TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

JANI MÄKINEN

FROM LIGHT FIELDS TO WAVEFIELDS: HOLOGRAM GENERATION USING MULTIPERSPECTIVE IMAGES

Master of Science thesis

Examiners: Prof. Atanas Gotchev,
Dr. Erdem Sahin
Examiner and topic approved by the
Faculty Council of the Faculty of
Computing and electrical engineering
on 7th September 2016

ABSTRACT

JANI MÄKINEN: From Light Fields to Wavefields: Hologram Generation Using Multiperspective Images
Tampere University of Technology
Master of Science thesis, 52 pages, 0 Appendix pages
May 2017
Master's Degree Programme in Information Technology
Major: Signal Processing
Examiners: Prof. Atanas Gotchev, Dr. Erdem Sahin
Keywords: light field, computer generated holography, wavefield propagation

In this thesis, the link between the ray-optics and wave-optics formalisms of light propagation modeling is studied through light field (LF) and holography. Multiperspective images, such as captured by multicamera arrays, are utilized to obtain the discrete LF information. Three different computer generated hologram (CGH) representations are discussed in the thesis: holographic stereogram (an example for incoherent CGH), phase-added stereogram and diffraction specific coherent panoramagram (examples for coherent CGH). Comparative analysis of these three different holographic representation techniques is carried out through experiments simulating the viewing process of the holograms by the human eye. In particular, reconstructed image quality is compared for different scenes at different viewpoints. The accommodation responses of each technique is also evaluated via changing the focal length of the lens in the human eye model to focus the eye at different distances.

The prominent issue of speckle noise apparent in hologram reconstruction process is particularly addressed in detail, since it heavily affects the quality of the reconstructed images. In addition to existing solutions analyzed in the thesis, random averaging and pixel separation, a speckle suppression method based on pixel separation for coherent holograms is proposed. The proposed method is shown to further enhance the reconstructed image quality with respect to existing speckle reduction techniques. Besides the perceived image quality, another topic that is seen to be critical in the context of the thesis is simplifying the capture process of LF. In this aspect, the strict camera sampling requirements in LF capture for holographic stereograms are shown to be relieved considerably through the use of shearlet-based LF reconstruction algorithm. This enables utilization of more appropriate capture devices, e.g. multi-camera arrays, instead of conventionally used camera rigs.

TIIVISTELMÄ

JANI MÄKINEN: Valokentistä aaltokentiksi: hologrammien generointi perspektiivisistä kuvista
Tampereen teknillinen yliopisto
Diplomityö, 52 sivua, 0 liitesivua
Toukokuu 2017
Tietotekniikan koulutusohjelma
Pääaine: Signaalinkäsittely
Tarkastajat: Prof. Atanas Gotchev, Erdem Sahin
Avainsanat: valokenttä, digitaalinen holografia, aaltokentän propagaatio

Tämän työn tavoitteena on tarkastella valon säde- ja aalto-optiikkaa valokenttien ja holografian kautta. Moniperspektiivisiä kuvia käytetään tallentamaan diskreetin valokentän informaatio. Kolme eri digitaalista hologrammiesitystä valittiin tähän työhön vertailtavaksi: holographic stereogram (esimerkkinä inkoherenteista hologrammeista), phase-added stereogram ja diffraction specific coherent panoramagram (esimerkkeinä koherenteista hologrammeista). Näiden hologrammiesitysten välisiä eroja analysoidaan ihmisnäköä numeerisesti simuloivien kokeiden avulla. Erityisesti eri hologrammitallenteista saatujen rekonstruktiokuvien visuaalista laatua vertaillaan simuloimalla katsojaa eri näkökulmista.

Holografiseen rekonstruktioprosessiin liittyvää pilkkuhäiriötä käsitellään yksityiskohtaisesti, sillä se heikentää havaittujen kuvien laatua huomattavasti. Nykyisten ratkaisujen, kuten satunnaiskeskiarvottamisen ja pikseliseparaation lisäksi johdetaan pikseliseparaatioon pohjautuva pilkkuhäiriötä vähentävä menetelmä koherenteille hologrammeille. Kokeiden perusteella tämän menetelmän osoitetaan parantavan rekonstruktiokuvien laatua. Havaitun kuvanlaadun lisäksi kriittinen aihe tämän työn kontekstissa on valokentän tallentamisen helpottaminen. Tiukkoja näytteistämisvaatimuksia tähän liittyen voidaan keventää huomattavasti shearletmuunnokseen pohjautuvan valokentän rekonstruktioalgoritmin avulla, mahdollistaen perinteisesti käytettyjen järjestelmien sijaan käytännöllisempien kameraryhmien käytön.

PREFACE

The work for this thesis was done with the 3D media group in the laboratory of Signal Processing at the Tampere University of Technology. The aim of the work was to study holographic recording and display through various numerical simulations, thus enabling the comparison of different holography related methods in a controlled manner.

I would like to thank my supervisor Prof. Atanas Gotchev for his guidance and for giving me the opportunity to work on this particular topic. I would also like to thank Dr. Erdem Sahin for his continuous support and feedback, as well as the entire 3D media group for providing a pleasant working environment. Special thanks to Dr. Suren Vagharshakyan for his help with the shearlet-based LF reconstruction algorithm.

Further thanks to all of the teaching staff in the SGN laboratory for sparking the interest to study signal processing and providing the necessary foundation of knowledge to build upon for future endeavors.

Finally, I am deeply grateful to my parents for their unending support and encouragement throughout my studies.

Tampere, 24.5.2017

Jani Mäkinen

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

3D	three-dimensional
CGH	Computer generated hologram
DIBR	Depth-image based rendering
DOF	Depth-of-field
DSCP	Diffraction specific coherent panoramagram
DSLF	Densely sampled light field
EPI	Epipolar-plane image
FFT	Fast Fourier transform
FOV	Field-of-view
HPO	Horizontal-parallax-only
HS	Holographic stereogram
HVS	Human visual system
IFFT	Inverse fast Fourier transform
LF	Light field
PAS	Phase-added stereogram
PSF	Point-spread function
PSNR	Peak signal-to-noise ratio

$\mathbf{a}$	vectorized version of a densely sampled EPI
α	acceleration coefficient
$\mathbf{b}$	vectorized version of a decimated EPI
C	speckle contrast
Δ_s	camera baseline
Δ_u	sensor pixel size
Δ_x	hologram plane spatial sampling
Δ_x^{HVS}	HVS lateral acuity
$\Delta\theta_{out}$	angular sampling limit according to HVS
D	disparity between adjacent views
d	hologram-camera plane distance
d_{focus}	distance of focus
$\mathcal{F}_d$	Fresnel transform by distance d
f	lens focal length
f_x	spatial frequency

$\mathbf{H}$	masking matrix
I	viewer perceived image
L	discrete light field
l	sensor-aperture distance
λ	wavelength
λ_n	threshold for hard threshold operator
$O(x, y)$	complex object field on (x, y) plane
r	Euclidean distance between point and hologram plane
$\Delta S_{x,y,z}$	sampling steps of the object point grid
$\mathbf{S}$	shearlet synthesis transform matrix
$\mathbf{S}^*$	shearlet analysis matrix
T	aperture (pupil) diameter
$T(s, t)$	lens transfer function
$\mathbf{T}_{\lambda_n}$	hard threshold operator for the threshold λ_n
U	complex wavefield
$\mathbf{U}$	discrete complex wavefield
W_0	the spatial extent of a wavefield
W_z	the spatial extent of a Fresnel diffracted wavefield at distance z
X_x	hologram plane pixel size
z_0	distance between camera aperture and recentering plane
z_b, z_f	scene depth limits of a DSLF
z_{eye}	distance between the hologram and the simulated eye
$\mathbb{Z}$	the set of integers
$\mathbb{Z}^+$	the set of positive integers
θ	light ray incidence angle

1. INTRODUCTION

Ever since the birth of photography, researchers have been in search of realistic 3-dimensional (3D) imaging containing the proper perception of depth, spatial relations and accommodation cues. Digitalization of imaging and advances in signal processing have led to several approximations, such as stereoscopic imaging and volumetric displays, although each of them have compromised some visual cues. Holography provides the ultimate way of 3D scene reconstruction, even though its original aim was not actually in achieving realistic 3D imaging. Dennis Gabor is considered the inventor of holography, whose aim was to record the wave shape emitted by point-like objects in electron microscopy and correct the distortion caused by the magnetic lenses with optical waves. However, recording the wavefront shape and amplitude was considered at the time unsolvable. Gabor managed to overcome this by proposing a two-beam recording method in 1948 [7], although the requirements of the method, coherent object and reference beam, limited it to small objects. The name of the recording method, holography, was inspired by the Greek words *holos* ("whole") and *graphos* ("message") as it was able to record both the amplitude and phase of the field as opposed to only the intensity. Later in the 1960s, as the coherent light requirement for the method in the form of lasers was made available, work on practical holography advanced rapidly [20]. This paved the way for the invention of CGHs [5], in which the physical processes of holography are simulated numerically.

The numerical calculation of the object and reference waves in CGH enabled hologram generation from synthetic 3D scenes or incoherently illuminated real-life scenes. The holographic information is numerically recorded as diffraction patterns obtained from the exact representation of the scene (e.g. as point sources) in accurate holographic representations, such as the Fresnel hologram [44]. However, the calculation of such diffraction patterns requires a large amount of data and computational effort, which has resulted in different compromises to reduce computational requirements and bandwidth. These include for example horizontal-parallax-only (HPO) holograms as well as multiview stereograms. Creating a hologram from multiview

images is a widely studied field, with several different proposed methods. One of the most common examples of such representations, the holographic stereogram (HS), is composed of several holographic elements (hogels), which encode the view-dependent intensity distribution of the scene employing planar wavefront segments [31, 8]. Phase-added stereogram (PAS) proposed by Yamaguchi introduces a more accurate approach by utilizing also the 3D positional information of the scene to solve the problem of discontinuous wavefront approximation in HS [47]. Senoh *et. al.* included depth maps in addition to the intensity images achieving several enhancements, such as phantom imaging elimination and occlusion-hole filling process, to produce high quality holographic images. Another improvement from multiview stereograms [37], known as diffraction specific coherent panoramagram (DSCP), extends the wavefield approximation by introducing controllable curvature elements, thus providing better visual depth cues, though still relieving the computational requirements in comparison to Fresnel holograms. In addition to multiview images, integral imaging has been researched as input data for digital holograms. Xiao *et. al.* utilized dense ray sampling and resampling of the 2D integral imaging images to create a digital hologram [45].

Depending on the type of data to be utilized in obtaining the holographic information, CGHs can be divided into model-based and image-based methods. In model-based CGH, the recorded object is described by a model, where it is represented e.g. as a collection of independent light sources (point-cloud model) or by planar segments (polygon-based model). In the point-cloud model, the object is described by a set of point sources emitting light, each contributing to the hologram. Due to the exact object model description, the density of the point sources can be adapted to that required by the human visual system acuity. Furthermore, holograms from point-based models can achieve high quality reconstructions of arbitrary shaped objects at high spatial and angular resolutions. That is, an ideal coherent hologram model, e.g. Fresnel or Rayleigh-Sommerfeld, can be obtained from model-based scene representations. However, the computational demands of obtaining such holograms are extremely high, as the number of arithmetic operations is relative to the total number of point sources and number of pixels on the hologram plane. Moreover, additional methods are required to handle occlusions in the recorded scene. Alternatively, the recorded object can be described as a polygonal representation where the surface of the object is modeled with more complex primitives than point sources. In this case, a mesh of polygonal shaped, e.g. triangle, planar surface light sources define a non-planar object. This time, the field at the hologram plane is

the superposition of the LFs emitted by each polygon (i.e. polygon field) as a field-oriented approach. Compared to the point-based model, computing the contribution of a single polygon field is more cumbersome than for a point source, though the number of polygons required to describe an object is in return significantly smaller than the number of point sources, leading often to a faster CGH synthesis. The downsides of the polygon-based approach, however, are that remapping consumes a notable portion of computational time [29] and that the number of pixels on the hologram is restricted due to the need for storing complex functions for numerical propagation [30].

Image-based CGH methods provide an alternative approach to holography by obtaining the holographic information through a set of light rays as multiview images. As such, they provide varying levels of approximation from the ideal coherent holograms due to lacking the exact model available for hologram generation. The implication of this is that often the computational burden is relieved and efficient computer graphics rendering techniques can be utilized [32]. Furthermore, image-based methods are more suitable for practical capture setups than model-based methods as the information for the CGH can be obtained from a set of images, thus enabling capture by multicamera setups or scanning camera rigs. These reasons motivate the choice of CGH methods to include in this thesis.

The aim of this thesis is to analyze the transformation from a discretely captured light field to a wavefield (realized as a hologram) through simulation of the hologram viewing process using wave optics principles. This allows the comparison of different holographic representations, as well as adjusting the parameters to obtain improvements in the visual quality of the results. Additionally, the coherent imaging issue of speckle noise degradation in reconstructed views is addressed with existing solutions. Furthermore, the comparative analysis of existing techniques have led us to come up with the following critical contributions. Namely, by relieving the strict requirements regarding the capture of light field for HSs more practical capture setups are enabled. Additionally, a method for reducing speckle noise in coherent CGH reconstructions is proposed.

The thesis is structured as follows. Chapter 2 discusses the theoretical background by discussing the basics of light modeling as rays and waves. This includes introducing the concepts and formulations of light field and how it can be expressed as a set of multiperspective images along with holography and wavefield propagation. Gen-

erating a digital hologram from a discrete LF is introduced in Chapter 3, along with different holographic representations and speckle reduction methods. In Chapter 4, the different CGHs and speckle reduction methods are assessed through numerical simulations. Finally, the thesis is concluded in Chapter 5.

2. BASICS OF LIGHT MODELING AND PROPAGATION

As the title of the thesis suggests, the aim of this work is to analyze the transformation of 3D visual content from one representation to another. Thus, the theoretical background for these representations of light should be considered. In this chapter, light modeling is discussed in terms of rays and waves. First, the concept of light field is introduced along with considerations of its capture and sampling. The second section examines holography and wavefields, whereas their propagation is analyzed in Sec. 2.3.

2.1 Light field

The human visual system (HVS) is a complicated system that interprets visual stimuli depending on different properties of the light entering the eye, e.g. wavelength, intensity and angle of incidence. Consequently, this raises the question of mathematically defining and quantifying the light in a region of space. The concept of LF was first introduced by Gershun in 1936, describing it as the amount of light traveling in every direction through every point in space using light vectors [9]. Let us consider geometrical optics, inferring rays as the fundamental light carrier. As such, any region of space is interpreted as a collection of light rays. The intensity distribution of such rays is described by the so-called the plenoptic function [21].

In the most general case, the plenoptic function is a 7-dimensional function of Cartesian coordinates (x, y, z), propagation direction of the light rays as horizontal and vertical angles (θ, ϕ), wavelength (λ) and time (t). To better comprehend the plenoptic function, one can consider a space filled with idealized pinhole apertures at every location recording the intensity of the light rays at every angle passing through it for each value of wavelength and time. As described by Adelson in [2], the plenoptic function should be considered more as an idealized concept, which cannot be fully

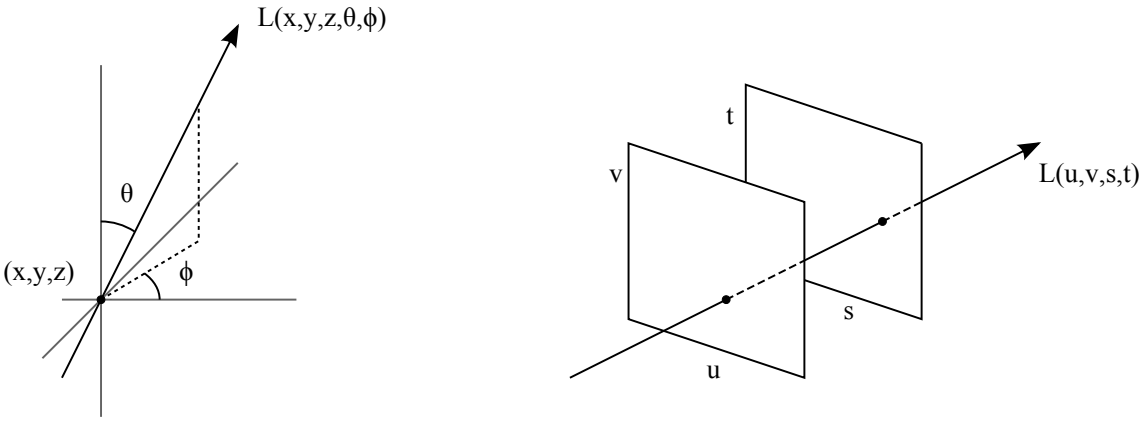


Figure 2.1 *The 5D plenoptic function ray parametrization (left) and an alternative 4D version (right).*

specified for a natural scene. However, as 3D objects in the natural world are viewed by an observer through sampling the pattern of light rays filling the space around the objects, the plenoptic function can be regarded as a communication link between objects and the perceived retinal images. Though by simplifying the presentation of the function, it can be more realistically measured, represented and approximated for 3D image processing purposes. For example, considering only single color (i.e. coherent light, single wavelength) and stationary scenes, the plenoptic function is simplified to 5 dimensions. Furthermore, adopting a two-plane parametrization under the assumption that only the set of rays propagating e.g. toward $+z$ direction are considered, each ray can be parametrized by two coordinates on two different planes resulting in a 4D function (with the exception of rays parallel to the planes if the two planes are parallel). These parametrizations are visualized in Fig. 2.1. For further discussions regarding light fields in the thesis, the 4D formalism is adopted.

