total number of iterations at 300. Hyperparameters of the Adam optimiser include a 0.05 learning rate and default exponential decay rates of $\beta 1$ = 0.9 and $\beta 2$ = 0.999. Images in the DIV2K dataset are stored as a Dataset class in PyTorch and loaded in the DataLoader module. Since holography reconstructs light without gamma correction, target images are linearised from sRGB to the linear intensity and sequentially converted to image amplitude. The sRGB to linear conversion relationship is given by [29,43]:

$$\begin{aligned} I_{lin} &= \frac{1}{12.92} I_{sRGB} && 0 \leq I_{sRGB} \leq 0.04045 \\ I_{lin} &= \left(\frac{I_{sRGB} + 0.055}{1.055}\right)^{2.4} . && 0.04045 < I_{sRGB} \leq 1 \end{aligned} \tag{5.6}$$

The input sRGB intensity target images are converted to linear space intensity images first and then to linear space amplitudes consecutively:

$$A_{target,lin} = \sqrt{I_{target,lin}} \approx I_{target,sRGB}{}^{1.1}, \tag{5.7}$$

The DataLoader module then indexes and loads images to establish an iterable dataset. To ease computational load, we can precalculate and store the quadratic phase term in the forward and backward propagation so that it will not be calculated repeatedly for different images. For each image, we forward propagate with the same constant initial phase to generate a reconstructed image amplitude, compare it to the target amplitude, and then backwards propagate to obtain the gradient for the loss function, which is used by the Adam optimiser to find the optimal phase hologram iteratively. During each iteration, we normalise the amplitude of the replay field.

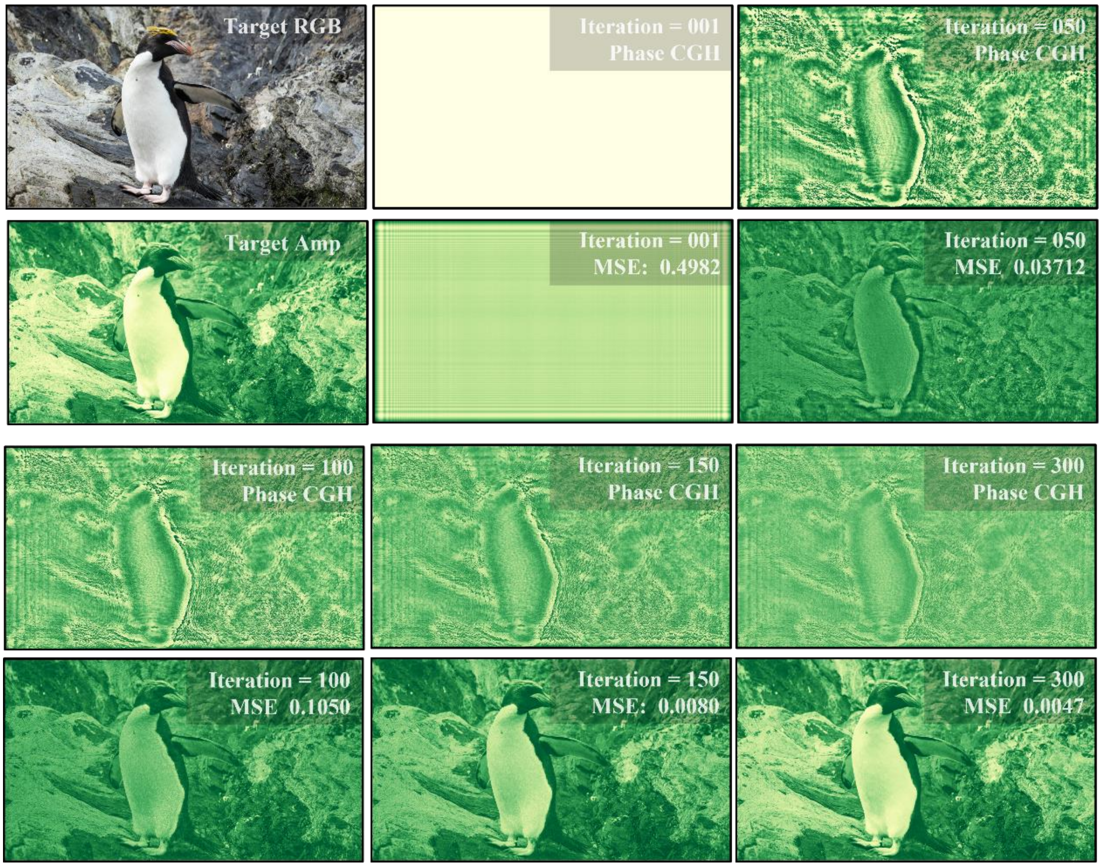

Figure 5.4. The CGH optimisation process over iterations using the gradient descent method.

The CGH generation is done on a machine with an Intel i7-8700 CPU @ 3.20GHz and a GeForce GTX 1080 GPU. PyTorch 1.9.0 and CUDA 10.2 are used to implement complex-amplitude gradient descent optimisation on the GPU. Computation takes 30 GPU seconds to generate the holograms. Figure 5.4 shows the reconstructed images for a target image during the optimisation process using the proposed method. The reconstructed image quality gradually becomes better with more iterations.

5.3 Quantitative Comparison to Other Algorithms

We further conduct a quantitative comparison to evaluate the performance of 2D CGH algorithms. The algorithms tested in the experiment include Gerchberg-Saxton (GS) algorithm, the double phase hologram (DPH) method, the One-Step-Phase-Retrieval algorithm (OSPR), and the gradient descent (GD) method. The experiment aims to optimise phase patterns using different algorithms and evaluate their performance using the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) metrics.

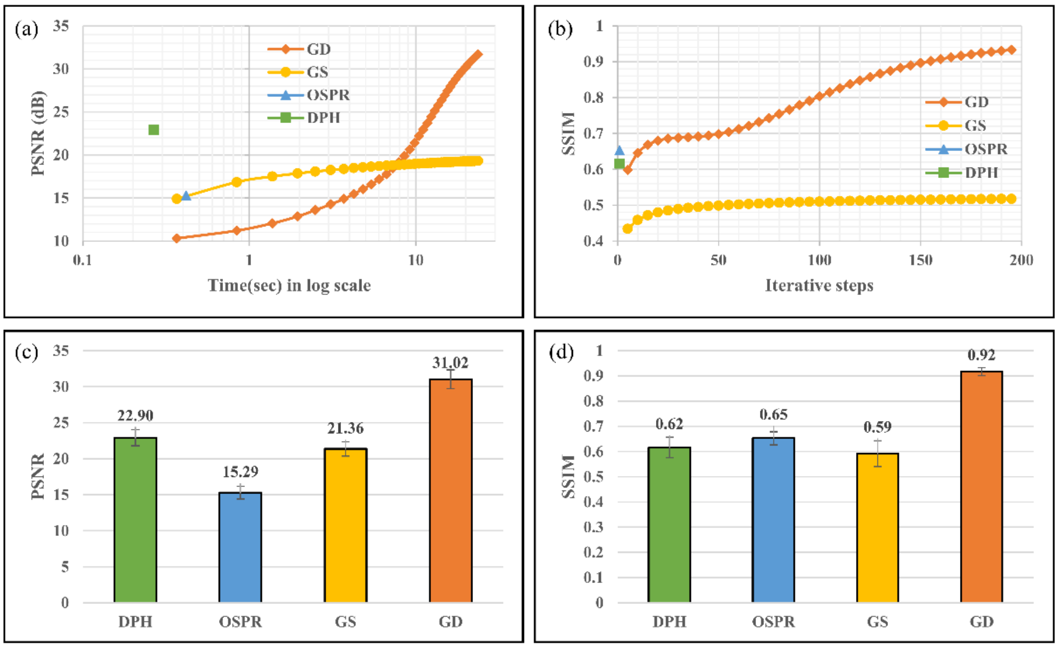

Figure 5.5. CGH algorithm performance comparison. (a) PSNR versus computational speed (sec) shown in log scale for DPH, OSPR, GS and GD methods. (b) SSIM versus optimisation steps for all methods. (c) and (d) evaluate the PSNR and SSIM for all simulated reconstructed images, respectively.

The experiment uses a phase-only SLM with a resolution of 1,920 × 1,080 pixels and a pixel size of 6.4 × 6.4 μm. The target image plane is set at 18cm away from the SLM, and simulated light sources are used with wavelengths of 638, 532, and 450 nm for the red, green, and blue colour channels, respectively. Iterative algorithms, including the Gerchberg-Saxton algorithm and gradient descent method, run 200 iterations for each colour channel until convergence. The OSPR algorithm uses 20 phase-only holograms for reconstructed image optimisation. The angular spectrum method is used as the wave field propagation operator in all cases for consistency in the comparison of the performance of the different CGH algorithms. The dataset of target images used in the experiment consisted of 50 test images randomly selected from the DIV2K dataset, each with a resolution of 1,920 × 1,080 pixels. To ensure a fair comparison, the resulting amplitude of all methods is scaled such that the mean amplitude of the scaled reconstructed image matches the mean amplitude of the target image. Subsequently, the scaled amplitude is cropped to the centre at a resolution of 1,680 × 960 to reduce the impact of ringing artefacts near the image edges. All algorithms are executed on an NVIDIA RTX 2080 graphics processing unit with Intel i7-9700K @ 3.60GHz using Pytorch.

Figure 5.5(a) displays the computational time required by each algorithm to generate phase-only holograms on a logarithmic scale. Figure 5.5(b) illustrates the performance of each algorithm based on the SSIM metric. Non-iterative algorithms such as DPH and OSPR are represented by single points on the charts and have the potential to achieve real-time frame rates. Iterative algorithms, including the GS and GD algorithms, can converge in approximately 20 seconds while trading off between the number of iterations and the resulting quality. However, the GS algorithm is prone to stagnation after around 20 iterations. Figure 5.5(c) and (d) present a quantitative comparison of selected algorithms, evaluating the PSNR and SSIM for all reconstructed images in the simulation. The PSNR and SSIM values for DPH, OSPR, GS, and GD are found to be 22.9 dB (0.62), 15.29 dB (0.65), 21.36 dB (0.59), and 31.02 dB (0.92), respectively. The reported values are obtained by taking the mean values for 50 reconstructed images for each algorithm. Error bars are also calculated to represent the standard deviations between scenes. These error bars indicate the variability in the results and suggest that the performance of each method is consistent across the scenes.

Figure 5.6 presents simulated results for DPAC, OSPR, GS, and GD, qualitatively comparing nine images for each experiment selected from 50 reconstructed images. For each image, two zoomed-in details are indicated with rectangular boxes and are presented side-by-side with the image. As can be seen, the reconstructions calculated using the gradient descent method outperform others significantly, providing much additional detail in the image and lowered speckle noise. The comparison demonstrates that the gradient descent method achieved the best results among the tested algorithms. Overall, this comparison provides valuable insights into the performance of different CGH algorithms and their potential applications in improving image quality for holographic displays.

Figure 5.6. For qualitative comparison, the simulated results are selected from 50 reconstructed DPH, OSPR, GS, and GD images. Two zoomed-in details are indicated for each image, with rectangular boxes presented side-by-side with the image.

5.4 IQM as Loss Functions

5.4.1 Motivation

As we see in previous sections, the gradient descent method predefines a loss function and uses its gradient to update the hologram phase at each iteration. The specific loss function selected is essential since its gradient drives the hologram phase to the optimal state. A standard choice of the loss function is the mean squared error due to its simplicity of use and clear physical meaning. It also remains the standard criterion for evaluating reconstructed image quality in CGH generation.

Though MSE objectively quantifies the per-pixel error in the reconstructed image, it is widely criticised for its poor correlation with perceptual quality [152–155]. Two distorted images may have completely different types of errors while having the same MSE, and some of the errors may be considerably more visible than others. An illustrative example is shown in Figure 5.7, where the gradient-based CGH optimisation reconstructs a target image with FSIM (b) and MS-SSIM (c) as the loss function. The MSE and the SSIM are given for both reconstructed images. Note that both reconstructed images yield nearly identical MSE but are highly different in perceived quality.

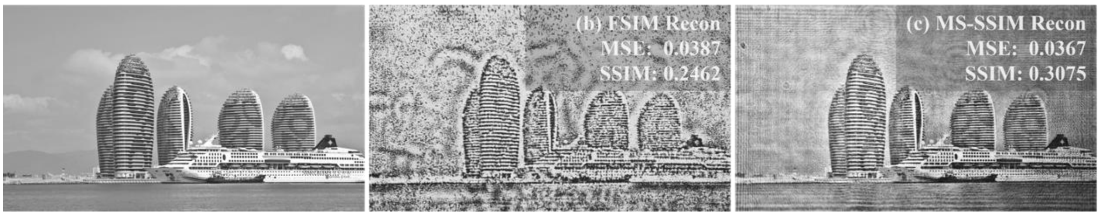

Figure 5.7. The gradient-based CGH optimisation reconstructs a target image with FSIM (b) and MS-SSIM (c) as loss functions.

Moreover, CGH optimisation involves numerically simulating an interference pattern whose pixels carry the target image information and are highly spatial correlated. When employed as a loss function to optimise the CGH algorithm, the MSE could potentially ignore the spatial correlation and other image features between the target and the reconstructed images. It is thus worthwhile to carefully examine the performance of MSE as the loss function, especially its suitability for the design of CGH optimisation.

A promising but less exploited approach is using image quality metrics (IQMs) in the phase-only CGH optimisation process. IQMs play a vital role in developing and optimising image processing and restoration algorithms. In digital holography, the traditional role of IQMs is to monitor the optimisation process dynamically and to evaluate the perceptual quality of obtained images [156–159].

The human visual perception system can extract features from images and identify the difference between the target and the distorted images. Modern IQMs model assesses visual quality based on a priori knowledge regarding the human visual system or uses learned models trained with large datasets, replicating human behaviour to perform tasks. They use image features in appropriate perceptual spaces [155,160] for image quality evaluation but have not yet been fully exploited in the CGH optimisation process.

Here, we focus on using IQMs as an alternative to the ubiquitous MSE for the training loss, with the intention of using the gradient of these perceptual metrics to strive for a better CGH optimisation algorithm. The use of perceptual-motivated loss functions has recently gained attention in foveated CGH [161,162], focusing specifically on speckle suppression in the foveal region and peripheral perception. Other non-holographic image restoration applications have also explored perceptual losses, though it is observed that no single loss function outperforms all others across different applications [163–165].

5.4.2 Selected IQMs for CGH Optimisation

Generally, IQMs can be classified into full-reference, reduced-reference, and no-reference methods according to the availability of the original reference image. Since the target image is available in the CGH optimisation, we only consider full-reference methods as loss functions.

However, the direct use of IQMs as loss functions is complex and depends on many unpredictable parameters, and different IQM implementations can yield significantly different results, further complicating the interpretation of our experiment. We, therefore, consider ten differentiable full-reference IQMs from existing libraries IQA [164] and PIQ [166], benchmarked on common databases, which we believe include a wide range of state-of-art full-reference IQMs. We also include MAE and MSE as standards for comparison. Therefore, this IQM collection includes three error visibility methods: MSE, MAE and NLPD [160], six structural similarity methods: SSIM [167], MS-SSIM [168], FSIM [169], MS-GMSD [170], VSI [171], HaarPSI [172], one information-theoretical method: VIF [173], and two learning-based methods: LPIPS [152] and DISTS [174].

Error visibility methods calculate the image error on a pixel-by-pixel basis. The NLPD method first subtracts the local luminance using the Laplacian pyramid construction and normalises the image contrast by removing an estimate of local amplitude at the second stage.

Structural similarity methods consider the perceived variation, including luminance, contrast, and structure, often using correlation measures to assess image distortion. The MS-SSIM method downsamples images in different scales by low-pass filters, while the Feature Similarity Index Metric (FSIM) uses the phase congruency and the image gradient magnitude as complementary HVS features. Phase congruency assumes that the points where Fourier components are maximal in phase are the most informative. The gradient magnitude can be computed by convolving images with a linear filter to extract discernable structural and contrast differences in the gradient magnitude domain. The MS-GMSD method extends the Gradient Magnitude Similarity Deviation, or GMSD, to a multiscale version, measuring standard deviation based on a pixel-wise gradient magnitude similarity map without using additional features to yield accurate quality prediction. The Visual Saliency-Induced Index (VSI) utilises the salient visual map to detect the visually discernable region and the gradient magnitude to compensate for the contrast sensitivity in the measurement. Similar to FSIM, the Haar Wavelet-Based Perceptual Similarity Index (HaarPSI) constructs Haar wavelet filters to compute feature maps and reduce the computational complexity induced by the calculation of phase congruency maps.

Information-theoretic methods measure some approximation of mutual information between the perceived reference and distorted images and quantify the amount of information loss in the distorted images. In the Visual Information Fidelity (VIF) metric, the image source is statistically modelled using a Gaussian scale mixture and the image distortion is modelled using signal attenuation and additive noise in the wavelet domain. The additive white Gaussian noise models the visual noise from the human visual system in the wavelet domain. Image fidelity is quantified as the mutual information from the target and the reference images from these models.

Learning-based methods propose neural networks trained with numerous pictures to learn a metric and assess image quality. The Learned Perceptual Image Patch Similarity, or LIPIPS, utilises the existing VGG network and evaluates the Euclidean distance between extracted features of two images. The network is trained from the proposed BAPPS image patch dataset, which perceptually calibrates the network weights in the feature space. The Deep Image Structure and Texture Similarity, DISTS, also modifies the VGG network and combines texture similarity measurements based on a parametric texture model with feature maps extracted from the VGG in an SSIM-like structure. Several IQMs that evaluate visual quality based on colour information are not included in the test since our holograms are reproduced in monochrome.

Table 5.1 summarises the library of the IQMs considered, as well as the underlying principle. The IQM is reformulated where necessary so that a lower score indicates higher predicted quality. For example, if the selected IQM is $SSIM$, then $\mathcal{L}$ is rewritten as $\mathcal{L}_{SSIM} = 1 - SSIM$.

Table 5.1. The utilised underlying principle of IQM losses for CGH optimisation.

IQM losses	Library	Underlying principle
MAE	Pytorch	Pixel-based absolute error with average pooling
MSE	Pytorch	Pixel-based squared error with average pooling
NLPD	IQA	Root MSE in the weighted Laplacian pyramid decomposition domain
SSIM	IQA	A weighted combination of measures: luminance, contrast, and structure
MS-SSIM	IQA	The Multi-Scale representation of the SSIM
FSIM	PIQ	A weighted combination of the phase congruency feature and the gradient magnitude feature
MS-GMSD	PIQ	The Multi-Scale representation of GMSD, measuring standard deviation based on pixel-wise gradient magnitude similarity map
VSI	PIQ	Similarities in the gradient magnitude and the visual saliency
HaarPSI	PIQ	local similarities and the relative importance of image areas based on Haar wavelet
VIF	PIQ	Model the image source using Gaussian scale mixtures on wavelet coefficients and quantify mutual information
LPIPS	IQA	Evaluate the Euclidean distance between image patches based on feature maps.
DISTS	IQA	Combination of SSIM-like structure and texture similarity measurements based on the VGG network

Having selected the IQMs as losses for hologram generation, we generate CGHs for each IQM for 100 high-resolution images in the DIV2K dataset, and we, therefore, generate a hologram dataset with a total of 1200 holograms. In each case, we forward propagate, compare to the target, and then backwards propagate to obtain the gradient for the IQM loss, which the Adam optimiser uses to find the optimal phase hologram iteratively. In all cases, we use the Adam optimiser with a 0.05 stepsize and default exponential decay rates of $\beta 1 = 0.9$ and $\beta 2 = 0.999$. The total number of iterations is empirically set to 1000 with the initial 15 iterations using MSE as the loss function. We apply this basic preprocessing step since initial predictions can significantly impact the performance of some IQMs. This step is necessary to yield acceptable optimisation results and reduce the training time for learning-based IQMs.

The CGH generation is done on the same machine in section 5.2.4. Computation takes roughly 190 GPU hours to generate the 1200 holograms to assess all 12 IQMs. Training details and computational time for each IQM loss are included below in Table 5.2 and Figure 5.8.

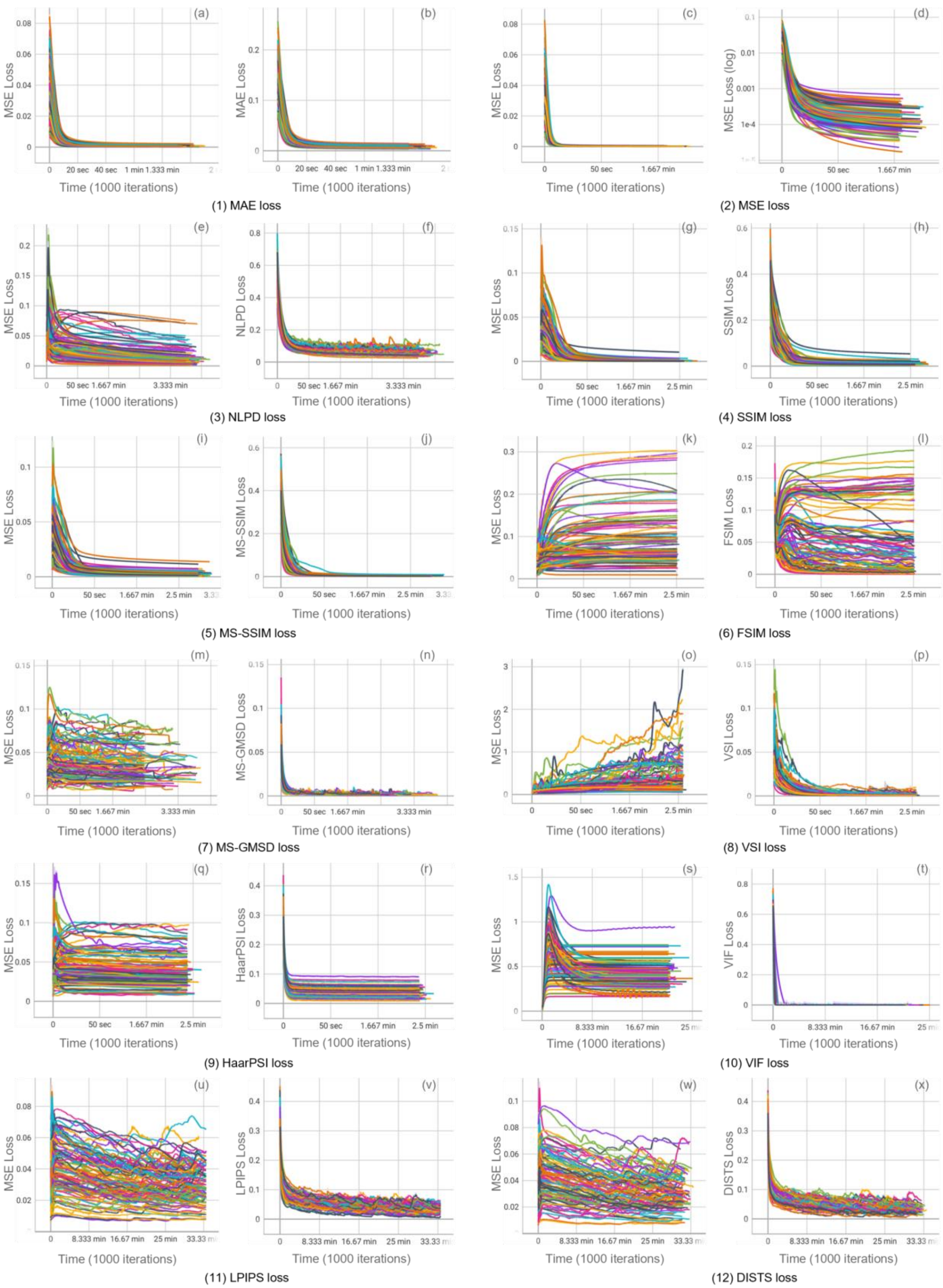

Figure 5.8. The computational time of each image and the training details of IQM losses for CGH optimisation. We plot all runs of images for each IQM loss function, showing how the MSE loss and its own metric loss change with each iteration.

Table 5.2. The computational time for each IQM loss.

	MAE	MSE	NLPD	SSIM	MS-SSIM	FSIM	MS-GMSD	VSI	HaarPSI	VIF	LPIPS	DISTS
Time per image(min.)	1.5	1.78	3.72	2.50	2.88	2.52	2.48	2.58	2.40	21.30	33.33	33.33
Total time (hr.)	2.50	2.97	6.20	4.17	4.80	4.20	4.13	4.30	4.00	35.50	55.56	55.56

5.5 Optical Reconstruction Setup

In order to verify our image quality from experiments, we develop a new physical optical display prototype to acquire an optical reconstruction dataset of IQM optimisation phase holograms using a camera. To enable quantitative comparison, we introduce the Canon DLSR camera into the holographic display system using python to capture reconstructed images. We calibrate captured images to target images for quantitative comparison. The introduced camera to the holographic system can provide adaptive feedback to the holographic display system. With precise control over the phase of the incident wavefront, holography with cameras can further provide aberration correction capabilities to compensate for the errors introduced by the imperfect optics or other deviations from the ideal wave propagation in an adaptive feedback loop fashion.

Compared with the alternative option that is directly displaying the experimental reconstructed images on a screen for human perceptual judgments, this experimental choice of using a camera first to record and then display the reconstructed images for subjective experiment may suffer from distortions produced during image acquisition and the replay process, which could adversely affect subjective evaluations on the reconstructed image quality. However, since all IQM methods go through the same experimental procedure and suffer from the same distortions, this experimental choice should have the same effect for all methods and have little impact on IQM method comparison. As referenced in [157,175,176], establishing the subjective evaluation of experimental CGHs is challenging. First, there is no widely accepted testing methodology for objectively evaluating the CGH reconstructed image quality. A common practice for subjective evaluation is to numerically reconstruct the CGH and display the reconstruction on high-end 2D monitors. Second, there is no widely accepted configuration of high-end holographic displays for subjective pair comparison benchmarking. Most holographic displays are operated under customized laboratory conditions with no standard procedure for calibrating, characterising, and testing holographic displays.

Moreover, holographic displays suffer from practical issues, including limited FoV, overall size, eyebox, laser speckle, eye safety issues with laser illumination, and optical alignment for the same testing condition per test subject. Those practical issues add another layer of complexity to holographic data benchmarking. Finally, objective and subjective assessments of CGH data should be taken under the same visual condition for a fair comparison. Objective quality metrics cannot evaluate a direct holographic projection for subjective evaluation without recording.

5.5.1 Holographic Display System Setup with Camera

The proposed holographic projection system is shown in Figure 5.9. Our system uses an 8-bit phase-only SLM (FSLM-2K55-P) with a pixel pitch of 6.4 µm and a resolution of 1920 × 1080. The SLM is made by the Xi'an Institute of Optics and Precision Mechanics company and is factory pre-calibrated in reflection mode. The first arm comprises a 532nm laser source (Thorlabs CPS532), a half waveplate, a 4F lens system, and a polariser. The 4F lens system comprises two lenses (lenses 1 and 2) with focal lengths of 13mm and 75mm, respectively, used to expand the beam. The expanded beam is then linearly polarised and illuminates the SLM. The second arm comprises a beam splitter and a 4F lens system with a spatial filter to reduce the DC component of the replay field and other unwanted higher diffraction orders. The focal lengths of these lenses (lenses 3 and 4) are 30 mm and 50 mm. The second arm is adjusted to relay the reconstructed images onto the camera sensor. A neutral density filter is inserted in the second arm to reduce the replay field intensity. Reconstructed images are captured using a Canon EOS 6D camera without a camera lens attached. The camera output resolution is 5472 by 3648 with a gain setting of ISO 125 to minimise amplifier noise. All reconstructed images are averaged across three captured images in sRGB, the camera's native colour space. We further applied the image linearisation process that converts the captured image from sRGB intensity into monochromatic linear space amplitude mentioned in section 5.2.4.

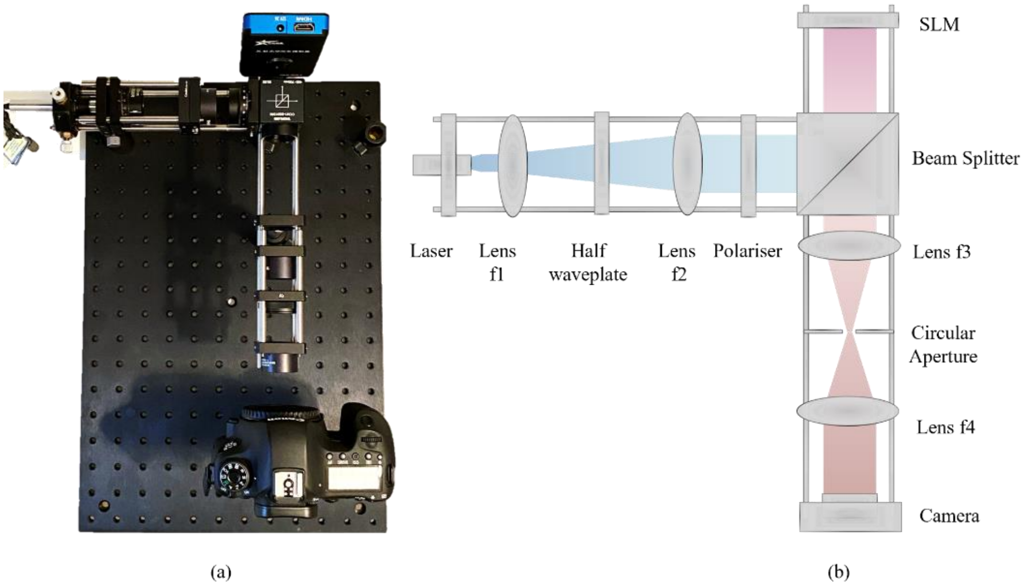

Figure 5.9. Holographic display system setup. (a) Our hardware display prototype with a Canon camera for image acquisition. (b) Optical system schematic diagram.

