NORTHWESTERN
UNIVERSITY

Computer Science Department

Technical Report
Number: NU-CS-2022-02

December, 2021

Quantitative Analysis of Holography Interpolation With NeRF

Jiwon Choi

Abstract

The presented work provides a way of preserving holography images through rendering the holography images with NeRF and NeRF--, a neural radiance field representation method. This paper explores the feasibility of interpolating photographs of holograms by using the neural radiance field to synthesize novel views. Through quantitative analysis, it demonstrates that the NeRF model is able to interpolate photographs of holograms.

NeRF, neural radiance field, holography interpolation, holography view synthesis

NORTHWESTERN UNIVERSITY

Quantitative Analysis of Holography Interpolation With NeRF

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

MASTER OF SCIENCE

Field of Computer Science

By

Jiwon Choi

EVANSTON, ILLINOIS

June 2022

ABSTRACT

Quantitative Analysis of Holography Interpolation With NeRF

Jiwon Choi

The presented work provides a way of preserving holography images through rendering the holography images with NeRF and NeRF--, a neural radiance field representation method. This paper explores the feasibility of interpolating photographs of holograms by using the neural radiance field to synthesize novel views. To analyze the performance of the view synthesis method, quantitative comparisons between captured hologram images and interpolated views are done with the following metrics: MSE, PSNR, SSIM, and LPIPS. The calculation was done by comparing the ground truth and the interpolated images at a particular view point. The quantitative analysis demonstrates that the NeRF model is able to interpolate photographs of holograms.

Acknowledgements

It has been an honor for me to work with my advisors Dr. Oliver Cossairt, Dr. Jack Tumblin, Dr. Florian Willomitzer, and Dr. Marc Walton. I would like to express my gratitude for all the guidance, dedication, patience, and care. The advisement that I have received is the best experience that I have ever had in my research career. I believe that this work experience will certainly benefit throughout my life.

In addition, I would like to express my appreciation to Dr. Pengxiao Hao, for providing the holography image dataset as well as the contribution that she has made to explain these datasets.

The completion of this achievement would not have been possible without my former advisor, Dr. Nils Napp, and former laboratory colleagues, Dr. Maíra Saboia da Silva, Vivek Thangavelu, Yifang Liu, and Mehdi Seyed. They taught me the value of discipline, work ethics, and urged me to be prepared to be in academia.

Finally, I am grateful to my parents Wonback Choi and Sukja Park for providing me with endless support, especially in my education, and encouragement throughout my life.

List of Abbreviations

BA: Bundle Adjustment. 25

CGH: Computer-Generated Holograms. 13

COLMAP: An improved SfM algorithm, especially in robustness, accuracy, completeness, and optimized runtime. 23