The analytical expression of the LF is rarely available, hence it is often approximated from measurement data, such as perspective views or integral imaging. In the case of perspective views, the scene is sampled by capturing a set of 2D images to form a 4D array of pixels representing the discrete version of the LF. This capturing method is fairly simple, as it requires e.g. a moving camera setup or an array of cameras. However, dense sampling of the LF requires a large number of images captured close to each other, thus encumbering the capturing process of dynamic scenes. The capture parameters, e.g. camera spacing, sensor size and number of captured pixels, determine the sampling of the LF as is later discussed in Sec. 2.1.1. The alternative, integral photography [24] inspired solution of replacing the multiple

camera setup with a single camera and an array of lenses, is applicable if the range of views is relatively short. The array of small lenses (i.e. lenslets) is placed in front of a sensor. Furthermore, the perspective views are recorded by the lenslets, each corresponding to a single perspective view. Thus, the LF is captured and the resolution on the two planes (see Fig. 2.1) is determined by the number of lenslets (u,v) and number of pixels behind each lenslet (s,t). If an additional field lens is placed in front of the lenslet array so that the scene is focused on the array, the LF is transposed, effectively switching the resolutions of the planes [21]. Though the system is physically thinner without the field lens, the resolution of the computed views in this setup is low, making the latter arrangement preferred if the thickness is not an issue. Additionally, the arrangement requires only the field lens to be corrected for aberrations instead of each lenslet. The advantages of this setup and development of microlenses has inspired the plenoptic camera, where a microlens array is added between then sensor and lens of a camera [1].

Once the LF is obtained, and if its sampling is dense enough, it can be used for example to generate arbitrary views from the scene within the boundaries of the LF. Specifically, this can be executed by extracting an appropriate slice from a 4D array of pixels representing the LF (from 2D images) in a process known as light field rendering [22]. Another use scenario is synthetic aperture photography (or digital refocusing), which allows to refocus after capturing a snapshot. The light field can also be utilized to generate different projections, such as orthographic and crossed-slits, and panoramic images by extracting slices from it. For further discussion and applications of light field the reader is referred to [21].

2.1.1 Discrete light field as multiperspective images

Utilizing the previously mentioned two-plane parametrization, the LF captured by a camera array can be defined. Let us define two parallel planes where the cameras and their sensors are located. To simplify the analysis and visualizations, a 2D cross-section of the 3D space is considered. The planes are denoted by s and u respectively and the distance between the two planes by l. The two planes are sampled by their respective sampling steps: the distance between adjacent cameras (often referred to as baseline), i.e. the camera plane sampling Δ_s and sensor sampling (i.e. pixel size on the sensor) Δ_u. Assuming a simple pinhole camera model, the rays going through the pinhole towards the sensor plane form a discrete sampled version of the LF between the two planes and the discretization is defined by their sampling

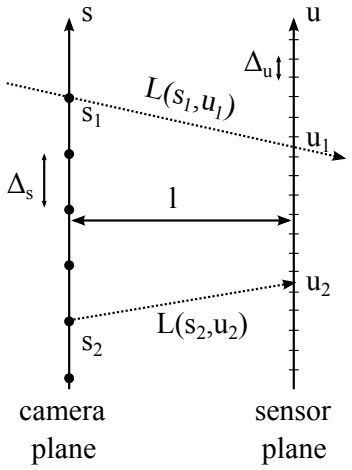


Figure 2.2 *Discrete light field between the camera and sensor planes.*

steps. Fig. 2.2 visualizes the parametrization of this LF. As a further note, each pixel in each camera captures the intensity of an individual light ray. Additionally, depending on the captured orientation of parallax, the LF can be considered as horizontal-parallax only (HPO) or full parallax (horizontal and vertical).

Certain applications, such as holograms (as explained later in Sec. 2.2), mostly require capturing a densely sampled light field (DSLF). The DSLF is defined by the limitations it imposes on the disparity range of the scene between adjacent views, more specifically, it should be in $[-1, 1]$ pixels with respect to the recentered scene plane. A continuous LF can be reconstructed from a sampled LF fulfilling this criterion by utilizing linear interpolation [23]. However, capturing such a LF usually requires extremely dense camera plane sampling, making multi-camera arrays an unfeasible solution. A highly accurate scanning camera rig could be used as an alternative way of obtaining a very small sampling step on the camera plane, though it would restrict the process to static scenes. Another option is to capture a sparse set of views and interpolate the intermediate views to achieve a densely sampled LF. Different methods have been proposed to generate intermediate views, such as the novel view synthesis approaches based on depth-image based rendering (DIBR) [38] which have been utilized to reduce the number of capture views also in CGH generation [17]. However, these methods are highly scene dependent, as the quality of the depth estimation determines their results. The LF reconstruction algorithm based on shearlet domain tiling on the other hand [43], can reconstruct a dense LF from highly sub-sampled LFs without explicitly dealing with depth estimation. As the reconstruction method does not rely on depth information, the results are

significantly less reliant on the captured scene. This view interpolation approach is considered in Sec. 3.2.2 to relieve capture requirements and enable more realistic practical capture setups [33].

2.1.2 Depth field

The two-plane parametrization of a LF captures by default only the direction and intensity of the rays within the LF boundaries. However, some applications require more detailed information about the LF regarding the point-of-origin of the captured rays, i.e. the location of the light source in the scene. This required depth information can be either captured by special sensors, such as time-of-flight, or estimated by post processing the LF.

In this thesis, the information about the location of scene points is assumed to be presented in the form of depth maps. A depth map is paired with a corresponding image to provide additional information about the 3D space captured by the discrete LF. For each captured intensity value, the depth map contains a numerical value representing the perpendicular (regards to the sensor) distance from the camera to the area captured by the corresponding pixel. By acquiring the depth value and by utilizing the known capture parameters, the point-of-origin for each ray can be solved by trigonometry, thus providing additional information of the scene for later calculations.

2.1.3 Recentering camera model

Let us consider a discrete LF as captured by cameras following the recentering model. That is, a perpendicular recentering image plane is defined at distance d from the camera array and images are captured such that the center of projection for each camera crosses the center of the recentering plane, as seen in Fig. 2.3. This can be achieved by either shifting the sensor behind the aperture or by appropriately resampling and cropping regular wide field-of-view (FOV) images. The advantage of using a recentering camera model is that less of the FOV is wasted to capture rays irrelevant to the scene. Furthermore, depending on the location of the edge-most cameras and the size of the scene, a regular camera model can require extremely wide FOV or long distance from the scene to capture the required rays, making such capture setup cumbersome. The magnification ratio of the cameras define the

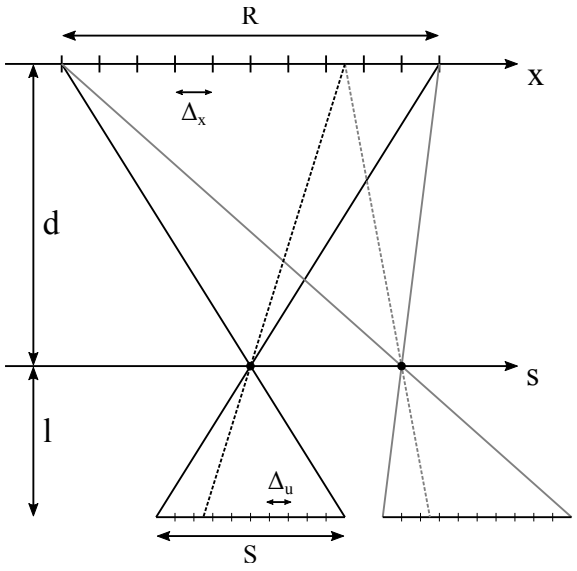


Figure 2.3 *The recentering camera model and its parameters.*

size of the recentering plane R and the pixel size on the plane Δ_x as $R = dS/l$ and $\Delta_x = d\Delta_u/l$ from the camera sensor size S and the sensor plane sampling step Δ_u respectively. Due to the properties of the recentering model, each pixel at the image plane corresponds to the same pixel at the captured image, thus capturing the angular distribution of the LF defined on the recentering plane.

Assuming a set of recentering camera model images and depth maps, i.e. discrete LF and depth, the locations of the captured scene points can be resolved. Let us examine an image captured by a camera at location s, at distance d from the recentering plane and a single pixel of the particular image. By knowing the camera parameters and the pre-defined recentering plane, each pixel has a known position at the image plane. For the pixel center at x, a depth map value z is also known. As demonstrated in Fig. 2.4, this results in three right-angled triangles from which the location of the corresponding scene point x_p can be obtained as

$$x_p = x + \frac{(d-z)(s-x)}{z}. \tag{2.1}$$

Similarly, the incidence angle θ for each ray can be resolved directly from the capture geometry

$$\theta = \tan^{-1}\left(\frac{s-x}{d}\right). \tag{2.2}$$

In this thesis, the multiperspective images are assumed to be captured by the re-

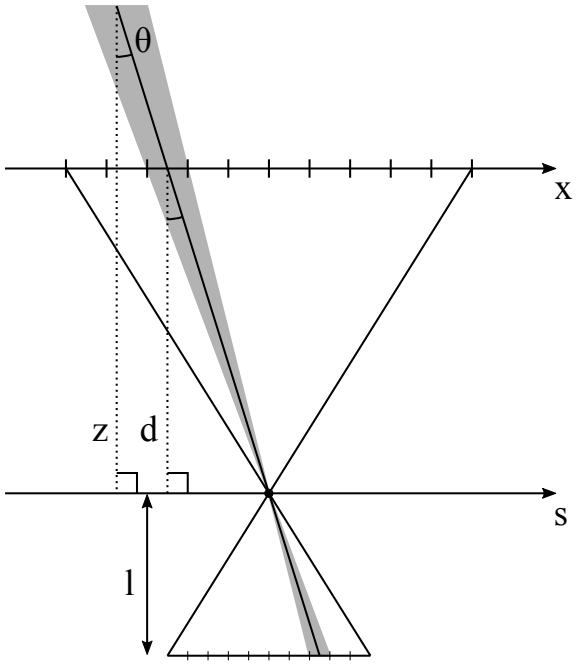


Figure 2.4 *Obtaining the location of a scene point from capture geometry and a depth map.*

centering camera model.

2.1.4 Densely sampled light field

As was previously discussed, a densely sampled light field is limited by the disparity range of adjacent views to be in $[-1, 1]$ pixels with respect to the recentered scene plane. Thus, the sampling requirements for a DSLF can be defined for a given scene. Let us define the z-axis such that it is 0 at the camera plane and that the value of z increases towards the scene and the recentering plane. Assuming a recentering plane at distance z_0 from the cameras and a scene that is bounded by z_f and z_b from the front and the back, respectively, w.r.t the camera plane as seen in Fig. 2.5. In order to fulfill the disparity range requirement, the distance between adjacent cameras has to be chosen according to

$$\Delta_s = \min\left\{\frac{\Delta_x z_b}{z_b - z_0}, \frac{\Delta_x z_f}{z_0 - z_f}\right\}. \tag{2.3}$$

That is, both scene boundaries define their separate limitation for the camera baseline, of which the stricter one is chosen. If the scene is entirely either in front of

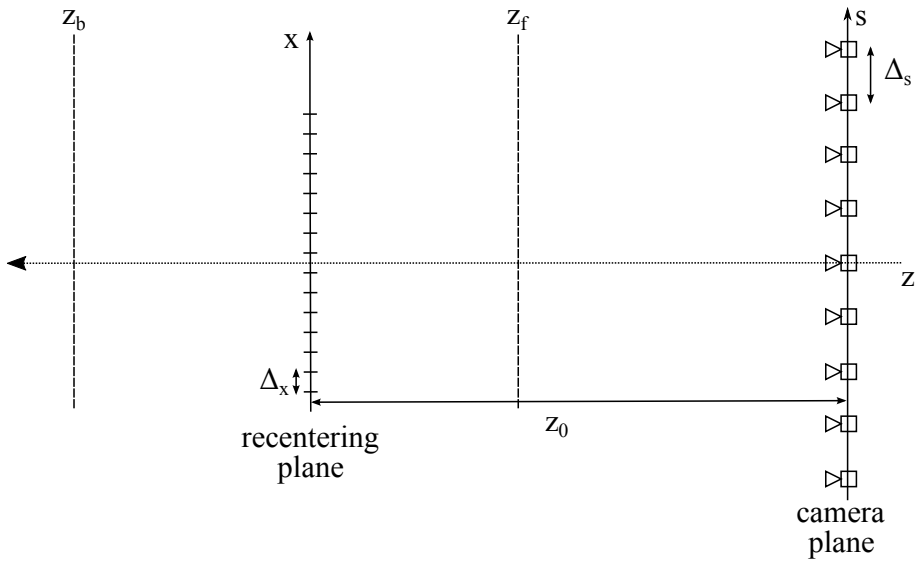


Figure 2.5 *The parameters and capture of a densely sampled light field.*

or behind the recentering plane, the back or front limit produces a negative value for the camera sampling distance and can be omitted, respectively, and Δ_s can be chosen in accordance with the remaining depth limit. This approach can be used to define the dense LF sampling for a fixed scene.

Alternatively, the camera sampling distance Δ_s can be first chosen and then the scene can be limited to meet the DSLF requirement. Let us define the depth limits of the scene z_b and z_f for the set of cameras sampled at the fixed distance Δ_s as

$$z_b = \frac{\Delta_s z_0}{\Delta_s - \Delta_x}, \tag{2.4}$$

$$z_f = \frac{\Delta_s z_0}{\Delta_s + \Delta_x}. \tag{2.5}$$

In applications where the camera sampling and distance from the recentering plane are fixed, this definition can be utilized to limit the scene such that the captured LF is densely sampled. This is especially useful in hologram generation from discrete LF, as is discussed in terms of light field and hologram sampling in Ch. 3.

In conclusion, the LF imposed by multiperspective images provides a solution for capturing and representing a 3D scene by the light rays emitted through a space defined by the capture properties. The array of images can be considered as a discrete 4D representation of the scene. This can be further extended by including

a set of depth maps to gain information of the scene points captured in the discrete LF. Combining a sufficiently dense discrete LF and depth maps, the capture 3D scene is described well enough to consider alternative representations to store and display it. In the next section such an alternative is discussed in the form of digital holography, simultaneously transitioning from expressing light as rays to light as waves.

2.2 Holography and wavefields

Optical holography is based on the physical phenomenon interference and diffraction, which are used to record and reconstruct a 3D image [13]. A hologram is generated from the interference between the light scattered from the recorded object and a mutually coherent reference beam. The holographic medium records both the magnitude and phase of the incident object wavefront in coded form. Thus, the variations in intensity and the direction of the light are recorded as fringe patterns on the hologram. By illuminating the fringes with the same reference beam as during the recording process, the hologram reconstructs the recorded 3D scene [20]. On the other hand, the discrete computational version of a hologram, known as computer-generated hologram (CGH), is calculated through simulating the physical processes numerically [6].

Let us examine the optical holography process of recording an object by examining the electrical fields. The object to be recorded on the hologram scatters light E_O and interferes with the light from the reference source E_R. Assuming a normalization of the electric field amplitude such that the squared magnitude equals to optical intensity, the total electric field I_T incident on the hologram is defined as [26]

$$I_T = |E_O + E_R|^2 = |E_O|^2 + |E_R|^2 + 2Re\{E_O \cdot E_R^*\}, \tag{2.6}$$

i.e. the interference of the light from the object and reference source. The expanded equation includes three components, known as object self-interference $|E_O|^2$, reference bias $|E_R|^2$ and useful fringes $2Re\{E_O \cdot E_R^*\}$. The first two terms are both unwanted as they mostly affect the reconstructed images negatively, as the object self-interference can produce image artifacts and reference bias adds a constant intensity along the hologram, thus wasting the dynamic range of the holographic medium.

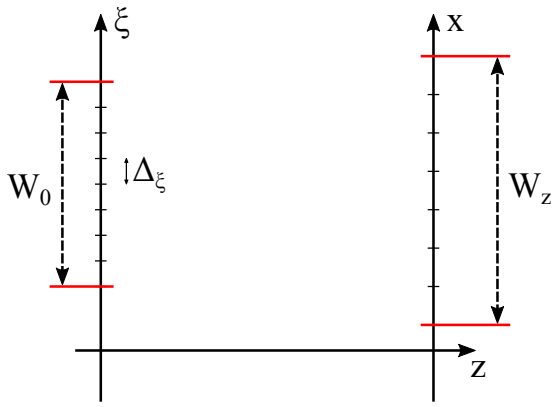


Figure 2.6 *Parameters of the wavefields along the input axis x and the propagated location s at distance z.*

In the context of this thesis, the complex object wave is considered as the hologram. as it contains the necessary information. This choice is further motivated by the fact that the conjugate object (reference) wave would lead to additional noise in the reconstruction step. By avoiding this, different approaches for hologram generation can be evaluated more reliably against each other. The object wave is represented as a complex-valued wavefield function defined on a certain plane perpendicular to the hologram. The notation to be used in this thesis for the object field function at depth z is $O_z(x)$. Furthermore, the terms hologram and wavefield are used interchangeably in further discussions.