5.5.2 Python Control over Cannon Camera

By integrating camera control into the hologram generation process, we can manage the camera to capture images automatically and instantly transfer the captured data to the local computer. This end-to-end streamline enables quantitative comparison, aberration correction and other computational imaging tasks. Canon provides The EOS Digital Software Development Kit (EDSDK) to enable the image capture pipeline, controlling both the camera hardware as well as underlying software algorithms. Though low-cost off-the-shelf machine learning cameras can capture reconstructed images and provide SDKs for applications in Python, the image quality is comparatively low. The Canon SDK enables users to program the camera to take a series of exposures with configurable settings. A Python wrapper is written to access functions from EDSDK in C/C++ language, connecting and controlling a Canon EOS 6D camera in Python via USB to the local computer. With the additional python wrappers, programming control of the camera is allowed, including taking pictures and changing camera, focus, exposure time and ISO.

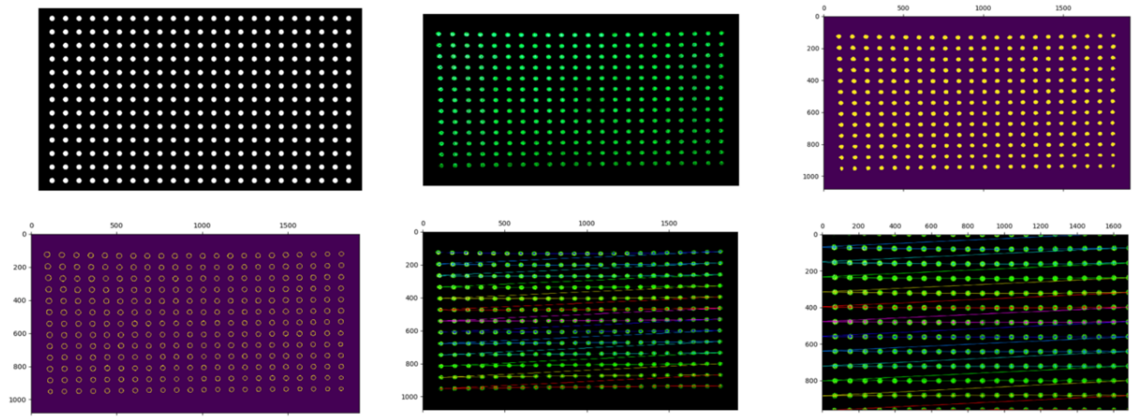

Figure 5.10. The calibration procedure for holographic reconstruction acquisition. (a) A circle grid pattern is a target for calibration. (b) The raw reconstructed image. (c) The binary-thresholded binary image using a morphological ellipse kernel with a gaussian blur. (d) Circle contours of the thresholded binary image. (e) Obtain the centre of all circles to compute the homography matrix. (f) The Undistorted image from the homography matrix.

5.5.3 Image Calibration

As captured images contain inaccurate measurements due to camera translation, we perform a geometric camera calibration to remap the captured images to the target images and minimise other distortion factors in the image-capturing pipeline. Figure 5.10 indicates the calibration procedure with a circle grid pattern as the target image.

A widely used pattern for camera calibration is the circle grid pattern. We use a binary symmetric circle grid pattern as the target image to generate a phase hologram. The pattern features 22×13 circles arranged evenly in rows and columns with an 80-pixel gap between the circle centres, measured from the centres of the first and last circle, and the pattern has a region of interest with 1,680×960 resolution. We then generate a CGH from the target circle grid pattern and display the hologram on the SLM with the proposed holographic projection system. The camera captures the reconstructed circle grid pattern, and we calibre the captured image. Ideally, we can directly detect the circle grid in the captured image and calculate a 3×3 homography matrix that relates the transformation between the captured image and the target image. The homography matrix can then transform the captured image for image undistortion. However, the circle detection technique is very sensitive, and circles may not be well reconstructed, especially at the edges of the holographic reconstructed image. We can create predefined circles using a morphological ellipse kernel to close small holes inside and apply gaussian blur for a loose threshold. We can further fill the circle contours of the thresholded binary image for better detection of the centres of all circles. Having a reasonably good augmented binary circle grid pattern, we can establish a planar homography matrix and apply the matrix to calibrate the original image accurately. Note that this operation can be executed before the start of every image capture or once to store the planar homography matrix.

The simulated reconstruction results based on IQM optimisation models are shown in Figure 5.11. Corresponding phase holograms and the experimental captured results in sRGB space are shown in the second and third rows, respectively.

5.6 Results Comparison and Discussion

5.6.1 Qualitative Comparison

We first make a qualitative comparison across all IQM-optimised methods for experimental results. As shown in Figure 5.12 and Figure 5.13, most IQM-based optimisation models converge on a reasonable visual quality. We observe that MAE, MSE, NLPD, SSIM, and MS-SSIM perform well but have undesirable local noise, which can be observed in the image patches selected from the reconstructed images. FSIM and VIF amplify high-frequency information, leading to structural over-enhancement. VSI, MS-GMSD and HaarPSI preserve the overall structures with a smooth appearance but artificially reduce local contrast with noticeable artefacts. Models based on deep-learning methods such as LPIPS and DISTS can recover the target image details but superimpose textures on the image. The optically reconstructed images exhibit laser speckle noise and are subject to optical aberrations, resulting in some noticeable common artefacts across all IQMs, including ghost and ripple effects. The dynamic range of the camera is limited, and captured images are prone to photometric distortions, including reduced contrast and saturation. Complementary qualitative results are provided in Figure 5.14, Figure 5.15 and Figure 5.16.

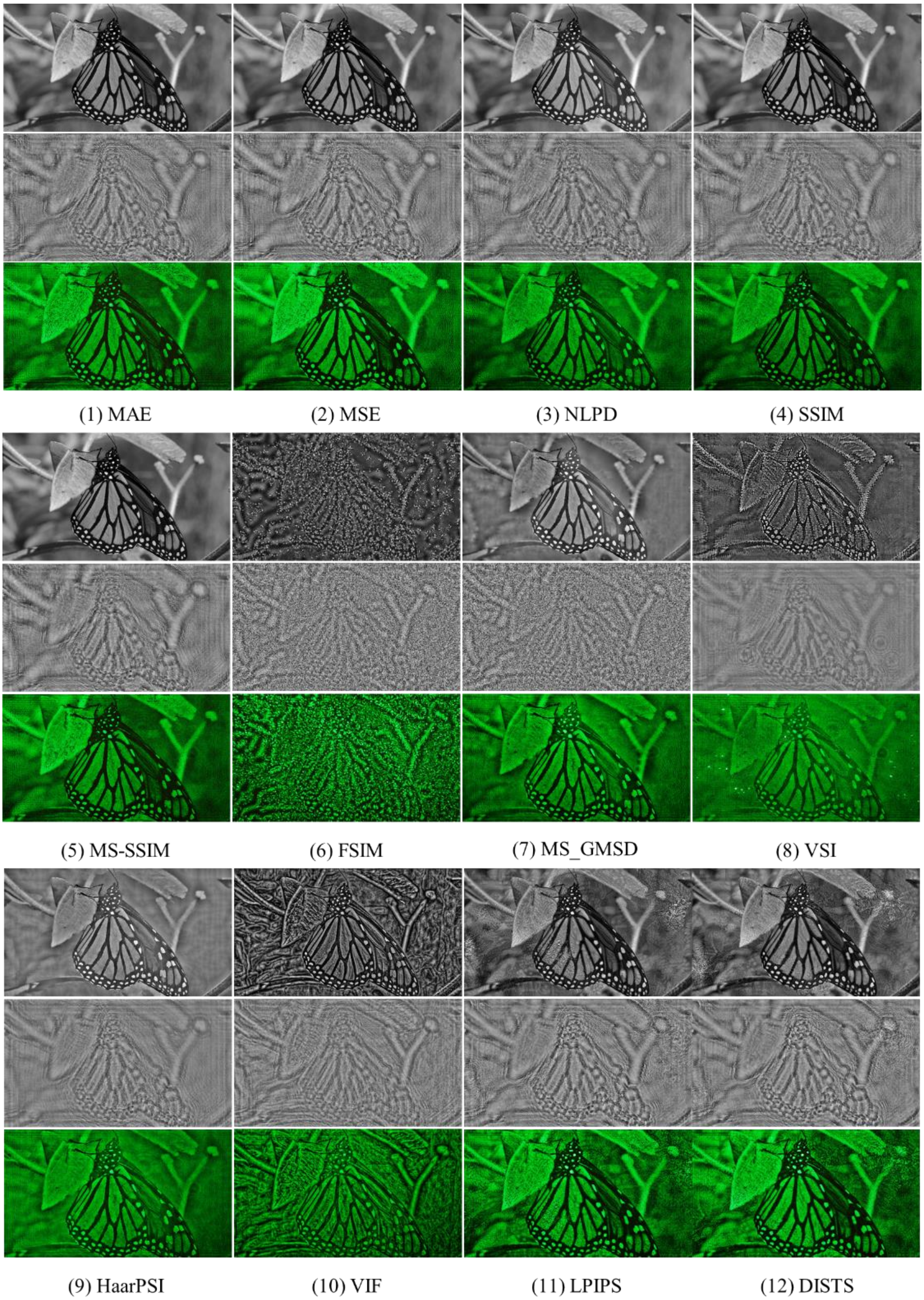

Figure 5.11. Simulated and captured results for CGH optimisation using twelve different IQM losses. We show the reconstructed image at the top for each loss, with the phase hologram in the middle and its corresponding captured results at the bottom.

Figure 5.12. Captured reconstruction results. For target images, we display phase holograms optimised by IQM losses. Reconstruction results of IQM losses are captured with our holographic display prototype for image quality comparison.

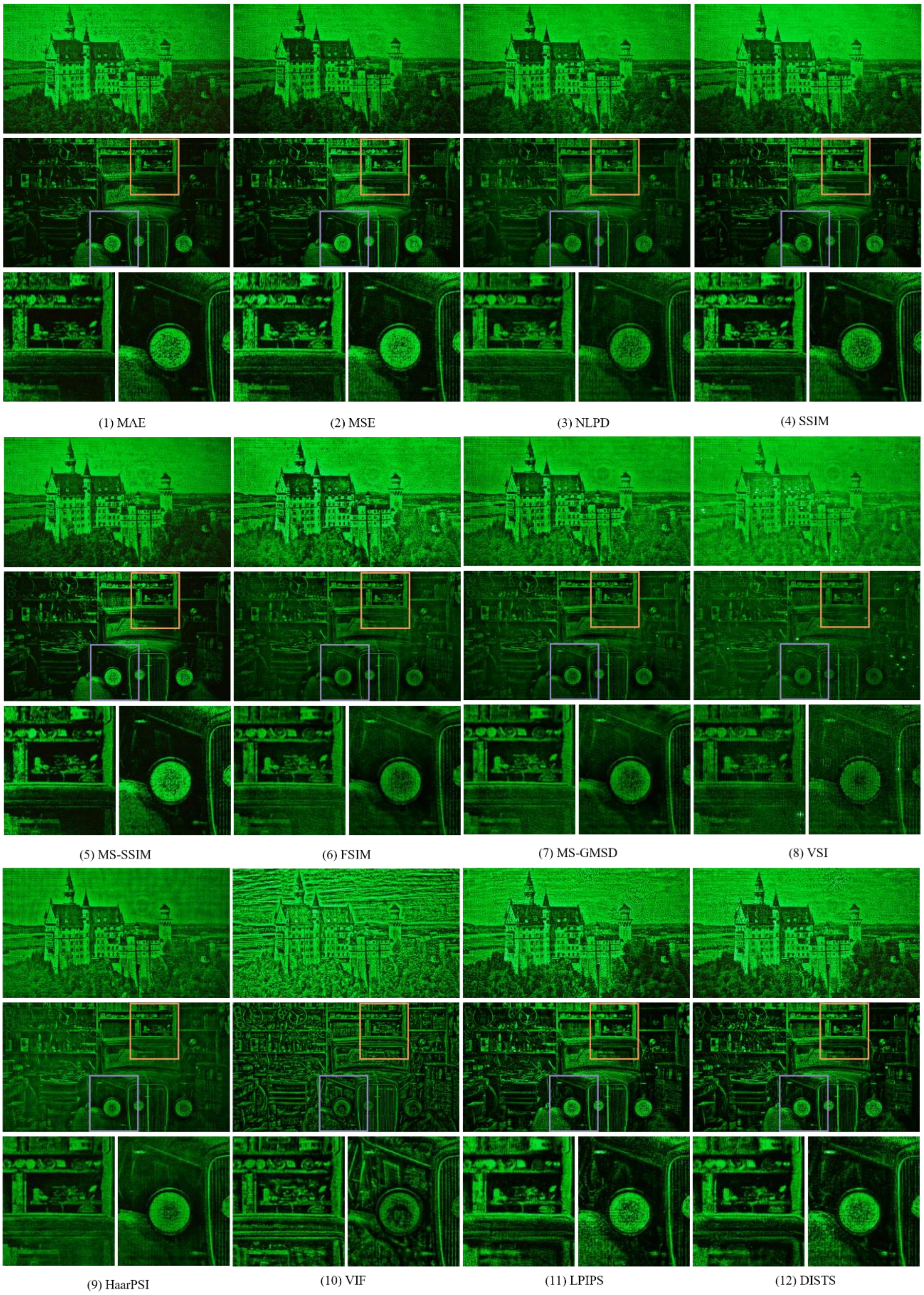

Figure 5.13. Captured reconstruction results with zoom-in details. We display phase holograms optimised by IQM losses. Reconstruction results of IQM losses are captured with our holographic display prototype for image quality.

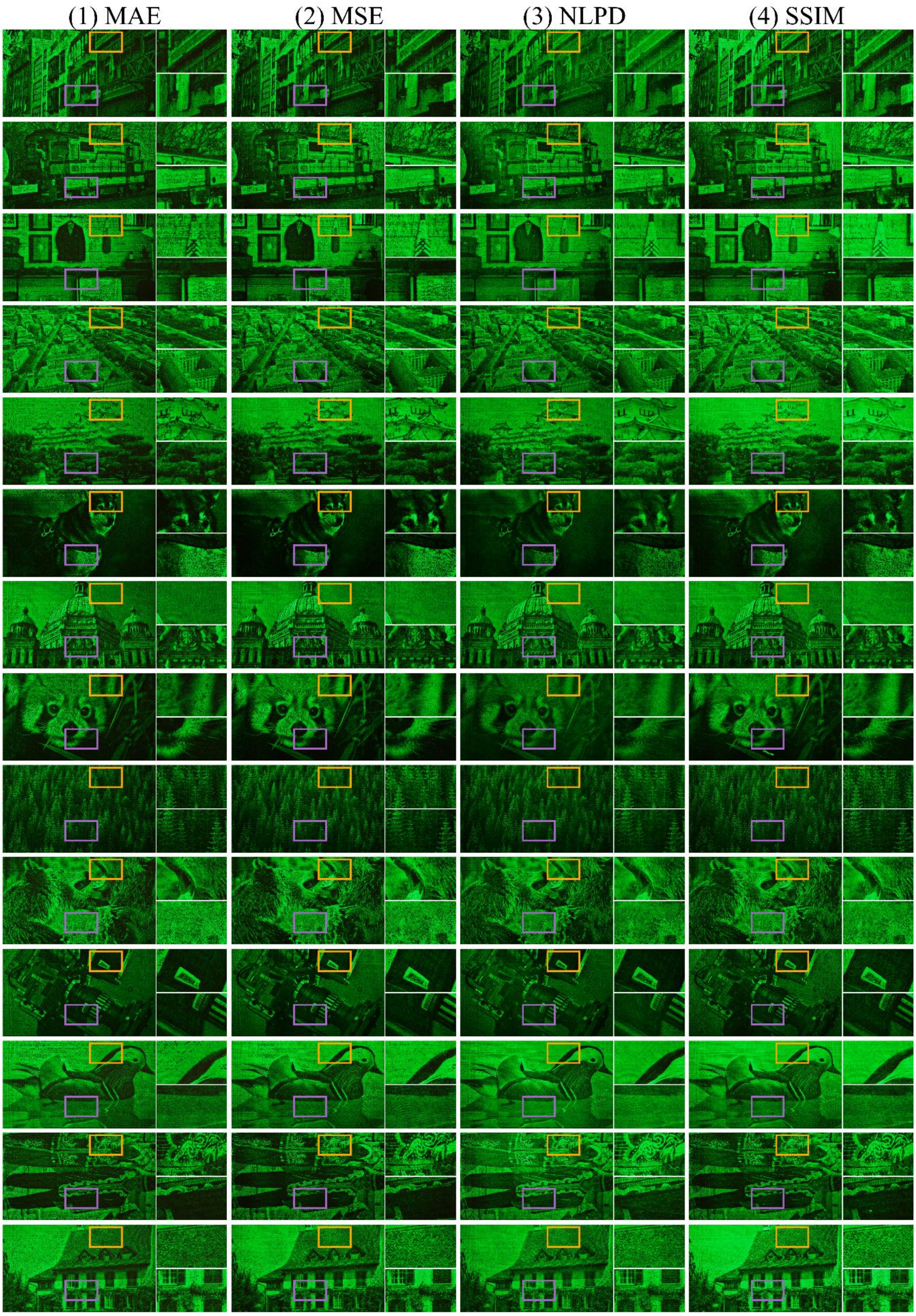

Figure 5.14. Additional experimentally captured results of MAE, MSE, NLPD, and SSIM with zoomed-in details.

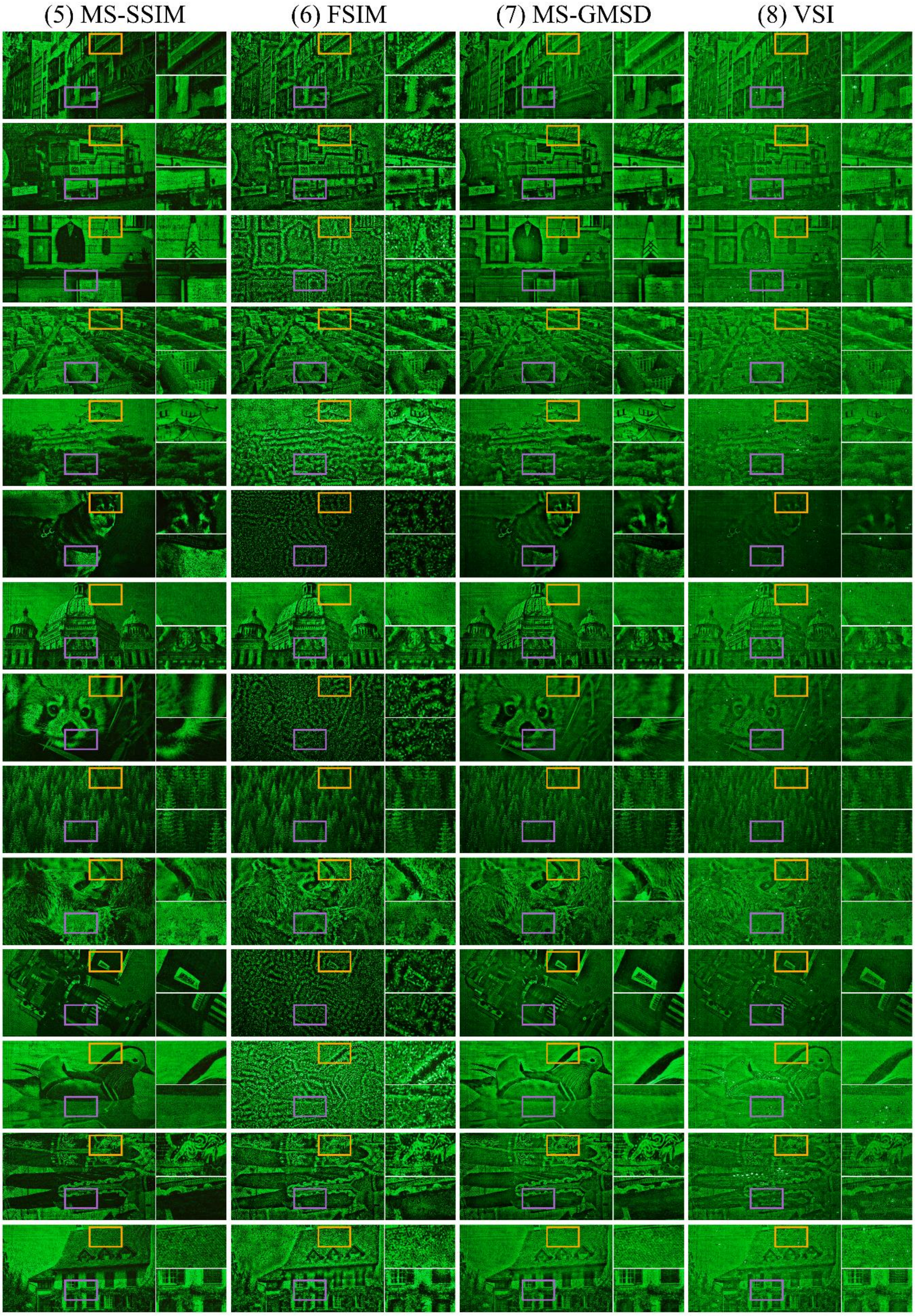

Figure 5.15. Additional experimentally captured results of MS-SSIM, FSIM, MS-GMSD, and VSI with zoomed-in details.

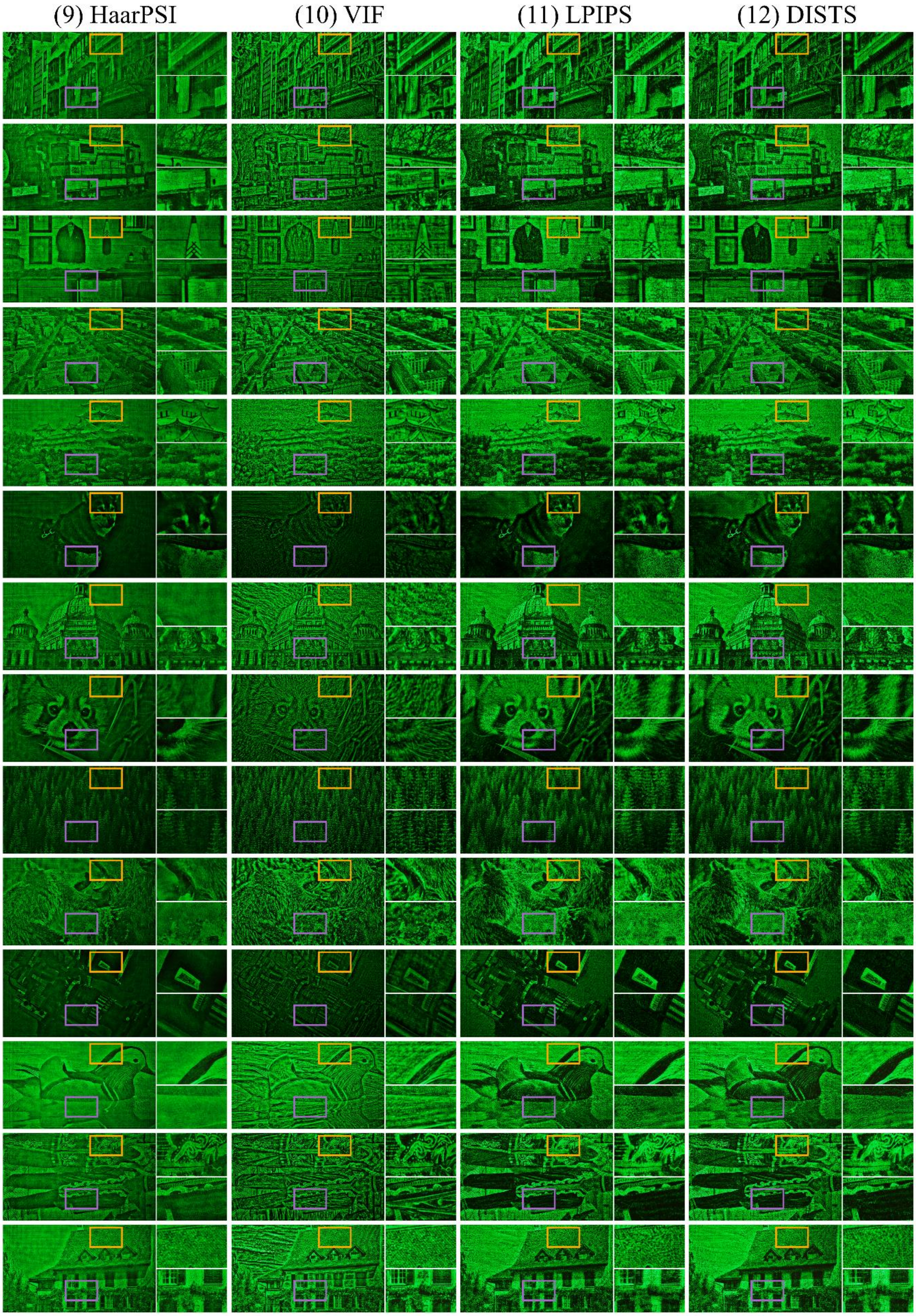

Figure 5.16. Additional experimentally captured of HaarPSI, VIF, LPIPS, and DISTS with zoomed-in details.

Table 5.3. Objective performance of IQM-based model evaluated by IQMs as quality metrics

IQM losses	Objective image quality metrics											
	MAE	MSE	NLPD	SSIM	MS-SSIM	FSIM	MS-GMSD	VSI	HaarPSI	VIF	LPIPS	DISTS
MAE	0.104	0.021	0.754	0.382	0.568	0.770	0.270	0.893	0.264	0.118	0.600	0.265
MSE	0.120	0.028	0.862	0.315	0.458	0.720	0.276	0.877	0.234	0.078	0.618	0.273
NLPD	0.118	0.024	0.717	0.365	0.566	0.783	0.258	0.905	0.287	0.117	0.601	0.271
SSIM	0.107	0.021	0.739	0.371	0.563	0.779	0.262	0.904	0.279	0.112	0.604	0.272
MS-SSIM	**0.096**	**0.018**	**0.696**	**0.414**	**0.610**	**0.795**	0.256	**0.913**	0.296	**0.133**	0.589	0.253
FSIM	0.185	0.058	1.083	0.219	0.305	0.648	0.294	0.795	0.187	0.067	0.664	0.387
MS-GMSD	0.153	0.040	0.833	0.328	0.451	0.744	0.258	0.879	0.274	0.098	0.608	0.283
VSI	0.158	0.040	0.816	0.299	0.430	0.761	0.256	0.894	0.276	0.079	0.628	0.406
HaarPSI	0.145	0.035	0.748	0.380	0.526	0.783	**0.245**	0.901	**0.313**	0.121	0.589	0.272
VIF	0.171	0.051	0.895	0.338	0.413	0.633	0.294	0.790	0.200	0.197	**0.580**	0.314
LPIPS	0.127	0.029	0.896	0.288	0.430	0.696	0.289	0.852	0.216	0.084	0.635	0.247
DISTS	0.130	0.030	0.911	0.279	0.415	0.690	0.289	0.852	0.212	0.077	0.636	**0.246**

5.6.2 Quantitative Comparison

To present a comprehensive analysis of employing contemporary CGH optimisation using the gradient descent method, we further introduce a rigorous procedure for evaluating the perceptual quality of holographic images based on extensive objective quality assessments as well as subjective comparisons informed by human perceptual judgments.

5.6.2.1 Objective Comparison

We can use the proposed IQMs as quality measures to evaluate the performance of gradient descent based CGH optimisation using different IQM losses. All IQMs are used to evaluate the captured results objectively. Scores are averaged over all 100 images for each metric and each IQM-based loss shown in Table 5.3. Each element indicates the score of an IQM loss evaluated using another IQM as a quality predictor.

By inspecting each row of the metric table, we find that MAE, NLPD, SSIM, and MS-SSIM maintain the best performance among all IQM losses, as previously predicted by the qualitative comparison. MS-SSIM loss produces superior reconstruction quality and objectively ranks as the best-performing IQM-based CGH optimisation model on most evaluation metrics, while

FSIM ranks as the least preferred method. Several other IQM losses, including NLPD, MAE, SSIM, HaarPSI and MS-GMSD, also outperform the MSE loss, objectively validating the use of IQMs for CGH optimisation.

Table 5.4. Objective performance of IQM-based model evaluated on different libraries.

IQM losses	Objective image quality metrics			
	SSIM	MS-SSIM	SSIM (piq)	MS-SSIM (piq)
MS-SSIM	0.414	0.610	0.619	0.641
NLPD	0.365	0.566	0.567	0.601
HaarPSI	0.380	0.526	0.550	0.591
MAE	0.382	0.568	0.577	0.602
SSIM	0.371	0.563	0.568	0.596
MS-GMSD	0.328	0.451	0.463	0.505
MSE	0.315	0.458	0.446	0.484

We can further convert Table 5.3 into a 2D ranking plot to give a well-defined and more illustrative comparison depicted in Figure 5.17. The horizontal axis indicates IQMs as quality measures used to evaluate the objective performance, and the vertical axis indicates IQMs used as loss functions for CGH optimisation. The rank order is colour coded from green to red with numbers 1–12 to indicate performance from best to worst.

Since the PIQ library implements its own SSIM and MS-SSIM metrics for image quality assessment, we can further evaluate our top-performing models using these metrics, as shown in Table 5.4. Though both the IQA and PIQ libraries have been benchmarked on a set of common databases and have nearly consistent ranking results in model evaluation, there is disagreement with the actual values of performance evaluation, with the IQM library generally obtaining lower scores. Hence, in the absence of a standard IQM implementation, it becomes more challenging to compare the performance of different algorithms.