LPIPS: Learned Perceptual Image Patch Similarity. 39

MLP: Multi-Layer Perceptron. 27

MSE: Mean Squared Error. 37

NeRF: Representing Scene as Neural Radiance for View Synthesis. 12, 26

NeRF--: NeRF Without Known Camera Parameters, same as NeRFMM. 12, 30

NeRFMM: NeRF Without Known Camera Parameters, same as NeRF--. 30

NVS: Novel View Synthesis. 30

PnP: Perspective-n-Point Problem. 25

PSNR: Peak Signal-to-Noise Ratio. 38

ReLU: Rectified Linear Unit. 27

SfM: Structure-from-Motion. 22

SSIM: Structured Similarity Indexing Method. 38

Table of Contents

ABSTRACT 3

Acknowledgements 4

List of Abbreviations 5

Table of Contents 7

List of Tables 9

List of Figures 10

Chapter 1. Introduction 12

1.1. Related Work 13

1.2. Thesis Organization 15

Chapter 2. Holography 16

2.1. Introduction to Holography 16

2.2. Holographic Stereograms 18

2.3. Related Work: Improvements In Holography 20

2.4. Limitations In Preservation of Holograms 21

Chapter 3. Overview of 3D Rendering Methods 22

3.1. COLMAP: Structure-from-Motion 22

3.2. NeRF: View Synthesis With Neural Radiance Fields 26

3.3. NeRF--: NeRF Without Known Camera Parameters 30

Chapter 4. Quantitative Analysis On Holography Interpolation 34

4.1. Experimental Setup 34

4.2. Quantitative Metrics 37

4.3. Results of Quantitative Measurements 39

4.4. Limitations 46

4.5. Future Work 50

Chapter 5. Conclusion 51

References 52

Appendix A. Holography Renderings With Different Methods 56

Appendix B. Holography Image Datatsets 58

List of Tables

4.1	Open-Source Rendering Software	37
4.2	Quantitative Analysis On Rendering at Center View Point	43
4.3	Quantitative Analysis On Interest-Area at Center View Point	45
4.4	NeRF Quantitative Measurement Results [**23**]	46
B.1	List of Holography Image Dataset	58

List of Figures

3.1	COLMAP Incremental SfM Pipeline	23
3.2	Overview of NeRF Pipeline	27
3.3	MLP F_Θ Network Architecture	28
4.1	Image Dataset Capture Settings	35
4.2	Lion Cubs Datasets (All, Col, Row) With COLMAP	40
4.3	Ground Truth Images: (a) Original; (b) Cropped	41
4.4	NeRF & NeRF-- Rendering Results With Different Datasets	41
4.5	Homorgraphy Transformation	44
4.6	Cropped NeRF & NeRF-- Rendering With Different Dataset	44
4.7	COLMAP from Raw Train Dataset & NeRF With Cropped Train Dataset	47
4.8	NeRF-- With Raw, Calibrated, and Cropped Train Dataset	47
4.9	Aperture Occlusion In Polaroid Dataset	49
4.10	NeRF & NeRF-- Rendering With Polaroid Datasets	49
A.1	Holography Rendering Result by Lucente et al. [**20**], Nishi et al. [**24**], and Matsushima et al. [**22**]	56

A.2 Holography Rendering Result by Chen et al. [**6**], Zhang et al. [**40**], and Walton et al. [**36**] 57

CHAPTER 1

Introduction

Preservation of holography images is a longstanding problem within the field of computer vision and computer graphics. The appearance of holograms change based on one's viewing location. It is also challenging to preserve holograms recorded on photographic film as their image-carrying emulsions gradually deteriorate over years and decades. Another challenge occurs in capturing all the information in hologram. In the capturing process, all the fringes, or diffraction patterns, has to be captured with high resolution. This capturing procedure is not an optimal way of preserving a hologram since it is not a practical to capture images densely enough to get good results for any viewing position by a simple linear interpolation of nearby captured images. This paper demonstrates a way to preserve holograms through the view synthesis field of neural radiance field. By utilizing a fully-connected neural network, the spatial location and the viewing directions are optimized and result in a volume density and radiance. Afterwards, a classical volume rendering technique will render this volume density and radiance. This novel view synthesis method is a scene representation as neural radiance fields for view synthesis, shortly called NeRF [**23**], and there is another method called NeRF-- which does not require the camera parameters of each image [**37**]. This work tries to answer the main question: "Is it possible to recreate images at different viewpoints from hologram by using NeRF to interpolate between captured hologram images?" To further address this question and support the claim, the quantitative analysis is provided with the following image quality

metrics: MSE, PSNR, SSIM, and LPIPS. The quantitative analysis supports the claim that NeRF is able to reconstruct and interpolate photographs of holograms.

To summarize, the main contributions are:

- Proposed holography preservation method with a neural radiance field synthesis, and overcome the challenges in conventional methods.
- NeRF and NeRF--, a radiance field novel view synthesis methods, are quantitatively analyzed with holography images.

1.1. Related Work

Rendering the restored (visualized) holography image is demonstrated to be a challenging task. The rendering approach have proposed starting from the early age. Lucente et al. (1995) proposed a method to render interactive holographic images [**20**]. The proposed method creates the holographic pattern by merging a series of view images rendered with a recentering shear-camera geometry. The method firstly renders the view with the conventional camera interpolation method (i.e. linear interpolation method), then computes the fringes to diffract light in specific directions. Finally, merge these stereogram components into the fringe pattern. Here, a stereogram component is a 2D representation of hologram in discrete way, and this representation allows to simplify the fringe calculation. In the rendered view, it exemplifies how the light should be diffracted or scattered by the fringes.