2.3 Wavefield propagation

Considering a CGH, the discrete hologram has certain parameters which should be examined. Let us define the hologram parameters as the hologram size W_0 and pixel size Δ_ξ. It should be noted, that the resolution of the hologram N can be determined from these values. Additionally, the axes are defined in the following manner: (ξ, η) as the hologram plane, (x, y) as the plane of the secondary field and z-axis as the propagation direction, such that the hologram is located at $z = 0$. By propagating a wavefield (or a hologram) with these parameters, the properties of the field at alternative locations along the z-axis can be examined. Specifically, the analysis is realized in both spatial and spectral domain. The hologram discretization is visualized in Fig. 2.6.

Analyzing the propagation and diffraction of light as an electromagnetic wave pro-

vides a foundation for studying the effect of different hologram parameters on the field properties at the assumed viewer location. Depending on the desired model or the problem at hand, different diffraction kernels can be used in the field propagation models. Assuming the Fresnel diffraction kernel, the unknown field $U_z(x, y)$ at axial distance z from a known field $U(\xi, \eta)$ at $z = 0$ can be obtained through Fresnel transform [10] defined as

$$\begin{aligned} U_z(x,y) &= \mathcal{F}_z\{U(\xi,\eta)\}(x,y) \\ &= \frac{\exp(j2\pi z/\lambda)}{\sqrt{j\lambda z}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U(\xi,\eta) \exp\left\{\frac{j\pi}{\lambda z}\left[(x-\xi)^2+(y-\eta)^2\right]\right\} \mathrm{d}\xi\mathrm{d}\eta. \end{aligned} \quad (2.7)$$

The analytical solution of the Fresnel transform integral can be solved in some cases, though mostly such cases are used to evaluate the accuracy of numerical solutions of the transform and to gain further insight into the propagation process. One such important analysis result, as noted in [18], is how the spacial extent of the field behaves during propagation. Assuming a field with finite spatial extent in the original location and known spatial frequency components, the extent of the propagated field is relative to both of these factors. The spatial extents and the maximum spatial frequencies of the fields are approximated in terms of their energy bandwidth. That is, the range for which 99.9 % of the signal's power is located, is considered as the spatial extent of the field [25]. Similarly, the maximum spatial frequency value is approximated as the frequency limit for which 99.9 % of the power in spectral domain is located. The extent of the field increases as it propagates according to

$$W_z = W_0 + 2\lambda z f_\xi, \quad (2.8)$$

where W_z is the spatial extent of the field at distance z and f_ξ is the maximum spatial frequency value of the original field.

In practical applications, however, numerical approaches for propagation are required. Thus, let us define the finite extent discrete field uniformly sampled at distance Δ_ξ and Δ_η with $N \times M$ total number of samples as

$$\mathbf{U} = \begin{bmatrix} U_{1,1} & U_{1,2} & \cdots & U_{1,N} \\ U_{2,1} & U_{2,2} & \cdots & U_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ U_{M,1} & U_{M,2} & \cdots & U_{M,N}. \end{bmatrix} \quad (2.9)$$

Replacing the continuous field in the Fresnel transform with this, the diffraction integral of Eq. 2.7 is obtained through summation as

$$U_z(x,y) = \frac{\Delta_\xi \Delta_\eta}{\sqrt{j\lambda z}} \sum_{m=1}^{M} \sum_{n=1}^{N} U_{n,m} \exp\left\{ \frac{j\pi}{\lambda z} \left[(x-\xi_n)^2 + (y-\eta_m)^2 \right] \right\}, \tag{2.10}$$

where $n, m \in \mathbb{Z}^+$ and $U_{n,m} = U(\xi_n, \eta_m)$. This transform solution is known as the direct calculation approach. A careful analysis of the approach provided in [18] yields a sampling rule of $W_z < \lambda z / \Delta_\xi$ to guarantee no overlapping replicas of the diffraction field. As an alternative solution, the Fresnel transform can be performed with the aid of Fourier transforms resulting in a spectral calculation approach. Let us define the forward Fourier transform for a complex 2D signal $f(x,y)$ as

$$\hat{f}(u,v) = \mathrm{FT}\{f(x,y)\}(u,v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) \exp\left[-2\pi j(ux+vy)\right] \mathrm{d}x\mathrm{d}y, \tag{2.11}$$

where u and v are the spatial frequency coordinates, and the inverse Fourier transform for $\hat{f}(u,v)$ as

$$f(x,y) = \mathrm{IFT}\{\hat{f}(u,v)\}(x,y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}(u,v) \exp\left[2\pi j(ux+vy)\right] \mathrm{d}u\mathrm{d}v. \tag{2.12}$$

Thus, Eq. 2.7 can then be expressed as

$$U_z(x,y) = \mathrm{IFT}\{\exp(-2\pi j \lambda z u^2 v^2) \mathrm{FT}\{U(\xi,\eta)\}(u,v)\}(x,y). \tag{2.13}$$

Similarly to the direct calculation approach, a sampling requirement for this approach can be obtained through further analysis to reduce the effect of replicas in the propagated field as a function of the spatial frequency sampling Δ_v to be $W_z < 1/\Delta_v$. Importantly, in both cases the sampling requirement can be connected to the properties of the original field through spatial extent in Eq. 2.8. In this thesis, the format in Eq. 2.13 is used to implement the Fresnel transform for wavefield propagation simulations.

Both the direct and spectral calculation approach are desirable to implement as the control over the output spatial variable x is retained. In order to achieve efficient realizations of the approaches, fast Fourier transform (FFT) based solutions are apparent. Due to the characteristics of the FFT algorithm, the properties of the

output window can be controlled by adding zeros to sampled field vector $\mathbf{U}$. For example, by padding the vector with zeros the number of samples can be increased without altering the sampling distance Δ_ξ in the original vector, while still keeping the spatial frequency extent constant, thus increasing the number of samples in the output window of FFT over the same frequency domain. Alternatively, zeros can be inserted between the samples of $\mathbf{U}$ resulting in an increased output window extent. This is achieved without changing the Fourier distribution, thus allowing to examine the distribution for its replicas over a larger spatial frequency extent.

3. FROM LIGHT FIELD TO HOLOGRAM

In the previous chapter two different representations of the light in a region of space have been discussed in the form of LF and hologram, which model light using rays and waves, respectively. However, the link between these two is yet to be established. For this purpose, this chapter discusses the transformation from a 4D discrete LF to a 2D complex wavefield corresponding to a hologram. For analysis of the inverse transform from hologram to light field the reader is referred to [50]. The wavefield can be estimated in several different ways regarding the shape of the wavefronts contributing to the total field, as well as by introducing segmentation and other approximations. Since generating an ideal coherent hologram from discrete LF data without the exact model information is not feasible, three different approximations are considered in this thesis. The hologram calculations are divided into incoherent and coherent methods based on the requirements they impose on the illumination (or reconstruction), or in other words, the data they utilize. In incoherent case, only LF (i.e. multiview images) is used, whereas in the coherent case the depth values of rays are also utilized (to obtain the necessary phase information).

Before considering any of the holographic representations, it is important to define the notation to be used in further discussions regarding the capturing setup, the discrete LF and the hologram. In this thesis, the LF is assumed to be captured by an array of cameras as multiperspective images. Thus, we define three parallel planes, which parametrize two 4D light fields. The hologram plane is the location of the (recorded) hologram that is placed accordingly relative to the 3D scene. The camera array is located at the camera plane and behind it along the direction of the z-axis is the sensor plane. The following notation is adopted: the hologram plane is denoted as (x, y), the camera plane as (s, t) and the sensor plane as (u, v). Most of the figures and equations consider a cross-section of the 3D space for simplicity, however, the expansion to the full 3D case is usually straightforward. Such a cross-section, illustrating the relative positions of x, s and u axes is shown in Fig. 3.1. The discretizations on the planes are Δ_x, Δ_s and Δ_u accordingly. Please note, the

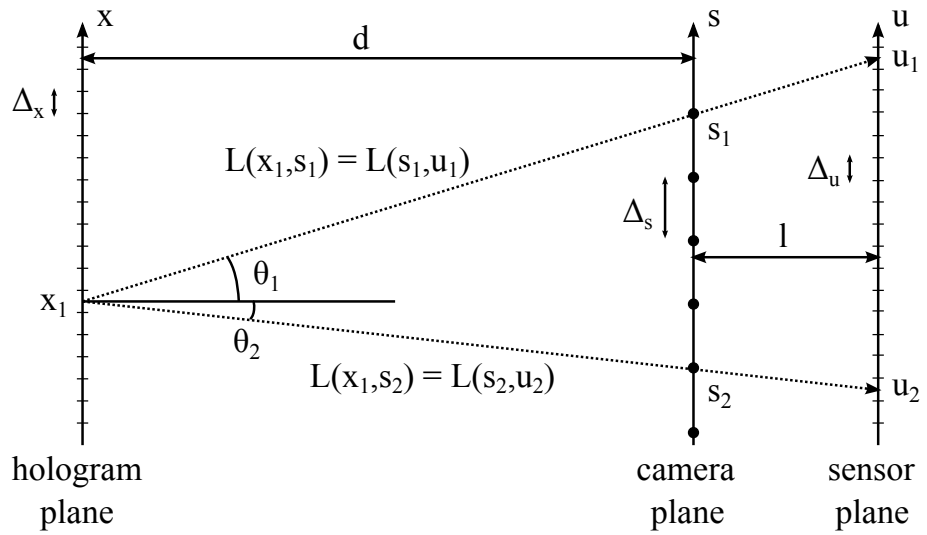


Figure 3.1 *Sampled light field and how the rays correspond to the hologram plane, the capture setup and the sensor plane (i.e. captured pixels).*

difference between hologram plane sampling Δ_x and the hologram pixel size X_x. Additionally, the distance between the hologram and camera planes and the distance between the camera and sensor planes is denoted by d and l respectively. Further notations are explained as they are introduced.

Let us consider the light fields parametrized between the hologram and camera planes as well as between the camera and sensor planes, defined as $L_1(x, s)$ and $L_2(s, u)$ respectively. Using the following notation, the light ray propagating at an angle θ_i from the hogel at x_i is captured by the camera at s_i and on the sensor plane at u_i. The relation between the two LFs is

$$L_1(x, s) = L_2(s, u_x), \tag{3.1}$$

where $u_x = s + l(s - x)/d$. The discrete LFs $L_1[m, i]$ and $L_2[i, k]$ have a one-to-one correspondence, if the following criteria are satisfied: the magnification equation $\Delta_x = \Delta_u d/l$ holds and the baseline is chosen such that the disparity D between adjacent views is an integer amount of pixels $D = \Delta_s l/d \in \mathbb{Z}$ [33]. As a result, the light field $L_1[m, i]$ is obtained directly from the captured images (see Fig. 3.1).

As the rays emitted from the hologram plane and the ones captured by the perspective views are symmetrically permutated along the horizontal and vertical axis, the captured rays should be reordered for processing purposes. Assuming that the

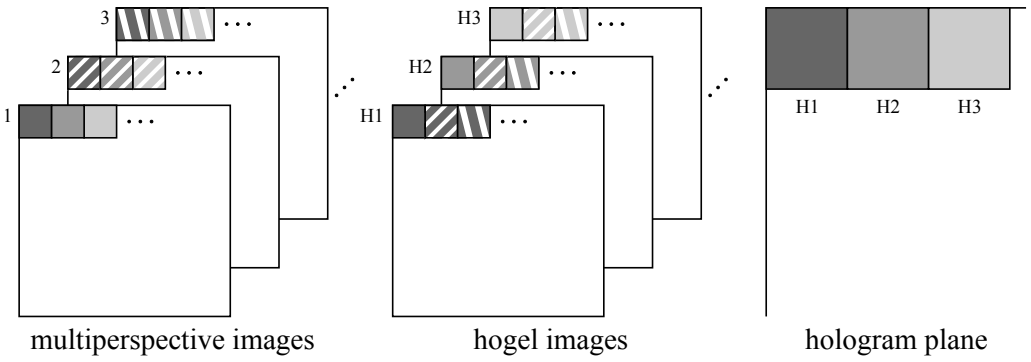


Figure 3.2 *Permutating the captured pixels from multiperspective images to the hologram plane.*

capture parameters have been chosen properly, each captured image contains a single ray (in the form of a pixel value) corresponding to a particular segment on the hologram plane. For example, assuming recentering camera model, the segment at the left-most edge on the hologram plane (viewed from the front) contains rays captured by the right most pixel of each view as seen in Fig. 3.2. Reordering the rays results in a spectral representation of the hologram, as each ray is considered as a spatial frequency component as will be discussed in Sec. 3.2.1.

3.1 Sampling requirements

As a discrete system, the light field and hologram are both subject to sampling requirements. The sampling of the LF is highly dependent on the sampling of the hologram, as the sampled LF is utilized as the input data for the transformation to a hologram. The hologram is sampled in two different ways: Let us consider the single plane space-angle representation of the light field on the hologram plane. The spatial sampling corresponds to the hologram plane segmentation size, whereas the angular sampling corresponds to a set of propagation directions for the waves to be emitted in reconstruction. Usually, it is a good practice to determine those parameters based on the requirements of the human visual system (HVS) [26]. Alternatively, the optimum sampling of stereograms can be analyzed through other means, such as the modulation transfer function characteristics for HS [14] and the spatial frequencies of adjacent segments for PAS [16].

Even though the lateral and depth spatial resolution of HVS are dependent on the viewing conditions, such as brightness and motion, the general properties of HVS

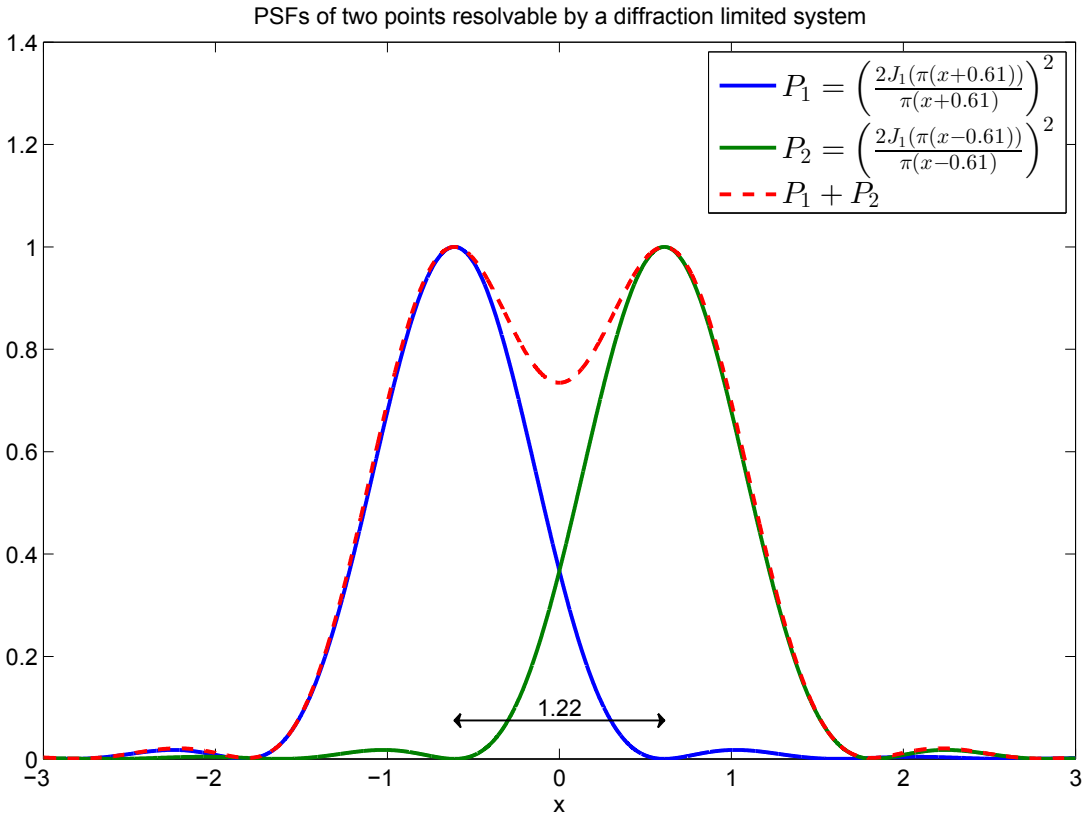


Figure 3.3 *The Rayleigh criterion for two resolvable points with equal intensity.*

accuracy can be applied to derive certain minimum sampling requirements for the hologram. For example, the lateral acuity dictates the minimum distinguishable distance between two points at a certain viewing distance. A common metric for this is the Rayleigh resolution criterion which states that two equal intensity point sources can be distinguished when the point spread function (PSF) maximum of one point overlaps with the first minimum of the other point [36], as visualized in Fig. 3.3. The Rayleigh criterion defines the lateral acuity at viewing distance d for a diffraction limited imaging system as

$$\Delta_x^{HVS} = \frac{1.22\lambda d}{T}, \tag{3.2}$$

where T is the aperture size (pupil size in HVS) [10]. By limiting the spatial sampling of the hologram plane to be less than or equal to the diffraction limit, the perceived image resolution will be maximized, i.e. Δ_x should be chosen to be less than Δ_x^{HVS} in Eq. 3.2 for an assumed viewing distance d and pupil diameter T. Similarly, the pupil size is the limiting factor for spectral sampling. Let us consider the rays captured by