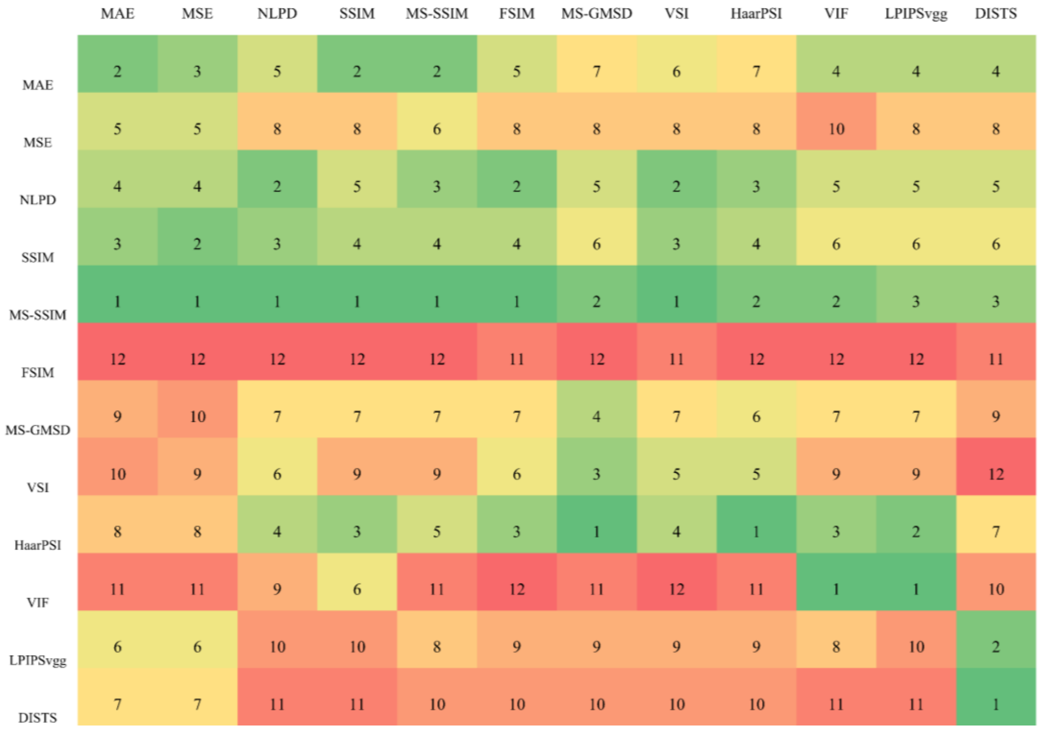

Figure 5.17. The objective ranking plot of the IQM-based model is evaluated by IQMs as quality metrics.

5.6.2.2 Subjective Comparison

To subjectively differentiate quality variations of tested models, we gather human perceptual judgments by employing a 2-alternative forced choice (2AFC) method. The experiment asks subjects to indicate which one of two distorted images is perceptually closer to the reference image. Figure 5.18 illustrates the interface for this experiment: an image triplet with a pair of experimentally captured images and the corresponding reference image are simultaneously presented. Subjects are asked to select the better image between two distorted ones. After the selection, two new experimentally captured images optimised according to different IQM losses appear on the upper screen in randomized left-right order. Progress is indicated, and a pause function is available to reduce visual fatigue. The screen has a 1920 × 1080 pixels resolution, with the displayed image resolution at 875 × 500. The user interface supports a zoom function for careful inspection of image details.

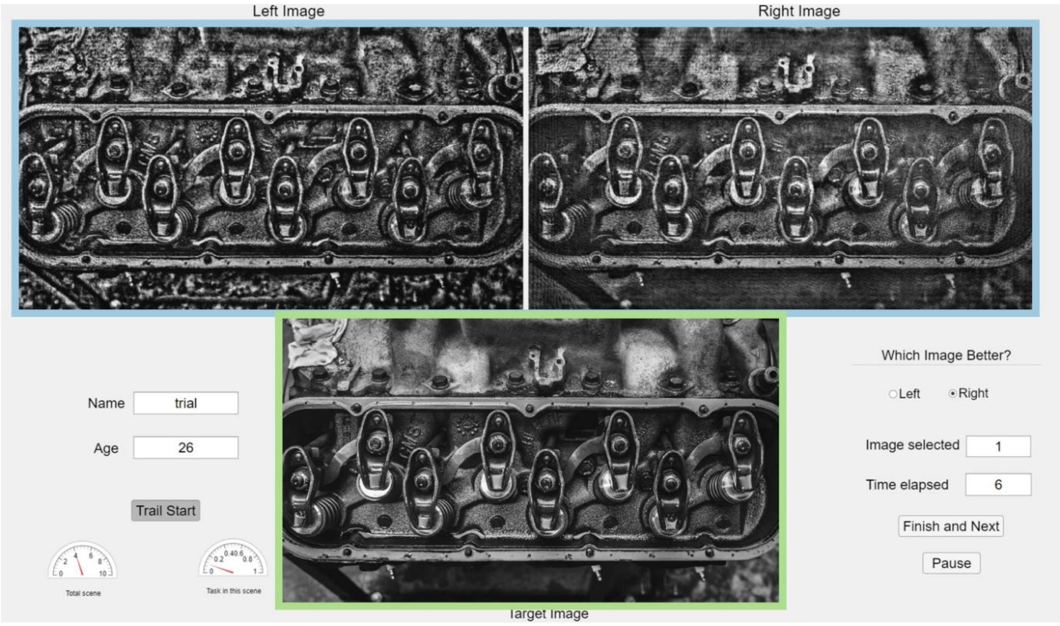

Figure 5.18. The user interface for collecting human judgments on IQM-based CGH optimisation. The experimentally captured image pair from two IQM losses and the corresponding reference image are shown in the blue and green boxes, respectively.

Participants are mainly university students and are provided with appropriate instructions, including an explanation of the experimental procedure as well as a demonstration session. To avoid fatigue, we pause the user interface every 15 mins and allow subjects to take a break at any time during the experiment. Experiments are performed at a normal indoor light level with reasonably varying ambient conditions according to the recommendations of ITU-R BT 500 [177]. This subjective experiment was approved by the Cambridge Engineering Research Ethics committee and carried out according to the Declaration of Helsinki. We obtained informed consent and gathered paired comparisons from 20 subjects. Each subject responded to all possible combinations of generated images for a pair of target images, doing so for ten pairs of target images, yielding $\binom{12}{2} \times 10 = 660$ stimuli. Data is saved for analysis, including time spent for each judgment, the paired-image display order and the results of pairwise comparisons. The preferred image of the displayed pair contributes one point to the score of its IQM loss. Therefore, for the selected 10 sample images, each paired comparison could receive 0 to 10 points as the subjective score from the subject. In order to exclude abnormal results, we check several sentinels in each observation data that consist of pairs with obvious visual quality contrast. Overall, we received 13200 judgments across 12 IQM losses, and each loss is ranked 1100 times. The average time for one judgment is approximately 3 seconds.

Table 5.5. Subjective ranking results from participants. Each column indicates the ranking of IQM losses evaluated by a subject. Numbers 1 to 12 denote the rank order from the best to the worst.

Subjects	MAE	MSE	NLPD	SSIM	MS-SSIM	FSIM	MS-GMSD	VSI	Haar PSI	VIF	LPIPS	DISTS
1	5	4	2	3	1	10	6	11	7	12	9	8
2	5	2	4	3	1	10	6	11	8	12	9	7
3	5	2	4	3	1	10	8	11	9	12	7	6
4	6	2	4	3	1	10	8.5	12	7	11	8.5	5
5	7	4	3	1	2	10.5	5	10.5	6	12	9	8
6	5	2	3	4	1	11	6.5	10	6.5	12	8	9
7	6	4	1	2	3	11	5	10	7	12	9	8
8	5	3	4	2	1	9	6	10	7	12	8	11
9	2	1	3.5	5	3.5	10	6	11	7	12	9	8
10	5	2	4	3	1	10	6	11	7	12	9	8
11	2	1	5	4	3	11	6	12	7.5	10	9	7.5
12	5	1	4	2	3	10	6.5	11	8	12	9	6.5
13	5	2	4	3	1	10	7.5	11	6	12	7.5	9
14	5	1	3.5	3.5	2	10	6	11	7	12	9	8
15	3	2	5	4	1	10	6	12	8	11	7	9
16	5	4	2.5	2.5	1	10	6	12	7	11	9	8
17	3	2	4.5	4.5	1	10	6.5	11	9	12	8	6.5
18	5	4	3	1	2	10	8	11	6	12	9	7
19	4	3	5	1	2	10	6	11	7	12	9	8
20	5	2	3	4	1	10	6	11	7	12	9	8

We employ the Bradley-Terry model [178,179] to aggregate pairwise comparisons and obtain a global ranking of IQM losses for CGH optimisation based on subjective data. From partial orderings provided in the data, we wish to infer the ranking order of tested losses and the subjective visual quality scores associated with them. If we denote $s = [s_1, s_2, s_3, \ldots s_m]$ as subjective scores of the evaluated IQM losses, the Bradley-Terry model assumes that the probability of choosing loss i over loss j is:

$$p_{ij} = \frac{e^{s_i}}{e^{s_i} + e^{s_j}}. \qquad 5.8$$

Given the observed number of times that IQM loss i is favoured over IQM loss j as w_{ij}, We then can obtain the likelihood of i over j as $p_{ij}^{w_{ij}}$.Thus, assuming outcomes of each paired comparison are statistically independent, the likelihood function of all (i, j) pairs is defined by:

$$P = \prod_{i=1}^{M} \prod_{\substack{j=1 \\ j \neq i}}^{M} p_{ij}^{w_{ij}}. \quad 5.9$$

The subjective score for IQM loss s_i can then be jointly estimated by maximising the log-likelihood of all pairwise comparison observations:

$$\ell(s_i) = \sum_{i=1}^{M} \sum_{\substack{j=1 \\ j \neq i}}^{M} \left(w_{ij} s_i - w_{ij} \log\left(e^{s_i} + e^{s_j}\right)\right). \quad 5.10$$

We implement the Bradley-Terry model in R to iteratively solve the given equation Eq.(8) and obtain the optimal estimate s_i for each model. The Bradley-Terry model scores are normalised by shifting to zero means, resulting in a global ranking of optimisation performance.

Table 5.6. Subjective winning matrix voted by all participants.

	MAE	MSE	NLPD	SSIM	MS-SSIM	FSIM	MS-GMSD	VSI	Haar PSI	VIF	LPIPS	DISTS
MAE	0	78	104	81	53	178	138	190	131	197	167	162
MSE	122	0	124	118	71	197	152	198	171	194	174	182
NLPD	96	76	0	91	93	196	167	200	156	193	168	158
SSIM	119	82	109	0	85	189	161	196	173	193	177	162
MS-SSIM	147	129	107	115	0	194	170	199	174	197	190	178
FSIM	22	3	4	11	6	0	25	115	35	132	63	58
MS-GMSD	62	48	33	39	30	175	0	198	105	183	132	124
VSI	10	2	0	4	1	85	2	0	8	158	25	42
HaarPSI	69	29	44	27	26	165	95	192	0	186	120	113
VIF	3	6	7	7	3	68	17	42	14	0	25	27
LPIPS	33	26	32	23	10	137	68	175	80	175	0	89
DISTS	38	18	42	38	22	142	76	158	87	173	111	0

We converted the pairwise comparisons of generated images from each subject using the B-T model to obtain the ranking order of IQM losses shown in each column in Table 5.5. Numbers 1 to 12 denote the rank order from the best to the worst. Table 5.6 shows the winning matrix voted by all participants. Each element in the table indicates the number of votes that the column method preferred to the row method. We received an overall of 13200 judgments across 12 IQM losses.

Table 5.7. Bradley-Terry scores and p-values of the t-test by comparing adjacent methods.

	MS-SSIM	MSE	SSIM	NLPD	MAE	MS-GMSD	HaarPSI	DISTS	LPIPS	FSIM	VSI	VIF
B-T scores	1.861	1.578	1.409	1.298	0.993	0.146	-0.007	-0.407	-0.553	-1.625	-2.103	-2.591
P-value (adjacent)	N/A	1.787e-02	1.368e-01	7.909e-02	4.900e-03	1.782e-05	5.714e-03	1.009e-01	2.227e-02	3.521e-10	6.100e-06	3.561e-04

Table 5.7 indicates the B-T scores calculated from the winning matrix and independent two-sample t-tests with two-tailed distribution to investigate whether the differences between the subjective performance of IQM losses are statistically significant. Specifically, we consider that the obtained observations from participants are normally distributed under the null hypothesis and compare the ranking scores for any of the two losses. If the comparison cannot reject the null hypothesis of no difference at the standard significance level $\alpha = 0.05$, we put the evaluated losses in the same group as they are statistically indistinguishable.

Figure 5.19 shows the scatter plot of the combined subjective and objective performance of tested IQM losses for CGH optimisation. Scatter points with the same colour are in the same statistical significance group for subjective tests. The objective global ranking score for each IQM loss can be obtained by adding ranking orders from all quality metrics derived from Table 5.3 and normalising them to zero-mean. Scores have been reformulated to ensure that higher scores indicate higher predicted quality.

The scatter plot indicates that the MS-SSIM is the top-ranking loss function, as agreed upon by both subjective and objective evaluations. NLPD and SSIM losses are statistically indistinguishable from the MSE loss for subjective performance. The MSE loss unexpectedly achieves higher performance in the subjective test than HaarPSI, and MAE losses, despite performing far worse in objective performance. A similar trend also occurs in VSI and VIF losses versus the FSIM loss. This disagreement is due to different objective and subjective weighting strategies on image structure similarity, image smoothness, luminance, and contrast.

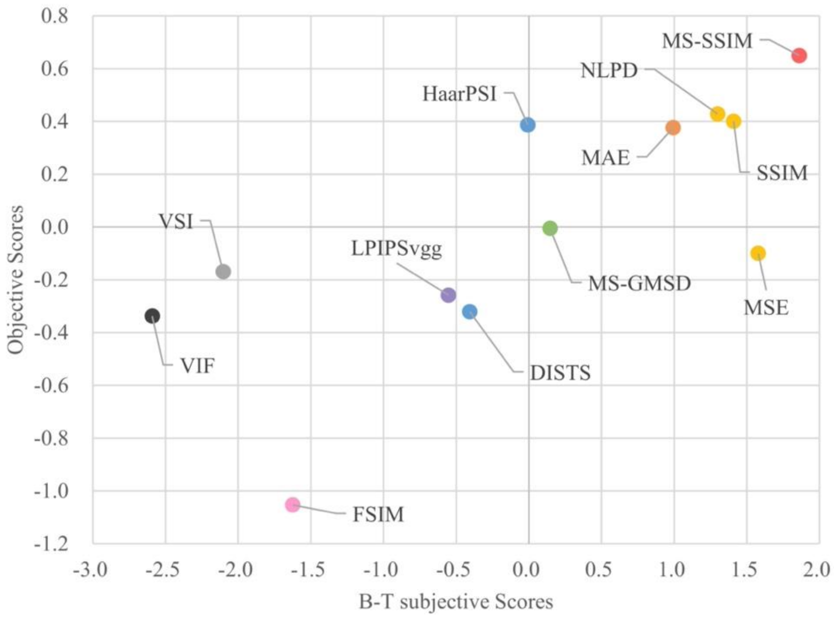

Figure 5.19. Quantitative comparison of IQM-based CGH optimisation. Scatter points represent the losses for CGH optimisation. Points with the same colour are statistically indistinguishable for subjective results. Vertical and horizontal axes indicate the objective performance and the subjective performance of each loss, respectively.

We further calculate Spearman's rank order correlation coefficient (SRCC) between objective and subjective scores, as shown in Table 5.8. Higher SRCC scores indicate a better correlation of a metric with subjective ratings. Although most modern image quality metrics show superior performance in existing image databases, we observe that, for the CGH optimisation task, they have less correlation than pixel-error-based metrics to human judgments. This may be because the most common image databases for benchmarking, such as LIVE [180], TID2008 [181] and TID2013 [182], comprise source images with synthetically distorted images. The synthetic distortion types, including White Gaussian Noise, JPEG2000 compression, and Gaussian Blur with varied distortion levels, attempt to reflect various image impairments in image processing. Experimental CGH reconstructed images, such as those seen here, can be rather more complex, with more types of distortions produced during the optical reconstruction and image acquisition. Furthermore, CGHs are predominantly tainted by noise, whereas some IQMs were developed for recognising blurry objects, inferring details in deblurred objects, or super-resolution imaging tasks. Current IQMs are not specifically benchmarked well for those real-world and CGH distortions. For partial coherent light illumination in the holographic optical system that could bring a more blurry effect and contrast reduction in the replay field [29,183], modern

IQMs may take advantage of inferring blurry and contrast-reduced information. Therefore, the use of IQMs may potentially have better performance in partially coherent holographic displays.

Table 5.8. SRCC between objective scores and subjective scores of IQM-based CGH optimisation

Objective image quality metrics	SRCC
MAE	0.846
MSE	0.825
NLPD	0.657
SSIM	0.587
MS-SSIM	0.839
FSIM	0.692
MS-GMSD	0.434
VSI	0.678
HaarPSI	0.566
VIF	0.189
LPIPS	0.266
DISTS	0.427

5.7 Discussion

One of the best-performing IQMs in this study is MS-SSIM for the 2D CGH optimisation, which outperformed all the other IQMs, including LPIPS. This superiority can be attributed to its characteristics and the well-defined gradient for CGH optimisation.

MS-SSIM considers the structural similarity of the reconstructed image with the original image across multiple scales and incorporates the sensitivity of the human visual system to contrast and luminance changes. MS-SSIM is also robust to compression and noise, making it a reliable metric for measuring image quality in various contexts. In comparison, characteristics of some IQMs may be more sensitive to the choice of image, including phase congruency, saliency map, and mutual information considered by FSIM and HaarPSI, VSI and VIF. The idea of visual saliency is that certain regions in an image are more important than others for human perception, and therefore the quality of an image should be evaluated based on how well those important regions are preserved. Phase congruency measures phase consistency across different scales and orientations in an image to indicate the change in local image structure. In contrast, The VIF computes the mutual information between the natural scene statistics features of the original and distorted images to measure the similarity of the information extracted by the HVS from the two images. These methods rely on the analysis of visually discernable features or properties to perform evaluations, and the effect of such features may be compromised due to the unique nature of the distortions introduced by the CGH optimisation process. Some imperceptible image distortions may cause the CGH optimisation algorithm to generate less plausible solutions.

Calculating the gradient from the loss function is critical in gradient descent optimisation. Ideally, IQMs used as loss functions should be injective, meaning that distinct inputs should map to distinct outputs. However, certain IQMs may not have unique optima to guarantee that images close to optimal, resulting in a lack of convergence or poor convergence during optimisation. In addition, some IQMs have complex compositions of different features that may not have a well-behaved gradient to steer the CGH optimisation process in the right direction efficiently. Furthermore, certain IQMs heavily rely on specific features while underweighting or even ignoring other perceptual features. This underweighting problem may be exacerbated during gradient calculation, resulting in less plausible images during CGH optimisation. For instance, MS-GMSD, HaarPSI, VSI, and NLPD are IQMs discard local luminance or contrast information essential to human perception of image quality. Therefore, choosing IQMs for loss functions considering all perceptually important features is crucial to avoid image degradation in CGH optimisation.

LPIPS is a recent IQM that utilises deep learning techniques to measure perceptual similarity between two image patches. While LPIPS has demonstrated promising results in the context of perceptual metrics, its performance in CGH optimisation is relatively less plausible. This can be attributed to the fact that LPIPS is carefully trained and fine-tuned on a specific dataset consisting of common distortions evaluated for four image processing tasks: denoising, deblurring, super-resolution, and image compression. However, LPIPS that relies on pre-trained deep networks for image processing tasks may not necessarily optimise the relevant features for the specific CGH task being performed. The explicit representation of fine textures in LPIPS may adversely lead to overweighting such features during optimisation, ultimately resulting in an undesirable over-enhancement of textures in reconstructed images. Additionally, the pooling operations in the hidden layers of the network during feature extraction can lead to non-bijective functions, which means that different inputs can result in identical latent representations [184]. This can result in poor convergence and instability during optimisation. Last, the high computational complexity of two DNN-based models (LPIPS and DISTS) and lack of interpretability may hinder their use for CGH optimisation.

5.8 Conclusion

In this chapter, we have validated the gradient descent method to phase-only CGH optimisation in simulation. Furthermore, we have comprehensively studied the real-world performance of using IQMs as loss functions in the CGH optimisation process. By benchmarking with a standard optical reconstruction dataset, we have collected the results of applying 12 distinct IQMs as loss functions in both objective and subjective ratings. The results from the comparison study show that IQM losses can achieve better image quality than the MSE loss in generating holograms, with the MS-SSIM loss outperforming all the other losses. This extensive comparison reveals that the perceived image quality improves considerably when the appropriate IQM loss function is used, highlighting the value of developing perceptually-motivated loss functions for hologram optimisation.

Beyond this study, individual IQM losses can be further combined based on their complementarity to incorporate the specific CGH distortions. We recognise that our analysis is limited to 2D hologram reconstruction. For 3D holographic applications, we believe several extensions to the work conducted in this study, such as the use of blurring distortion, which could be a significant perceptual factor to be considered in hologram optimisation. This method can be combined with a camera as a feedback optimisation strategy to eliminate optical artefacts in experimental setups [43,141].

Chapter 6 Natural Defocused Computer-Generated Holography

6.1 Introduction

The previous chapter presented using the gradient descent method in phase-only hologram optimisation and demonstrated that the perceived quality of reconstruction results improves considerably should the gradient descent method with the appropriate IQM loss be selected for 2D hologram generation. This chapter focuses on extending the gradient descent method to 3D hologram generation, enabling high-quality 3D CGH reconstruction. Although recent algorithms [29,43–45,138,141,185,186] have made progress in improving the image quality and computational speed with the introduction of deep learning techniques, simulating the natural defocus blur effect and occlusion effect at depth discontinuities remain the main issues in these algorithms; reconstructed scenes often have a large depth-of-field. Additionally, these algorithms only optimise the all-in-focus target images to calculate the loss function, leading to image quality degradation in the out-of-focus area.

This chapter first introduces the gradient descent method for 3D hologram generation and validates its effectiveness in numerical simulation and optical experiments. Next, a 3D hologram generation method is proposed using an incoherent propagation model to generate target images. The generated incoherent images can accurately describe natural defocus blur and are directly used as target images to compensate for the unrealistic defocus effects. The proposed method is validated experimentally, demonstrating its capability of reproducing naturally defocused images highly similar to physically rendered 3D scenes and real objects.

6.2 Gradient Descent Method for 3D CGH Generation

6.2.1 Method Description

Figure 6.1 demonstrates the process of using the gradient descent method as a forward-backward optimisation for 3D CGH generation. In the forward pass, rather than being propagated to a single replay plane as in the 2D case, the phase hologram is propagated to replay planes for a set of distances $z\{i\}, i = 1,2,3 \ldots N$. The reconstructed images from complex-amplitude replay fields at various distances are then binary masked to extract the in-focus regions, which are then summed to form an all-in-focus image. The reconstructed focal image is then compared to the target image for calculating the loss function. In the backward pass, the optimiser calculates the derivatives of the loss function with respect to the current phase hologram and updates the hologram to minimise the loss.

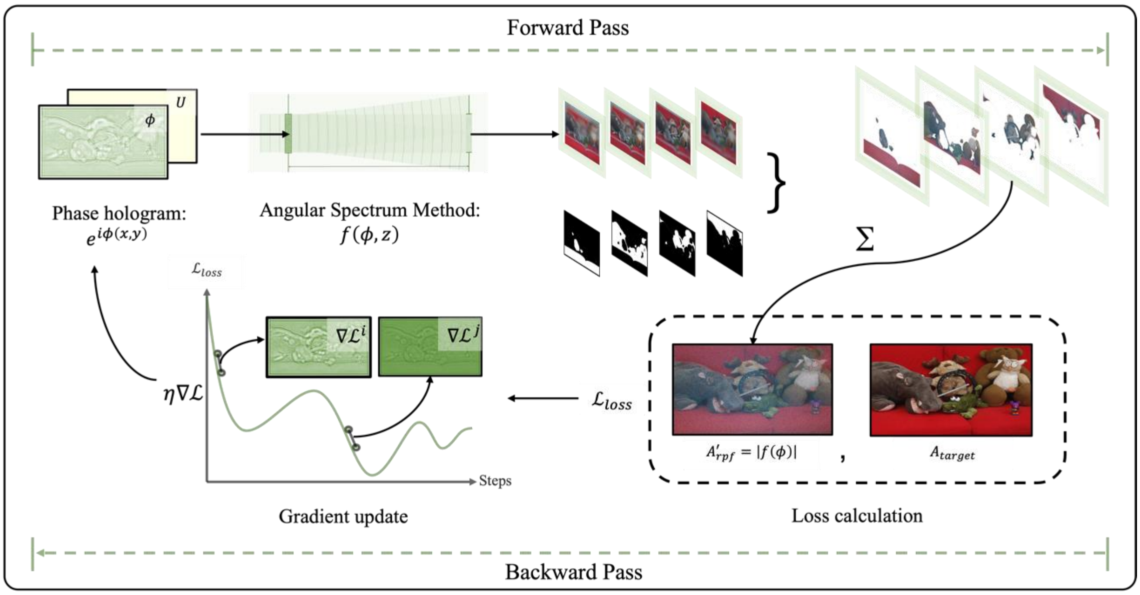

Figure 6.1. The gradient descent method for 3D CGH generation algorithms.

In the forward pass, the angular spectrum method propagates the phase hologram $e^{i\phi(\mu,\nu)}$ to a distance z_i with a planar illuminating wave can be expressed as:

$$f(\phi, z_i) = \mathcal{F}^{-1}\left\{\mathcal{F}\left\{e^{i\phi(\mu,\nu)}\right\} \times exp\left[j2\pi z_i\sqrt{\frac{1}{\lambda^2} - f_x^2 - f_y^2}\right]\right\}, \quad 6.1$$

where $f(\phi, z_i)$ is the resulting complex replay field, and f_x and f_y are spatial frequencies. The reconstructed image intensity at a distance z_i is $I(x, y, z_i) = |f(\phi, z_i)|^2$. As discussed in the previous chapter, the amplitude of the complex replay field $A_{rpf}(x, y, z_i) = |f(\phi, z_i)|$ with the target image amplitude is used for optimisation. Due to the fact that the replay field $A_{rpf}(x, y, z_i) = |f(\phi, z_i)|$ not only reconstructs the target image at a distance z_i but also reconstructs the blurred images corresponding to other distances, the reconstructed amplitudes are masked according to the corresponding depth layers. The formulated all-in-focus reconstructed amplitude by binary masks is then compared to the target image amplitude. RGBD images are used as target images for hologram generation, and the depth images can be directly quantised and used as thresholding binary masks to extract the signal area of reconstructed amplitudes:

$$BM(x, y, i) = \begin{cases} 1, & if\, |z_i - D(x, y, i)| < |z_j - D(x, y, i)|, \forall i \neq j \\ 0, & otherwise \end{cases} \quad 6.2$$

Intuitively, the thresholding binary mask $BM(x, y, i)$ is set to 1 if the value of the depth map $D(x, y, i)$ matches with the distance z_i. It is necessary to ensure that the energies of all depth images are consistent inside each colour channel.

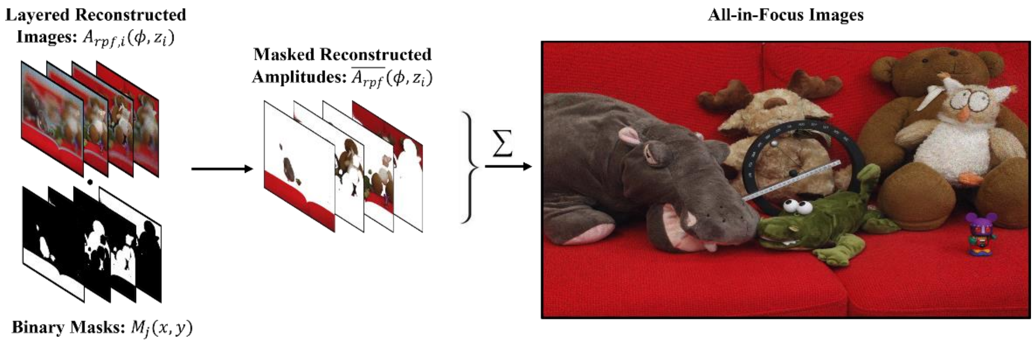

Figure 6.2. The all-in-focus reconstructed image is obtained from reconstructed images at various replay planes.