Nishi et al. (2011) proposed a novel rendering method to create computer-generated holograms (CGH) from 3D polygonal meshes used in computer graphics [**24**]. CGH digitally reconstruct interference and diffraction patterns ('fringes') of holograms. The

proposed novel Phong reflection model determines the spectral structure of the light reflected from simulated specular surfaces. Here, the spectral surfaces of the reflected light are modified to fit into a spectral shape.

Another polygon-based approach is proposed by Matsushima et al. (2012) [**22**]. This method is called "wave-field rendering method", and this computes the optical wave-field of virtual 3D scenes specified as collections of polygonal meshes. Analogous to the computer graphics' polygon-based method, smooth shading and texture mappings are applied to the rendered surfaces of holography images. The computer-generated high-definition holograms are composed of billions of pixels with depth reconstruction, making them impractical for most historical preservation tasks.

Chen et al. (2014) proposed a rapid hologram generation through a layer-based graphical rendering with angular tilting method [**6**]. A point cloud description of 3D objects are sliced into constant-depth layers or "billboards," and incorporates clear depth cues, occlusion, and shading in the rendered result. Here, angular tiling enables merging of multiple adjacent views to form a continuous image. The proposed algorithm strongly outperforms previous methods by its reduced computational time.

An enhanced layer-based rendering approach is proposed by Zhang et al. (2017) [**40**]. The conventional layer-based approach gives limited accuracy in simulating occlusion effects; this may cause complex surfaces such as leaves on trees to appear disconnected from their branches in the reconstructed result despite the continuous, connected form of the original object. The approach of Zhang et al. tackles the occlusion effect by applying slab-based orthographic projection. The proposed method is capable of rendering small hidden primitives for occlusion processing, and by generating shading information

in each layer. The resulting CGH captures complex 3D scenes with more accurate depth information, with fewer small occlusion errors.

These related rendering approaches are attempts to re-create holograms' abilities synthetically, by discretized CGH method. To the best of my knowledge, ours is the first published hologram interpolation approach. It is not a reconstruction of holography on computer, but instead utilizes the neural radiance field view synthesis to recreate the appearance of a hologram from any viewpoint, yet captured from a sparse, evenly-spaced set of photographs. Rendering with neural radiance field does not require any information of interference patterns (or "fringes"), unlike proposed CGH approaches. The rendered result of the CGH images from each author is provided in Appendix A.

1.2. Thesis Organization

The remainder of the thesis is structured as follows: §2 discusses more on holography and its limitations in preservation, along with various applications. §3 introduces the view synthesis methods with neural radiance field, and §4 covers more on experiment and the quantitative analysis on holography interpolation. Finally, §5 discusses strengths and challenges of the proposed method, and further improvements that can be taken.

CHAPTER 2

Holography

This chapter introduces holographic projection and the display characteristics of holograms. Unlike all previous imaging methods, holograms provide a visual display in a novel 3-dimensional format, one that recreates the actual wavefronts of coherent light reflected from a 3D scene, and made possible only by the 1960s-era advent of lasers and extremely high resolution photographic film [**17**]. Holography projection is widely used in various fields, including business, education, science, art, and healthcare [**8**]. This chapter discusses holographic recording and playback methods, along with the proposed holography enhancement methods proposed by several other researchers. This chapter concludes with discussion of the limitations of preserving existing holograms recorded on photographic film emulsions.