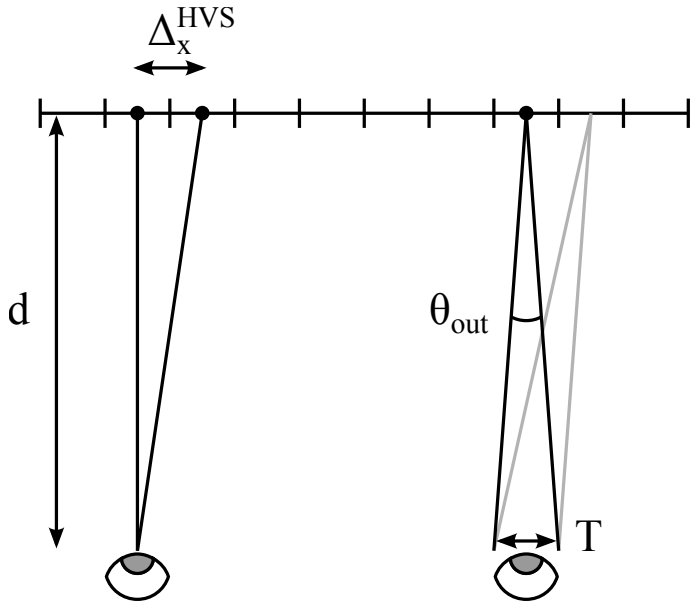


Figure 3.4 *Sampling requirements of the human visual system.*

the LF within a single hologram segment. In order to perceive continuous motion parallax, at least two rays should be entering the pupil at any location of the viewer within the viewing zone (as visualized in Fig. 3.4) [26]. Thus, the upper boundary for the angular sampling $\Delta\theta_{out}$ is defined as a function of pupil size and viewing distance

$$\Delta\theta_{out} < \tan^{-1}\left(\frac{T}{d}\right). \tag{3.3}$$

Depending on the holographic representation, the angular sampling of the rays diffracted by a discrete hologram is in certain cases determined by the fixed spatial frequency values. In these cases, both the hologram plane pixel size X_x and the parameters defining the spatial frequency grid should be chosen accordingly. Otherwise, the CGH records the rays captured by the discrete LF, and thus, the angular sampling needs to be considered through the capture parameters.

As the spatial and angular sampling of the hologram are fixed accordingly, the discrete LF capture parameters are defined strictly by these values. Thus, it is good practice to choose the camera sampling distance Δ_s to capture at least two rays within the assumed pupil diameter. That is, for an assumed viewing distance d_{view} from the hologram, the camera sampling is limited to be less than Td/d_{view}. For a recentering camera model, the recentering plane should be placed at the location of the hologram plane to capture the necessary light rays, i.e. the correct LF, to

generate the CGH. As the discrete LF capture parameters are fixed accordingly, the recorded scene must be chosen such that the discrete LF fits within the limits defined by the scene for a DSLF (see Eq. 2.4 – 2.5) and can be then resampled to obtain the required light rays for the CGH calculations as was explained in Sec. 2.1.4.

However, if the scene cannot be limited in this manner, denser sampling of the LF should be considered. That is, by reducing the camera sampling distance in accordance with the DSLF theory, the LF would be oversampled w.r.t the HVS requirements. In this case pre-filtering will be necessary before resampling in order to avoid aliasing. This can be achieved e.g. by increasing the sampling accuracy of the hologram beyond the capabilities of the HVS. Alternatively, if this is not possible e.g. due to limitations in display equipment, anti-aliasing of the discrete LF data is required. By blurring scene regions outside the DSLF range, the highest frequencies are filtered and thus, anti-aliasing is accomplished. In this thesis, it is assumed that the scene satisfies the DSLF limits for the requirements set by the HVS.

3.2 Incoherent hologram

The holographic representations examined in this thesis are divided into two categories based on their illumination requirements. First, holograms characterized as incoherent are discussed. In terms of imaging systems, incoherency is observed as completely uncorrelated wavefields at any two points in the scene [36]. Consequently, all imaged points are summed in intensity instead of amplitude. In holography, incoherency allows removing the prerequisite for a spatially coherent reference beam, thus enabling hologram acquisition for passively illuminated scenes. In terms of CGH generation, incoherent holograms can be characterized so that they utilize only the intensity of light. These types of holograms can be generated from photographs (i.e. multiperspective images), a solution often applied in digital holograms, or by utilizing incoherent optical equipment for analog holograms. As an example of incoherent holography, the holographic stereogram (HS) is discussed and its generation from LF data in the context of multiperspective images is presented. Moreover, an inherent issue with hologram reconstructions in the form of speckle noise as well as reducing its effects are examined.

3.2.1 Holographic stereogram

Holographic stereogram constitutes the most common incoherent holographic representation technique. The hologram is commonly generated from a large number of multiperspective images of a 3D scene, which are processed to form parallax-related images (i.e. permutated elementary images). These provide the directional information on the hologram plane and can be considered as the spectral representation of the hologram through a collection of spatial frequency components [12].

Utilizing the notations introduced earlier in this chapter, the calculation of the CGH from discrete LF is presented. The object field for a holographic stereogram O_{HS} is obtained as a superposition of (rectangularly) windowed plane waves with amplitudes defined by the corresponding discrete LF samples [49]. As shown in Fig. 3.5, the plane waves are emitted from the holographic elements (hogels) to different directions. For the sake of simplicity, the formulas from this point onwards are given in 1D. Please note that the extension to 2D is straightforward. The object field at the hologram plane is given by

$$O_{HS}(x) = \sum_m \mathrm{rect}\left(\frac{x - m\Delta_x}{\Delta_x}\right) \times \sum_i \sqrt{L_1[m,i]}\exp\left(j2\pi f_x^{mi} x\right) \tag{3.4}$$

where f_x^{mi} is the spatial frequency on the x-axis (for hogel m and ray i). Assuming perpendicular reference beam (in relation to the hologram plane), the frequency is related to the incidence angle θ_x^{mi} of a ray (along the x-axis) according to the grating equation as [34]

$$f_x^{mi} = \frac{\sin\theta_x^{mi}}{\lambda}, \tag{3.5}$$

where λ is the wavelength of the monochromatic light. The inner sum in Eq. 3.4 is the spatial pattern to-be-written in the corresponding hogel, which can be obtained by inverse Fourier transform of the permutated perspective image segments (see Fig. 3.2). In practical terms, efficient implementations can be realized by utilizing inverse fast Fourier transform (IFFT) to obtain the spatial pattern inside each hogel. However, it should be noted, that in this case the discrete spatial frequency values are fixed, thus requiring resampling of the continuous set of spatial frequencies. Furthermore, the size of the calculated FFT signal can be controlled to vary the density of the spatial frequency grid. Increasing its size achieves a denser frequency grid at the cost of more computational time.

Due to the simplicity of the wavefield approximation, only the discrete LF is required for generating an HS and it can be generated efficiently by utilizing FFT algorithms. Furthermore, assuming that the sampling requirements according to the HVS have been fulfilled during the discrete LF capture, the HS provides proper parallax cues and the HVS perceived resolution is maximized. However, since the ability to reconstruct a point in the scene is limited by the hogel size for the HS, it cannot represent deep scenes well [12]. This also affects the accommodation cue negatively, thus limiting the ability to focus on object deep in the scene. Due to its popularity both in research and practical applications, HS serves as an adequate baseline for holographic 3D scene reconstruction.

3.2.2 Light field capture by sparse set of cameras

The LF capture requirements for HSs in terms of multiview images are usually strict, that is, the set of view points is quite dense due to the requirements imposed by the HVS. Due to such sampling requirements, the discrete LF has to be captured by a highly accurate scanning camera, which is a cumbersome process and limits the capture process to static scenes. Thus, it is critical to relieve this requirement for practical applications so as to enable the use of more convenient multi-camera setups, with which it will become possible to capture dynamic scenes. This can be achieved by the use of light field reconstruction algorithms, which can reconstruct dense light fields from a sparse set of views. In our study [33], we propose that through the use of shearlet transform, one can significantly relieve the sampling requirements of HS, which enables utilization of multi-camera setup during the capture of the light field.

In shearlet transform based LF reconstruction, the LF is analyzed through an epipolar-plane image (EPI) representation, which can be formed by taking slices of the 4D LF. The dense LF is then reconstructed from the sparse set of samples by reconstructing each densely sampled EPI slice [43]. The reconstruction is solved by using regularization in the shearlet domain, as the LFs have sparse representation in this domain [42]. Moreover, the frequency domain of the EPI is tiled by shearlet atoms so that the directions in the tiling represent each disparity value in the EPI. The unknown EPI samples are reconstructed by an estimation model for the vectorized versions of the densely sampled EPI $\mathbf{a}$ and decimated EPI $\mathbf{b}$ as $\mathbf{b} = \mathbf{Ha}$, where $\mathbf{H}$ is the masking matrix for known sample positions. An iterative hard thresholding

procedure is utilized for obtaining the reconstruction [42], presented as

$$\mathbf{a}_{n+1} = \mathbf{S}^* \left\{ \mathbf{T}_{\lambda_n} \left\{ \mathbf{S} \left[\mathbf{a}_n + \alpha(\mathbf{b} - \mathbf{H}\mathbf{a}_n) \right] \right\} \right\}, \tag{3.6}$$

where $\mathbf{S}$, $\mathbf{S}^*$ are the shearlet analysis and synthesis transform matrices, respectively, α is an acceleration coefficient and $\mathbf{T}_{\lambda_n}$ is the hard threshold operator for the threshold λ_n. The solution is acquired as $\mathbf{a}_n$ after a sufficient number of iterations.

The LF reconstruction algorithm is shown through experiments in [33] to be able to reconstruct the dense LF for HS generation from a sparse set of views decimated by factor as much as 8. The HS reconstruction images from such set of images achieves comparable visual quality to a originally captured dense LF, as well as performing similarly as when using a DIBR reconstructed LF. The advantage, however, of using the shearlet-based method is that it does not require depth estimation which can suffer from artifacts such as misregistration. The LF reconstruction algorithm is shown in [43] to be capable of reconstructing from even sparser (than factor of 8) set of views, which in terms of practical capture setups suggests that using this method enables the use of wide baseline capture. For further discussion and details of the solution, the reader is referred to [33].

3.2.3 Speckle noise reduction in incoherent CGH

Reconstructing simulated views from CGHs is not entirely straight forward, as certain optical issues need to be considered. Importantly, random phase needs to be added to the object light (i.e. hologram) diffusing it, and thus avoiding its concentration on the hologram. However, adding random phase introduces another problem in the reconstructed image in the form of speckle noise. As this noise heavily degrades the visual quality of the simulated views, it needs to be addressed properly. Several different solutions have been proposed to reduce speckle noise, such as decreasing the spatial coherence by diverging light with a diffuser [28, 46], correlation fringe averaging [15], reducing temporal coherence with LED illumination [48], the random phase-free CGH [35] and iterative phase retrieval algorithms [27]. Two, of such methods, have been chosen to be evaluated in this thesis. Namely, the random averaging approach and a pixel separation method.

The first speckle reduction method we consider is random averaging [3]. Its basic principle is computing uncorrelated CGHs, which are propagated in sequence. Due

to the added random phase, the speckles display spatial randomness. The average contrast of the summed speckle noise patterns is lowered, thus the quality of the final reconstructed image can be improved by averaging reconstructed images from multiple CGHs. To obtain statistically independent holograms for the method to be effective, however, the random phase distributions of the different CGHs must be statistically independent. For N CGHs the speckle reduction capability of the method is proportional to $\sqrt{N}$. Practical application of random averaging requires a high-speed (refresh rate) display device to be able to display the different CGHs in sequence.

Pixel separation [39] is considered here as an alternative speckle reduction solution. A considerable source of speckle noise in the reconstructed images is the interference between adjacent hogels. Particularly, one needs to consider the point-spread functions (PSF) of the pixels perceived from those hogels, as the side lobes of the PSFs of nearby pixels overlap on the sensor plane and cause interference. To reduce this interference, and thus speckle noise, the pixels within a certain neighbourhood are separated from each other and multiple CGHs are created from the sparse pixel sets. In incoherent holography, the separation has to be done on the hologram plane over hogels, as explicit information about the location of point sources in the scene is not available. Similarly to random averaging, the speckle reduced reconstruction image is obtained by displaying the hogel sub-sampled CGHs at high-speed refresh rate, or in computational simulations by propagating them in sequence and incoherently summing the reconstructed images.

3.3 Coherent holograms

Coherent holograms form the second category of holograms. As opposed to incoherent holography, coherent holograms require spatially and temporally coherent reference beam which, in practice, means illuminating the hologram with a high-power laser. In CGH, the coherency refers to the continuity of the wavefronts, i.e. for a given scene primitive element (e.g. point source) the corresponding wavefield should be continuously recorded throughout the hologram pixels or segments. Achieving this requires knowledge of either the positions of points sources relative to the hologram or other point sources in the scene, depending on how the wavefield is modeled. Thus, additional information of the scene supplementary to the discrete LF is needed in CGH in the form of depth values. This is assumed to be given as a set of depth maps corresponding to the multiperspective images as explained in

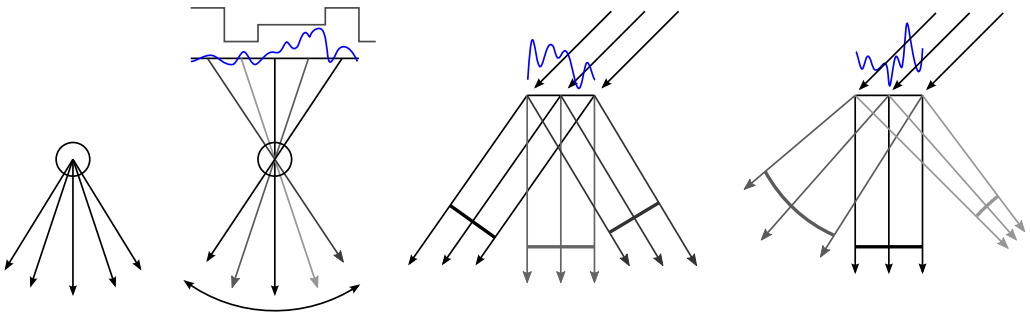


Figure 3.5 *Different display elements, from left to right: picture element (pixel), directional element (direl), holographic element (hogel) and wavefront element (wafel). Adapted from Figure 1 in [37].*

Sec. 2.1.2. As mentioned before, in real scenarios, such depth maps can be either directly sensed via depth sensors or calculated by post-processing the light field, as done in [50]. In computer graphics, on the other hand, depth maps can be easily obtained using a renderer (ray tracer engine, e.g. Blender). The coherent hologram representations are presented starting from the most accurate one in comparison to an ideal coherent hologram, simplifying step-by-step the representation towards the later ones. The equations for obtaining the complex-valued wavefield at the hologram plane are presented assuming the discrete LF and depth field as the input. Furthermore, solutions for speckle noise reduction are considered similarly to incoherent holography.

3.3.1 Diffraction-specific coherent panoramagram

The requirement of capturing the LF at hologram resolution can be relaxed in such a way that a single camera pixel (i.e. single captured ray) corresponds to several hologram pixels, effectively resulting in a spatially under-sampled hologram. The segment consisting of several hologram pixels is referred to as a hogel or a wavefront element (wafel). Instead of a constant value emitted to all directions as in a regular pixel, these elements emit a set of wavefronts with controllable intensity to different directions. A hogel employs planar wavefronts, whereas the wafel can also control the curvature of the wavefront. Fig. 3.5 visualizes the differences of these display elements.