The phase hologram is propagated to all replay planes and uses the corresponding masks to extract the in-focus region of the reconstructed images. The reconstructed focal image amplitude is the sum of masked reconstructed amplitudes $\overline{A_{rpf}}(x, y, z_i)$ at all distances:

$$A_{out} = \sum_{i}^{N} \overline{A_{rpf}}(x, y, z_i) = \sum_{i}^{N} BM(x, y, i) \cdot |f(\phi, z_i)|. \quad 6.3$$

The bar over the reconstructed amplitudes denotes that amplitudes are binarily masked. The generation process of the all-in-focus reconstructed image from reconstructed images at various replay planes is shown in Figure 6.2. As such, the goal of the gradient descent CGH generation algorithm is to find the optimal quantised phase hologram $\hat{\phi}$ that can minimise the loss function in describing the visual quality between the target all-in-focus target image amplitude $A_{target,AIF}$ with the reconstructed focal stacked image amplitude:

$$\hat{\phi} = \underset{\phi}{argmin}\, \mathcal{L}\left(\sum_{i}^{N} \overline{A_{rpf}}(x, y, z_i)\, , A_{target,AIF}\right). \quad 6.4$$

With this objective in hand, we can use the standard loss function MSE by comparing the per-pixel errors between the reconstructed focal amplitude and the target all-in-focus amplitude. The overall MSE loss function is:

$$\mathcal{L}_{MSE} = \frac{1}{mn}\sum_{m,n}\left[\sum_{i}^{N} BM(x, y, i) \cdot |f(\phi, z_i)| - A_{target}\right]^2. \quad 6.5$$

In the backward pass, we can compute the gradient of the loss function and update the next estimate of the phase hologram. Again, deep learning libraries, including PyTorch and TensorFlow, have implemented auto-differentiation packages using Wirtinger derivatives to calculate the complex amplitude gradients for the gradient descent methods.

Figure 6.3. The sample RGBD image as the input object.

6.2.2 Method Implementation

To validate the suggested method for producing CGHs for 3D models, we use an RGBD image as the input object, illustrated in Figure 6.3. We first slice the RGBD images into discrete layers according to the depth values of the depth image. In this implementation, we linearly slice the depth images into four layers with corresponding distances $z = \{12cm, 13cm, 14cm, 15cm\}$ and resolution at 1920 × 1080. This step involves locating an index matrix of out-of-range depth values and then using this index matrix to set these depth values to zero for this layer. The RGB image intensities are converted into linear space amplitude as the target image for optimisation. We store the sliced RGBD image as a tensor with a shape of $[N, C, H, W]$, where N is the number of layers; C is the number of channels; H is the height of the image; W is the width of the image. Therefore the 4-layer sliced RGB image is a $[4, 3, 1080, 1920]$ tensor, and the sliced depth image is stored as a $[4, 1, 1080, 1920]$ tensor. The mean values of each layer per colour channel are recorded to rectify optimised image means. The sliced target RGB images are shown in Figure 6.4.

Figure 6.4. The input target RGB image is sliced into different layers.

The RGBD images are sliced into only four layers due to the computational capacity of the GPU. Since the phase hologram optimisation executes in RGB channels simultaneously, the GPU manages large amounts of data and could have limited memory available. Therefore, we limit the layer number so that the optimisation process would not run out of GPU memory. Alternatively, we can optimise the phase hologram per single colour channel for multiple layers individually and combine all three colour channels to formulate the RGB phase hologram.

The initial random phase holograms of RGB channels are propagated to replay fields in different depths with the forward transfer functions by the angular spectrum method. To improve computational efficiency, the transfer functions are precalculated per depth layer at wavelengths $[465nm, 532nm, 620nm]$ of the light source in RGB channels, with the SLM pixel pitch at $6.4\mu m$. The amplitudes of replay fields are then normalised with the mean values of RGB layers and compared with target images.

We use MSE as the loss function for demonstration purposes to evaluate the image quality at different depths. The reconstructed images are optimised over 500 iterations using the Adam optimiser at a 0.05 learning rate and default exponential decay rates of $\beta 1$ = 0.9 and $\beta 2$ = 0.999. CGHs are generated using a computer with an i7-8700 CPU@3.20GHz and a GeForce RTX3060 GPU. PyTorch 1.9.0 and CUDA 10.2 are used to implement complex-amplitude gradient descent optimisation on the GPU.

6.2.3 Simulation Validation

Figure 6.5 shows the reconstructed all-in-focus image as well as its residual when compared with the target image. The residual image is the difference between the noisy reconstructed image and the target image, which can be directly calculated by subtracting both images. The residual image, therefore, contains the per-pixel error compared with the target image.

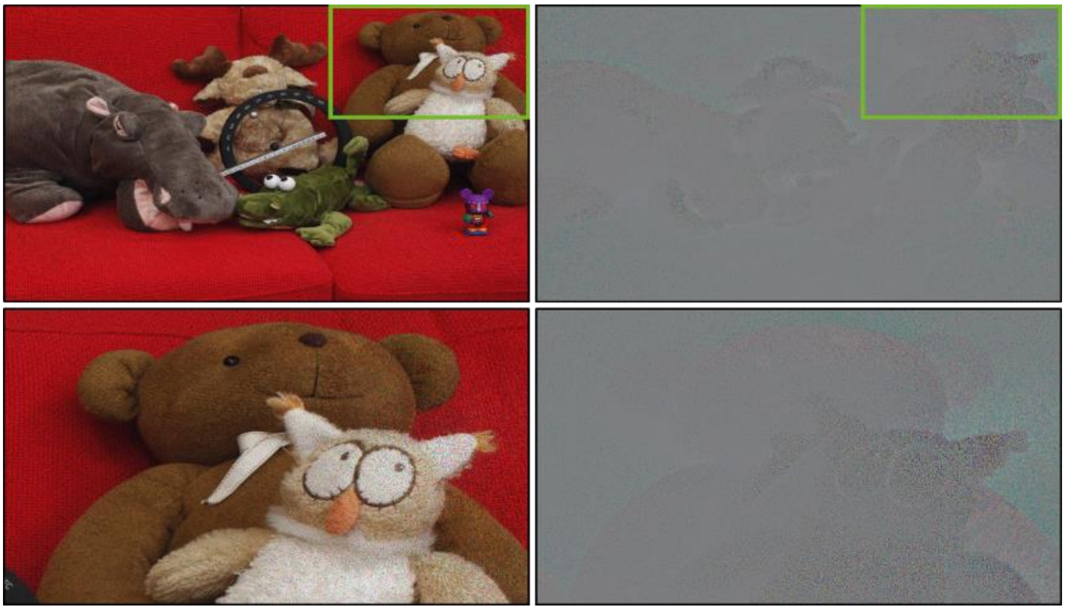

Figure 6.5. The reconstructed all-in-focus image with its residual image.

The proposed method can optimise the initial phase hologram and result in a good-quality reconstructed image. The reconstruction result provides a nearly ideal all-in-focus image with well-preserved fine details. The noise in the in-focus region of the target image is substantially reduced during the optimisation, though some noticeable noise perturbations are shown in the zoomed-in image. We can see significant image quality degradation from the residual image at the depth continuities. This degradation can be further confirmed by directly simulating the amplitude distribution of the reconstructed image at different planes with details of the out-of-focus blur effect shown in Figure 6.6. As shown in the figure, though the depth information is preserved with a clear defocus effect when viewed from other focused planes, image quality is sharply degraded at depth discontinuities, suffering from a severe edge-enhancing artefact. Moreover, the proposed method has an unsmooth amplitude distribution in the blurred area, with noticeable chromatic artefacts.

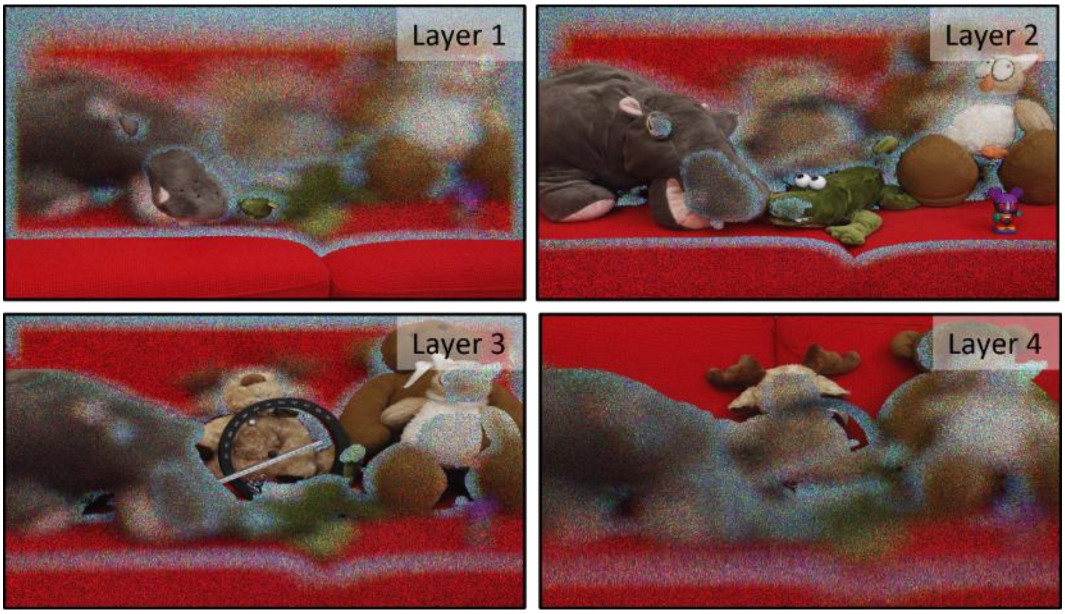

Figure 6.6. Simulated amplitude distributions of the reconstructed image at different planes.

Depth discontinuities contain sharp intensity variations, representing high spatial frequency parts of the image. As we only preserve the phase in the CGH encoding process, the amplitude information is lost. Though the phase information carries the majority of the spatial information, amplitude information loss could result in high-frequency variation during the optimisation. Additionally, the proposed method applies amplitude constraints to the in-focus parts of each depth layer, resulting in a uniform amplitude distribution in the focus area. However, as there are no constraints in the out-of-focus area, the area maintains the initial random profile in the reconstructed plane and, on the other side, reserves the degrees of freedom for the in-focus area during the optimisation process. Noise in the out-of-focus area is left with no constraints and thus cannot be suppressed during the optimisation. Therefore, the most affected region is the depth discontinuities between the optimised and the unoptimised regions of the focal stacked reconstructed images. Moreover, the in-focus and defocused areas of the image are distinctly separated by binary masks, which may exacerbate image quality issues at depth discontinuities. This abrupt separation can introduce edge-enhancing artefacts and amplify noise, leading to degraded image quality at the boundary between the two areas.

Quantitatively, we use SSIM MSE and PSNR metrics to evaluate the reconstructed image quality only on the focus-stacked images. We plot the average values of 10 runs, showing how the MSE loss changes with each iteration. Figure 6.7(a) shows the optimisation curve and the calculation time during the hologram optimisation process. We demonstrate that this gradient descent-based 3D CGH optimisation converges exponentially with a few tens of seconds for a 4-layer RGB image in 500 iterations with RGB channels. As this method is an iterative approach and optimises CGH over multiple reconstruction planes, it is hard to compute an optimised 3D CGH in real-time. However, recent research [43–45] demonstrated that neural network based approaches could potentially be used for real-time phase hologram generation. The neural network-based method explicitly separates the optimisation process into the training and inference phases and could potentially compute the 3D phase CGH during the inference phase in real-time. Figure 6.7(b) demonstrates the optimisation details of PSNR and SSIM metrics during the CGH optimisation. The higher PSNR and SSIM mean the reconstruction result is closer to the target all-in-focus image. The proposed method confirms the improvements over iterations for all-in-focus reconstructed images.

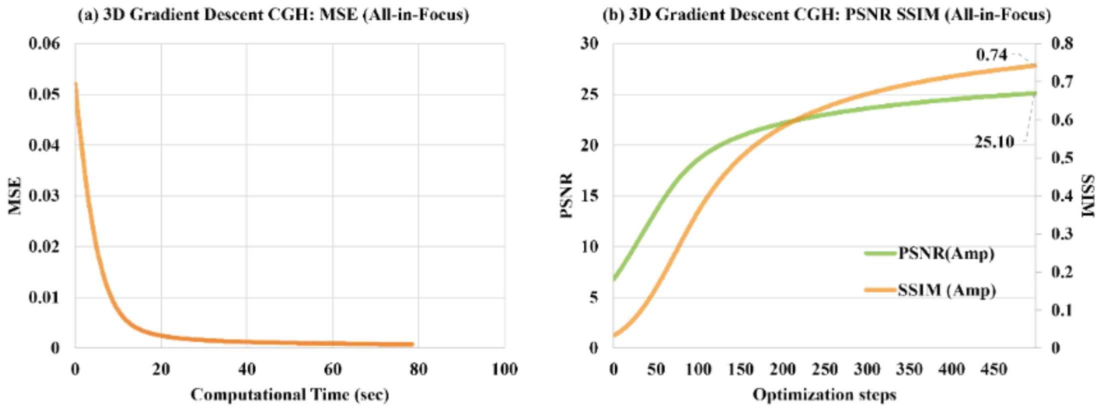

Figure 6.7. (a) The optimisation curve and the calculation time during the hologram optimisation process. (b) PSNR and SSIM metrics in the CGH optimisation.

6.2.4 Optical Validation

We use the monochrome holographic prototype introduced in the last chapter to demonstrate the realisation of 3D holography. We only use the green colour channel holograms for demonstration, but the system can be upgraded to colour displays by sequentially displaying each colour channel of CGHs with RGB laser illumination. The system includes a 532nm laser source and an SLM with 8-bit depth, a $6.4\ \mu m$ pixel pitch and a 1920×1080 resolution. The SLM modulates the beam after it has been expanded and collimated by a 4-f optical system. A polariser is inserted to match the polarisation direction of the SLM. We update the relay optics by using a Canon 35 mm f/2 EF lens with a Canon EF 50mm f/1.4 EF lens to photograph and record video of the modulated beams. The image sensor is a Cannon 6D camera to photograph the reconstructed images. We mount the camera on a linear translation stage to capture the 3D volumes at different distances. Figure 6.8 shows the experimentally captured results of several multiplane 3D scenes focused on various distances.

Apart from algorithm-induced noise in replay planes, other reasons lead to noise and poor image quality in the physical holographic display systems, including coherent laser speckle, quantisation errors, fill factor and nonlinearities of the SLM, and optical aberration during the beam propagation. The experimental implementation shows reasonably good overall quality for in-focus regions of the scenes, although the speckle at out-of-focus regions is quite significant due to the unconstrained behaviour of out-of-focus regions during the optimisation.

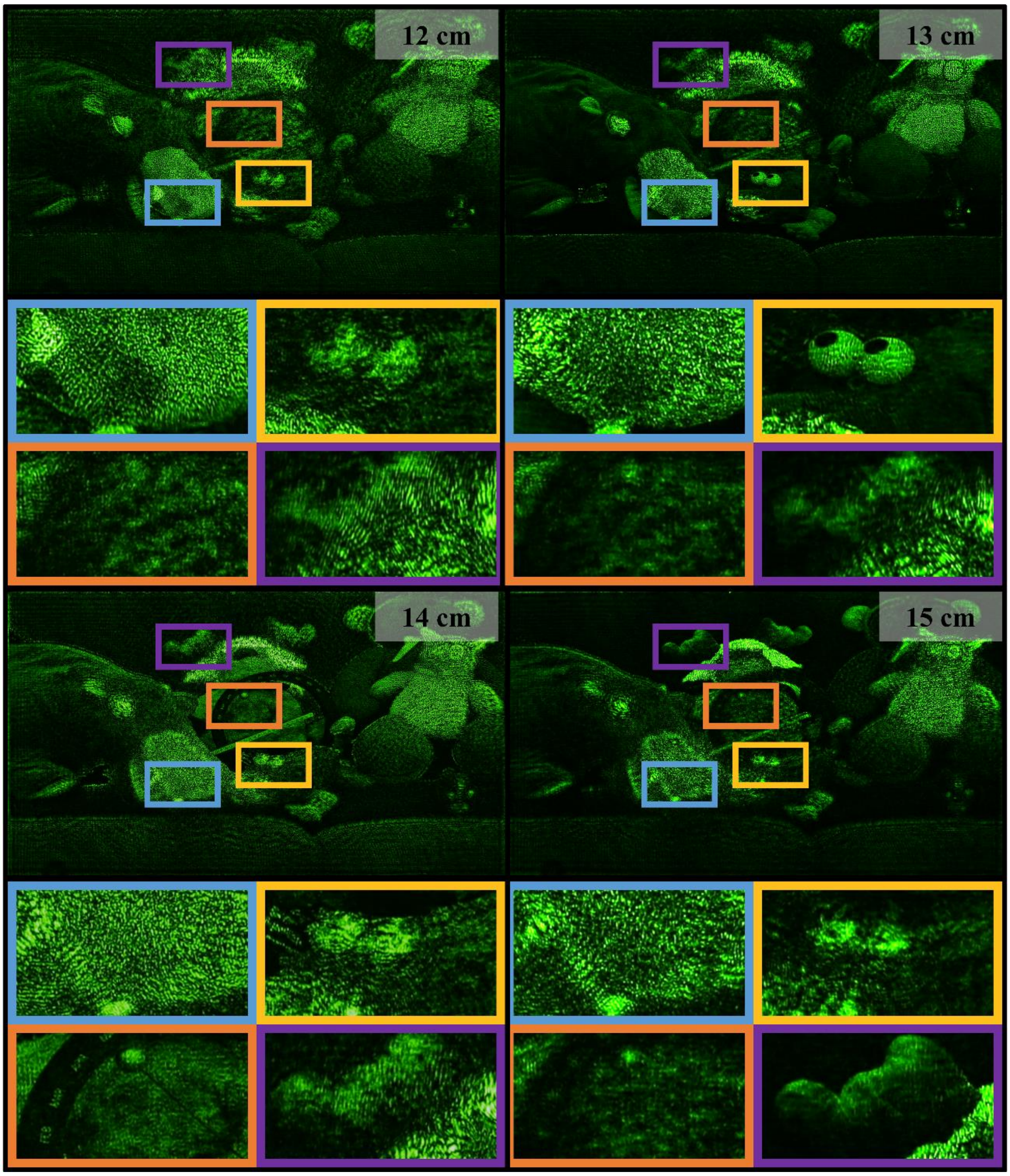

Figure 6.8. In-focus and out-of-focus visualisation of images reconstructed at four depths using naïve gradient descent method for 3D CGH optimisation. The scene is sliced into four layers and placed at 12cm, 13cm, 14cm and 15cm. Rectangular boxes with different colours highlight image patches where the scene target distance matches the focused distance.

6.3 Incoherent Imaging Module

As demonstrated above, though the gradient descent method can achieve good image quality on the focal-stacked images, there are still barriers to achieving realistic visual quality in the defocused areas at depth-varying regions. The current 3D CGH optimisation method only forces the focal-stacked reconstructed images at different planes to match the target images, leaving the out-of-focus area unconstrained. The ambiguity and complexity of optimising image quality in the out-of-focus area remain problematic. On the other hand, the depth of field generated by the CGH relies on the coherent imaging model, creating different defocus behaviour compared with incoherent illumination in real scenes that are more familiar to human perception.

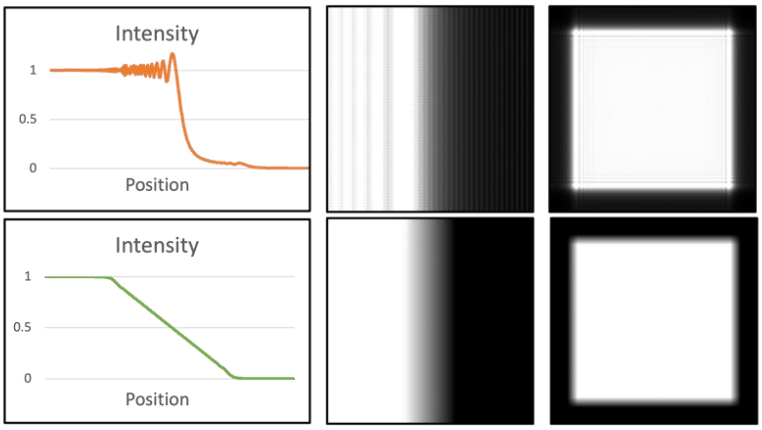

Figure 6.9. Simulated reconstruction for defocusing a rectangle under coherent (first row) and incoherent (second row) illuminating conditions, reproduced from [187].

The response of an incoherent system to a sharp edge is drastically different from that of a coherent system. Figure 6.9 is the simulated reconstruction for defocusing a rectangle under two different illuminating conditions, demonstrating the visual difference in defocus blur, especially at the sharp edges of the image [187]. We can see that there exhibits a rather pronounced ringing effect in the coherent system with steep discontinuities. In contrast, the incoherent system has a significantly smooth drop at the edges. The cut-off frequency of diffracted coherent-system aperture was elucidated by the experiments and wave theories in [187–189] and [5] in section 7.5.3. The limiting coherent-system aperture leads to a sharp cut-off for a coherently illuminated lens.

Therefore, this section proposes an incoherent rendering model using depth-dependent point spread functions (PSFs) to generate natural defocused blur target images. These target images can provide natural defocus blur that accounts for incoherent illumination and offer the possibility to mitigate the defocus visual quality limitation under the constraint of the coherent holographic propagation for multiplane 3D CGH optimisation.

Recent advancements in computational imaging enable coded phase apertures to encode information for monocular depth estimation [190–193]. The coded phase aperture can generate depth-dependent 3D PSFs, which can accurately realise defocused images using image formation models. As the defocus effect depends on the complex transmissive function of the aperture, we can directly simulate the aperture to control depth-dependent incoherent PSFs. The generated PSFs with occlusion-aware image formation models can generate realistic incoherent defocused images. The defocused images generated from incoherent imaging can be used in 3D CGH generation with coherent light sources to improve image quality, fundamentally tackling the coherent CGH optimisation problem.

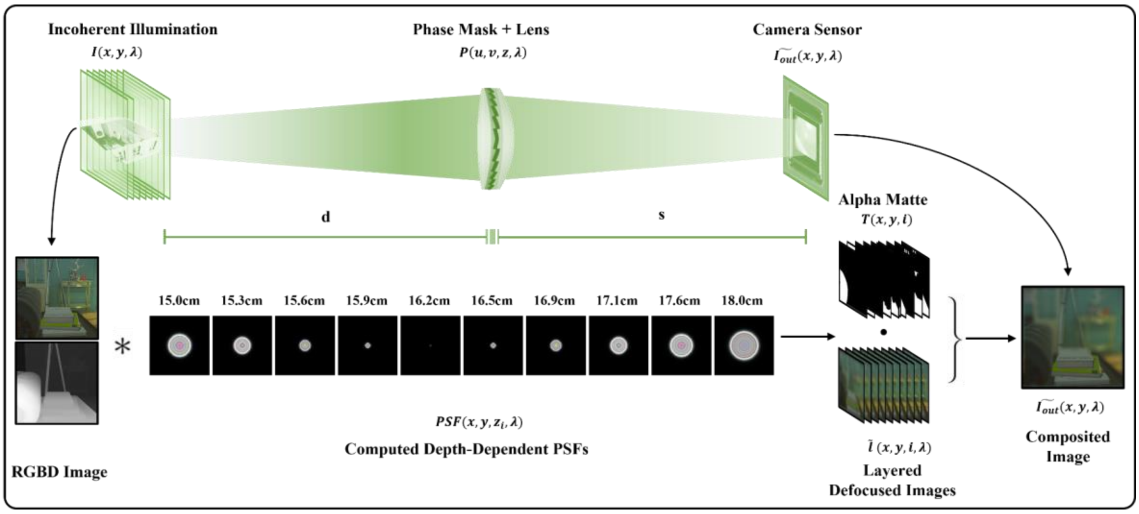

Figure 6.10. A simple incoherent imaging system capable of reproducing 3D scenes at different focal planes.

6.3.1 Method Description

We consider a simple incoherent imaging system capable of reproducing 3D scenes at different focal planes shown in Figure 6.10. This system takes an all-in-focus RGB image with its associated depth map as input and produces reconstruction images with physically natural defocus blur at each plane as output. The system consists of a camera with an imaging lens focusing the scene on the camera. A customised phase mask can be inserted into its aperture plane to control the depth-dependent PSFs of the imaging system. The input RGBD image representing the 3D scene is sliced into multiple layers according to the object distances. Layers are then convoluted with the PSFs to generate defocused images at different distances. The final captured image is the composition of these convoluted layered defocused images. The major components to be simulated in the suggested system are 1) PSFs, which are dependent on the wavelength of the light and depth, and 2) an image formation model that correctly renders the defocus blur images given an RGBD image as input.

6.3.2 Point Spread Function

Based on the Fourier optics [5], for a diffraction-limited imaging system, the image shown on the camera is the convolution of the original image with a PSF. In the incoherent case, the PSF is proportional to the squared magnitude of the Fourier transform of the pupil function in the Cartesian coordinate:

$$\begin{aligned} PSF(x,y,z,\lambda) &= \left|\frac{1}{\lambda s}\iint_{-\infty}^{\infty} P(u,v,z,\lambda)e^{-\frac{i2\pi}{\lambda s}(ux+vy)}dudv\right|^2 \\ &= \left|\frac{1}{\lambda s}\iint_{-\infty}^{\infty} P(u,v,z,\lambda)e^{-i2\pi(uf_x+vf_y)}dudv\right|^2 = \left|\frac{1}{\lambda s}\mathcal{F}\{P(u,v,\lambda)\}\right|^2, \end{aligned} \tag{6.6}$$

where the 2D spatial coordinates are defined as (u,v) and (x,y) at the aperture plane and the camera sensor planes, respectively. The distance between the lens and the camera sensor is s, and the wavelength is λ. The pupil function $P(u,v,\lambda)$, also called the aperture function, is a complex-amplitude function describing the relative amplitude and phase change of the incident light upon transmission through the optical imaging system on the aperture plane. The pupil function can be defined as a complex transmittance function consisting of a defocus factor and a complex modulation due to the phase aperture:

$$P(u,v,z,\lambda) = A(u,v,z,\lambda)e^{i\varphi(u,v,z,\lambda)} = D(u,v,z,\lambda)M(u,v,\lambda). \tag{6.7}$$

The defocus factor $D(u,v,z,\lambda)$ models the defocus variation caused by the mismatch between the actual focusing depth d and the focal length f of the lens. The thin lens equation gives the distance relation $\frac{1}{f} = \frac{1}{d} + \frac{1}{s}$. The defocus factor is given by [5,191]:

$$D(u,v,z,\lambda) = \frac{z}{\lambda(u^2+v^2+z^2)}e^{i\frac{2\pi}{\lambda}\left(\sqrt{u^2+v^2+z^2}-\sqrt{u^2+v^2+d^2}\right)}. \tag{6.8}$$

The complex modulation due to the aperture is modelled by the factor $M(u,v,\lambda) = A_M(u,v)e^{i\phi_M(u,v,\lambda)}$. The amplitude $A_M(u,v)$ of the aperture can be regarded as a circ function since there is no amplitude attenuation within the aperture. The phase delay $\phi_M(u, v, \lambda)$ is typically caused by the height map $h(u, v)$ of the coded phase aperture with a refractive index $n(\lambda)$:

$$\phi^M(u, v, \lambda) = \frac{2\pi}{\lambda}(n(\lambda) - n_{air})h(u, v), \tag{6.9}$$

where n_{air} is the reflective index of air. However, in our case, we can consider the refractive index as a constant and regard the height map as the surface profile of an imaging lens:

$$\phi^M(u, v, \lambda) = -\frac{\pi}{\lambda f}(u^2 + v^2). \tag{6.10}$$

Since both the defocus factor and the aperture modulation are circular symmetric functions, the calculation of the PSF can also be achieved by Fourier-Bessel functions, expressing the PSF calculation in polar coordinates [191,193]. This alternative expression is especially helpful for learning the height map of the coded phase aperture for depth estimation to save computational memory and reduce the complexity of the optimisation process.

6.3.3 Image Formation Model with Natural Defocus Blur

We can then reproduce the captured image of a 3D scene on the camera sensor with these simulated PSFs. As demonstrated in the 3D CGH data preparation process, the 3D scene is represented as an RGBD image and sliced into multiple discrete layers. The RGB image is decomposed into a set of $i = 1 \dots N$ discrete depth layers, with each layered image as $I_{in}(x, y, i, \lambda)$ for a colour channel λ. To simulate the blur effect due to the depth of field, we convolve each layer with its corresponding depth-dependent PSF: $\tilde{l}(x, y, i, \lambda) = PSF(x, y, z_i, \lambda) * I_{in}(x, y, i, \lambda)$. The $*$ denotes the 2D convolution operation, and the tilde denotes the result from incoherent imaging. The final image is a combination of the in-focus and blurred images. The standard image formation model directly uses a simple linear convolution of the PSF with corresponding layered images:

$$\widetilde{I_{out}}(x, y, \lambda) = \sum_{i=1}^{N} \tilde{l}(x, y, i, \lambda) + \eta, \tag{6.11}$$

where $\widetilde{I_{out}}(x, y, \lambda)$ is the captured image on the camera for a single wavelength λ. The additive noise η is typically simulated as Gaussian noise to mimic the noise during the capture. The standard linear model can correctly reproduce the defocus effect for most regions. However, the image formation model is invalid near the depth discontinuities since the model does not consider occlusion [192].

The conventional linear model directly cascades blurred images together. When adding all blurred images, background layers are partly occluded near the borders of foreground layers. Therefore, the nonlinear differentiable image formation model has been proposed to overcome the inaccurate defocus blur at depth discontinuities with a marginal expanse of computational load [191]. The nonlinear differentiable image formation model combines the alpha compositing technique with the PSF to generate a more realistic defocus at depth boundaries.