2.1. Introduction to Holography

Holography was invented by Dennis Gabor in 1948, who later received the Nobel Prize in Physics in 1971 for this work [**5**]. Holograms are a 2D record of the interference patterns or "fringes" formed by combining coherent light directly from a light source (usually a laser) and that same coherent light reflected from a 3D scene [**8**]. Holograms rely on recording the interference produced by the wavefront reflected from an object, while conventional photography relies on the projection of light rays incident onto a photosensitive surface without any significant interference effects [**36**]. Holography has

two steps: recording and playback. The desired output must be illuminated onto the object along with the intended playback light configuration, and these configurations will result in an interference pattern. This diffraction pattern will then be recorded by photo-chemical processes: either as intensity variations on a surface (e.g. high resolution photographic emulsions) or as changes in refractive index within a volume hologram [**11**].

2.1.1. Needs of Holographic Projection

The interest in 3D viewing is one of the factor that film-based holography persisted at the cutting-edge through the 1990s, and has had a revival with the advent of white-light holographic display techniques for head-mounted displays for virtual reality/augmented reality display systems [**21**]. Holography became less interested in these days, due to the cost of holographic data disks and holotechnology drives (i.e. holographic data disks), where holographic data gets encoded and stored. Another issue is that these devices require an expert to manipulate [**8**]. Yet, holograms are used in a wide array of fields.

2.1.2. Application of Holography

There are multiple applications in holography: security uses, advertisement, artwork, medical and military usage, and etc. Each of these applications is discussed below.

2.1.2.1. Security. A pioneer in holography, Stephen Benton, has a patent in rainbow hologram which is delivered with a white light transmission. This embossed rainbow holograms can be found on credit cards, banknotes, stamps, and other security uses [**4**].

2.1.2.2. Advertisement. Coca-Cola gave a sales conference presentation in Prague, in 2009. During the presentation, senior directors of the company were beamed into the stage

as 3D holograms. The content of their presentation, including the product information, was also visualized with holograms [**8**].

2.1.2.3. Art. Holography collections are displayed in museums as artworks. For instance, Stephen Benton and Polaroid Staffs' Engine no. 9 (1975) -the train hologram- is in MIT Museum Collection [**36**]. Benton's other collections including "Portrait of Stephen Benton and Yuri Denisyuk" are also exhibited in MIT Museum Collection, and archived in the web museum (`https://webmuseum.mit.edu`). There is another collection at The Movieum of London Museum. The holography collection displays their large scale show and event, and medium and small size exhibitions [**16**].

2.1.2.4. Medical. 3D Medical Animation Studio -3d medical illustrations (2008), proposed a holography display of medical animations, along with the option of interactivity [**33**].

2.1.2.5. Military. Zebra Imaging (2008) proposed a 3D holographic map for military usage, and reported a higher effectiveness in 3D holography map than using a traditional 2D map [**13**]. Under military settings, this holography map can be used to familiarize terrain, planning raids, debriefing after incidents, and etc.

2.2. Holographic Stereograms

To render a holography scene, a closely-spaced series of discrete perspective views is captured, through indexing a camera on a rectangular grid or sliding camera in front of the scene. These films will then be fed into a laser-illuminated optical printer, which merges images and yield a synthetic hologram. This procedure is called as a "holographic

stereogram." The holographic stereogram diffracts a fraction of the light to several different viewing locations upon illumination. Then it modulates beam with the corresponding perspective view, which allows to be visible to an eye at the location corresponding to that point of view. Depending on the parallex, these viewing perspectives can be located differently, such as centered on a rectangular grid, or along a line. The different views in left and right eye (stereoscopic pair) yield an impression of depth. When the observer moves up-down or left-right, he will sweep across the closely-spaced perspective views and be able to observe a continuous and realistic view of a holographic object [**3**].

There are two ways of generating and viewing holograms: laser-viewed, and white-light holographic stereograms. As the name indicates, laser-viewed method uses the emission of the laser beam, while the white-light method uses white-light sources such as halogen lamps [**3**]. Each of these methods will be briefly discussed.

2.2.1. Laser-Viewed Holographic Stereogram

In the early stage of holographic stereogram, methods utilize the monochromatic light from the source. Types of monochromatic lights include lasers and mercury arc lamps. The downside of this laser-viewed method is the limitation of the holographic uses in darkened environments [**3**]. "Denisyuk" reflection holography is one of the laser-viewed holographic stereogram methods. This method uses the laser beam that is emitted from the source, passes through a beam expander and this will reach to the holographic plate and the object [**36**].