The diffraction-specific coherent panoramagram (DSCP) introduced in [37] is a holographic representation utilizing wafels. In order to obtain the correct wavefront

curvatures in addition to the intensities, the positions of the captured points in the scene are needed. Utilizing the set of depth maps corresponding to the discrete LF, each ray corresponds to a point in the scene and its location can be solved as explained in Sec. 2.1.3. The DSCP object field O_{DSCP} at the hologram plane for such data is defined as

$$O_{DSCP}(x) = \sum_m \text{rect}\left(\frac{x - m\Delta_x}{\Delta_x}\right) \times \sum_i \frac{\sqrt{L_1[m,i]}}{r^{mi}} \exp\left[\frac{j2\pi}{\lambda}\left(\sqrt{(x - x^{mi})^2 + z_{mi}^2} - z_{mi}\right)\right], \tag{3.7}$$

where x^{mi} and z_{mi} are the Cartesian coordinates (in x and z respectively) of the point in the scene captured by the discrete light ray at $[m, i]$. The exponential function defines the curvature and alignment of the segmented wavefronts and the intensity is defined by the corresponding LF sample. All points seen through a wafel by the discrete LF are recorded as a segment of the complete wavefront, or inversely examining, the complete spherical wavefront emitted by a point is segmented by the wafels and recorded on the segments that can see the point. As the discrete LF is obtained from multiperspective views, any occlusions in the recorded scene are intrinsically handled, and thus recorded correctly to the appropriate wafels.

The advantage of segmenting the wavefield in comparison to recording it entirely at the hologram resolution is that the capture requirements are relieved significantly, while still preserving most of the wavefield approximation accuracy. Furthermore, as a coherent hologram it can provide accommodation cues. However, generating a DSCP is relatively demanding as a complex exponential function needs to be evaluated for each captured ray, unlike HS which can utilize Fourier transforms for fast calculations. The DSCP is included in the thesis as the most complex holographic representation that can be acquired from a discrete LF and depth.

3.3.2 Phase-added stereogram

The Fresnel hologram and DSCP are both accurate holographic representations, though they also impose strict limitations on the capture process and even more on the computational side. To relieve some of the limitations, the complexity of the model can be further reduced by replacing the wafels with hogels, i.e. approximating the wavefront with plane wave segments. The plane waves are still aligned by using

the depth information to keep the approximated total wavefront as accurate as possible. This holographic representation is known as the phase-added stereogram (PAS) [47].

Since the wavefront segments are required to be aligned to achieve a continuous approximation of the wavefield, the hologram is generated from discrete LF and a set of depth maps. The object field at the hologram plane for PAS is defined as a superposition plane waves aligned according to the distance from the hologram to the captured points, i.e.

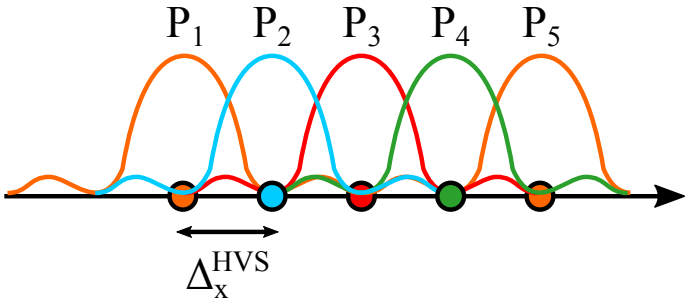
$$O_{PAS}(x) = \sum_m \text{rect}\left(\frac{x - m\Delta_x}{\Delta_x}\right) \times \sum_i \frac{\sqrt{L_1[m,i]}}{r^{mi}} \exp\left[j2\pi\left(f_x^{mi}x + \frac{r^{mi}}{\lambda}\right)\right], \quad (3.8)$$

where f_x^{mi} is the spatial frequency of a ray and r^{mi} is the Euclidean distance from the hogel center at $[m, i]$ to the point captured by the corresponding ray. As was noted previously in Sec. 3.2.1, each hogel can be obtained as a inverse Fourier transform of the permutated image segments. In comparison to the HS, the planar wavefront segments are now aligned using the relative phase factor derived from the Euclidean distance between the point source captured by the ray and the corresponding hogel center r^{mi}.

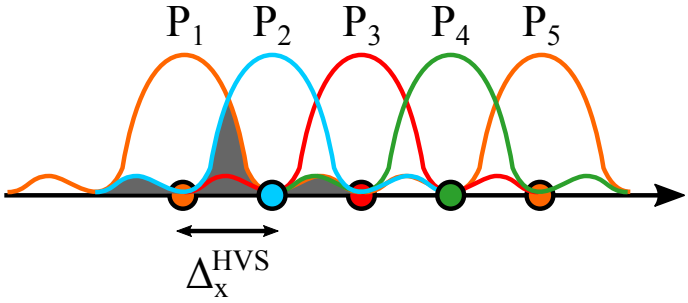
Utilizing PAS as a holographic representation allows for efficient implementations through the use of FFT algorithms, however, the accuracy compared to classical PAS is reduced, as FFT introduces discretization in frequency domain and as a result, the continuity of the wavefronts cannot be ensured [16]. Nevertheless, it is included in the thesis as a compromise between simple stereogram models and more complex coherent holograms.

3.3.3 Speckle noise reduction in coherent CGH

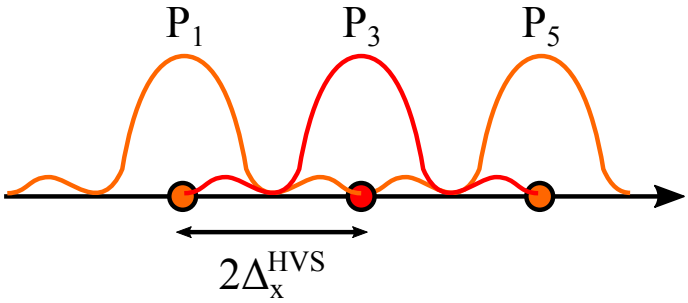
Speckles are an inherent physical phenomenon related with coherent imaging, and thus are required to be addressed in coherent holography as well. Similar speckle noise reduction solutions as introduced in Sec. 3.2.3 are utilized in coherent holography. As previously mentioned, random averaging and pixel separation are evaluated in the thesis. Random averaging is applied in a similar fashion as in incoherent holography, i.e. by summing the propagated reconstruction images of different random phase distribution CGHs. However, care must be taken to properly add the



(a) Points separated by Rayleigh resolution limit



(b) PSF overlap between adjacent points



(c) Increased distance between points

Figure 3.6 *The concept of pixel separation methods in speckle reduction. The PSFs of adjacent points overlap causing additional interference. If the points are separated by a larger distance as in (c), the overlap, and the amount of speckle noise, is reduced.*

random phase to the hologram such that different rays originating from the same point in the scene have the same phase value.

Utilizing the pixel separation solution of reducing speckle noise in coherent holography is slightly different to its incoherent counterpart. As the positions of point sources in the object space are known within certain accuracy, the separation should be done in the object space instead of the hologram plane. Moreover, the most significant percentage of speckle noise is due to points on the recorded scene separated by less than the Rayleigh resolution limit (see Sec. 3.1) [41]. Thus, the point sources

of the recorded scene should be separated by at least that distance to reduce noise in the reconstruction images. The solution presented in [41] is chosen to be included in this thesis. Specifically, in [41] the problem is solved by quantizing the location of each captured point (corresponding to each captured ray) on a uniform grid to obtain a new set of point sources which are referred to as object points. The object points are then separated in the object space and from each sparse set of object points, a separate CGH is calculated. These CGHs are then propagated in sequence and the reconstruction images summed to obtain the speckle-suppressed image. The concept behind this can also be observed in Fig. 3.6 as the PSFs of points separated by exactly the Rayleigh criterion designated limit still overlap to certain degree. By increasing the distance between the nearest points in the recorded scene in CGH generation, this effect can be reduced, further improving the speckle suppression. The structure of this speckle reduction method is as follows. From the set of captured points S_{cap}, a new set of object points on the quantized grid is found by finding the nearest points from the set of all object points S_q. Thus, for a captured point (x_{cap}, y_{cap}, z_{cap}) the corresponding object point (x_{ob}, y_{ob}, z_{ob}) is defined as

$$(x_{ob}, y_{ob}, z_{ob}) = \underset{(x,y,z)\in S_q}{\arg\min} \left\{ \sqrt{(x - x_{cap})^2 + (y - y_{cap})^2 + (z - z_{cap})^2} \right\}. \tag{3.9}$$

The uniform object point grid is defined such that the horizontal and vertical sampling step (ΔS_x and ΔS_y) is the Rayleigh diffraction limit, whereas the depthwise sampling step is determined by the average estimated speckle size in axial (z-axis) direction to be [11]

$$\Delta S_z = 8\frac{\lambda l^2}{T^2}. \tag{3.10}$$

The intensity of a ray emitted from an object point through a particular hogel is obtained as the sum of all ray intensities that were quantized to the object point within the hogel. This preserves the angular variations in intensity for the point sources, including occlusions and lighting effects.

As will be demonstrated via experiments in Sec. 4.2–4.3, neither method is without its flaws. The random averaging speckle noise reduction method is limited in its ability to reduce speckle noise, requiring a large amount of holograms to be propagated to achieve sufficient speckle suppression. The second speckle reduction method, pixel separation by sparse object points, is more efficient in suppressing the noise, though it suffers from other problems. As the intensities of the object points are obtained as the sum of captured point intensities within the quantization volume, it

is highly susceptible to uneven distribution of rays within different quantization volumes due to varying pixel disparity values (depending on depth and object shape). Furthermore, the approach effectively considers the intensities as nearest neighbour values. These problems result in large intensity variations across different points and can produce stripe patterns on the reconstructed object. Thus, a better speckle reduction method tackling these issues is needed.

3.3.4 Proposed speckle noise reduction method for coherent CGH

The speckle reduction method that we propose is inspired by the physical process of camera capture, where the intensity of each pixel is obtained through integration of values inside its area. In the context of pixel separation based speckle suppression, this affects the relation between LF capture and quantization of the scene. Ideally, quantizing the continuous LF emitted by the scene to the object point grid (as defined by the hologram properties) would uniformly sample the LF values resulting in a uniform distribution of quantized LF samples to be recorded on the hologram. Due to the discrete nature of the light field capture, a camera pixel is usually treated to be correspondent to a single ray crossing the center point of the pixel and the center of projection of the camera. This results in an uneven distribution of quantized samples on the grid causing the issues discussed in Sec. 3.3.3. Thus, the issues encountered by the previous methods are approached through means of traditional signal processing, namely resampling a discrete signal to obtain new samples. By this way, a more even distribution of rays on the quantization grid is obtained.

The method relies on the DSLF capture, which ensures that the continuous LF can be reconstructed from such discrete LF by linear interpolation. Thus, the number of rays within a pixel can be arbitrarily increased. For example, by increasing the number of rays by a factor of 3, a bundle of 3×3 rays is obtained for each original pixel (and thus hogel). Having increased the number of LF intensity samples, the number of depth values is also required to be increased in order to derive their corresponding position in the scene. Thus, the depth maps should be oversampled. In this thesis we use the simple bilinear interpolation for obtaining the oversampled depth values. As a result of these operations, a denser set of rays is utilized to obtain the quantized samples on the object point grid and the continuous nature of LF is better approximated. This also means that the distribution of rays in relation to the

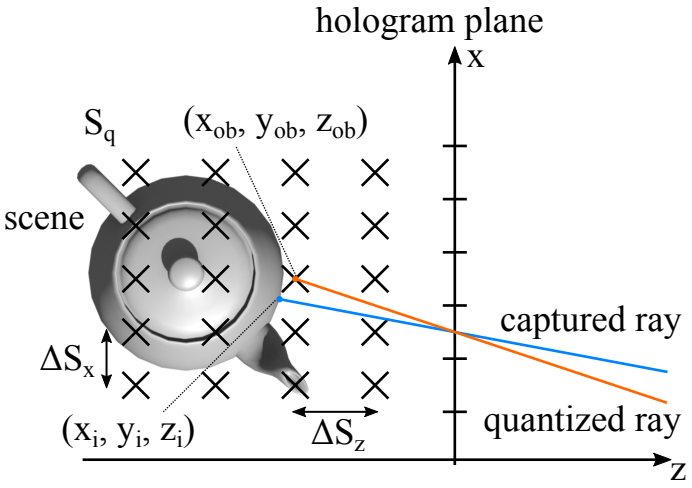


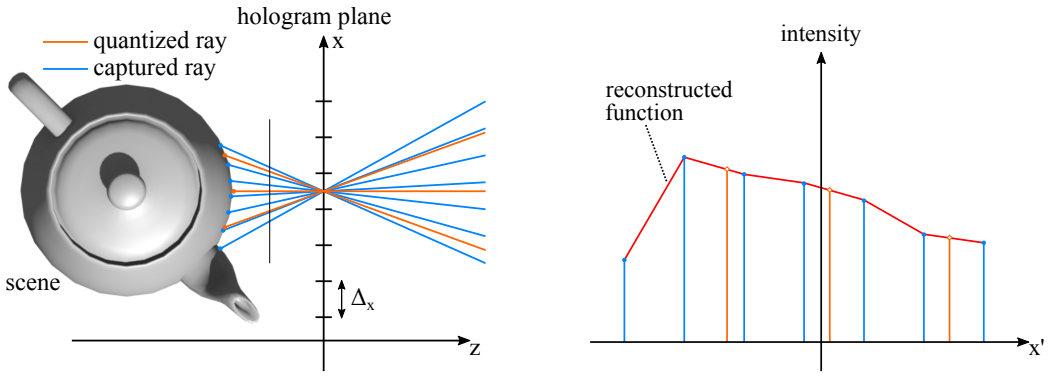
Figure 3.7 *The acquisition of object points from the set of captured and oversampled points. The set of all possible object point coordinates S_q is visualized by the cross symbols.*

quantization grid is more even, approximating the ideal case of continuous LF. In the following discussions of the method, points refer to the locations of the rays in the scene space resolved from the corresponding depth values. Let us define the set of points captured by the original set of cameras as S_{cap} and the set of oversampled points as S_{os}. A new set of points is found from the combined set of S_{cap} and S_{os} by quantizing these points to the object point grid, i.e. for each captured and oversampled point (x_i, y_i, z_i) an object point (x_{ob}, y_{ob}, z_{ob}) is found from the set of all possible object points S_q as

$$(x_{ob}, y_{ob}, z_{ob}) = \underset{(x,y,z)\in S_q}{\arg\min} \left\{ \sqrt{(x-x_i)^2 + (y-y_i)^2 + (z-z_i)^2} \right\}. \tag{3.11}$$

This process is also visualized in Fig. 3.7. For each hogel, the unique set of these object points is taken to be recorded on the hologram.

After the new set of points has been found, the next step is to estimate the intensity of those points. The dense LF sampling through multiperspective images provides a foundation for approaching this problem via interpolation. As such, the problem can be formulated in the following manner. Given the discrete captured and oversampled data at S_{cap} and S_{os}, i.e. the discrete LF from the captured points and the oversampled LF rays, obtain the resampled signal at S_{ob}. Since the LF is captured by a uniform grid of cameras, the light rays of a hogel form a uniform grid on any constant depth. By considering the intensity function on a certain (x,y)



(a) The sets of rays (b) Interpolation

Figure 3.8 *Obtaining the intensity values for the quantized rays from the densely sampled light field.*

plane by taking the cross-section of captured rays and rays traveling through the object points and the hogel center, the horizontal and vertical coordinates on that plane form a uniform input grid and a scattered grid of desired output values. The new intensities can be obtained by resampling the reconstructed continuous function, obtained from the original set of LF samples (i.e. input grid). This process is visualized in Fig. 3.8. Due to the strict LF sampling requirements imposed by the DSLF capture, linear interpolation kernel is assumed to be sufficient in obtaining the new intensity values.

The pixel separation methods can also suffer from issues not discussed in [41]. Particularly, by quantizing the captured point sources to the object point grid S_q it is possible to introduce point sources into locations which would be originally occluded. Including such points in the hologram generation would be erroneous and can cause different problems in the reconstructed images, such as increasing the amount of noise on parts of the scene where the object points change from one depth level to another. Thus, additional steps for handling this issue have to be considered. Due to the nature of the multiperspective images and how the holograms have been defined, a reasonable approximation can be done such that of all object points with the same horizontal and vertical coordinates within a hogel, only the frontmost point source is included. More accurate solutions could be derived, though the presented approach is deemed sufficient for the purpose of this thesis.

4. EXPERIMENTAL RESULTS

As the process of generating a hologram from a discrete LF has been presented, the next step is to find suitable methods to evaluate and compare the different holographic representations against each other. Furthermore, the validity of the presented sampling requirements should be assessed. Comparing or extracting information directly from the complex wavefield would be cumbersome and provide little relevance to the visual reconstructions seen by the viewer of the hologram. Hence, the viewing process by a human eye is simulated to determine how well each method can reconstruct the original scene.