Rather than directly cascading blurred image layers, the nonlinear image formation model generates an alpha channel for soft-edge image composition [194–196]. The alpha channel can be used as a matte to control the image composition of generated layered images. The Alpha channel is an 8-bit grayscale channel to record the transparency information in the image, defining transparent and opaque areas. An alpha of 0, represented as black, indicates full transparency, and it is zero for fully occluded points. Fractions are represented in grey, corresponding to partial occlusion. Calculating the alpha matte of different layers is possible by using PSFs with binary masks that are obtained from the quantised depth layers:

$$T(x, y, i) = \prod_{i=i'+1}^{N} [1 - BM(x, y, i) * PSF(x, y, z_i, \lambda)]. \qquad 6.12$$

Intuitively, we calculate the *cumulative occlusion* from defocused layers with convoluted binary masks rather than directly cascading the entire convoluted blurred images. The alpha matte for each layer calculates the weight contribution from each layer to composite the rendered image on the camera sensor. These weights given by the alpha mattes can model the thin lens geometry. Compared with binary masks used in the linear model, the alpha matte is not knife-sharp, as shown in Figure 6.11. We then obtain the rendered image by alpha compositing in back-to-front order with the algebra matte to model the imaging system accurately:

$$\widetilde{I_{out}}(x,y,\lambda) = \sum_{i=1}^{N} \tilde{l}(x,y,i,\lambda) \cdot T(x,y,i) + \eta. \tag{6.13}$$

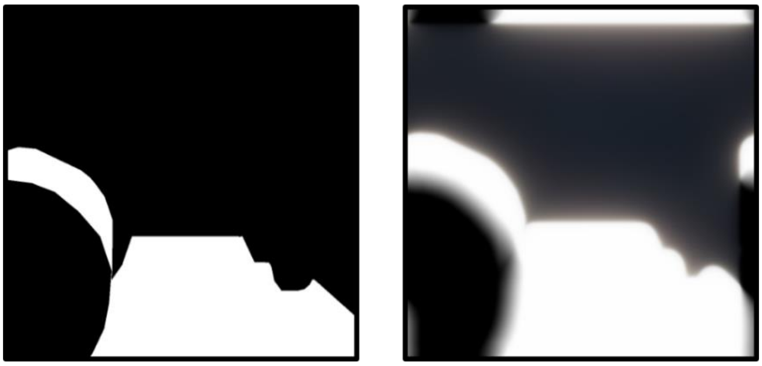

Figure 6.11. The Binary mask in the linear image formation model and the alpha matte used in the nonlinear image formation model.

We further apply normalisation to the convoluted image $\tilde{l}(x,y,i,\lambda)$ and the alpha matte $T(x,y,i)$ for each layer to compensate for the energy reduction during convolution with the PSFs. Figure 6.12 shows that compared to the standard linear models, the nonlinear image formation model generates a more natural-looking defocused picture with fewer ringing effects in the vicinity of depth discontinuities from RGBD input.

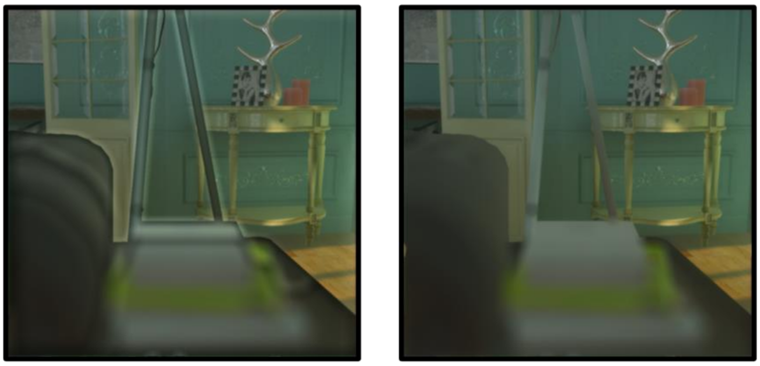

Figure 6.12. The generated images are from the standard linear image formation models and the nonlinear image formation model.

6.4 Natural Defocus CGH

6.4.1 Method Description

The realisation of incoherent natural defocus blur in 3D scenes motivates us to train our CGH generation model using coherent illumination for both in-focus and out-of-focus regions. The most straightforward method to integrate the rendered images with the CGH training process is to reformulate the loss function to directly use these incoherent defocused images as target amplitudes for coherent CGH optimisation. As such, the 3D CGH optimisation process using rendered incoherent images can be shown in Figure 6.13. The schematic diagram comprises a coherent CGH optimisation module and an incoherent image rendering module.

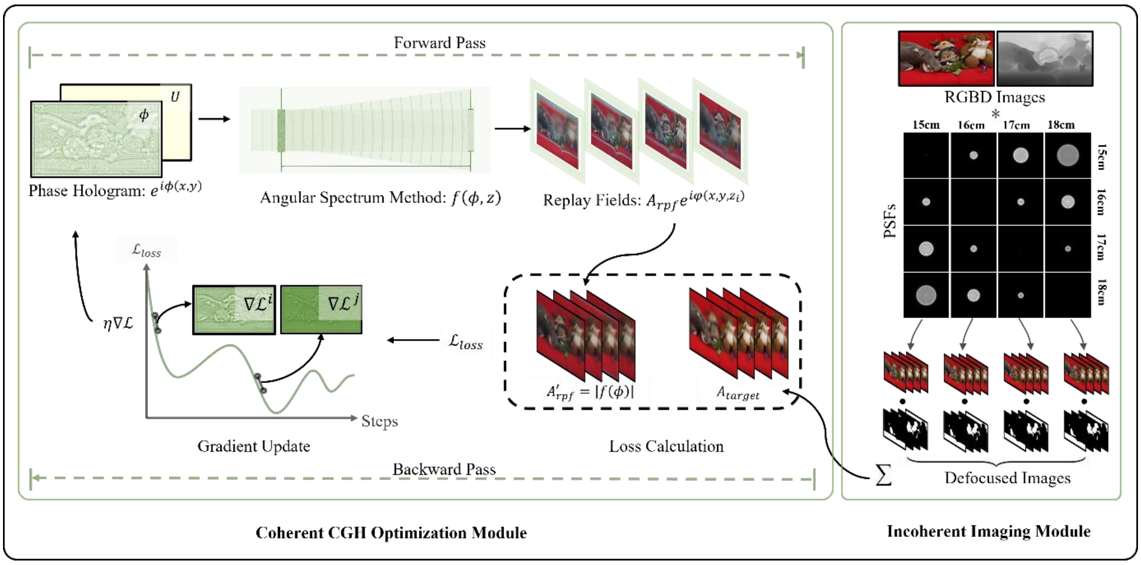

Figure 6.13. The gradient descent method for the 3D CGH optimisation using an incoherent imaging module.

In the forward pass, the coherent CGH optimisation module propagates the initial phase hologram $e^{i\phi(\mu,\nu)}$ by the angular spectrum method at a set of distances z_i with a coherent planar wave. The propagated wave is reconstructed at multiple replay planes. In the backward pass, the coherent CGH optimisation module computes the gradient of the loss function to update the next estimate phase hologram.

The incoherent module calculates the wavelength and depth-dependent PSFs and convolutes the PSFs with the input RGBD images using the nonlinear image formation model. The convolution produces reconstructed images with physically natural defocus blur for out-of-focus regions. We iterate this process to simulate incoherent reconstructed images at different depths so that the rendered images can be used as target images for coherent CGH optimisation. Therefore, the reformulated MSE loss function using rendered incoherent images for evaluation can be expressed as:

$$\mathcal{L}_{MSE} = \frac{1}{mn}\sum_{m,n}\left[\sum_{i}^{N} m_i|f(\phi, z_i)| - \widetilde{A_\iota}\right]^2, \qquad 6.14$$

6.4.2 Result and Discussion

Simulation Validation

The proposed method is implemented using PyTorch on an NVIDIA RTX 3060 GPU with Adam optimiser. Most parameters remain the same as the naïve 3D CGH generation introduced in section 6.2. We optimise the CGH for 1000 iterations with a 0.02 learning rate and exponential decay rates of $\beta1$ = 0.9 and $\beta2$ = 0.999. This modification increases the iteration number and reduces the learning rate, enabling close monitoring of the optimisation process. The defocused images are rendered at the target depth range $z = \{12cm, 13cm, 14cm, 15cm\}$, and the camera is focused accordingly with a pixel pitch at $6.55\mu m$. The simulated aperture has a $35mm$ focal length lens with an f-number of 5.0. Parameters of the incoherent imaging module can be easily modified to suit different viewing conditions, rendering defocused images in various ways. The proposed two-stage CGH optimisation method jointly reproduces 3D holograms with natural defocus blur at different depths.

In our experiments, we first generate the defocused images using the incoherent imaging modules, as shown in Figure 6.14. Compared with the coherent reconstruction, the incoherent reconstruction simultaneously achieves smooth defocus and speckle-free effects. Using the rendered defocused images from the incoherent imaging module, we optimise an initial random phase hologram in RGB channels with the proposed method. Benefiting from the rendered defocused images, our simulation results in Figure 6.15 show that our method provides the most appealing image quality with realistic looking defocus blur.

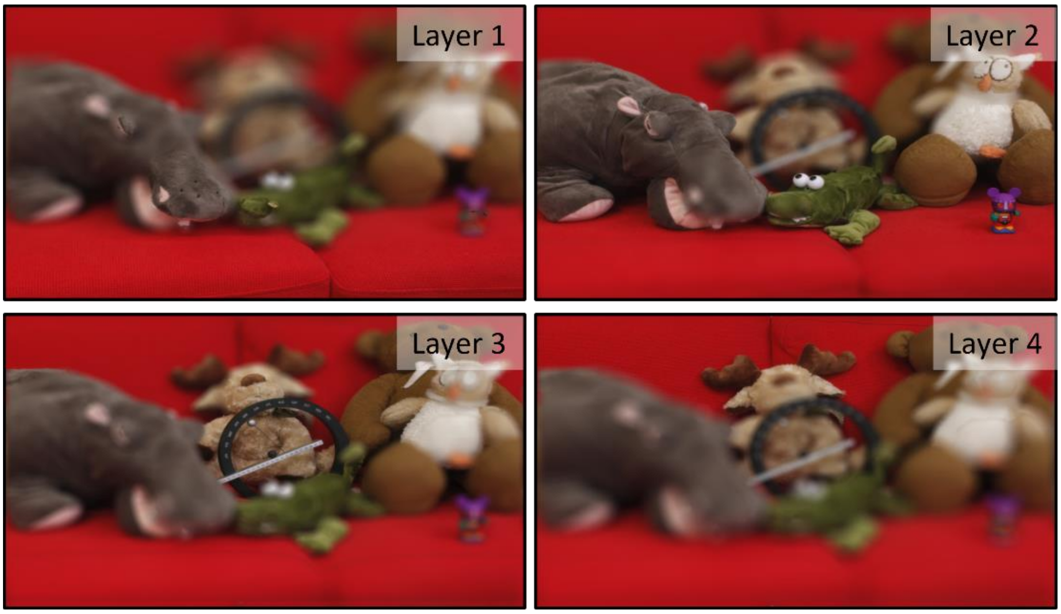

Figure 6.14. Defocused target images generated from the incoherent imaging module show smooth defocus and speckle-free effects.

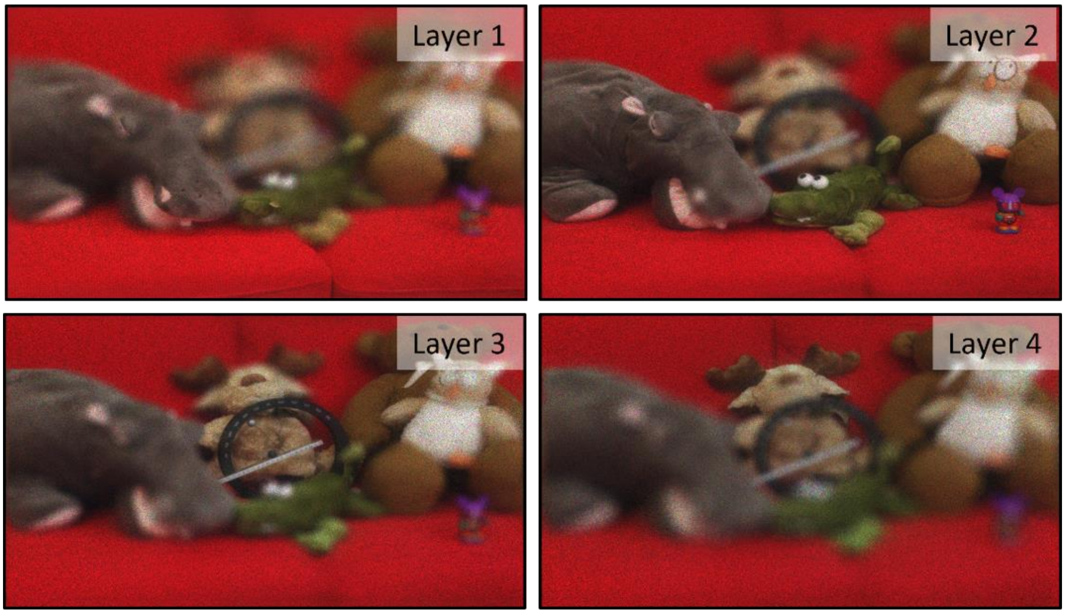

Figure 6.15. Simulated reconstructed image results of CGHs using the incoherent defocused images at targets.

Figure 6.16 compares the simulated reconstructed images by the naïve 3D CGH optimisation described in section 6.2 and the proposed method. The naïve method leads to strong artefacts at the occlusion boundaries and produces incorrectly defocused images. On the other hand, benefiting from the natural defocus estimation of an incoherent imaging module, the proposed method can estimate the depth discontinuities between depth layers and effectively removes the artefacts. This is due to the fact that the incoherent imaging module can seamlessly combine images at different depth layers using alpha composition and render out natural defocused images. Instead of using masks at in-focus regions of reconstructed images, the coherent CGH optimisation module of the proposed natural defocus CGH method avoids combining images at different layers, which could easily lead to the occlusion problem at depth discontinuities.

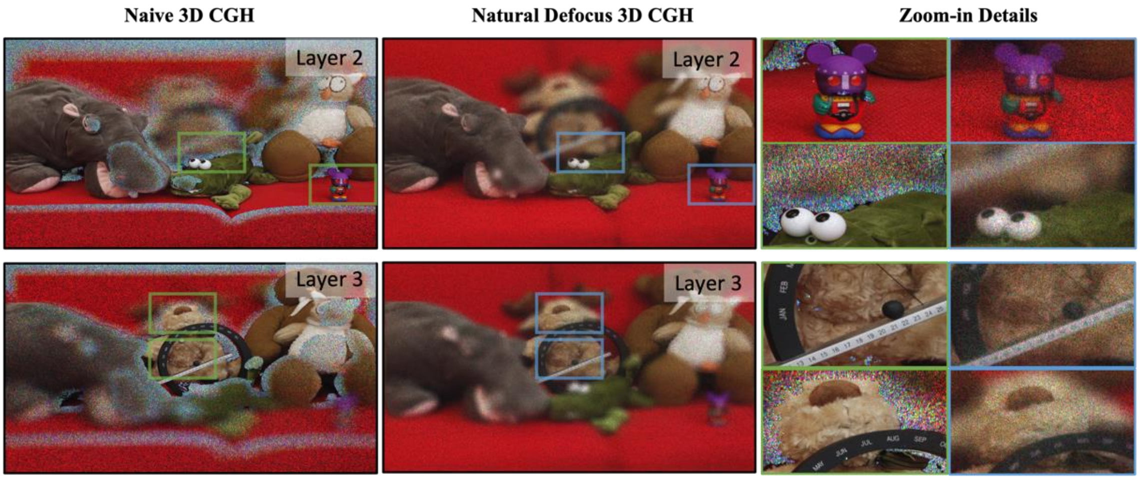

Figure 6.16. Simulated reconstructed image comparison between the naïve 3D CGH method and the proposed natural defocus CGH method.

Optical Validation

We further validate our proposed method experimentally using the same holographic display. We upload the 3D holograms of the above scene to the SLM and photograph the reconstructed images at different replay planes with the Canon camera.

Figure 6.17 shows the experimentally captured results of 4 multiplane 3D scenes, with square boxes indicating image patches where the scene target distance matches the focused distance. The captured results can closely resemble the simulation in Figure 6.15. In contrast to the naïve method, the proposed method generates reconstructed images that are more naturally defocused and mitigate the out-of-focus speckle behaviour with suppressed artefacts at depth discontinuities. As shown in the zoomed-in details, while the proposed natural defocus CGH largely enhances the natural defocus effect at occlusion boundaries, the resolution of the in-focused image is degraded. Specifically, at layer 3, characters on the scale are in-focus for this layer and can be easily identified with the naïve gradient descent method; however, it is difficult to recognise the characters with the suggested method using incoherent target images.

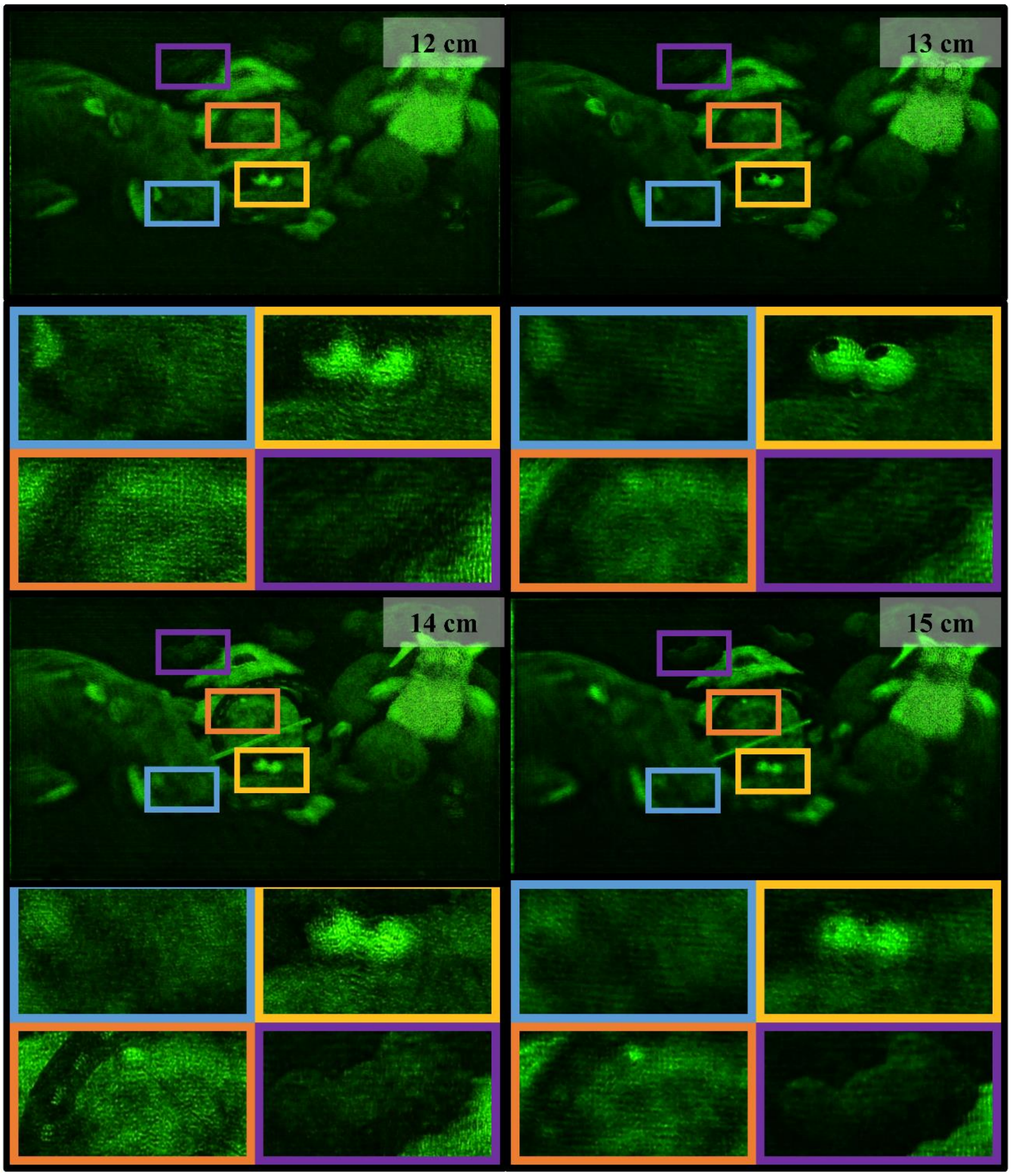

Figure 6.17. In-focus and out-of-focus visualisation of images reconstructed at four depths using Natural Defocus CGH. The experimentally captured results of 4 multiplane 3D scenes, with square boxes indicating image patches where the scene target distance matches the focused distance.

Discussion on Image Quality Degradation of Proposed Method

The image degradation in the in-focus region could result from different optimisation mechanisms. Firstly, the attention loss in the proposed method may result in low image quality and resolution in in-focus regions of the reconstructed image. While both the proposed method and the naïve gradient descent method optimise target images, the latter exclusively focuses on in-focus regions and discards out-of-focus regions during optimisation. Attention loss refers to the reduction of focus and resources devoted to certain parts of an image during optimisation. This can occur when the CGH optimisation process allocates too much attention to the out-of-focus regions of the reconstructed image, at the expense of enhancing the in-focus regions.

To address the attention loss issue in the proposed method, a binary weighting parameter can be incorporated to focus the optimisation process on the relevant areas of the reconstructed image, implicitly introducing visual attention mechanisms. Specifically, the loss function can be reformulated manually by specifying weights between the in-focus and out-of-focus regions of the reconstructed images. Alternatively, soft attention mechanisms can be used in conjunction with the CGH optimisation process, where the weight is learned through forward and backward propagation via gradient descent. This approach can maximise the image quality in the in-focus regions while minimising the loss of image quality in the out-of-focus regions, resulting in an overall improvement in the quality of the reconstructed images.

Secondly, the direct supervision over multiple layered images could potentially overconstrain the optimisation problem, exceeding the available degree of freedom of current SLMs. The number of degrees of freedom of a given phase-only SLM is determined by its resolution and bit depth, which are finite and directly impact the quality of the reconstructed images. Supervision on out-of-focus regions or additional layered images could potentially exceed the available degree of freedom, limiting the performance of the proposed natural defocus CGH algorithm. However, recent research has demonstrated the feasibility of optimising CGH over more than five mutually independent images with limited degradation in overall image quality, as shown in [29]. This research indicates that the degree of freedom of current SLMs may be sufficient to handle multiple layered images with only a limited degradation in the overall image quality (SSIM from 0.8 to 0.72).

A potential solution to overcome the limited degree of freedom issue is to reduce the resolution of the reconstructed image. Previous research has utilised this approach, such as [43], primarily to calibrate the resulting image within a confined region of interest. By reducing the resolution of the reconstructed images, the number of pixels that need to be optimised is decreased, thereby increasing the degree of freedom available for CGH optimisation. However, it is essential to note that this increase in image quality comes at the expense of reducing the image resolution in the optimised region. Therefore, it is necessary to carefully balance the trade-off between image quality and resolution to ensure that the reduced resolution does not significantly affect the overall visual experience of the reconstructed images. An alternative approach to increasing the degree of freedom in CGH optimisation is to use an SLM with a larger resolution. This approach has been demonstrated in a recent study in [197], where a high-resolution SLM with a resolution of 3840×2160 was used to generate full high-definition holograms. However, using a higher-resolution SLM can also significantly increase the computational time and complexity of the optimisation process.

The third reason is that the generated incoherent images could lead to noise in the in-focus regions of the reconstructed image. Since the generated PSFs are band-limited functions, when convoluted with the angular frequencies of the input images, high frequencies of the generated incoherent images could be attenuated. This band limitation means that details in the signal that contains high-frequency information, such as sharp edges or fine textures, can be lost or blurred. As a result, the convolved signal can have reduced sharpness and clarity compared to the original signal.

Several approaches can be adopted to mitigate the noise caused by incoherent images. One option is to use a smaller diameter lens in the incoherent imaging module. The smaller diameter lens can help increase the depth of field and reduce the defocus effect of the incoherent images, resulting in better image quality and reduced noise in the in-focus regions of the reconstructed image. Additionally, generating multiplane images with reduced maximum scene depth is another approach. Limiting the depth of the scene can reduce the range of distances over which the defocus effect occurs, which can lead to sharper and clearer images in the in-focus regions of the reconstructed image. These two methods offer effective solutions to mitigate the noise issue caused by incoherent images and can significantly improve the quality of reconstructed images.

Quantitative Result

To investigate the image quality degradation in the all-in-focus regions of the reconstructed image by the proposed method, we mask each layer of the reconstructed defocused images individually and composite them together to produce the all-in-focus reconstructed image. We plot image quality metrics such as MSE, PSNR and SSIM to check the image quality changes over iterations quantitatively. Though the image quality of all-in-focus regions is increasing over iterations, compared with the result, we can see that the quality of the proposed method evaluated at all-in-focus regions drops significantly, from 0.74 to 0.53 for the SSIM metric and from 25.10 to 20.66 for the PSNR metric, indicating that the proposed method though visually increases the quality of out-of-in-focus regions, lead to quality degradation in all-in-focus regions of the reconstructed images.

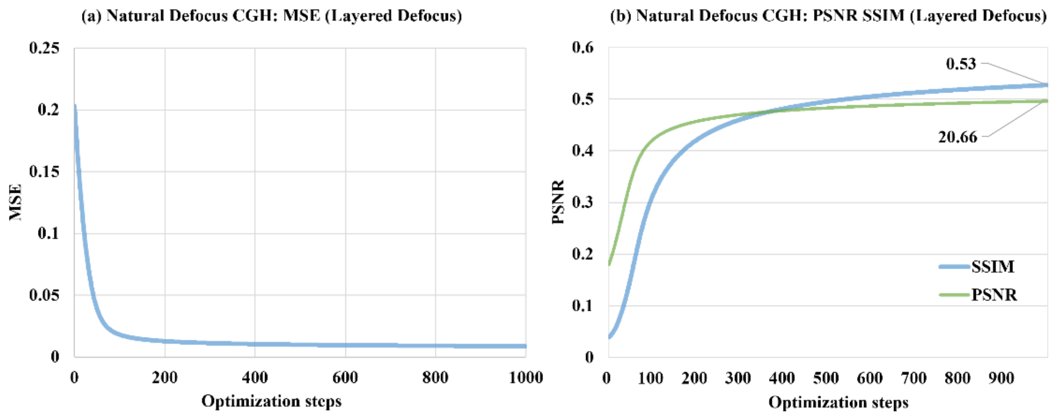

Figure 6.18. (a)The optimisation curve and the calculation time of Natural Defocus CGH. (b) PSNR and SSIM metrics in the Natural Defocus CGH optimisation.

6.4.3 Improvement with Attention Mechanism

We apply the hard attention mechanism in the 3D natural defocus CGH optimisation process, reformulating the loss function to introduce more weight on the in-focus regions over the out-of-focus regions of the reconstructed images. As demonstrated in [187], the weight of the out-of-focus regions is empirically set to $m_0 = 1$, and the in-focus weight $m_1 = 2.1$ for the $L2$ loss function under statistical criterion. We separate different regions of the in-focus and out-of-focus regions of the reconstructed images using the binary mask and apply weights correspondingly:

$$\mathcal{L}_{MSE,total} = m_0 \mathcal{L}_{MSE,in\_focus} + m_1 \mathcal{L}_{MSE,out-of-focus} \qquad 6.15$$

In Figure 6.19, we intuitively validate the effectiveness of the proposed method with an attention mechanism to boost the image quality at all-in-focus regions of the reconstructed images at various distances.

Reconstructing at 14cm, the newly proposed natural defocus CGH method with an attention mechanism can resolve the scale more clearly than the direct Natural Defocus CGH method, owing to the introduced weights. In all results, the proposed method consistently produces less speckle for out-of-focus regions compared to the naïve gradient descent CGH method and has more realistic depth boundaries using rendered defocused images.

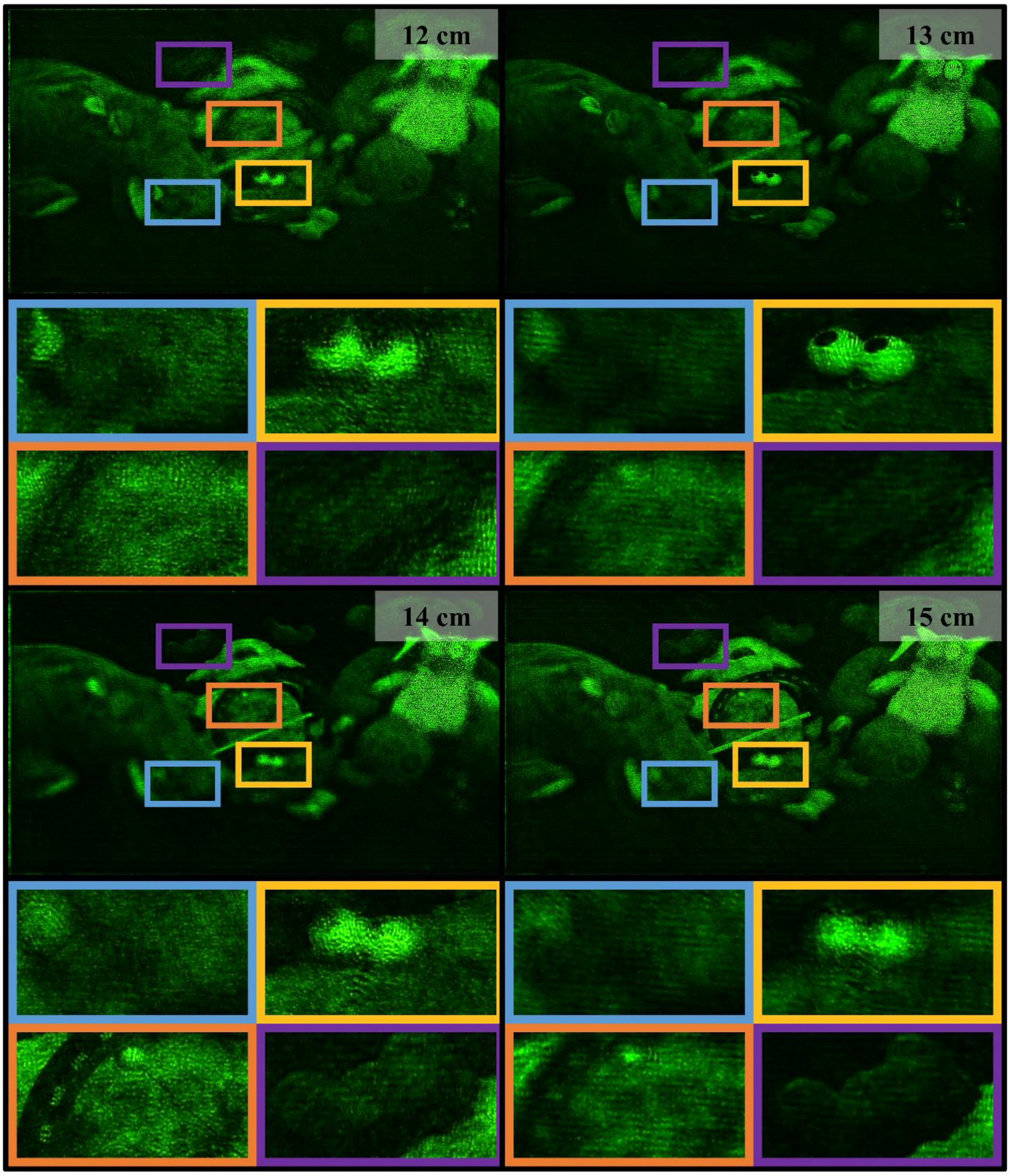

Figure 6.19. In-focus and out-of-focus visualisation of images reconstructed at four depths using Natural Defocus CGH with an attention mechanism. The experimentally captured results of 4 multiplane 3D scenes, with square boxes indicating image patches where the scene target distance matches the focused distance.