2.2.2. White-Light Holographic Stereogram

Utilizing a single point-like white light, such as small bright overhead halogen lamps will allow the holographic images to be viewed in common environments, and not limited to the dark environments. Unlike monochromatic lights, white-lights provide strong, broad-spectrum illumination, so that the holograms can be viewed in much brighter surroundings [**3**].

2.3. Related Work: Improvements In Holography

A promising recent direction in holography imaging, several works have proposed to make amendments in conventional holography methods. These enhancements include circumventing the scattering effect, efficient simulation of holographic process, and reservation of recorded holography.

Willomitzer et al. (2021) proposed a method which exploits spectral correlations in scattered wavefronts, in order to reduce detrimental 'speckle' effects of scattering [**38**]. The scattering presence in the imaging path between an object and observer, and this critically limits the visual acuity. The proposed Synthetic Wavelength Holography method recovers a holographic representation of hidden targets, over a nearly hemispheric angular field of view.

Ballester et al. (2021) proposed an efficient simulation of holography [**1**]. The proposed method applies propagation to the free-space Helmholtz Green's functions and the Born approximation assumption to enhance the efficiency in its computational time.

Peixeiro et al. (2016) performed a quantitative analysis on compression methods for holographic data [**25**]. The quantitative measures indicate that the most appropriate reconstruction method varies with applications, and this depends on the holography content generation method and reconstruction distances of holography. Yet, their 'HEVC Intra' method outperforms others as an encoding method due to its improved representation formats: Phase Shifted Distances and Real-Imaginary formats. This yields only a minute difference in overall compression performance.

Huebschman et al. (2003) proposed a preservation of dynamic holographic image projection with digital micromirror devices (DMD) [**15**]. They first calculated fringes from the object into a 3D scene, and then record the 2D digital hologram into the device.

2.4. Limitations In Preservation of Holograms

Walton et al. (2021) claimed that holograms are difficult to be preserved [**36**]. Since holograms recorded on photographic film emulsions are an active material, they require user interaction to observe and appreciate fully: viewers must move as they watch the hologram to see occluded portions. But, as the holograms deteriorate over time there are technological challenges to preserve them. For instance, the deterioration of the film emulsions and film base are slowly destroying the Stephen Benton's pioneering and historic holography collections. Thus, it is important to capture the visual appearance of these holograms to save in permanent digital form. This will allow accurately recreating the appearance of these historic holograms long after the original film emulsions are gone. To the best of my knowledge, the experiment discussed in §3 is the first hologram preservation approach that utilizes the neural radiance field method.

CHAPTER 3

Overview of 3D Rendering Methods

This chapter introduces various view synthesis methods in order to interpolate and reconstruct the shapes depicted in given images. First of all, COLMAP, an open-source software package that reconstructs 3D models from photos using structure-from-motion techniques, will not only render views of its own triangle-mesh reconstructions, but also computes the camera positions that captured the input photos. Two other approaches to 3D capture and viewing, the NeRF and NeRF-- software packages, use deep neural nets to build estimates of the entire radiance fields of the scene. While the NeRF package requires camera lens parameters (intrinsic calibration) and camera poses (extrinsic calibration) obtained via COLMAP, the NeRF-- does not –it forms its own camera parameter estimates from the supplied photographs. Both methods have a same baseline approach of NeRF, which is a representation of novel scenes with neural radiance fields. In this chapter, each of these rendering methods will be introduced, and these methods will later be used in the following chapter to perform the quantitative analysis on the photograph of hologram image dataset.