The entire pipeline for the simulations is as follows. The discrete LF is first captured as full-parallax multiperspective views along with the depth maps. A hologram is generated from the data as presented in Sections 3.2 – 3.3 depending on the chosen presentation to utilize. Since the LF and holograms are assumed as monochromatic for simplicity, only the green colour channel of the multiperspective views is considered, thus fixing the wavelength of the light λ in the simulations to 534 nm. The HVS viewing process is simulated using the Fresnel diffraction model to achieve the image perceived by a viewer $I(u, v)$ as

$$I(u,v) = |\mathcal{F}_l\{T(s,t)\mathcal{F}_{z_{eye}}\{O(x,y)\}\}|^2, \tag{4.1}$$

where $T(s,t)$ is the lens transfer function of the human eye [10]. The lens transfer function describes mathematically how the simulated lens affects different spatial frequencies, and thus, how a wavefield propagates through the lens. The eye is considered as a camera with a circular aperture and a thin lens placed at a distance z_{eye} from the hologram plane. The aperture diameter T is chosen for each scene accordingly. The distance between the lens and sensor (i.e. pupil and retina) l is fixed at 25 mm, whereas the focal length of the lens f is controlled to focus at

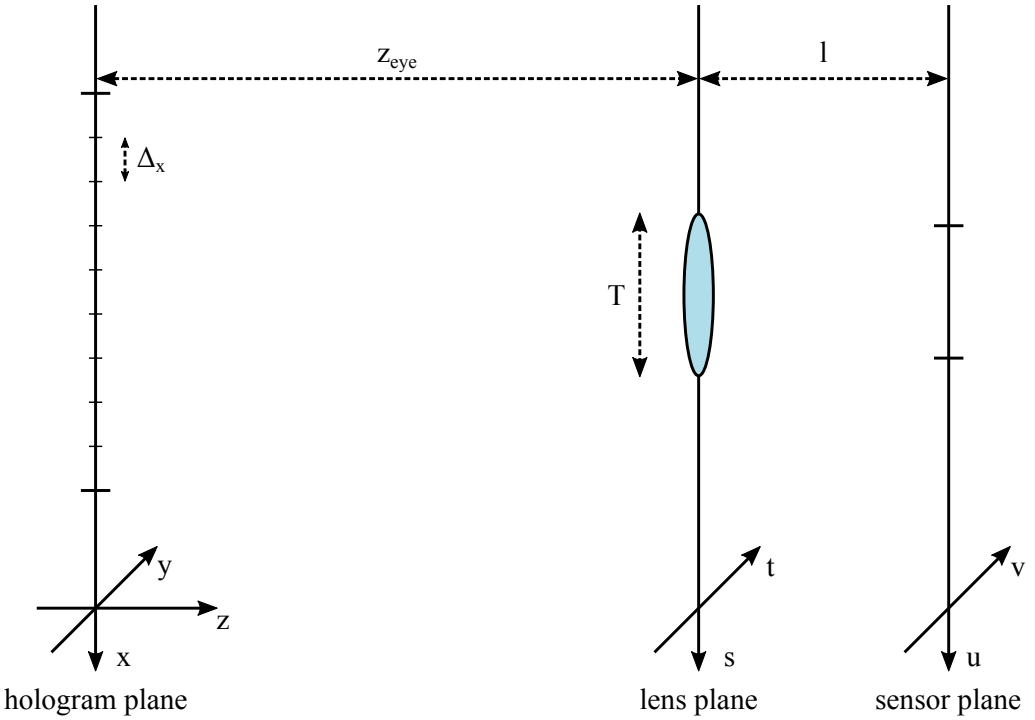


Figure 4.1 *Simulating the viewing process of a human eye. The field at the hologram plane $O(x, y)$ is propagated towards the positive z direction.*

different distances. The focal length is obtained as

$$f = \left(\frac{1}{d_{focus}} + \frac{1}{l} \right)^{-1} \tag{4.2}$$

for an eye focusing at distance d_{focus}. Thus, for a lens with the previously specified parameters placed at the location (s_{eye}, t_{eye}) on the lens plane, the transfer function is defined as

$$T(s,t) = \begin{cases} \exp\left(-\frac{j\pi}{\lambda f} r_{eye}^2\right) & \text{, if } r_{eye} \leq T/2 \\ 0 & \text{, if } r_{eye} > T/2, \end{cases} \tag{4.3}$$

where $r_{eye} = \sqrt{(s - s_{eye})^2 + (t - t_{eye})^2}$. The simulation process and its parameters are also visualized in Fig. 4.1. In order to avoid aliasing in various parts of the simulation process, the extent of the field at the hologram plane can be increased by zero padding as was discussed in Sec. 2.3. This way the original field can be modified to be within the necessary sampling requirements, e.g. the resolution condition to avoid aliasing in sampling the lens transfer function. Failing to meet these criteria can be observed on the final sensor plane image as phantom replicas of the recorded

scene.

4.1 Single point source

In order to properly study the effects of the hologram generation process, first a simple scene is considered. It also allows to demonstrate in practice how the sampling values for the hologram can be derived from the properties of the HVS. The scene in question consists of a single point source of light at z_{obj} depth from the hologram plane. By controlling this value and the camera spacing, the pixel disparity between adjacent views can be altered to be either an integer or a fractional value. Both of these cases are considered to demonstrate the effects that would be present in a discretely captured LF of a general continuous 3D scene.

First, let us derive the sampling requirements. The hologram is assumed to be viewed from a distance of 200 mm by a 2 mm diameter pupil. The Rayleigh criterion in Eq. 3.2 states that for such diffraction limited system the minimum resolvable distance Δ_x^{HVS} is 65.15 μm. Thus, the hologram plane spatial sampling Δ_x must not exceed this value if the perceived resolution is to be maximized. However, reducing the sampling distance further would not improve the perceived quality and as such, the value 64 μm is chosen for Δ_x. The hologram plane pixel size X_x is selected as 2 μm which results in 32×32 hologram pixels within each hogel, corresponding to the same number of propagated rays per each hologram segment. In the case of HS and PAS, the fixed frequency grid of the FFT calculations dictates the spectral sampling, and as such can be connected to the angular sampling by way of the grating equation in Eq. 3.5. If the size of the FFT grid is chosen to be equal to the number of rays (32), the spatial frequencies are between $-1/(2 \cdot 2\mu\text{m}) = -250000\mu\text{m}^{-1}$ and $1/(2 \cdot 2\mu\text{m}) - 1/(32 \cdot 2\mu\text{m}) = 234375\mu\text{m}^{-1}$ with a frequency sampling step of $1/(32 \cdot 2\mu\text{m}) = 15625\mu\text{m}^{-1}$. These correspond to the minimum and maximum incident angles of -7.67 and 7.19 degrees for the propagated rays, respectively. The linear frequency sampling results in non-uniform angular sampling, where the maximum angle between adjacent rays as approximately 0.48 degrees. The angular sampling limit for the assumed viewer properties according to Eq. 3.3 is 0.57 degrees, and is thus well fulfilled.

In the first test, the variables z_{obj} and Δ_s are set such that the pixel disparity between the perspective views is an integer amount. Thus, the captured rays correspond exactly to the single point in the scene. To eliminate any additional sources of error

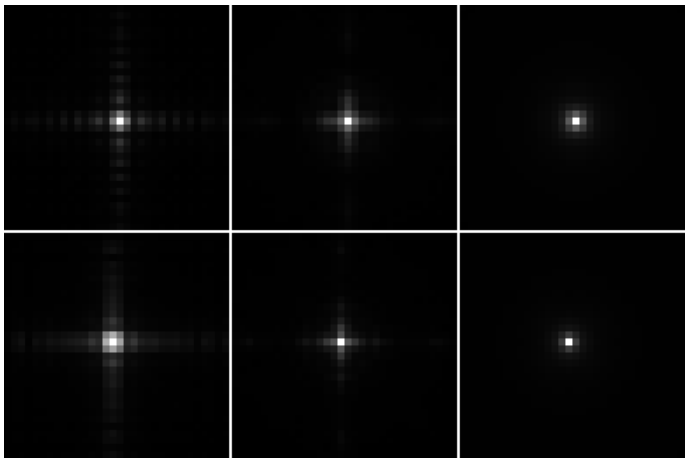


Figure 4.2 *Simulated views for a single point source at integer pixel disparity focused at hologram (upper row) and at the object (lower row). The holograms used from left to right: HS, PAS and DSCP.*

from the capturing process, the holograms are generated directly from the point cloud information (i.e. exact point location and intensity). The viewer observed image is simulated as previously explained by the Fresnel diffraction model, focusing the simulated eye at the object and the hologram, resulting in two images for each hologram. The eye is placed at the center of the lens plane, i.e. $(s_{eye}, t_{eye}) = (0, 0)$ mm. The resulting images of these simulations are presented in Fig. 4.2. As is expected, the more complex representations provide better reconstructions, and in the case of coherent holograms, the blur around the point is reduced by focusing the simulated eye at the depth of the point instead of the hologram.

The second test considers the fractional pixel disparity case, i.e. adjacent views are captured in such a way that rays emitted from the point through the hogel centers are not aligned on the sensor grid. Considering an idealized case of capturing the exact rays through each hogel center, the fractional pixel disparity would cause the rays to cross the sensor plane offset to the defined sampling grid. In a more realistic situation, i.e. capturing with a camera (even pinhole), the captured pixel at the sensor plane is an integrated value within the area defined by the magnification of the capturing camera, thus resulting in blur and additional pixels capturing intensity from a single point. This can be seen also in the reconstructed images in Fig. 4.3. Since the point-of-origin location for each ray is resolved from depth and camera positions, the captured light rays correspond to several nearby points instead of a single point. These points are separated by distances smaller than the Rayleigh

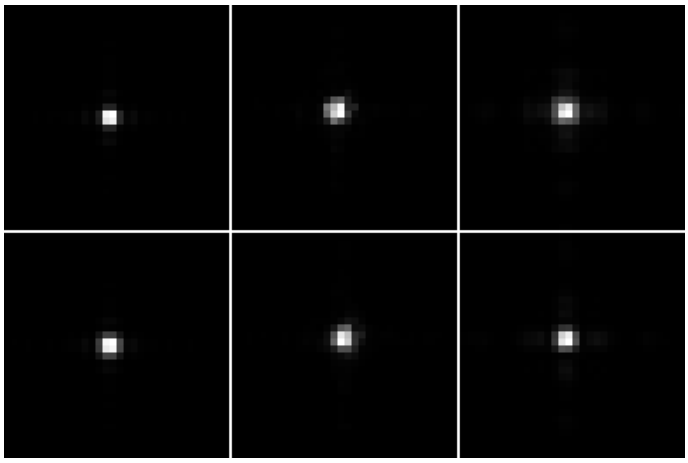


Figure 4.3 *Simulated views for a single point source at fractional pixel disparity focused at hologram (upper row) and at the object (lower row). The holograms used from left to right: HS, PAS and DSCP.*

resolution limit and will overlap in terms of their PSFs. As was discussed in Sec. 3.2.3 and Sec. 3.3.3, this can lead to issues in the reconstructed images especially for more complex scenes where the use of random phase is required.

From these simulation results it can be seen that the points (objects) at a depth with fractional pixel disparity can be problematic for more complex scenes. The underlying effect is due to interpreting such pixel-depth combinations as point sources with extremely small distance between each other, thus introducing additional interference as pointed out previously. To properly observe the effects and solutions to reduce them, scenes with multiple point sources with variable spacings are examined to determine the distance at which the effect is present.

4.2 Set of point sources

Considering the previous results, the effects of nearby point sources should be considered in more detail. Thus, the scene is expanded from a single point source to several point sources. In addition to the previously used notations, a new parameter for the sampling distance of the point sources Δx_{obj} is introduced. To investigate the effects of this sampling distance in the perceived reconstructions, holograms generated from scenes with different values of Δx_{obj} are considered. The values are chosen as multiples of the human eye diffraction limit Δ_x^{HVS} at distance $d + z_{obj}$. Specifically, four different scenes with values of Δx_{obj} as $8\Delta_x^{HVS}$, $4\Delta_x^{HVS}$, $2\Delta_x^{HVS}$

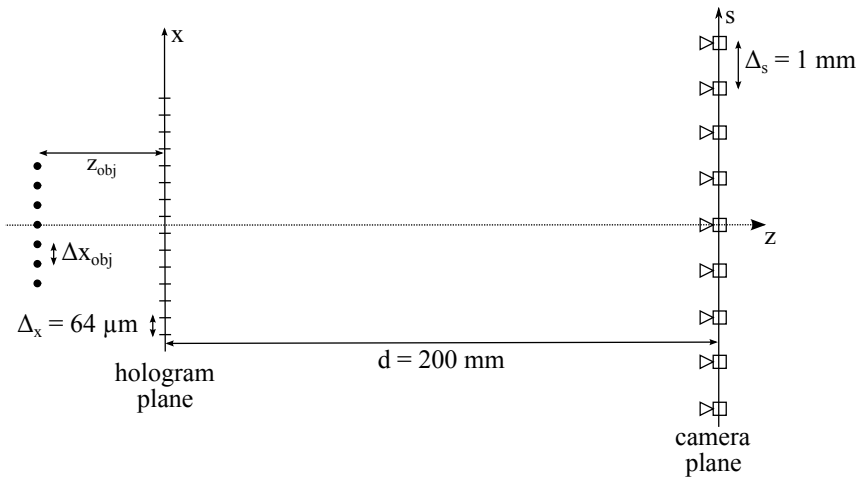


Figure 4.4 *The parameters and arrangement of the set of point sources.*

and Δ_x^{HVS} are examined. The capture parameters, however, are left as previously derived according to the properties HVS. That is, the hologram plane spatial sampling step Δ_x is chosen as 64 μm, camera sampling step Δ_s as 1 mm and the eye with pupil aperture T of 2 mm is placed 200 mm from the hologram plane at the center of the lens plane. The scene placement and capture setup properties are visualized in Fig. 4.4. As was mentioned in Sec. 3.2.3, random phase needs to be added to the hologram and is thus used in all following simulations.

The holograms for all 4 scenes are generated without speckle suppression to properly observe the speckle phenomenon. Additionally, both random averaging and pixel separation are utilized to evaluate their speckle suppression capabilities for a simple scene. In both cases, 16 separate CGHs are generated and propagated in sequence, that is, for random averaging each hologram is generated with a separate set of random phase values and for pixel separation every fourth hogel (incoherent) or object point (coherent) is included in each separate CGH. Let us discuss first the visual appearance of the reconstructed result images before any numerical evaluations. Observing the results in Fig. 4.5 for the HS, it can be seen that the amount of speckle noise in the images without speckle suppression methods is substantial. There is a slight amount of improvement as the distance between the points in the scene is increased, though better improvement is obtained though the speckle suppression methods, both seemingly performing similarly. The PAS reconstructions in Fig. 4.6 also have severe speckle degradation, however, random averaging performs well for all four scenes. The pixel separation method provides very little improvement

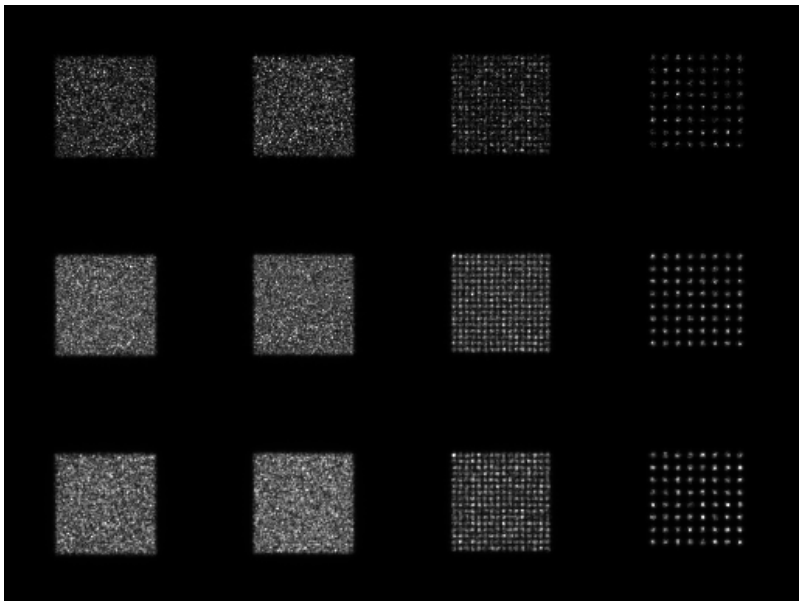


Figure 4.5 *Simulated views for a set of points, focused at the hologram plane of the HS. The object point distance is increased by a factor of 2 for each scene from left to right. No speckle suppression on the top row, random averaging on the center row and pixel separation on the bottom row.*

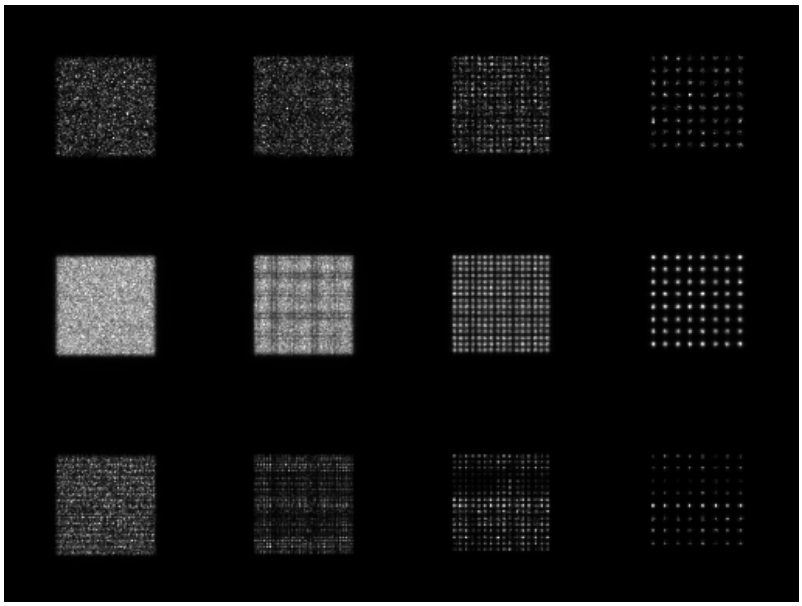


Figure 4.6 *Simulated views for a set of points, focused at the hologram plane of the PAS. The object point distance is increased by a factor of 2 for each scene from left to right. No speckle suppression on the top row, random averaging on the center row and pixel separation on the bottom row.*