6.5 Comparison to Other Work

We further present experimental results of the proposed method and compare it with several existing approaches, including a multiplane stochastic gradient descent optimisation with angular spectrum wave propagation (SGD-ASM) [185], a multiplane 3D model that incorporates an alternating direction method of multipliers (ADMM) solver enforcing piecewise smooth phase constraints of the in-focus multiplane images (ADMM-ASM) [185], and realistic defocus holography (RDH) [187]. To evaluate the performance of these methods, we gather 20 2K RGBD images from diverse publicly available datasets [187,198–200] and optimise the initial phase of the models at five different target planes. Multiplane 3D images are captured and can be optically reconstructed at these target planes within a range of 14cm to 16cm from the SLM. Within this range, we uniformly distribute a total of five planes at equal physical distances. The models are trained using the same procedure outlined in the original paper, and their parameters are set to their original values. Hologram optimisation is conducted for 1000 iterations with a learning rate of 0.02 for all models. The RDH uses a target blur size of 40 with a blur ratio of 5. Experiments are performed using the same optical system, with a 532nm laser source, an 8-bit phase-only SLM with $6.4\mu m$ pitch and 1920×1080 resolution. The reconstructed images are captured using an EOS6D camera attached with an f1.4, 50mm EF lens.

The proposed improved method, natural defocus holography with attention mechanism, referred to as NDH, is then compared using experimental results with our implementation of the conventional multiplane SGD-ASM, the multiplane ADMM-ASM and RDH. Figure 6.20 shows experimentally captured results of two multiplane 3D scenes, focused on a near, an intermediate, and a far distance. In-focus regions of the images in various areas are enlarged using colour-coded rectangular boxes to illustrate image details.

The results demonstrate that all methods are capable of reconstructing scenes at different distances. However, due to its unconstrained out-of-focus behaviour, the SGD-ASM approach exhibits more significant noise in the out-of-focus image regions. This unconstrained behaviour produces significant speckles in the out-of-focus parts of the image as the wave field propagates away from the constrained in-focus areas. The ADMM-ASM method improves the out-of-focus speckle by imposing constraints on the phase of in-focus regions of reconstructed images, implicitly to smooth the in-focus phase and mitigate the out-of-focus speckle. Although the out-of-focus speckle behaviour of ADMM-ASM is superior to that of SGD-ASM, the resulting images still lack the natural blur present in the defocused regions of the reconstructed scene. NDH and RDH use incoherent defocused images as targets, leading to a visually more pleasing defocused appearance. Both proposed models can adequately achieve excellent image quality with significantly reduced speckle and better image quality in both in-focus and out-of-focus parts of reconstructed images. However, because of using the depth-dependent PSFs with the alpha channel blending technique, the proposed method exhibits better visual quality.

Figure 6.21 presents experimentally captured results of a multiplane 3D scene reconstructed at all planes, with additional results shown in Figure 6.22, Figure 6.23 and Figure 6.24. Specifically, the top row shows a composite image, which combines only the in-focus parts of all captured images alongside images captured at various distances for one scene in corresponding rows. The columns compare the proposed method with others, including SGD-ASM, ADMM-ASM, and RDH. The last column provides ground truth images generated using the proposed incoherent imaging module for qualitative and quantitative (PSNR/SSIM in boxes) comparison. Although the first two methods, which disregard optimisation of the out-of-focus regions, may yield higher image quality in the in-focus areas depicted in the composite all-in-focus images, the reconstructed image at various planes shows incorrect defocus blur. While RDH can produce smooth defocused images, the proposed method can provide a more pleasing defocus appearance with higher image quality in the in-focus regions.

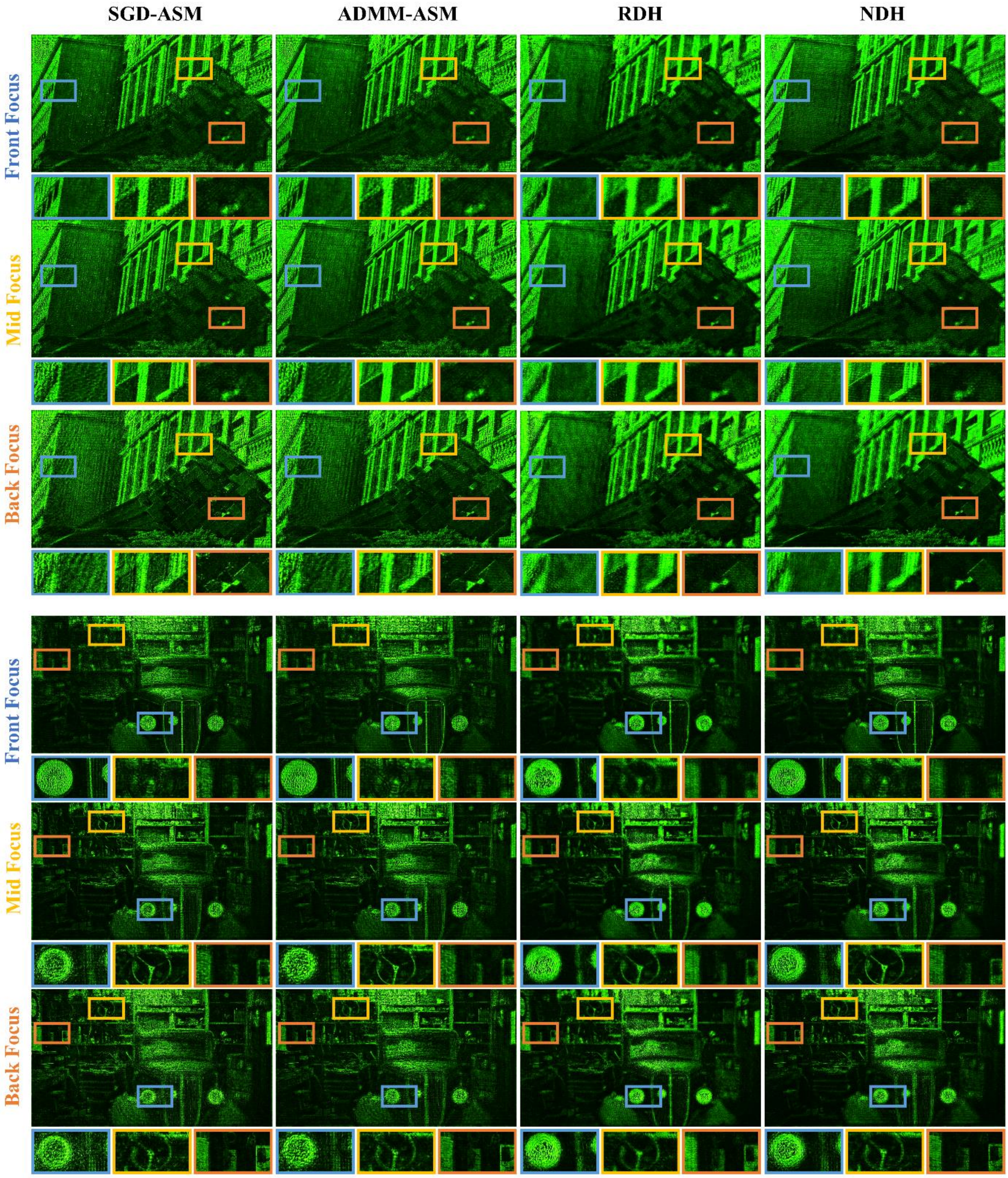

Figure 6.20. Comparison of 3D CGH methods with experimental captured results. Methods include SGD-ASM, shown on the left; ADMM-ASM, shown in the centre left; RDH and the proposed method (NDH), shown in the centre right and right, respectively. While the first two methods constrain only the in-focus areas resulting in good image quality in those regions, they produce significant out-of-focus speckle artefacts. On the other hand, the RDH and NDH methods use incoherent defocused images to smooth out-of-focus noise in the image.

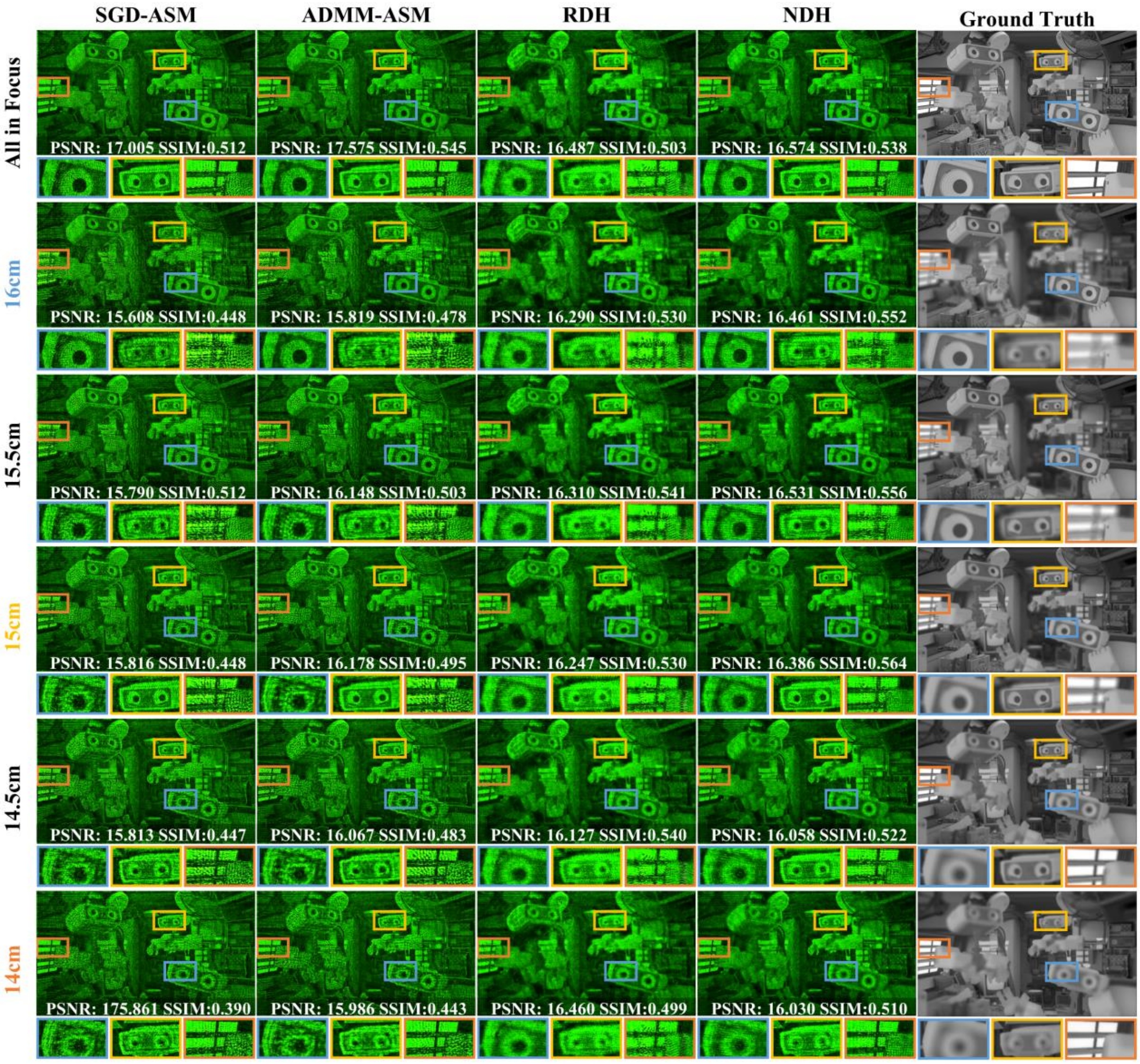

Figure 6.21. The experimentally captured results of a multiplane 3D scene reconstructed at all planes with the all-in-focus images in the first row. The ground truth images are generated using the incoherent module and the all-in-focus in-focus parts of several different scenes. The all-in-focus image is a composite image that combines only the in-focus parts from all planes. PSNR/SSIM metrics indicates quantitative comparison at various distance of each method to the ground truth.

Figure 6.22. Additional experimentally captured images of SGD-ASM, ADMM-ASM, RDH and NDH with zoomed-in details at different focal planes. An outdoor scene with characters and ‘primitives’ from [187].

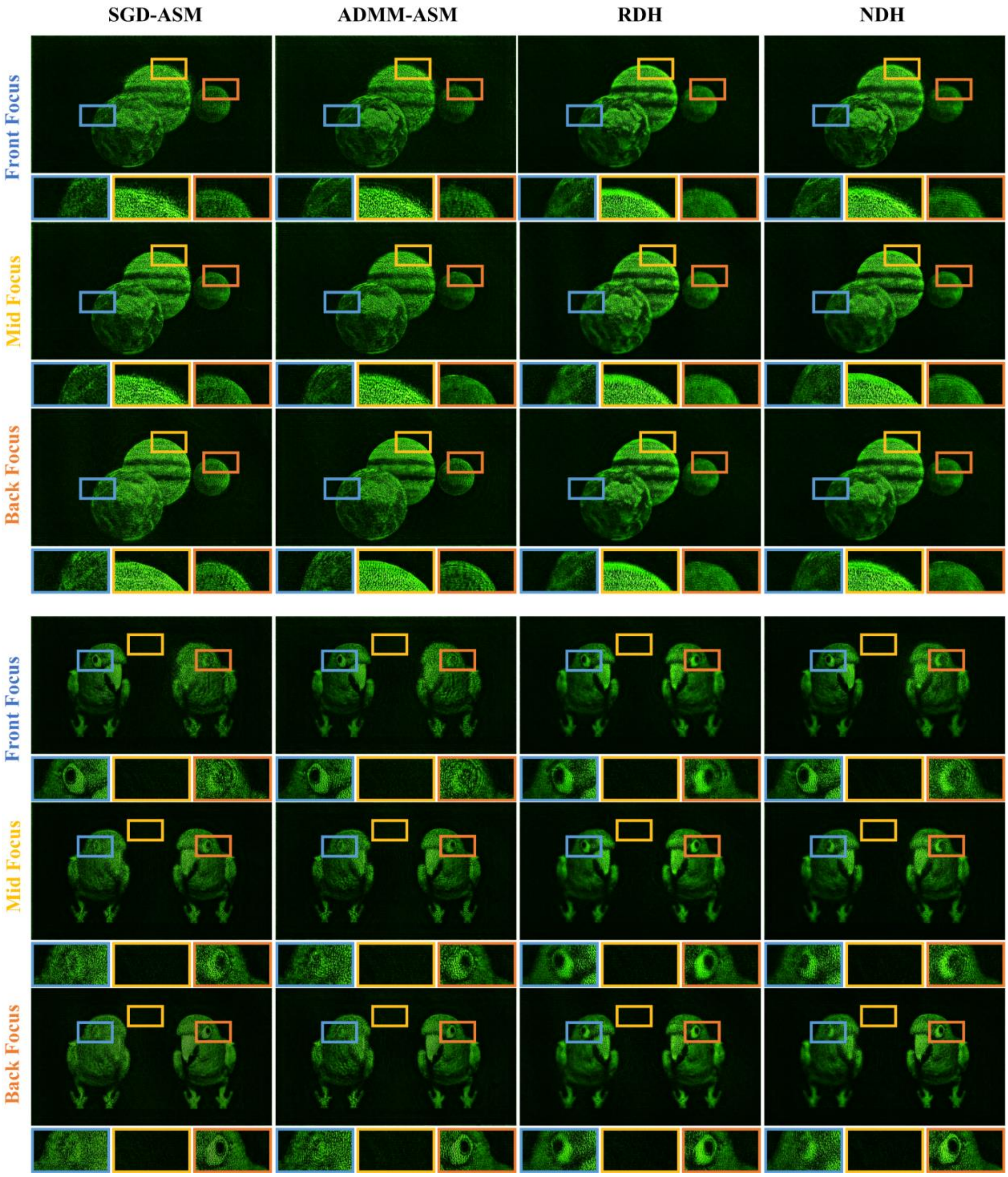

Figure 6.23. Additional experimentally captured images of SGD-ASM, ADMM-ASM, RDH and NDH with zoomed-in details at different focal planes. ‘gezegenler’ and ‘birds’ from [187].

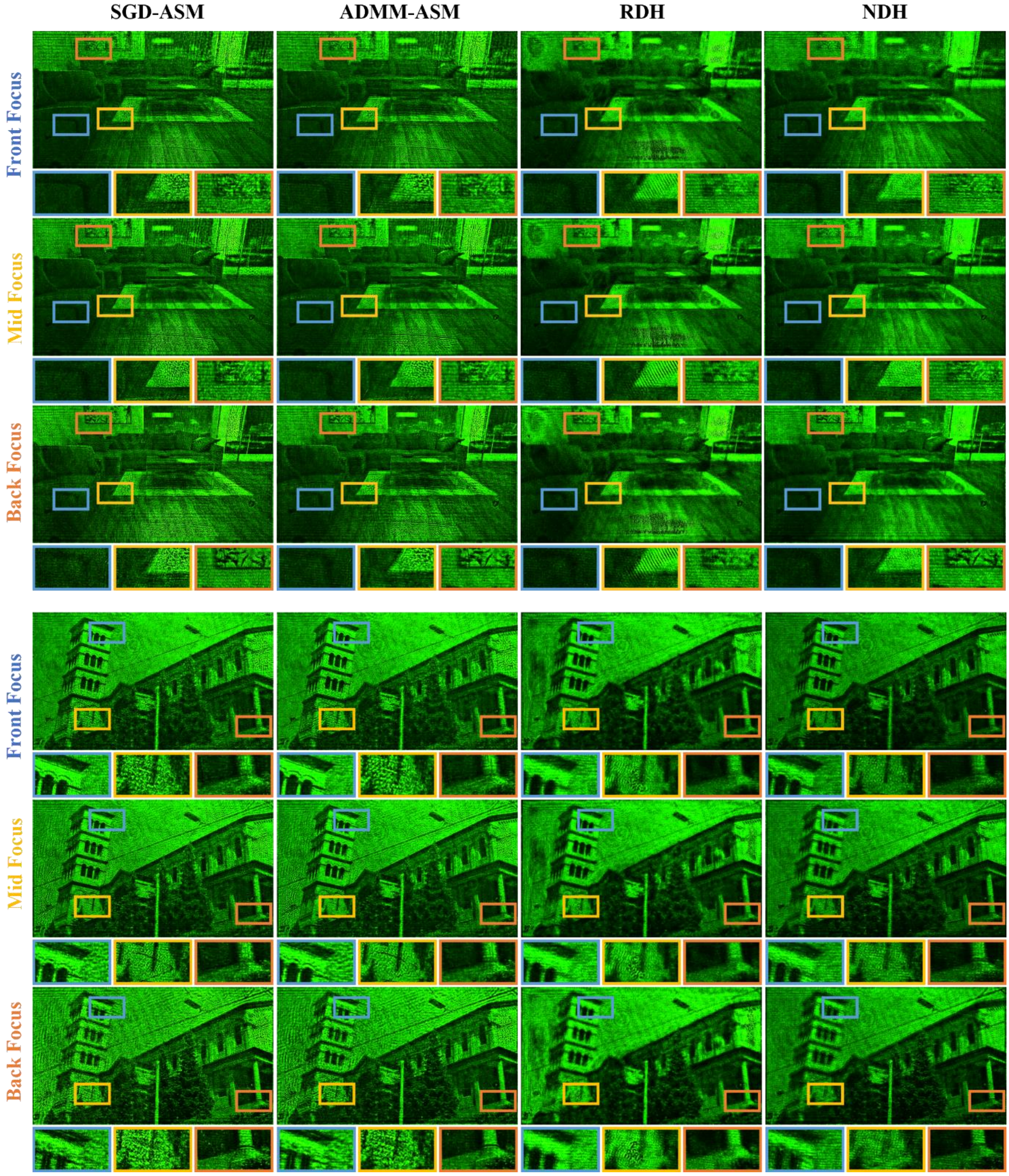

Figure 6.24. Additional experimentally captured images of SGD-ASM, ADMM-ASM, RDH and NDH with zoomed-in details at different focal planes. A living room scene from [198] and a ‘statue’ from [199].

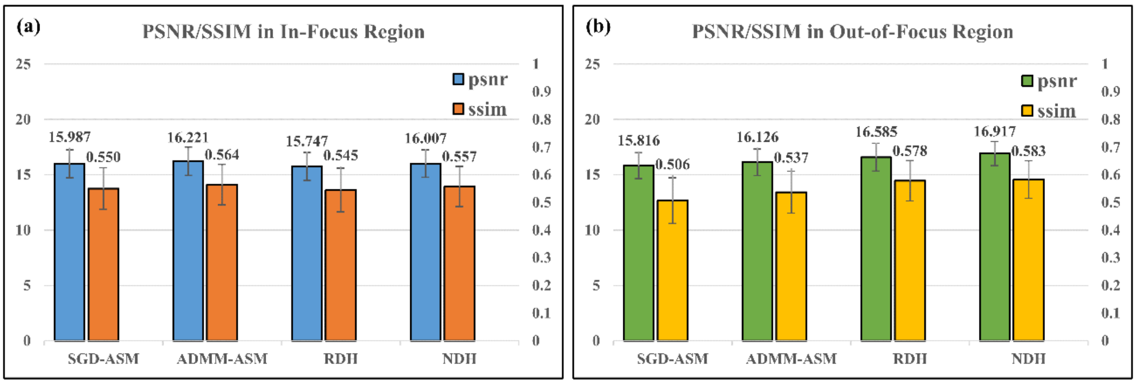

Figure 6.25. Quantitative performance of experimental captured images from SGD-ASM, ADMM-ASM, and RDH methods in in-focus and out-of-focus regions based on PSNR/SSIM metrics.

In Figure 6.25, we present quantitative assessments of the proposed method as well as other methods, including SGD-ASM, ADMM-ASM, and RDH, based on PSNR/SSIM metrics in both in-focus and out-of-focus regions using experimentally captures from the 20-image dataset. Specifically, the metrics for in-focus regions are evaluated on the all-in-focus images and compared to the target image for each experimental reconstruction using each method. The metrics for out-of-focus regions are evaluated on each depth plane and compared to its ground truth defocused image, and then averaged across all five planes to indicate the overall image quality in the defocused regions. Finally, all metrics are averaged across the image dataset to indicate the mean values of the comparison results, thereby reflecting both the smoothness of the defocus blur and the sharpness of the focused object.

Based on the PSNR/SSIM values in both in-focus and out-of-focus regions, the proposed method NDH outperforms the other three methods (SGD-ASM, ADMM-ASM, and RDH) in terms of image quality. Specifically, NDH achieved higher PSNR and SSIM values (16.92 and 0.58, respectively) in the out-of-focus regions compared to the other three methods, indicating better performance in reconstructing the defocus blur. NDH also maintain reasonably good PSNR and SSIM values (16.01 and 0.56, respectively) in the in-focus regions, demonstrating its ability to capture the sharpness of the focused object. These results also suggest that explicitly using incoherent defocused images to smooth out-of-focus part image noise, as in the RDH and NDH methods, can lead to better image quality in defocused regions. Overall, these qualitative and quantitative experiments demonstrate that the proposed method maintains good image quality in in-focus regions while exhibiting better image quality for out-of-focus regions, indicating its superiority over the other methods.

Note that this work differs from recent works on 3D holography by Shi et al. [44] and Choi et al. [185]. Both works train neural networks based on large datasets to approximate the analytic wave propagation model and constrain the phase values of the replay field for better image quality. The reproduced images are compared only with in-focus images using binary masks, and the out-of-focus regions are implicitly optimised by employing the smooth phase constraint. Instead of constraining the phase, this work explicitly considers the out-of-focus regions, employing an incoherent imaging module to simulate the natural defocus in the replay field. The proposed method does not employ neural networks for approximation and imposes phase constraints for phase regularisation, potentially alleviating the smooth phase problem discussed later. Similar work from Kavakli et al. [187] also aims to improve the image quality of out-of-focus regions by applying Gaussian kernels to simulate the defocus blur with additional phase constraints. This work, however, simulates the depth-dependent point spread functions to generate physic-based incoherent defocus blur. The proposed method combines the alpha-matte-based image formation model to render natural defocused images as target images.

6.6 Further Discussion

6.6.1 Further Image Quality Improvement

The experimental results presented in this study incorporate the use of captured images that have undergone further enhancement to achieve a higher image quality, as shown in Figure 6.26. The implementation details of the imaging techniques used are carefully examined from the literature, specifically from Shi et al. [45], Kavaklı et al. [201], and Choi et al. [202], to ensure the high quality of the results.

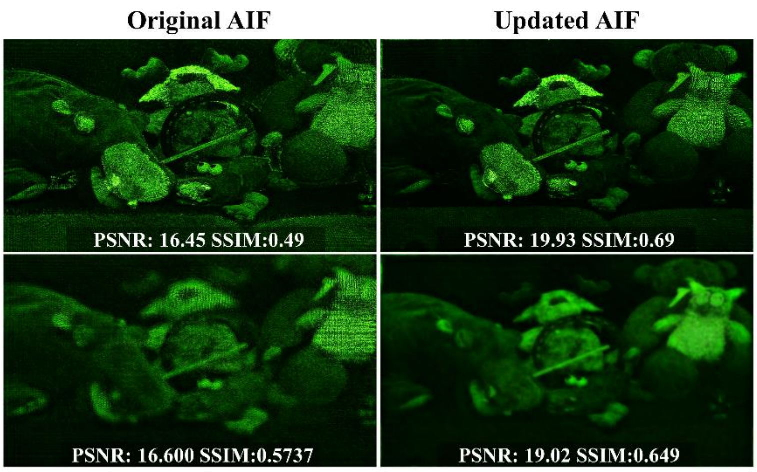

Figure 6.26. Comparison of all-in-focus (AIF) images of the original experiment and the updated experiments.

Improvement in image quality is achieved through the precise alignment of the laser source polarisation to the SLM and the insertion of an additional polariser to reduce the zeroth and higher-order diffraction artefacts. Proper polarisation alignment can maximise the diffraction efficiency and contribute to increased brightness and contrast levels in the images. It should be noted that the captured images were obtained under ordinary room conditions in the original experiment, which could potentially result in image degradation due to ambient light. As such, the new experimental images are captured under dark light conditions to ensure the quality of the images is not compromised by the artefacts induced by the ambient light condition.

The calibration process is further improved to carefully align the captured image to the target image and reduce the mismatch between simulations and experiments. Instead of capturing a single image to compute the homography matrix, the improved approach iteratively performs the calibration to find the optimal matrix for the system setup. Additionally, the exposure time of the camera is adjusted to improve visual quality. We optimise and evaluate the reconstructed images over a confined region of interest at the centre with a resolution of 1680×960. The primary focus of this approach is to calibrate the resulting images, which not only improves the calibration process but also can increase the degree of freedom required in optimisation since fewer pixels need to be adjusted. To further improve the image quality of the experiment, it is recommended to utilise a horizontal grating to eliminate undiffracted light.

Current state-of-the-art techniques utilise complex kernels or convolutional neural networks to learn and optimise the correlation between a hologram and its optical reconstruction to enhance image quality in the presence of optical aberrations and bridge the gap between holographic display simulations and physical displays. A dataset containing multiple phase-only holograms and their corresponding image reconstructions is typically necessary to train these techniques. For example, Kavaklı et al. [187] used a training dataset of 797 phase patterns and their corresponding captured intensity images, employing four million parameters (2x1080x1920 for amplitude and phase) during the training process. Similarly, Choi et al. [185] utilised ~65 million parameters in their CNNpropCNN model for 3D holography, trained on a dataset containing 8,800 phase patterns and their corresponding experimental captures. However, the need for numerous parameters and a sizable training dataset poses significant challenges in this field. Additionally, the approximation of the wave propagation model using convolutional neural networks may result in less obvious mapping onto physical representation.