3.1. COLMAP: Structure-from-Motion

Structure-from-Motion, often called SfM, is a reconstruction process of a 3D structure from its projections into data sets of photographed images, which are taken from many different camera positions. Multiple SfM strategies have been developed over decades of

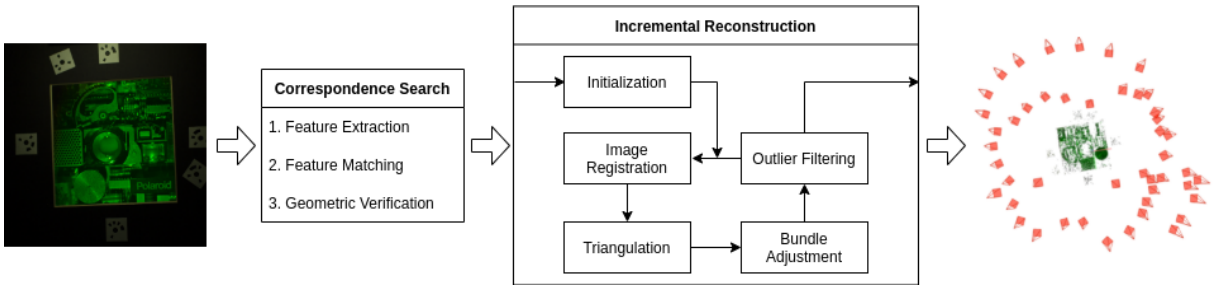

Figure 3.1. COLMAP Incremental SfM Pipeline

work, including incremental, hierarchical, and global approaches. Of these, the incremental SfM is now perhaps the most well-regarded strategy for accurate 3D reconstruction from unordered image datasets [**28**]. This SfM method optimizes the perspective camera which is parameterized by an eleven-parameter projection matrix. Yet, such optimization algorithms may fall into the local minima, especially if it is a large-scale dataset. Thus, it is critical to provide a good initial pair of images. Here, the incremental approach estimates and optimizes the parameters one-by-one, by adding one camera at a time, rather than estimating all the parameters with all the images all at once [**31**]. The hierarchical approach organizes the image dataset into a hierarchical cluster tree, and performs the reconstruction from the leaves to the root in the hierarchical way [**10**].

Global SfM requires local estimates of geometry, and from there solving for a global set of camera poses [**7**]. Note that hierarchical and global approaches are not used in the COLMAP SfM package, which uses the incremental approach instead.

The COLMAP SfM pipeline starts with feature extraction which identifies 3D point locations in the photographed scene, and feature matching that finds the 2D image location of a 3D feature in two or more photographed images, and then geometric verification that performs multi-camera tests to ensure the matched features describe a single 3D

location. After this, the incremental reconstruction involves an iterative procedure after the initialization, and the iterative process is as follows: image registration which chooses the (next) best view, triangulation that reconstructs from multiple images, bundle adjustment which solves extrinsic parameters of camera, and outlier filtering [**28**]. Details are discussed in the following paragraphs. Figure 3.1 visualizes the explained process with an example of the Polaroid dataset. The captured Polarioid hologram input image dataset is fed into the model, and the model applies searching and the incremental reconstruction, then finally outputs the 3D rendered that depicts camera position and aiming directions as tiny red pyramids –the peak of the pyramid is the camera's center of projection, the base is the camera's image plane, and the sides form its viewing frustum.

Here, COLMAP proposed a new SfM pipeline that improves the following challenges that the conventional SfM has: robustness, accuracy, completeness, and optimized runtime. The feature extraction step chooses features that are invariant under radiometric and geometric changes. These invariant features will later be uniquely recognized by SfM in multiple images [**39**]. The feature matching step searches for the most similar feature in one image for every feature in another image. Since the matching step is based on the appearance of the images, corresponding features in images might not actually map to the same 3D locations. Thus, geometric verification is introduced to verify whether the estimated transformation corresponds to the features between images using the geometry projection. This verification allows to improve the robustness of the initialization and triangulation [**28**].