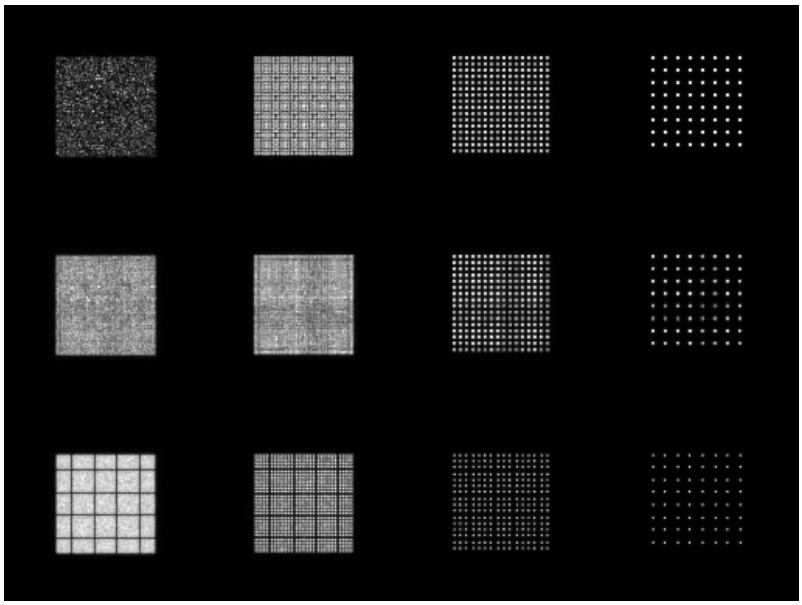


Figure 4.7 *Simulated views for a set of points, focused at the hologram plane of the DSCP. The object point distance is increased by a factor of 2 for each scene from left to right. No speckle suppression on the top row, random averaging on the center row and pixel separation on the bottom row.*

visually. Finally, the resulting perceived images of the most accurate CGH model in this thesis, DSCP, can be seen in Fig. 4.7. In contrast to HS and PAS, the speckle noise is already suppressed by increasing the distance between adjacent points in the scene even without any speckle suppression methods. This further exhibits the validity of the pixel separation method in coherent holography. Additionally, both speckle suppression methods reduce significantly the visual degradation in the images. However, the pixel separated result images have horizontal and vertical spaces within the object. This is caused by the difference in the distance between adjacent points in the original scene (approximately 68.4 μm) and adjacent points in the quantization grid (64 μm). Additionally, the quantization in depth in the object point grid can also relocate the points at a distance with fractional pixel disparity, which can cause problems especially if the disparity is increased to be larger in magnitude than 1 pixel. Furthermore, a portion of the difference in quality between DSCP and the two other holographic representations is likely caused by the fixed spatial frequency grid in FFT employed by both in the CGH generation.

In order to quantify the improvement regarding the speckle noise reduction, a metric known as speckle contrast is evaluated. For an object surface in the reconstructed

image, the speckle contrast C is defined as

$$C = \frac{\sigma}{\bar{I}}, \tag{4.4}$$

where σ is the standard deviation of the intensity values and $\bar{I}$ is the mean intensity [11]. As is apparent from the definition, lower speckle contrast indicates reduced amount of speckle noise. The speckle contrast is evaluated for the densest set of points ($\Delta x_{obj} = \Delta_x^{HVS}$) by including the values corresponding to the object surface from the reconstructed images. The resulting values are collected in Table 4.1. Certain initial conclusions can be made from these values. First, both random averaging and pixel separation reduced speckle noise in all cases. Secondly, the initial amount of speckle degradation is relatively equal for HS and PAS, though slightly lower for DSCP. Finally, the overall performance of the random averaging method is fairly good, whereas the pixel separation method is slightly hindered by problems with the fixed spatial frequency grid (PAS) and the sampling step on the object point grid (DSCP).

Table 4.1 *Speckle contrast values from the reconstructed images of the scene with the diffraction limit as the sampling distance between point sources.*

Speckle suppression	**HS**	**PAS**	**DSCP**
None	0.667	0.695	0.571
Random averaging	0.378	0.242	0.189
Pixel separation	0.377	0.572	0.284

The results of the point source simulations provide some necessary insight on the limitations and properties of the different holographic representations as well as how to improve the visual quality of the viewer perceived images through means of speckle suppression. Therefore, the following simulations consider more complex 3D scenes in order to estimate the validity of the methods for more realistic scenarios.

4.3 Synthetic 3D scene

The validity of the presented holographic representations and speckle reduction techniques has been confirmed for simplified cases consisting of constant intensity point sources. For the case of more realistic synthetic scenes, the simulations are divided into two parts, the first assessing the visual reconstruction quality of a 3D object from

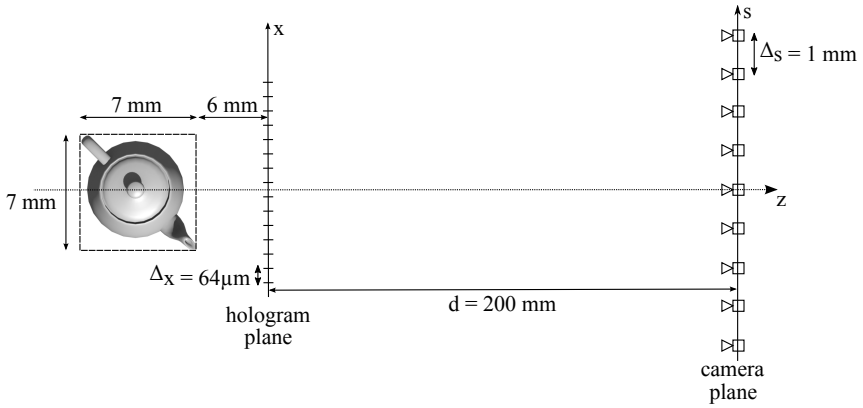


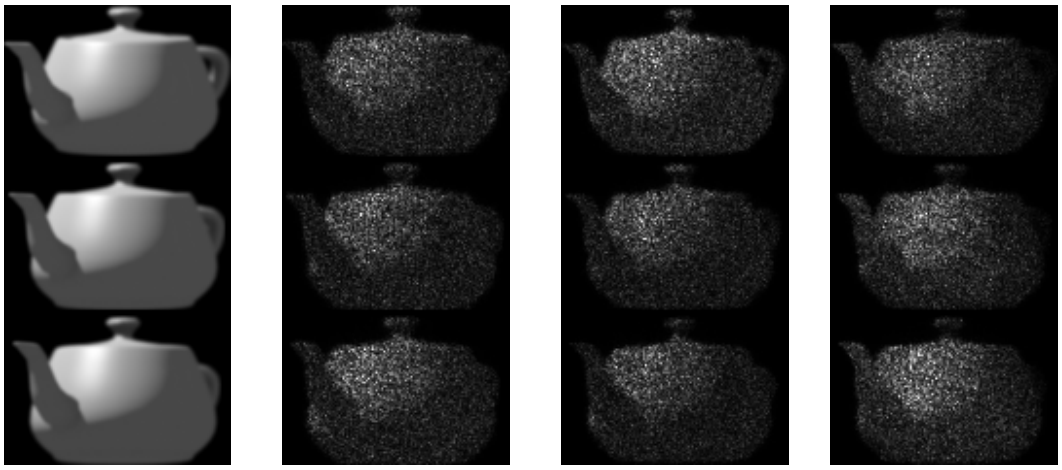
Figure 4.8 *The scene and capture setup for the Utah teapot scene.*

different viewer positions and the second examining the accommodation response. The 3D-modelling software Blender [4] is used to render the multiperspective images

4.3.1 Single object (Utah teapot)

The first scene consists of a single 3D object, the Utah teapot [40], without any additional texture. The object is single coloured, though the lighting on the scene produces slight variations in intensity along the surface of the object. Holograms for all three representations are produced from this scene both with and without the various speckle reduction methods, totaling in 10 different holograms. Different viewer positions in horizontal and vertical directions are simulated in order to evaluate the methods thoroughly. The reconstructed images obtained from the human eye viewing simulation are evaluated against a synthetic aperture image captured in Blender with the same human eye parameters as in the simulations. The setup for the scene and its capture is shown in Fig. 4.8. The camera spacing is chosen and the teapot is placed such that it is entirely within the range of a single pixel disparity, thus fulfilling the criterion for DSLF. The set of multiperspective images are obtained by rendering pinhole camera views in Blender.

In the first set of simulations, all holograms and simulations were carried out without any speckle reduction methods. All holograms have the same parameters where applicable to ensure equal comparison between the different representations. In order to properly evaluate the reconstruction quality, three different simulated views were obtained from each hologram. The following positions for the simulated viewer (s_{eye}, t_{eye}) are used: (-10,10), (0,0) mm and (10,-10) mm. These are referred to



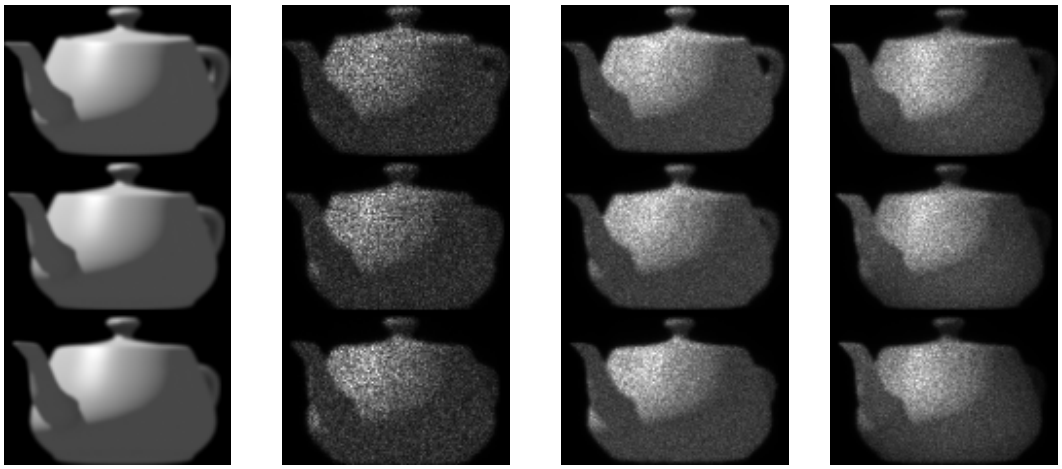
(a) Reference (b) HS (c) PAS (d) DSCP

Figure 4.9 *Simulated views for the Utah teapot without speckle reduction methods. The lens positions (-10,10), (0,0) mm and (10,-10) mm are used on the top, center and bottom row, respectively.*

Table 4.2 *PSNR values (dB) without any speckle reduction methods.*

View	No speckle suppression		
	HS	**PAS**	**DSCP**
(-10,10) mm	12.13	12.43	11.60
(0,0) mm	11.85	11.55	12.18
(10,-10) mm	12.44	11.60	12.76

as top, center and bottom view, respectively. The lens diameter is increased from previous simulations to 3 mm in order to capture more light rays within the extent of the lens. However, since the spatial sampling of the hologram remains at 64 μm, the sampling requirement for HVS is not fulfilled and the perceived resolution is lowered. The results are compared against a finite aperture view, that is, an image simulating the aperture effects of the human eye. As discussed in [19], such images can be obtained as a superposition of elementary apertures. In this case, the final finite aperture image is acquired as a sum of pinhole captured images within the extent of the simulated aperture. The set of simulated views are compared in Fig. 4.9. As indicated by previous experiment results in Sec. 4.2, the reconstruction images suffer from heavy speckle noise degradation in all hologram types and views. When compared to the reference views, it is clear both by visually inspecting them and by the PSNR values in Table 4.2 that addressing the speckle phenomenon in



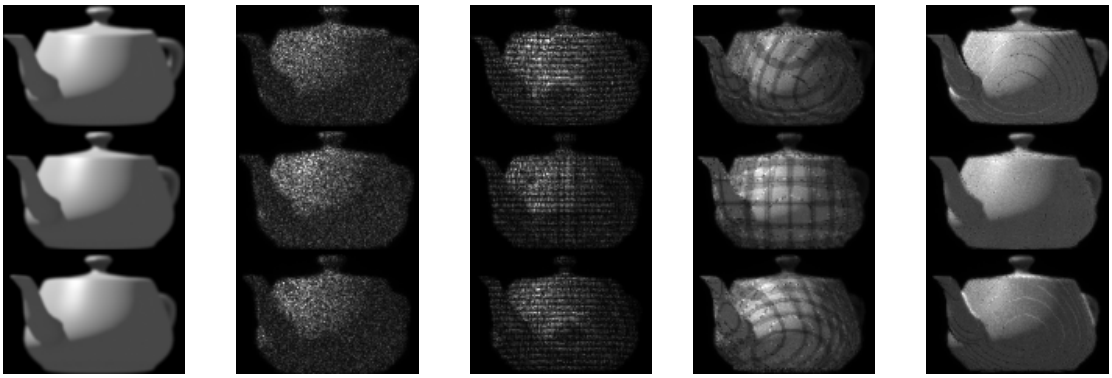
(a) Reference (b) HS (c) PAS (d) DSCP

Figure 4.10 *Simulated views for the Utah teapot with random averaging. The lens positions (-10,10), (0,0) mm and (10,-10) mm are used on the top, center and bottom row, respectively.*

digital holography is extremely important.

Next, the same set of simulated views is obtained from holograms with the two chosen speckle reduction methods employed. The random averaging method is implemented across 16 different holograms, requiring 16 CGHs to be propagated in succession. The reconstructed views in Fig. 4.10 show a significant improvement in visual quality due to the reduced speckle noise. The simulation results seem to indicate that the coherent holograms benefit more from this speckle suppression method when compared to the HS reconstruction images. However, as the theory of the random averaging method dictates, the speckle reduction is limited and further improvement would require a significant increase in the number of holograms, hindering practical implementations of such holograms. The PSNR values of the reconstructed images using the finite aperture views as reference are presented in Table 4.3. These values indicate that the PAS is the most suited for random averaging, although both DSCP and HS achieved adequate results as well.

For fair comparison, the pixel separation method is also tested for 16 holograms, thus including every fourth hogel (incoherent) or object point (coherent) in horizontal and vertical directions to obtain the sparse holograms. Comparing the reconstructed views in Fig. 4.11 to their random-averaged counterparts, it can be seen that the speckle reduction is improved in some cases. The HS reconstructions display



(a) Reference (b) HS (c) PAS (d) DSCP (e) DSCP*

Figure 4.11 *Simulated views for the Utah teapot with pixel separation. The images in column (e) are the result of the proposed speckle reduction method. The lens positions (-10,10), (0,0) mm and (10,-10) mm are used on the top, center and bottom row, respectively.*

Table 4.3 *PSNR values (dB) for the different speckle reduction methods.*

View	Random averaging			Pixel separation			Proposed
	HS	**PAS**	**DSCP**	**HS**	**PAS**	**DSCP**	**DSCP***
(-10,10) mm	15.79	22.17	19.05	13.14	12.22	15.23	23.44
(0,0) mm	15.87	21.25	19.27	14.03	11.58	17.06	25.03
(10,-10) mm	16.63	21.24	19.05	13.11	11.65	17.77	19.52

similar amount of speckle suppression, which is further reflected in the corresponding PSNR values seen in Table 4.3. In coherent holograms, however, the method has proven to be problematic due to the object points quantization approach. As in the experiments with several point sources, the pixel separation method does not perform well with PAS. In the case of DSCP, the pixel separation method exhibits the problems highlighted in Sec. 3.3.4. The improvements proposed in Sec. 3.3.4 have addressed these issues fairly well, though room for further improvement still exists as can be seen for example in the off-center views where the steps in depth quantization are clearly visible. Nevertheless, this method achieved the highest PSNR values in 2 out of the 3 views, with the exception of the bottom view where random averaged PAS acquired better PSNR.