Figure 6.27. Comparison of reconstructed images using conventional methods and the camera-in-the-loop (CITL) approach. The use of CITL can remove noise and optical system-dependent artefacts, as shown by the rectangular boxes.

On the other hand, one could also use the camera-in-the-loop technique (CITL) [43]to address phase-error compensation during backpropagation in the training process. In Figure 6.27, we perform the CITL approach and qualitatively demonstrate its superiority over conventional methods in improving the quality of reconstructed images. Specifically, CITL has been found to effectively smooth out noise in reconstructed images and remove optical system-dependent artefacts through phase compensation in the iterative process, as indicated by the rectangular boxes in Figure 6.27. The results demonstrate the ability of CITL to bridge the gap between experimental and simulated results, thereby enhancing the quality of experimental reconstructed images.

6.6.2 Is it a Hologram?

One of the most exciting characteristics of the hologram is that every piece contains an image of the whole object. An observer can still perceive the whole image through a small piece of the hologram with a limited perspective and a reduced image quality. It is equivalent to looking at an object through a small hole in a window. The object stays the same size with a limited viewing angle. CGHs that utilise the entire diffraction angle to modulate light are also referred to as diffusive holograms [197]. These holograms allow reconstructed 3D objects to be viewed from any position within a diffraction angle. In contrast, recent trends in computer-generated holography have shifted towards optimising CGH algorithms with smooth target phases, also known as non-diffusive holograms. These holograms are primarily focused on enhancing reconstructed image quality. However, non-diffusive holograms do not fully utilise the diffraction region of the spatial light modulator. This approach only enforces constructive interference at the central view perspective, which attenuates other views by destructive interference [197,202]. This phenomenon can be intuitively perceived by observing the optimised non-diffusive holograms, which contain a significant amount of residual target image at the front view. As pointed out in [29], Smooth phase holography has a smaller effective Fourier spectrum with concentrated energy in DC, leading to a lack of full parallax within the diffraction region of the SLM. Figure 6.28 illustrates the replay field from diffusive and non-diffusive holograms, depicting their respective achievable field of view within the replay field. The diffusive hologram can modulate the light up to the maximum angle $\theta_{\max}$ bounded by a pixel pitch to the replay field, allowing reconstructed objects to be seen from any position within a field of view. On the other hand, the FOV of non-diffusive holograms is limited to the front view, and image quality from other views is heavily attenuated.

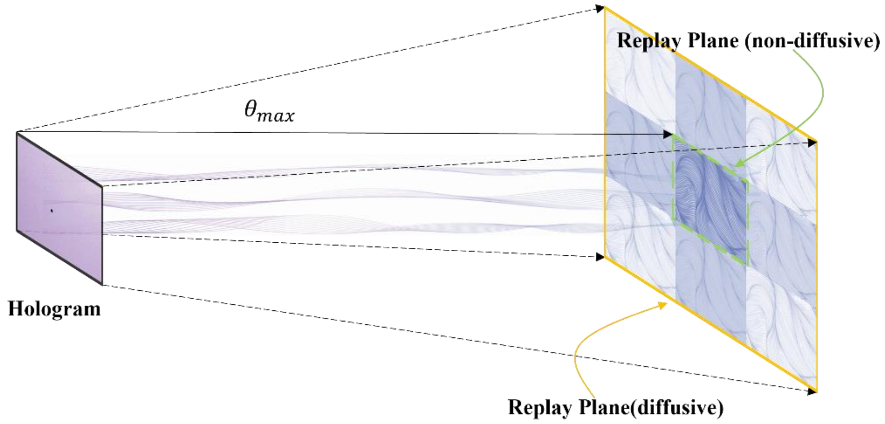

Figure 6.28. Comparison of the achievable FOV of diffusive and non-diffusive holograms. The diffusive hologram can reconstruct objects from any position within its maximum diffraction angle, resulting in a larger and more versatile replay field. In contrast, non-diffusive holograms are limited to the front view and suffer from significant attenuation in image quality from other views.

Furthermore, it should be noted that non-diffusive holograms have certain drawbacks. While they may produce higher-quality images, the reduction of the numerical aperture and the content-dependent defocus pattern can make the coherent properties of light more prominent. Specifically, interference patterns generated by Fresnel propagation in non-diffusive holograms can differ significantly from the defocus blur observed in real-world objects. This inconsistency in defocus patterns can disrupt the relationship between depth and blur, a critical factor in depth perception. Additionally, the presence of a clear boundary at the interface between objects with different depths, caused by interference, can distort the perception of relative depth between objects. Therefore, a combination of the advantages of both diffusive and non-diffusive holograms may be necessary to achieve high-quality holograms without distortion of depth perception.

Additionally, the updating strategy of non-diffusive hologram algorithms typically uses smooth target phases to update the subsequent estimation. Though resulting in better image quality, this strategy establishes a strong phase correlation between adjacent reconstructed points. This phase correlation is susceptible to disruption during experimental reconstruction and can cause inconsistent defocus patterns that disrupt the relationship between depth and blur as well. Therefore, a carefully calibrated system from simulation to optical reconstruction is often required to compensate accurately for the disturbing phase noise. Examples include using the camera-in-the-loop technique to compensate for the phase error during backpropagation or using specific neural networks that have more degrees of freedom in the propagation process to delicately compensate for the phase error during the training process. However, the approximation of the wave propagation model using convolutional neural networks may result in less obvious mapping onto physical representation; Target images that have less similarity to the trained dataset could not be able to produce CGHs through neural networks. Moreover, true holography is inherently fault-tolerant; even defective holographic pixels would not significantly degrade the reconstructed image quality.

6.6.3 ASM for Fraunhofer Region

Through Chapter 5 and Chapter 6, the gradient-based CGH algorithms are performed based on the angular spectrum method in the near-field region. Its ability to optimise CGH in the far-field region has not been validated. The far-field region requires long diffraction distances, which could result in a less accurate simulation from aliasing due to under-sampling [5,203]. The bandwidth of the hologram is related to the SLM characteristics that generally would not be changed, while the bandwidth of the spatial frequency transfer function in the ASM method increases with the increase of the propagation distance. Since the bandwidth of reconstructed objects is the sum of both, employing the same sampling frequency in the replay field could degrade the image quality. Experimental results in [142,186] also demonstrate that high frequencies of reconstruction are lost when the propagation distance becomes longer. One solution is to increase the size of the hologram resolution by zero-padding prior to the simulation [43,203,204]. Another solution is to band limit the spatial frequency transfer function of the ASM method as in [108].

Furthermore, we can calculate the diffraction at a long distance using the Fraunhofer diffraction function and at the Fresnel region using the Fresnel diffraction function. The far-field reconstruction is simply the Fourier transform of the CGH. The Fresnel transform translates the reconstruction plane along the optical axis using an additional quadratic phase factor as a virtual lens. Both methods can substantially reduce the computational load required for CGH optimisation and effectively accelerate the calculation process.

6.7 Conclusion

This chapter has extended the gradient descent method to 3D hologram generation, enabling high-quality 3D CGH reconstruction. The proposed method is validated by a monochrome holographic prototype and can reproduce reasonably good overall quality for in-focus regions of the scenes. However, the image quality of out-of-focus regions is degraded by the speckle due to unconstrained behaviour during the optimisation. Therefore, this work combines the gradient descent method with an incoherent imaging module to render natural incoherent realistic defocused images as targets to supervise the CGH training process. The incoherent imaging module controls the depth-dependent point spread function and simulates naturally incoherent defocused images with the nonlinear occlusion-aware image formation model. The proposed method can closely resemble reconstructed images with more naturally defocused regions; however, the image quality is slightly degraded in in-focus regions.

We further apply the hard attention mechanism, reformulating the loss function to set weights on both regions of the reconstructed images manually. In this way, the hard attention mechanism introduces more weight on the in-focus regions over the out-of-focus regions of the reconstructed images. The effectiveness of the newly proposed method has been experimentally demonstrated. In contrast to the directly natural defocus CGH method, this method can intuitively resolve more clearly for in-focus regions while reducing the speckle noise artefacts at out-of-focus regions and depth discontinuities.

Chapter 7 Conclusion and Future Work

7.1 Conclusion

Currently, 3D display technologies are still in the primary stage of development, and how to function as a "magic window" through which viewers can perceive truly natural 3D scenes is the key to 3D display technology. Among these technologies, holographic 3D display technology is one of the most promising 3D technologies that can perfectly reconstruct 3D scenes with all depth cues. However, its technical implementation imposes significant challenges in practice. This thesis addresses these challenges based on hardware system design and algorithmic development to improve reconstructed image quality.

This thesis has established an experimental holographic display from acquisition to optical reconstruction for generating CGHs from established 3D scenes or real-world scenes using commercially available RGB-D cameras. Though calculating a CGH from a real-world scene has been successfully achieved by depth completion, reconstruction errors of real-world scenes have been observed close to boundaries and near complex features. The reconstruction errors have been recognised due to the inaccurate depth information examined by comparing with a CGH calculated from a CG version of the same scene produced in Unity. Other experimental reconstruction issues such as zero-order, image quality degradation, inaccurate image defocus, and gamma correction have been evaluated.

Given that image quality degradation is one of the significant issues in holographic displays, the gradient descent method is introduced to phase-only CGH optimisation. However, this method is typically optimised using mean squared error, which is widely criticised for its poor correlation with perceptual quality. Therefore, contemporary image quality metrics (IQM) considering human visual systems are employed as loss functions to improve the reconstructed image quality. Extensive objective and subjective assessments of experimentally reconstructed images reveal that the perceived quality improves considerably with the selected MS-SSIM loss, highlighting the value of finding a specific perceptually-motivated loss function for CGH generation.

The gradient descent method is then extended to 3D hologram generation, validated in simulation and experimentally demonstrated on a holographic display prototype. While previous works have attempted to optimise 3D CGH generation only on the in-focus area, there are still barriers to achieving realistic visual quality in the defocused and depth-varying regions. Therefore, an incoherent imaging module is introduced to the gradient-decent-based 3D hologram generation process, simulating the natural defocus blur and occlusion effects at depth discontinuities. Additionally, the loss function is reformulated using an attention mechanism to maintain reconstructed image quality at the in-focus regions while reducing the speckle noise artefacts in the out-of-focus ones. The experimentally captured result presents its effectiveness of using Natural Defocus CGH with an attention mechanism for 3D CGH synthesis, demonstrating its potential to reconstruct realistic 3D images beyond the capabilities of existing 3D hologram generation algorithms.

7.2 Future work

7.2.1 Holographic Display System Advancement

Although this thesis has realised the construction of a holographic display system from acquisition to optical reconstruction, the prototype is still at the laboratory stage for demonstration purposes only and is not well calibrated to reduce optical aberration correction. A 2D holographic display system with a volume of around 57 cm^3 was made by Cable in 2006 using an SLM with a 13.62 μm pixel pitch and axillary optics [66]. Additional engineering effort system is needed to further miniaturise the system to an ultra-compact form that may support applications such as mobile phones or head-mounted displays. Although residual image artefacts in 2D holographic display systems due to optical aberration can be reduced using automated testbeds for factory-assembled holographic projectors [205,206] or recent hardware feedback systems [43,141], the aberration correction for 3D holographic display systems has not been systematically studied yet. As CGHs can directly support aberration correction from algorithmic perspectives, 3D holographic displays inherently are more favourable than any other display technologies, especially for near-eye AR applications where users with ametropia may need additional corrective lenses.

The proposed holography display system utilises stereoscopic RGB-D cameras for scene acquisition, which could be further replaced with holographic cameras in the near future. Ideal holographic cameras would potentially direct support real-time, real-world scene recording as digital holograms, which can be optical reconstructed using SLMs. Furthermore, the core device SLM is currently limited to modulating only the phase of light. A programmable fast-switching complex-amplitude spatial light modulation would reduce the computational complexity of phase-only CGH optimisation and fundamentally accelerate real-time CGH generation. Additionally, present SLMs only provide holographic displays with a relatively limited eyebox. Higher-resolution SLMs with a smaller pixel pitch could help mitigate this issue. Ultimately, SLMs are expected to function as magical surfaces that can switch modes to display on-screen images or floating 3D images behind the screen.

The holographic system can add a CGH compression and transmission module to support real-time CGH cloud computing. We can transmit the captured real-time 3D data on the cloud, and the calculated CGH can be compressed, encoded and transmitted to the local server and decoded for local holographic display. However, there are no generic CGH compression and transmission techniques. Dynamic high-resolution CGHs require high bandwidth for data transmission, which would not be feasible without compression techniques. Moreover, the statistical properties of the holographic data are different from the conventional photographic data. Standard compressing techniques such as JPEG and MPEG may not be applicable to CGHs.

7.2.2 Algorithmic Advancement

The algorithms in this thesis are all based on iterative algorithms, which are time-consuming with a high computational load. The study of high-quality non-iterative algorithms has always been one of the most exciting tasks for researchers in the CGH field. One potential research direction is to use deep learning techniques to calculate CGHs. The trained neural network could generate real-time CGHs while maintaining the same image quality. The trained neural network can be further combined with existing networks in head-up display (HUD) applications, supporting optical 3D object detection and annotation in augmented driving.

However, this technique may not be generic enough when CGH-related parameters vary in different conditions, including propagation distances, input data type, and CGH resolution. Analytic solutions that consider both the wave physical propagation process and the complex amplitude to phase-only encoding process are preferable to generate high-quality 3D CGHs.

Evaluating the CGH reconstructed image quality requires a widely accepted subjective testing methodology. The common practice for subjective evaluation is to utilise a camera first to record and display the reconstructed images on high-end 2D monitors. The alternative option is to display the hologram directly on a screen for subjective perceptual judgments on the experimental reconstructed images. Subjective benchmarking algorithms are preferable from a widely accepted configuration of high-end holographic displays with standard procedures for display calibrating, characterising, and testing. Furthermore, typical objective testing methodologies for image quality are based on 2D images. Objectively characterisation the quality of 3D CGH reconstructed images counting the defocus effect has not yet been accomplished. With the 3D holographic image characterisation, hologram generation techniques can be further improved by using numerical feedback systems.

The extension of IQMs to 3D CGH optimisation represents a promising research direction with the potential to enhance the quality of 3D reconstructed holographic images significantly. One potential approach is to directly utilise 2D IQMs as loss functions in 3D CGH optimisation. In using 2D IQMs as a loss function, the in-focus parts of the 3D reconstructed image are constrained to match the target RGB image through the application of binary masks at each plane, which is computed from the target depth map. An alternative approach that could fully utilise the potential of 2D IQMs in 3D CGH optimisation is extending their ability to access the out-of-focus regions of the reconstructed image. In contrast to binary masking the reconstructed image, IQMs can directly access both the in-focus and out-of-focus regions by comparing these regions to simulated incoherent target images. Using IQMs as loss functions, this approach has the potential to further improve the quality of the reconstructed holographic images, especially in out-of-focus regions. While the above approach can be applied to 2D image slices, it is important to note that CGH optimisation can also be performed on other data formats, such as polygon models or point clouds. These primitives can be transformed into 2D planes to apply 2D IQMs as loss functions, but other direct quality assessment methods may be more suitable for these primitives and result in better image quality in CGH optimisation.

Directly using IQMs as loss functions may also result in increased complexity and computational cost. Furthermore, combining IQMs with other optimisation techniques, such as regularization terms, may lead to even further improved results.

This thesis reconstructs a 3D scene using a preliminary algorithm by adding an extra incoherent imaging module with an attention mechanism to improve the defocus effect without sacrificing in-focus image quality. However, generating photorealistic defocused layer images may be challenging while maintaining real-time CGH optimisation. Incoherent image rendering is time-consuming, and it is therefore preferable to incorporate this process with modern rendering techniques such as ray tracing to accelerate the computational speed. Furthermore, rather than implicitly optimising the defocus areas using incoherent images as targets, alternative hologram-generating approaches that explicitly consider the defocusing effect and other human visual cues, such as occlusion and motion parallax, are desirable for holographic 3D displays in the future.

References

1. D. GABOR, "A New Microscopic Principle," Nature **161**, 777–778 (1948).
2. Y. N. Denisyuk, "Photographic Reconstruction of the Optical Properties of an Object in Its Own Scattered Radiation Field," Sov. Phys. Dokl. 543–545 (1962).
3. E. N. Leith and J. Upatnieks, "Reconstructed Wavefronts and Communication Theory*," J. Opt. Soc. Am. **52**, 1123 (1962).
4. E. N. Leith and J. Upatnieks, "Wavefront Reconstruction with Diffused Illumination and Three-Dimensional Objects*," J. Opt. Soc. Am. **54**, 1295 (1964).
5. J. W. Goodman, *Introduction to Fourier Optics (Macmillan Education, 2017)*, Fourth edition. (2017).
6. B. R. Brown and A. W. Lohmann, "Complex spatial filtering with binary masks," Appl. Opt. **5**, 967–969 (1966).
7. J. W. Goodman and R. W. Lawrence, "Digital image formation from electronically detected holograms," Appl. Phys. Lett. **11**, 77–79 (1967).
8. S. Reichelt, R. Hussler, G. Ftterer, and N. Leister, "Depth cues in human visual perception and their realization in 3D displays," in *PROC SPIE* (BELLINGHAM: Spie-Int Soc Optical Engineering, 2010), Vol. 7690, pp. 76900B-76900B–12.
9. N. S. Holliman, N. A. Dodgson, G. E. Favalora, and L. Pockett, "Three-Dimensional Displays: A Review and Applications Analysis," IEEE Trans. Broadcast. **57**, 362–371 (2011).
10. M. Martínez-Corral and B. Javidi, "Fundamentals of 3D imaging and displays: A tutorial on integral imaging, light-field, and plenoptic systems," Adv. Opt. Photonics **10**, 512–566 (2018).
11. J. Geng, "Three-dimensional display technologies," Adv. Opt. Photonics **5**, 456–535 (2013).
12. H. Urey, K. V. Chellappan, E. Erden, and P. Surman, "State of the Art in Stereoscopic and Autostereoscopic Displays," Proc. IEEE **99**, 540–555 (2011).

13. C. Wheatstone, "Contributions to the Physiology of Vision.–Part the First. On Some Remarkable, and Hitherto Unobserved, Phenomena of Binocular Vision," Philos. Trans. R. Soc. Lond. **128**, 371–394 (1838).
14. W. Rollmann, "Zwei neue stereoskopische Methoden," Ann. Phys. **166**, 186–187 (1853).
15. H. Kang, S. Roh, I. Baik, H. Jung, W. Jeong, J. Shin, and I. Chung, "3.1: A Novel Polarizer Glassestype 3D Displays with a Patterned Retarder," SID Int. Symp. Dig. Tech. Pap. **41**, 1–4 (2010).
16. S. S. Kim, B. H. You, H. Choi, B. H. Berkeley, D. G. Kim, and N. D. Kim, "31.1: Invited Paper: World's First 240Hz TFT-LCD Technology for Full-HD LCD-TV and Its Application to 3D Display," SID Int. Symp. Dig. Tech. Pap. **40**, 424–427 (2009).
17. J.-Y. Son, V. V. Saveljev, Y.-J. Choi, J.-E. Bahn, S.-K. Kim, and H. Choi, "Parameters for designing autostereoscopic imaging systems based on lenticular, parallax barrier, and integral photography plates," Opt. Eng. **42**, 3326–3333 (2003).
18. T. Okoshi, "Three-dimensional displays," Proc. IEEE **68**, 548–564 (1980).
19. D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, "Vergence-accommodation conflicts hinder visual performance and cause visual fatigue," J. Vis. Charlottesv. Va **8**, 33.1-3330 (2008).
20. P. A. Howarth, "Potential hazards of viewing 3-D stereoscopic television, cinema and computer games: a review: Potential health hazards of viewing 3-D stereoscopic displays," Ophthalmic Physiol. Opt. **31**, 111–122 (2011).
21. M. Urvoy, M. Barkowsky, and P. L. Callet, "How visual fatigue and discomfort impact 3D-TV quality of experience: a comprehensive review of technological, psychophysical, and psychological factors," Ann. Telecommun. - Ann. Télécommunications **68**, 641–655 (2013).
22. E. Chang, H. T. Kim, and B. Yoo, "Virtual Reality Sickness: A Review of Causes and Measurements," Int. J. Hum.-Comput. Interact. **36**, 1658–1682 (2020).
23. E. B. Goldstein, *Sensation and Perception.*, Tenth edition / E. Bruce Goldstein (University of Pittsburgh, University of Arizona), and James R. Brockmole (University of Notre Dame).; Student edition. (2016).
24. G. Lippmann, "Epreuves reversibles donnant la sensation du relief," J. Phys. **7**, 821–825 (1908).
25. Y. Takaki and N. Nago, "Multi-projection of lenticular displays to construct a 256-view super multi-view display," Opt. Express **18**, 8824–8835 (2010).
26. C. Chang, K. Bang, G. Wetzstein, B. Lee, and L. Gao, "Toward the next-generation VR/AR optics: a review of holographic near-eye displays from a human-centric perspective," Optica **7**, 1563–1578 (2020).
27. G. Kuo, L. Waller, R. Ng, and A. Maimone, "High resolution étendue expansion for holographic displays," ACM Trans. Graph. **39**, 66:1-66:14 (2020).
28. P. Chakravarthula, S.-H. Baek, F. Schiffers, E. Tseng, G. Kuo, A. Maimone, N. Matsuda, O. Cossairt, D. Lanman, and F. Heide, "Pupil-aware Holography," ACM Trans. Graph. TOG **41**, (2022).

29. B. Lee, D. Kim, S. Lee, C. Chen, and B. Lee, "High-contrast, speckle-free, true 3D holography via binary CGH optimization," Sci. Rep. 2022 121 **12**, 1–12 (2022).

30. S. A. Benton and V. M. B. Jr, *Holographic Imaging* (John Wiley & Sons, Ltd, 2008).

31. E. Hecht, *Optics*, Fifth edition, global ed. (Pearson Education, Limited, 2017).

32. T. D. Wilkinson, "Photonic Systems: 4B11 Lecture Notes,"Engineering Department, University of Cambridge (2017).

33. T. Shimobaba, T. Kakue, and T. Ito, "Review of Fast Algorithms and Hardware Implementations on Computer Holography," IEEE Trans. Ind. Inform. **12**, 1611–1622 (2016).

34. Y. Wang, D. Dong, P. J. Christopher, A. Kadis, R. Mouthaan, F. Yang, and T. D. Wilkinson, "Hardware implementations of computer-generated holography: a review," Opt. Eng. **59**, 102413 (2020).

35. T. Nishitsuji, Y. Yamamoto, T. Sugie, T. Kakue, H. Nakayama, T. Shimobaba, and T. Ito, "Dedicated computer for computer holography and its future outlook," in *PROC SPIE* (BELLINGHAM: SPIE, 2019), Vol. 10997, pp. 109970H-109970H–7.

36. A. Cable, E. Buckley, P. Mash, N. Lawrence, T. D. Wilkinson, and W. A. Crossland, "53.1: Real-time Binary Hologram Generation for High-quality Video Projection Applications," SID Symp. Dig. Tech. Pap. **35**, (2004).

37. R. W. Gerchberg and W. O. Saxton, "A Practical Algorithm for the Determination of Phase from Image and Diffraction Plane Pictures," Optik **35**, 237–246 (1971).

38. J. R. Fienup, "Phase retrieval algorithms: a comparison," Appl. Opt. Vol 21 Issue 15 Pp 2758-2769 **21**, 2758–2769 (1982).

39. H. H. Bauschke, P. L. Combettes, and D. R. Luke, "Hybrid projection–reflection method for phase retrieval," JOSA Vol 20 Issue 6 Pp 1025-1034 **20**, 1025–1034 (2003).

40. M. A. Seldowitz, J. P. Allebach, and D. W. Sweeney, "Synthesis of digital holograms by direct binary search," Appl Opt **26**, 2788–2798 (1987).

41. A. G. Kirk and T. J. Hall, "Design of binary computer generated holograms by simulated annealing: coding density and reconstruction error," Opt. Commun. **94**, 491–496 (1992).

42. M. H. Eybposh, N. W. Caira, M. Atisa, P. Chakravarthula, and N. C. Pégard, "DeepCGH: 3D computer-generated holography using deep learning," Opt Express **28**, 26636–26650 (2020).

43. Y. Peng, S. Choi, N. Padmanaban, and G. Wetzstein, "Neural holography with camera-in-the-loop training," ACM Trans. Graph. **39**, 1–14 (2020).

44. L. Shi, B. Li, C. Kim, P. Kellnhofer, and W. Matusik, "Towards real-time photorealistic 3D holography with deep neural networks," Nat. 2021 5917849 **591**, 234–239 (2021).

45. L. Shi, B. Li, and W. Matusik, "End-to-end learning of 3D phase-only holograms for holographic display," Light Sci. Appl. **11**, 247 (2022).

46. Z. Wen, C. Yang, X. Liu, and S. Marchesini, "Alternating direction methods for classical and ptychographic phase retrieval," Inverse Probl. **28**, 115010–115018 (2012).

47. S. Marchesini, Y. C. Tu, and H. T. Wu, "Alternating projection, ptychographic imaging and phase synchronization," Appl. Comput. Harmon. Anal. **41**, 815–851 (2016).

48. J. Zhang, N. Pégard, J. Zhong, H. Adesnik, and L. Waller, "3D computer-generated holography by non-convex optimization," Optica **4**, 1306–1313 (2017).

49. H. H. Bauschke, P. L. Combettes, and D. R. Luke, "Phase retrieval, error reduction algorithm, and Fienup variants: A view from convex optimization," J. Opt. Soc. Am. A Opt. Image Sci. Vis. **19**, 1334–1345 (2002).

50. A. Jesacher, A. Schwaighofer, S. Fürhapter, C. Maurer, S. Bernet, and M. Ritsch-Marte, "Wavefront correction of spatial light modulators using an optical vortex image," Opt. Express **15**, 5801–5808 (2007).

51. J. S. Liu and M. R. Taghizadeh, "Iterative algorithm for the design of diffractive phase elements for laser beam shaping," Opt Lett **27**, 1463–1465 (2002).

52. J. S. Liu, A. J. Caley, and M. R. Taghizadeh, "Symmetrical iterative Fourier-transform algorithm using both phase and amplitude freedoms," Opt. Commun. **267**, 347–355 (2006).

53. W. Qin and X. Peng, "Vulnerability to known-plaintext attack of optical encryption schemes based on two fractional Fourier transform order keys and double random phase keys," J. Opt. Pure Appl. Opt. **11**, 075402–075402 (2009).

54. C. Chang, J. Xia, L. Yang, W. Lei, Z. Yang, and J. Chen, "Speckle-suppressed phase-only holographic three-dimensional display based on double-constraint Gerchberg-Saxton algorithm," Appl Opt **54**, 6994–7001 (2015).

55. S. Tao and W. Yu, "Beam shaping of complex amplitude with separate constraints on the output beam," Opt. Express **23**, 1052–1062 (2015).

56. R. G. Dorsch, A. W. Lohmann, and S. Sinzinger, "Fresnel ping-pong algorithm for two-plane computer-generated hologram display," Appl Opt **33**, 869–875 (1994).

57. M. Makowski, M. Sypek, A. Kolodziejczyk, and G. Miku a, "Three-plane phase-only computer hologram generated with iterative Fresnel algorithm," Opt. Eng. **44**, 125805–125807 (2005).

58. D. Prongué, H. P. Herzig, R. Dändliker, and M. T. Gale, "Optimized kinoform structures for highly efficient fan-out elements," Appl. Opt. **31**, 5706–5711 (1992).

59. M. Johansson and J. Bengtsson, "Robust design method for highly efficient beam-shaping diffractive optical elements using an iterative-Fourier-transform algorithm with soft operations," J. Mod. Opt. **47**, 1385–1398 (2000).

60. D. Schäfer, "Design concept for diffractive elements shaping partially coherent laser beams," J. Opt. Soc. Am. A Opt. Image Sci. Vis. **18**, 2915–2922 (2001).

61. Z. Zalevsky, D. Mendlovic, and R. G. Dorsch, "Gerchberg-Saxton algorithm applied in the fractional Fourier or the Fresnel domain," Opt. Lett. **21**, 842–844 (1996).

62. B.-Z. Dong, Y. Zhang, B.-Y. Gu, and G.-Z. Yang, "Numerical investigation of phase retrieval in a fractional Fourier transform," J. Opt. Soc. Am. A Opt. Image Sci. Vis. **14**, 2709 (1997).

63. J. A. Rodrigo, H. Duadi, T. Alieva, and Z. Zalevsky, "Multi-stage phase retrieval algorithm based upon the gyrator transform," Opt. Express **18**, 1510–1520 (2010).

64. Z. Liu, L. Xu, Q. Guo, C. Lin, and S. Liu, "Image watermarking by using phase retrieval algorithm in gyrator transform domain," Opt. Commun. **283**, 4923–4927 (2010).