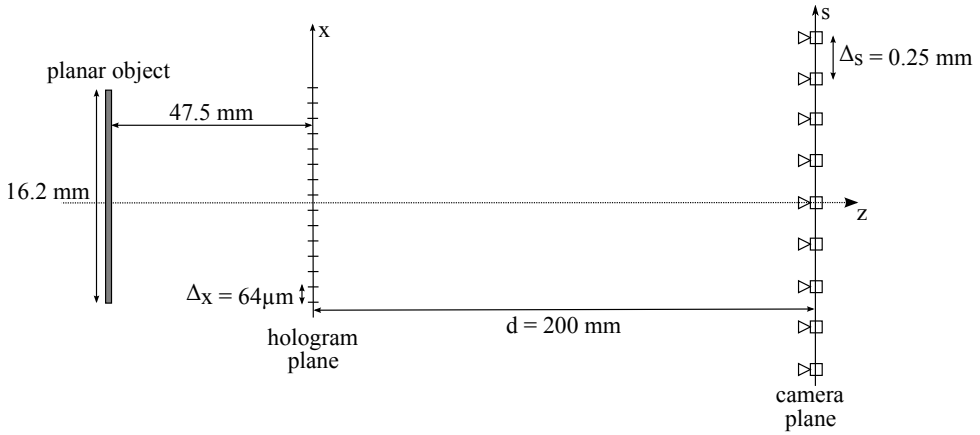


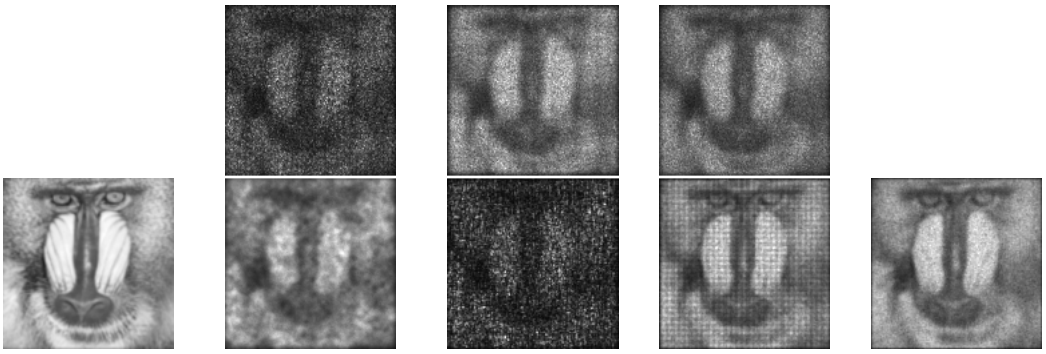
Figure 4.12 *The scene and capture setup for the accommodation experiment.*

4.3.2 Accommodation cues

In addition to the visual reconstruction quality, an important aspect of holograms is the accommodation cue they provide, i.e. the ability to focus on different parts of the scene. In object-based holography, the more complex holograms such as Fresnel and DSCP can recreate deep scene points with minimal blur. However, if the holograms are created from multiperspective images, the data to be recorded on the hologram is already sampled by the cameras, and thus cannot be reconstructed with perfect accuracy. This is especially true for deep scenes with high resolution textures.

The accommodation cues of each holographic representation are evaluated with the following setup. A planar object is placed behind the hologram far enough to represent a deep scene. The object is textured with a high resolution image in comparison to the capture and hologram sampling. The setup of the scene and the capture parameters are shown in Fig. 4.12. A human eye is simulated to view the hologram, focusing on the object. Thus it is expected that accurate holographic representations can resolve as much as possible of the available detail. The lens diameter is increased from 2 mm to 3 mm to observe accommodation response more realistically. The lens is placed at the center of the viewing zone at 200 mm distance from the hologram plane. As previously, the simulated views are compared against a finite aperture reference view. The result images and a reference view are shown in Fig. 4.13.

Since the surface is relatively far away from the hologram plane, the downside of incoherent holography can be seen well. Both the pixel separated and random averaged



(a) Reference (b) HS (c) PAS (d) DSCP (e) DSCP*

Figure 4.13 *Simulated views for the accommodation cue test. Random averaging used on the upper row, pixel separation on the lower row. Column (e) shows the result for DSCP calculated with the proposed speckle reduction method.*

Table 4.4 *PSNR values (dB) of the reconstructed images from the accommodation cue test.*

Speckle suppression	**HS**	**PAS**	**DSCP**	**DSCP***
Random averaging	8.42	13.22	12.17	–
Pixel separation	15.28	7.25	15.25	18.04

reconstruction images show the inability to resolve the higher resolution texture, and instead severe amounts of blur is visible. The coherent holograms achieve better results with the random averaging method, however, the leftover speckle noise means that a significant amount of details is still lost in the process. This is further reflected in the PSNRs of the reconstruction images as seen in Table 4.4. In this case, the pixel separation method provides the highest quality reconstructions. The problems highlighted in Sec. 3.3.3 can be seen in Fig. 4.13(d) as horizontal and vertical stripes are again present on the object surface, although to a lesser degree than in the previous simulation case. The improvements proposed in the thesis to this speckle reduction method have eradicated these issues, and as a result, the highest PSNR value is achieved with this method.

5. CONCLUSIONS

Holographic representation and display of 3D scenes that are captured in incoherent lighting conditions, e.g. via multiperspective cameras, is considered as the main topic of this thesis. Three different holographic representation techniques consisting of both coherent and incoherent type of computer generated holograms are examined and their ability to reconstruct the recorded scene are studied through numerical experiments. The holograms are divided into two different categories depending on whether they use only the light field intensities (incoherent) or both the light and depth fields (coherent). The holographic stereogram (example for incoherent hologram), phase-added stereogram and diffraction-specific coherent panoramagram (examples for coherent hologram) are investigated as three representative cases and their properties reviewed.

First, the capture of a discrete light field and how it can be represented by a set of multiperspective images is examined. Hologram generation from discrete light field is discussed extensively, beginning from the relations between the capture and hologram parameters and sampling requirements. The sampling requirements imposed on the capture of light fields are observed to be usually harsh. Based on this observation, it is demonstrated that by using shearlet-transform based light field reconstruction algorithm, the strict capture requirements for holographic stereograms can be relieved and more practical capture setups are enabled, i.e. cumbersome camera rigs can be replaced with multicamera arrays.

The CGH methods are compared in experiments via simulation of viewing process by human eye, which is implemented using wave optics principles. That is, the wavefield on the CGH plane is propagated towards the viewer and the viewer perceived image is obtained as the result of the simulation process. The issue of speckle noise inherent to coherent imaging is particularly discussed in detail, as it highly degrades the quality of the reconstructed images from holograms. Based on this analysis, a speckle reduction method for coherent CGH that improves the existing techniques

is proposed.

Initially the methods are verified for an elementary case involving constant intensity point sources as the recorded scene. The corresponding simulation results indicate issues with objects at depth ranges with fractional pixel disparity. Moreover, including random phase in the CGH generation to diffuse the object light introduces speckle noise to the reconstruction images. This issue is shown to be alleviated by utilizing existing speckle reduction methods that are based on random averaging and pixel separation. To better mimic a realistic scenario, experiments for a synthetic 3D object are carried out. Respectable results in terms of perceivable visual quality are obtained as the random averaging method performs well across all holograms. However, in the case of coherent holograms, the pixel separation method suffers from intensity variations on the reconstructed images in the form of stripe patterns due to its naive object point quantization approach. The proposed speckle suppression solution alleviates this issue by taking advantage of the DSLF properties.

In addition to the visual quality of the reconstructions, the ability to accommodate on object in the recorded scene is important for proper perception of 3D space. The experiments show that incoherent holograms cannot reproduce deep scenes well, whereas the coherent holograms perform better in this sense. Furthermore, generally higher accuracy reconstructions are observed by the pixel separation based speckle suppression methods, and thus, making them more suitable for accommodation response dependent applications.

In summary, the presented CGH generation process can be utilized to examine the sampling effects and quality of the holographic reconstructions. The presented work is, thus, especially useful in deciding the necessary system parameters for holographic displays or holographic printing devices together with tailored capture parameters to be used in content generation. The presented framework also enables studying various novel scenarios and their consequences both in capture and display stages, such as non-uniform camera sampling or varying sized segmentation throughout the hologram.

BIBLIOGRAPHY

[1] E. H. Adelson and J. Y. A. Wang, "Single lens stereo with a plenoptic camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 99–106, 1992.

[2] E. H. Adelson and J. R. Bergen, "The plenoptic function and the elements of early vision," in *Computational Models of Visual Processing*. MIT Press, 1991, pp. 3–20.

[3] J. Amako, H. Miura, and T. Sonehara, "Speckle-noise reduction on kinoform reconstruction using a phase-only spatial light modulator," *Applied Optics*, vol. 34, no. 17, pp. 3165–3171, Jun 1995.

[4] Blender Online Community, *Blender - a 3D modelling and rendering package*, Blender Foundation, Blender Institute, Amsterdam, 2015. [Online]. Available: http://www.blender.org

[5] B. R. Brown and A. W. Lohmann, "Complex spatial filtering with binary masks," *Applied Optics*, vol. 5, no. 6, pp. 967–969, 1966.

[6] W. J. Dallas, "Computer-generated holograms," *Digital Holography and Three-Dimensional Display: Principles and Applications*, pp. 1–50, 2006.

[7] D. Gabor *et al.*, "A new microscopic principle," *Nature*, vol. 161, no. 4098, pp. 777–778, 1948.

[8] N. George and J. T. McCrickerd, "Holography and stereoscopy - the holographic stereogram," in *Modulation Transfer Function*, R. R. Shannon and R. J. Wollensack, Eds., vol. 13, 1969, pp. 342–350.

[9] A. Gershun, "The light field," *Journal of Mathematics and Physics*, vol. 18, no. 1-4, pp. 51–151, 1939.

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

[11] ——, *Speckle Phenomena in Optics: Theory and Applications*. Roberts and Company Publishers, 2007.

[12] M. W. Halle, "Holographic stereograms as discrete imaging systems," *Proc. SPIE*, vol. 2176, pp. 73–84, 1994.

[13] P. Hariharan, *Optical Holography: Principles, Techniques and Applications*, 2nd ed., ser. Cambridge Studies in Modern Optics. Cambridge University Press, 1996.

[14] P. S. Hilaire, "Modulation transfer function and optimum sampling of holographic stereograms," *Applied Optics*, vol. 33, no. 5, pp. 768–774, Feb 1994.

[15] J. M. Huntley and L. Benckert, "Speckle interferometry - noise reduction by correlation fringe averaging," *Applied Optics*, vol. 31, no. 14, pp. 2412–2414, 1992.

[16] H. Kang, T. Fujii, T. Yamaguchi, and H. Yoshikawa, "Compensated phase-added stereogram for real-time holographic display," *Optical Engineering*, vol. 46, no. 9, p. 95802, 2007.

[17] B. Katz, N. T. Shaked, and J. Rosen, "Synthesizing computer generated holograms with reduced number of perspective projections," *Optics Express*, vol. 15, no. 20, pp. 13 250–13 255, 2007.

[18] D. P. Kelly, "Numerical calculation of the fresnel transform," *Journal of the Optical Society of America A*, vol. 31, no. 4, pp. 755–764, Apr 2014.

[19] D. Lanman, R. Raskar, and G. Taubin, "Modeling and Synthesis of Aperture Effects in Cameras," in *Computational Aesthetics in Graphics, Visualization, and Imaging*. The Eurographics Association, 2008, pp. 81–88.

[20] E. N. Leith and J. Upatnieks, "Reconstructed wavefronts and communication theory," *Journal of the Optical Society of America*, vol. 52, no. 10, pp. 1123–1128, 1962.

[21] M. Levoy, "Light fields and computational imaging," *Computer*, vol. 39, no. 8, pp. 46–55, Aug. 2006.

[22] M. Levoy and P. Hanrahan, "Light field rendering," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '96. New York, NY, USA: ACM, 1996, pp. 31–42.

[23] Z. Lin and H.-Y. Shum, "A geometric analysis of light field rendering," *International Journal of Computer Vision*, vol. 58, no. 2, pp. 121–138, 07 2004.

[24] G. Lippmann, "Épreuves réversibles donnant la sensation du relief," *J. Phys. Theor. Appl.*, vol. 7, no. 1, pp. 821–825, 1908.

[25] A. W. Lohmann, R. G. Dorsch, D. Mendlovic, Z. Zalevsky, and C. Ferreira, "Space-bandwidth product of optical signals and systems," *Journal of the Optical Society of America A*, vol. 13, no. 3, pp. 470–473, 1996.

[26] M. Lucente, "Diffraction-specific fringe computation for electro-holography," Ph.D. dissertation, Cambridge, MA, USA, 1994.

[27] M. Makowski, M. Sypek, A. Kolodziejczyk, and G. Mikuła, "Three-plane phase-only computer hologram generated with iterative fresnel algorithm," *Optical Engineering*, vol. 44, no. 12, p. 125805, 2005.

[28] M. Matsumura, "Speckle noise reduction by random phase shifters," *Applied Optics*, vol. 14, no. 3, pp. 660–665, 1975.

[29] K. Matsushima, "Performance of the polygon-source method for creating computer-generated holograms of surface objects," in *ICO Topical Meeting on Optoinfomatics/Information Photonics 2006*, 2006, pp. 99–100.

[30] K. Matsushima and S. Nakahara, "Extremely high-definition full-parallax computer-generated hologram created by the polygon-based method," *Applied Optics*, vol. 48, no. 34, pp. H54–H63, Dec 2009.

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

[32] P. W. McOwan, W. J. Hossack, and R. E. Burge, "Three-dimensional stereoscopic display using ray traced computer generated holograms," *Optics Communications*, vol. 82, no. 1-2, pp. 6–11, 1991.

[33] E. Sahin, S. Vagharshakyan, J. Mäkinen, R. Bregovic, and A. Gotchev, "Shearlet-domain light field reconstruction for holographic stereogram generation," in *2016 IEEE International Conference on Image Processing (ICIP)*, Sept 2016, pp. 1479–1483.

[34] B. E. A. Saleh and M. C. Teich, *Fundamentals of photonics*, 2nd ed. Hoboken, N.J: John Wiley & Sons, 2007.

[35] T. Shimobaba and T. Ito, "Random phase-free computer-generated hologram," *Optics Express*, vol. 23, no. 7, pp. 9549–9554, Apr 2015.

[36] D. Smith, *Field Guide to Physical Optics*, ser. SPIE field guides ; volume FG17. SPIE, 2013.

[37] Q. Y. J. Smithwick, J. Barabas, D. E. Smalley, and V. M. Bove, Jr., "Interactive holographic stereograms with accommodation cues," vol. 7619, 2010, pp. 761 903–761 903–13.

[38] W. Sun, L. Xu, O. C. Au, S. H. Chui, and C. W. Kwok, "An overview of free view-point depth-image-based rendering (dibr)," in *APSIPA Annual Summit and Conference*, 2010, pp. 1023–1030.

[39] Y. Takaki and M. Yokouchi, "Speckle-free and grayscale hologram reconstruction using time-multiplexing technique," *Optics Express*, vol. 19, no. 8, pp. 7567–7579, 2011.

[40] A. Torrence, "Martin newell's original teapot," ser. SIGGRAPH '06. New York, NY, USA: ACM, 2006.

[41] T. Utsugi and M. Yamaguchi, "Speckle-suppression in hologram calculation using ray-sampling plane," *Opt. Express*, vol. 22, no. 14, pp. 17 193–17 206, Jul 2014.

[42] S. Vagharshakyan, R. Bregovic, and A. Gotchev, "Image based rendering technique via sparse representation in shearlet domain," in *Image Processing (ICIP), 2015 IEEE International Conference on*, Sept 2015, pp. 1379–1383.

[43] ——, "Light field reconstruction using shearlet transform," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PP, no. 99, pp. 1–1, Jan 2017.

[44] J. P. Waters, "Holographic image synthesis utilizing theoretical methods," *Applied Physics Letters*, vol. 9, no. 11, pp. 405–407, 1966.

[45] X. Xiao, K. Wakunami, X. Chen, X. Shen, B. Javidi, J. Kim, and J. Nam, "Three-dimensional holographic display using dense ray sampling and integral imaging capture," *Journal of Display Technology*, vol. 10, no. 8, pp. 688–694, Aug 2014.

[46] M. Yamaguchi, H. Endoh, T. Honda, and N. Ohyama, "High-quality recording of a full-parallax holographic sterogram with a digital diffuser," *Optics Letters*, vol. 19, no. 2, pp. 135–137, 1994.

[47] M. Yamaguchi, H. Hoshino, T. Honda, and N. Ohyama, "Phase-added stereogram: calculation of hologram using computer graphics technique," vol. 1914, 1993, pp. 25–31.

[48] F. Yaraş, H. Kang, and L. Onural, "Real-time phase-only color holographic video display system using LED illumination," *Applied Optics*, vol. 48, no. 34, pp. H48–H53, 2009.

[49] T. Yatagai, "Three-dimensional displays using computer-generated holograms," *Optics Communications*, vol. 12, no. 1, pp. 43–45, 1974.

[50] R. Ziegler, S. Bucheli, L. Ahrenberg, M. Magnor, and M. Gross, "A bidirectional light field - hologram transform," *Computer Graphics Forum*, vol. 26, no. 3, pp. 435–446, 2007.