65. Q. Wang, Q. Guo, and L. Lei, "Multiple-image encryption system using cascaded phase mask encoding and a modified Gerchberg–Saxton algorithm in gyrator domain," Opt. Commun. **320**, 12–21 (2014).
66. A. J. Cable, "Real-time high-quality two and three-dimensional holographic video projection using the one-step phase retrieval (OSPR) approach," Thesis (Ph.D.), University of Cambridge (2007).
67. J.-H. Park, "Recent progress in computer-generated holography for three-dimensional scenes," J. Inf. Disp. **18**, 1–12 (2017).
68. D. Pi, J. Liu, and Y. Wang, "Review of computer-generated hologram algorithms for color dynamic holographic three-dimensional display," Light Sci. Appl. **11**, 231–231 (2022).
69. M. E. Lucente, "Interactive computation of holograms using a look-up table," J. Electron. Imaging **2**, 28 (1993).
70. P. W. M. Tsang, T.-C. Poon, and Y. M. Wu, "Review of fast methods for point-based computer-generated holography [Invited]," Photonics Res. Wash. DC **6**, 837 (2018).
71. T.-C. Poon, *Optical Scanning Holography with MATLAB*, 1st ed. (Springer New York, NY, 2007).
72. L. Ahrenberg, P. Benzie, M. Magnor, and J. Watson, "Computer generated holography using parallel commodity graphics hardware," Opt. Express **14**, 7636–7641 (2006).
73. R. H.-Y. Chen and T. D. Wilkinson, "Computer generated hologram from point cloud using graphics processor," Appl. Opt. **48**, 6841–6850 (2009).
74. R. H.-Y. Chen and T. D. Wilkinson, "Computer generated hologram with geometric occlusion using GPU-accelerated depth buffer rasterization for three-dimensional display," Appl. Opt. **48**, 4246–4255 (2009).
75. S.-C. Kim and E.-S. Kim, "Effective generation of digital holograms of three-dimensional objects using a novel look-up table method," Appl. Opt. **47**, D55 (2008).
76. S.-C. Kim, J.-M. Kim, and E.-S. Kim, "Effective memory reduction of the novel look-up table with one-dimensional sub-principle fringe patterns in computer-generated holograms," Opt. Express **20**, 12021 (2012).
77. S.-C. Kim and E.-S. Kim, "Fast computation of hologram patterns of a 3D object using run-length encoding and novel look-up table methods.," Appl. Opt. **48**, 1030–41 (2009).
78. T. Nishitsuji, T. Shimobaba, T. Kakue, and T. Ito, "Fast calculation of computer-generated hologram using run-length encoding based recurrence relation," Opt. Express **23**, 9852 (2015).
79. Z. Zhang, J. Liu, J. Jia, X. Li, J. Han, B. Hu, and Y. Wang, "Tunable nonuniform sampling method for fast calculation and intensity modulation in 3D dynamic holographic display," Opt. Lett. **38**, 2676–2679 (2013).
80. D. Pi, J. Liu, Y. Han, A. U. R. Khalid, and S. Yu, "Simple and effective calculation method for computer-generated hologram based on non-uniform sampling using look-up-table," Opt. Express **27**, 37337–37348 (2019).
81. S.-C. Kim and E.-S. Kim, "Fast one-step calculation of holographic videos of three-dimensional scenes by combined use of baseline and depth-compensating principal fringe patterns," Opt. Express **22**, 22513 (2014).

82. J. Jia, Y. Wang, J. Liu, X. Li, Y. Pan, Z. Sun, B. Zhang, Q. Zhao, and W. Jiang, "Reducing the memory usage for effective computer-generated hologram calculation using compressed look-up table in full-color holographic display," Appl. Opt. **52**, 1404 (2013).

83. Y. Pan, X. Xu, S. Solanki, X. Liang, R. B. A. Tanjung, C. Tan, and T.-C. Chong, "Fast CGH computation using S-LUT on GPU," Opt. Express **17**, 18543 (2009).

84. J.-H. Park, "Recent progress on fully analytic mesh based computer-generated holography," in (SPIE, 2016), Vol. 10022, pp. 100221G-100221G–9.

85. Y. Zhang, H. Fan, F. Wang, X. Gu, X. Qian, and T.-C. Poon, "Polygon-based computer-generated holography: a review of fundamentals and recent progress [Invited]," Appl. Opt. 2004 **61**, B363–B374 (2022).

86. D. LESEBERG and C. FRERE, "Computer-generated holograms of 3-D objects composed of tilted planar segments," Appl. Opt. **27**, 3020–3024 (1988).

87. T. TOMMASI and B. BIANCO, "Frequency analysis of light diffraction between rotated planes," Opt. Lett. **17**, 556–558 (1992).

88. T. TOMMASI and B. BIANCO, "Computer-generated holograms of tilted planes by a spatial frequency approach," J. Opt. Soc. Am. A **10**, 299–305 (1993).

89. K. Matsushima, H. Schimmel, and F. Wyrowski, "New creation algorithm for digitally synthesized holograms in surface model by diffraction from tilted planes," in *Proceedings of SPIE* (SPIE, 2002), Vol. 4659, pp. 53–60.

90. K. Matsushima, "Computer-generated holograms for three-dimensional surface objects with shade and texture," Appl. Opt. **44**, 4607 (2005).

91. K. Matsushima and A. Kondoh, "Wave optical algorithm for creating digitally synthetic holograms of three-dimensional surface objects," in T. H. Jeong and S. H. Stevenson, eds. (International Society for Optics and Photonics, 2003), Vol. 5005, p. 190.

92. H. Nishi, K. Matsushima, and S. Nakahara, "Rendering of specular surfaces in polygon-based computer-generated holograms," Appl. Opt. **50**, H245–H252 (2011).

93. Y. Pan, Y. Wang, J. Liu, X. Li, J. Jia, and Z. Zhang, "Analytical brightness compensation algorithm for traditional polygon-based method in computer-generated holography," Appl. Opt. 2004 **52**, 4391–4399 (2013).

94. H. Kim, J. Hahn, and B. Lee, "Mathematical modeling of triangle-mesh-modeled three-dimensional surface objects for digital holography," Appl. Opt. **47**, D117 (2008).

95. L. Ahrenberg, P. Benzie, M. Magnor, and J. Watson, "Computer generated holograms from three dimensional meshes using an analytic light transport model," Appl. Opt. **47**, 1567–1574 (2008).

96. Y.-P. Zhang, F. Wang, T.-C. Poon, S. Fan, and W. Xu, "Fast generation of full analytical polygon-based computer-generated holograms," Opt. Express **26**, 19206–19224 (2018).

97. J.-H. Park, S.-B. Kim, H.-J. Yeom, H.-J. Kim, H. Zhang, B. Li, Y.-M. Ji, S.-H. Kim, and S.-B. Ko, "Continuous shading and its fast update in fully analytic triangular-mesh-based computer generated hologram," Opt. Express **23**, 33893 (2015).

98. Y.-M. Ji, H.- Yeom, and J.-H. Park, "Efficient texture mapping by adaptive mesh division in mesh-based computer generated hologram," Opt. Express **24**, 28154 (2016).

99. K. Matsushima, "Exact hidden-surface removal in digitally synthetic full-parallax holograms," in T. H. Jeong and H. I. Bjelkhagen, eds. (International Society for Optics and Photonics, 2005), Vol. 5742, p. 25.

100. K. Matsushima, M. Nakamura, and S. Nakahara, "Silhouette method for hidden surface removal in computer holography and its acceleration using the switch-back technique," Opt. Express **22**, 24450 (2014).

101. M. Askari, S.-B. Kim, K.-S. Shin, S.-B. Ko, S.-H. Kim, D.-Y. Park, Y.-G. Ju, and J.-H. Park, "Occlusion handling using angular spectrum convolution in fully analytical mesh based computer generated hologram," Opt. Express **25**, 25867–25878 (2017).

102. F. Yang, A. Kaczorowski, and T. D. Wilkinson, "Fast precalculated triangular mesh algorithm for 3D binary computer-generated holograms," Appl. Opt. **53**, 8261 (2014).

103. F. Wang, T. Shimobaba, Y. Zhang, T. Kakue, and T. Ito, "Acceleration of polygon-based computer-generated holograms using look-up tables and reduction of the table size via principal component analysis," Opt. Express **29**, 35442–35455 (2021).

104. C. Chang, J. Wu, Y. Qi, C. Yuan, S. Nie, and J. Xia, "Simple calculation of a computer-generated hologram for lensless holographic 3D projection using a nonuniform sampled wavefront recording plane," Appl. Opt. **55**, 7988–7996 (2016).

105. J.-S. Chen, Q. Smithwick, and D. Chu, "Implementation of shading effect for reconstruction of smooth layer-based 3D holographic images," in A. J. Woods, N. S. Holliman, and G. E. Favalora, eds. (International Society for Optics and Photonics, 2013), Vol. 8648, p. 86480R.

106. J.-S. Chen, D. Chu, and Q. Smithwick, "Rapid hologram generation utilizing layer-based approach and graphic rendering for realistic three-dimensional image reconstruction by angular tiling," J. Electron. Imaging **23**, 023016 (2014).

107. K. Matsushima and T. Shimobaba, "Band-limited angular spectrum method for numerical simulation of free-space propagation in far and near fields," Opt. Express **17**, 19662–19673 (2009).

108. N. Okada, T. Shimobaba, Y. Ichihashi, R. Oi, K. Yamamoto, M. Oikawa, T. Kakue, N. Masuda, and T. Ito, "Band-limited double-step Fresnel diffraction and its application to computer-generated holograms," Opt. Express **21**, 9192 (2013).

109. H. G. Kim, H. Jeong, and Y. Man Ro, "Acceleration of the calculation speed of computer-generated holograms using the sparsity of the holographic fringe pattern for a 3D object," Opt. Express **24**, 25317–25328 (2016).

110. H. G. Kim and Y. Man Ro, "Ultrafast layer based computer-generated hologram calculation with sparse template holographic fringe pattern for 3-D object," Opt. Express **25**, 30418–30427 (2017).

111. H. Zhang, L. Cao, and G. Jin, "Computer-generated hologram with occlusion effect using layer-based processing," Appl. Opt. **56**, F138–F143 (2017).

112. G. Makey, Ö. Yavuz, D. K. Kesim, A. Turnalı, P. Elahi, S. Ilday, O. Tokel, and F. Ö. Ilday, "Breaking crosstalk limits to dynamic holography using orthogonality of high-dimensional random vectors," Nat. Photonics **13**, 251–256 (2019).

113. J.-S. Chen and D. P. Chu, "Improved layer-based method for rapid hologram generation and real-time interactive holographic display applications," Opt. Express **23**, 18143–18155 (2015).

114. J. Jia, J. Si, and D. Chu, "Fast two-step layer-based method for computer generated hologram using sub-sparse 2D fast Fourier transform," Opt. Express **26**, 17487–17497 (2018).

115. R. Piestun, J. Shamir, B. Weßkamp, and O. Bryngdahl, "On-axis computer-generated holograms for three-dimensional display," Opt. Lett. **22**, 922 (1997).

116. S.-C. Kim, D.-C. Hwang, D.-H. Lee, and E.-S. Kim, "Computer-generated holograms of a real three-dimensional object based on stereoscopic video images," Appl. Opt. **45**, 5669 (2006).

117. S. H. Lee, S. C. Kwon, H. B. Chae, J. Y. Park, H. J. Kang, and J. D. K. Kim, "Digital hologram generation for a real 3D object using by a depth camera," J. Phys. Conf. Ser. **415**, 12049–6 (2013).

118. M. Y. M. Yamaguchi, K. W. K. Wakunami, and M. I. M. Inaniwa, "Computer generated hologram from full-parallax 3D image data captured by scanning vertical camera array (Invited Paper)," Chin. Opt. Lett. **12**, 060018–060023 (2014).

119. G. Li, A.-H. Phan, N. Kim, and J.-H. Park, "Synthesis of computer-generated spherical hologram of real object with 360° field of view using a depth camera," Appl. Opt. **52**, 3567 (2013).

120. H. Yanagihara, T. Kakue, Y. Yamamoto, T. Shimobaba, and T. Ito, "Real-time three-dimensional video reconstruction of real scenes with deep depth using electro-holographic display system," Opt. Express **27**, 15662 (2019).

121. "Depth Quality Testing White Paper for Intel® RealSenseTM Camera," https://www.intel.co.uk/content/www/uk/en/support/articles/000026982/emergingtechnologies/ intel-realsense-technology.html, accessed on 08/10/2022.

122. M. Senthilvel, R. K. Soman, and K. Varghese, "Comparison of Handheld Devices for 3D Reconstruction in Construction," in *Proceedings of the 34th International Symposium on Automation and Robotics in Construction (ISARC)*, M.-Y. (National T. U. of S. Cheng, Technology), H.-M. (National T. U. of S. Chen, Technology), K. C. (National T. U. of S. Chiu, and Technology), eds. (Tribun EU, s.r.o., Brno, 2017), pp. 698–705.

123. L. E. C. de V. per Computador, V. E. Cabrera, and L. M. G. Goncalves, *Depth Data Error Modeling of the ZED 3D Vision Sensor from Stereolabs* (Centre de Visió per Computador, 2018), Vol. 17.

124. M. Carfagni, R. Furferi, L. Governi, M. Servi, F. Uccheddu, and Y. Volpe, "On the Performance of the Intel SR300 Depth Camera: Metrological and Critical Characterization," IEEE Sens. J. **17**, 4508–4519 (2017).

125. L. Keselman, J. I. Woodfill, A. Grunnet-Jepsen, and A. Bhowmik, "Intel(R) RealSense(TM) Stereoscopic Depth Cameras," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (IEEE, 2017), pp. 1267–1276.

126. I. Armeni, S. Sax, A. R. Zamir, and S. Savarese, "Joint 2D-3D-Semantic Data for Indoor Scene Understanding," CoRR **abs/1702.01105**, (2017) , accessed on 08/10/2022.

127. J. McCormac, A. Handa, S. Leutenegger, and A. J. Davison, "SceneNet RGB-D: Can 5M Synthetic Images Beat Generic ImageNet Pre-training on Indoor Segmentation?," in *2017 IEEE International Conference on Computer Vision (ICCV)* (2017), pp. 2697–2706.

128. S. Gupta, P. Arbeláez, and J. Malik, "Perceptual Organization and Recognition of Indoor Scenes from RGB-D Images," in *2013 IEEE Conference on Computer Vision and Pattern Recognition* (2013), pp. 564–571.

129. S. Song, S. P. Lichtenberg, and J. Xiao, "SUN RGB-D: A RGB-D scene understanding benchmark suite," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2015), pp. 567–576.

130. I. Armeni, O. Sener, A. R. Zamir, H. Jiang, I. Brilakis, M. Fischer, and S. Savarese, "3D Semantic Parsing of Large-Scale Indoor Spaces," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), pp. 1534–1543.

131. J. Xiao, A. Owens, and A. Torralba, "SUN3D: A Database of Big Spaces Reconstructed Using SfM and Object Labels," in *2013 IEEE International Conference on Computer Vision* (2013), pp. 1625–1632.

132. N. Silberman and R. Fergus, "Indoor scene segmentation using a structured light sensor," in *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)* (2011), pp. 601–608.

133. N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor Segmentation and Support Inference from RGBD Images," in *ECCV'12 Proceedings of the 12th European Conference on Computer Vision - Volume Part V* (Springer-Verlag Berlin, 2012), pp. 746–760.

134. A. Levin, D. Lischinski, and Y. Weiss, "Colorization using optimization," ACM Trans. Graph. **23**, 689–694 (2004).

135. I. Armeni, S. Sax, A. R. Zamir, and S. Savarese, "Joint 2D-3D-Semantic Data for Indoor Scene Understanding," ArXiv **abs/1702.01105**, (2017), accessed on 08/10/2022.

136. P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," IEEE Trans. Pattern Anal. Mach. Intell. **14**, 239–256 (1992).

137. Y. Chen and G. Medioni, "Object modelling by registration of multiple range images," Image Vis. Comput. **10**, 145–155 (1992).

138. M. H. Eybposh, N. W. Caira, M. Atisa, P. Chakravarthula, and N. C. Pégard, "DeepCGH: 3D computer-generated holography using deep learning," Opt. Express Vol 28 Issue 18 Pp 26636-26650 **28**, 26636–26650 (2020).

139. S. Liu and Y. Takaki, "Optimization of Phase-Only Computer-Generated Holograms Based on the Gradient Descent Method," Appl. Sci. 2020 Vol 10 Page 4283 **10**, 4283–4283 (2020).

140. P. Chakravarthula, Y. Peng, J. Kollin, H. Fuchs, and F. Heide, "Wirtinger holography for near-eye displays," ACM Trans. Graph. TOG **38**, (2019).

141. P. Chakravarthula, E. Tseng, T. Srivastava, H. Fuchs, and F. Heide, "Learned hardware-in-the-loop phase retrieval for holographic near-eye displays," ACM Trans. Graph. TOG **39**, (2020).

142. C. Chen, B. Lee, N.-N. Li, M. Chae, D. Wang, Q.-H. Wang, and B. Lee, "Multi-depth hologram generation using stochastic gradient descent algorithm with complex loss function," Opt. Express Vol 29 Issue 10 Pp 15089-15103 **29**, 15089–15103 (2021).

143. L.-H. Yeh, J. Dong, J. Zhong, L. Tian, M. Chen, G. Tang, M. Soltanolkotabi, L. Waller, J. Chung, X. Ou, R. P. Kulkarni, C. Yang, H. Yeh, M. Chen, J. Zhong, and L. Waller, "Experimental robustness of Fourier ptychography phase retrieval algorithms," Opt. Express Vol 23 Issue 26 Pp 33214-33240 **23**, 33214–33240 (2015).

144. L. Cao and Y. Gao, "Generalized optimization framework for pixel super-resolution imaging in digital holography," Opt. Express Vol 29 Issue 18 Pp 28805-28823 **29**, 28805–28823 (2021).

145. K. Kreutz-Delgado, "The Complex Gradient Operator and the CR-Calculus ECE275A - Lecture Supplement - Fall 2005,"University of California, San Diego (2009).

146. E. J. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval via wirtinger flow: Theory and algorithms," IEEE Trans. Inf. Theory **61**, 1985–2007 (2015).

147. A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "PyTorch: An Imperative Style, High-Performance Deep Learning Library," in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d Alché-Buc, E. Fox, and R. Garnett, eds. (Curran Associates, Inc., 2019), pp. 8024–8035.

148. J. Duchi, E. Hazan, and Y. Singer, "Adaptive Subgradient Methods for Online Learning and Stochastic Optimization," J. Mach. Learn. Res. **12**, 2121–2159 (2011).

149. D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Y. Bengio and Y. LeCun, eds. (2015).

150. E. Agustsson and R. Timofte, "NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study," IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshop **2017-July**, 1122–1131 (2017).

151. R. Timofte, S. Gu, L. Van Gool, L. Zhang, and M. H. Yang, "NTIRE 2018 challenge on single image super-resolution: Methods and results," 2018 IEEECVF Conf. Comput. Vis. Pattern Recognit. Workshop CVPRW **2018**, 965–96511 (2018).

152. R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric," 2018 IEEECVF Conf. Comput. Vis. Pattern Recognit. CVPR 586–595 (2018).

153. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, *Image Quality Assessment: From Error Visibility to Structural Similarity* (2004), Vol. 13.

154. Z. Wang and A. C. Bovik, "Mean squared error: Lot it or leave it? A new look at signal fidelity measures," IEEE Signal Process. Mag. **26**, 98–117 (2009).

155. J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual Losses for Real-Time Style Transfer and Super-Resolution," Lect. Notes Comput. Sci. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinforma. **9906 LNCS**, 694–711 (2016).

156. Y. Rivenson, Y. Zhang, H. Günaydın, D. Teng, and A. Ozcan, "Phase recovery and holographic image reconstruction using deep learning in neural networks," Light Sci. Appl. **7**, 17141–17141 (2018).

157. A. Ahar, T. Birnbaum, M. Chlipala, W. Zaperty, S. Mahmoudpour, T. Kozacki, M. Kujawinska, and P. Schelkens, "Comprehensive performance analysis of objective quality metrics for digital holography," Signal Process. Image Commun. **97**, 116361–116361 (2021).

158. A. Ahar, D. Blinder, T. Bruylants, C. Schretter, A. Munteanu, and P. Schelkens, "Subjective quality assessment of numerically reconstructed compressed holograms," Proc SPIE 9599 Appl. Digit. Image Process. XXXVIII 95990K 22 Sept. 2015 **9599**, 188–202 (2015).

159. D. Blinder, A. Ahar, S. Bettens, T. Birnbaum, A. Symeonidou, H. Ottevaere, C. Schretter, and P. Schelkens, "Signal processing challenges for digital holographic video display systems," Signal Process. Image Commun. **70**, 114–130 (2019).

160. V. Laparra, J. Ballé, A. Berardino, E. Simoncelli, U. de Valencia, N. York University, and H. Hughes Medical Institute, "Perceptual image quality assessment using a normalized Laplacian pyramid," Proc IST Int'l Symp Electron. Imaging Hum. Vis. Electron. Imaging 2016 **28**, 1–6 (2016).

161. P. Chakravarthula, Z. Zhang, O. Tursun, P. Didyk, Q. Sun, and H. Fuchs, "Gaze-Contingent Retinal Speckle Suppression for Perceptually-Matched Foveated Holographic Displays," IEEE Trans. Vis. Comput. Graph. **27**, 4194–4203 (2021).

162. D. R. Walton, K. Kavaklı, R. K. Dos Anjos, D. Swapp, T. Weyrich, H. Urey, A. Steed, T. Ritschel, and K. Akşit, "Metameric Varifocal Holograms," in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (2022), pp. 746–755.

163. Y. Jo, S. Yang, and S. J. Kim, "Investigating loss functions for extreme super-resolution," 2020 IEEECVF Conf. Comput. Vis. Pattern Recognit. Workshop CVPRW 1705–1712 (2020).

164. K. Ding, K. Ma, S. Wang, and E. P. Simoncelli, "Comparison of Full-Reference Image Quality Models for Optimization of Image Processing Systems," Int. J. Comput. Vis. **129**, 1258–1281 (2021).

165. A. Mustafa, A. Mikhailiuk, D. A. Iliescu, V. Babbar, and R. K. Mantiuk, "Training a Task-Specific Image Reconstruction Loss," 2022 IEEECVF Winter Conf. Appl. Comput. Vis. WACV 21–30 (2022).

166. S. Kastryulin, J. Zakirov, D. Prokopenko, and D. V. Dylov, "PyTorch Image Quality: Metrics for Image Quality Assessment," (2022).

167. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Trans. Image Process. **13**, 600–612 (2004).

168. Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," Thrity-Seventh Asilomar Conf. Signals Syst. Comput. **2**, 1398–1402 (2003).

169. L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A Feature Similarity Index for Image Quality Assessment," IEEE Trans. Image Process. **20**, 2378–2386 (2011).

170. B. Zhang, P. V. Sander, and A. Bermak, "Gradient magnitude similarity deviation on multiple scales for color image quality assessment," 2017 IEEE Int. Conf. Acoust. Speech Signal Process. ICASSP 1253–1257 (2017).

171. L. Zhang, Y. Shen, and H. Li, "VSI: A Visual Saliency-Induced Index for Perceptual Image Quality Assessment," IEEE Trans. Image Process. **23**, 4270–4281 (2014).

172. R. Reisenhofer, S. Bosse, G. Kutyniok, and T. Wiegand, "A Haar wavelet-based perceptual similarity index for image quality assessment," Signal Process. Image Commun. **61**, 33–43 (2018).

173. H. R. Sheikh and A. C. Bovik, "Image information and visual quality," IEEE Trans. Image Process. **15**, 430–444 (2006).

174. K. Ding, K. Ma, S. Wang, and E. P. Simoncelli, "Image Quality Assessment: Unifying Structure and Texture Similarity," IEEE Trans. Pattern Anal. Mach. Intell. **44**, 2567–2581 (2022).

175. A. M. G. Pinheiro, J. Prazeres, A. Gilles, T. Birnbaum, R. K. Muhamad, and P. Schelkens, "Definition of common test conditions for the new JPEG pleno holography standard," in *Optics, Photonics and Digital Technologies for Imaging Applications VII*, P. Schelkens and T. Kozacki, eds. (SPIE, 2022), Vol. 12138, p. 121380N.

176. A. Ahar, M. Chlipala, T. Birnbaum, W. Zaperty, A. Symeonidou, T. Kozacki, M. Kujawinska, and P. Schelkens, "Suitability analysis of holographic vs light field and 2D displays for subjective quality assessment of Fourier holograms," Opt. Express **28**, 37069–37091 (2020).

177. ITU-R, "Methodologies for the subjective assessment of the quality of television images," (2002), , accessed on 08/10/2022.

178. R. A. Bradley and M. E. Terry, "Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons," Biometrika **39**, 324–324 (1952).

179. H. Turner and D. Firth, "Bradley-Terry Models in R: The BradleyTerry2 Package," J. Stat. Softw. **48**, 1–21 (2012).

180. H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," IEEE Trans. Image Process. **15**, 3440–3451 (2006).

181. N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, J. Astola, M. Carli, and F. Battisti, "TID2008-A Database for Evaluation of Full-Reference Visual Quality Assessment Metrics," Adv. Mod. Radioelectron. **10**, 30–45 (2009).

182. N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C. C. Jay Kuo, "Image database TID2013: Peculiarities, results and perspectives," Signal Process. Image Commun. **30**, 57–77 (2015).

183. Y. Peng, S. Choi, J. Kim, and G. Wetzstein, "Speckle-free holography with partially coherent light sources and camera-in-the-loop calibration," Sci. Adv. **7**, 5040–5040 (2021).

184. Y. Blau and T. Michaeli, "The Perception-Distortion Tradeoff," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE, 2017), pp. 6228–6237.

185. S. Choi, M. Gopakumar, Y. Peng, J. Kim, and G. Wetzstein, "Neural 3D holography: learning accurate wave propagation models for 3D holographic virtual and augmented reality displays," ACM Trans. Graph. **40**, 1–12 (2021).

186. J. Lee, J. Jeong, J. Cho, D. Yoo, B. Lee, and B. Lee, "Deep neural network for multi-depth hologram generation and its training strategy," Opt. Express **28**, 27137–27154 (2020).

187. K. Kavaklı, Y. Itoh, H. Urey, and K. Akşit, "Realistic Defocus Blur for Multiplane Computer-Generated Holography," IEEE VR 2023 (2023).

188. P. S. Considine, "Effects of Coherence on Imaging Systems," J. Opt. Soc. Am. 1930 **56**, 1001 (1966).

189. M. Born, *Principles of Optics : Electromagnetic Theory of Propagation, Interference and Diffraction of Light.*, 7th ed. (Cambridge University Press, 2002).

190. J. Chang and G. Wetzstein, "Deep Optics for Monocular Depth Estimation and 3D Object Detection," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (2019), pp. 10192–10201.

191. H. Ikoma, C. M. Nguyen, C. A. Metzler, Y. Peng, and G. Wetzstein, "Depth from Defocus with Learned Optics for Imaging and Occlusion-aware Depth Estimation," IEEE Int. Conf. Comput. Photogr. ICCP (2021).

192. Y. Wu, V. Boominathan, H. Chen, A. Sankaranarayanan, and A. Veeraraghavan, "PhaseCam3D — Learning Phase Masks for Passive Single View Depth Estimation," in *2019 IEEE International Conference on Computational Photography (ICCP)* (2019), pp. 1–12.

193. X. Dun, H. Ikoma, G. Wetzstein, Z. Wang, X. Cheng, and Y. Peng, "Learned rotationally symmetric diffractive achromat for full-spectrum computational imaging," Optica **7**, 913–922 (2020).

194. X. C. Zhang, K. Matzen, V. Nguyen, D. Yao, Y. Zhang, and R. Ng, "Synthetic defocus and look-ahead autofocus for casual videography," ACM Trans. Graph. TOG **38**, 1–16 (2019).

195. T. Zhou, R. Tucker, J. Flynn, G. Fyffe, and N. Snavely, "Stereo Magnification: Learning View Synthesis Using Multiplane Images," ACM Trans Graph **37**, (2018).

196. S. W. Hasinoff and K. N. Kutulakos, "A Layer-Based Restoration Framework for Variable-Aperture Photography," in *2007 IEEE 11th International Conference on Computer Vision* (2007), pp. 1–8.

197. D. Yang, W. Seo, H. Yu, S. I. Kim, B. Shin, C.-K. Lee, S. Moon, J. An, J.-Y. Hong, G. Sung, and H.-S. Lee, "Diffraction-engineered holography: Beyond the depth representation limit of holographic displays," Nat. Commun. **13**, 6012–6012 (2022).

198. M. Roberts, J. Ramapuram, A. Ranjan, A. Kumar, M. A. Bautista, N. Paczan, R. Webb, and J. M. Susskind, "Hypersim: A Photorealistic Synthetic Dataset for Holistic Indoor Scene Understanding," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (2021), pp. 10892–10902.

199. C. Kim, H. Zimmer, Y. Pritch, A. Sorkine-Hornung, and M. Gross, "Scene Reconstruction from High Spatio-Angular Resolution Light Fields," ACM Trans Graph **32**, (2013).

200. L. Xiao, A. Kaplanyan, A. Fix, M. Chapman, and D. Lanman, "DeepFocus: Learned Image Synthesis for Computational Displays," ACM Trans Graph **37**, (2018).

201. K. Kavaklı, H. Urey, and K. Akşit, "Learned holographic light transport," Appl. Opt. **61**, B50–B55 (2022).

202. S. Choi, M. Gopakumar, Yifan, 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," (2022).

203. T.-C. Poon and J.-P. Liu, *Introduction to Modern Digital Holography: With Matlab* (Cambridge University Press, 2014).

204. F. Yang, A. Kadis, R. Mouthaan, B. Wetherfield, A. Kaczorowski, and T. D. Wilkinson, "Perceptually motivated loss functions for computer generated holographic displays," Sci. Rep. **12**, 7709–7709 (2022).

205. A. Kaczorowski, G. S. Gordon, A. Palani, S. Czerniawski, and T. D. Wilkinson, "Optimization-Based Adaptive Optical Correction for Holographic Projectors," J. Disp. Technol. **11**, 596–603 (2015).

206. A. Kaczorowski, G. S. D. Gordon, and T. D. Wilkinson, "Adaptive, spatially-varying aberration correction for real-time holographic projectors," Opt. Express **24**, 15742–15756 (2016).