After the searching step, initialization should be done in order to incrementally reconstruct an object. During the initialization, choosing a good initial pair is critical;

initializing from a dense location with many overlapping cameras results more robust and accurate reconstruction result [**2**]. Image registration step chooses the next best view, and thus minimizes the reconstruction error. This step solves the Perspective-n-Point (PnP) problem using corresponding triangulated points in the registered images [**9**]. PnP solves for an orientation and position of a fully-calibrated perspective camera, with $n \geq 3$ number of 3D points of the object framework and their corresponding 2D projections [**42**]. Note that solving this PnP problem depends on the number of observations and the distribution of those [**19**]. To minimize the uncertainty, choosing the image with the most triangulated points is critical [**30**].

Next step is triangulation. Triangulation is the intersection of two known rays in space from 2 or more known camera positions, and to an extent, refers to the reconstruction from several images in photogrammetry [**12, 29**]. Triangulation in SfM enables registration of new images by providing additional corresponding points, as well as increase the stability through maximizing the overlap [**34**]. COLMAP proposed an efficient and robust way of triangulation, by applying Kang et al.'s (2014) estimation of points via feature tracking [**18**]. Feature tracking may generate lots of outliers due to the poorly matched 3D position estimates, but COLMAP overcomes these by a recursive triangulation method that determines consistent trajectories for multiple points from faulty or inconsistently merged feature sets [**28**]. Lastly, further refinements are done by bundle adjustment (BA). BA solves for the joint non-linear refinement of camera extrinsic parameters P_c and 3D point-location parameters X_k. With the function π, which projects 3D scene point locations to each camera's 2D image space, compares them to the originally detected feature locations, and adjusts parameters incrementally to minimize their

differences, known as the reprojection error. The loss function ρ_j potentially allows to reduce the weight of outliers. The equation 3.1 refers to the explained BA equation.

$$E = \sum_j \rho_j(||\pi(P_c, X_k) - x_j||_2^2) \tag{3.1}$$

BA uses Levenberg-Marquardt optimization for bundle adjustment –a method that can solve non-linear problems in the least-squares sense. [**34**]. With these steps and implementations, COLMAP was able to make improvements in completeness, robustness, accuracy, and efficiency from the naive SfM.

The utilization of COLMAP open-source software is necessary to execute some of the view synthesis steps. The input of the view synthesis with neural radiance field and neural factorization (NeRF, NeRF--) requires the camera positions COLMAP can compute, along with the image dataset. These neural view synthesis methods will be further discussed in later sections.

3.2. NeRF: View Synthesis With Neural Radiance Fields

The open-source software for representing scene as neural radiance field, shortly called NeRF, synthesizes novel views of scenes through optimizing the continuous volumetric scene function. The required inputs are a set of images with known camera poses, intrinsic parameters and scene bounds as obtained by COLMAP. Afterwards, the fully connected deep neural network without convolutions is utilized with an input of a single continuous 5D coordinate. The 5D coordinate here is composed of spatial location (x, y, z) and viewing direction (θ, ϕ). This outputs the volume density and emitted radiance at the spatial location. The conventional volume rendering step projects the output colors and

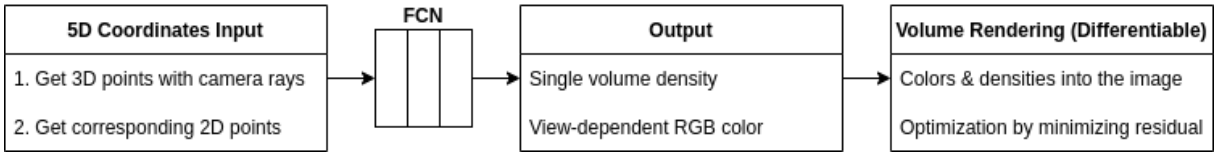

Figure 3.2. Overview of NeRF Pipeline

densities into an image by merging 5D coordinates with camera rays [**23**]. The overview of this NeRF process is illustrated in Figure 3.2.

NeRF proposed the novel view synthesis by directly optimizing a continuous 5D scene representation parameters, and this minimizes the error between rendered and ground truth observed images. Minimizing this residual across multiple views allows the network to predict a coherent model by assigning high volume densities with accurate scene colors. The first step, neural radiance field scene representation takes an input of 3D location of $x = (x, y, z)$ and 2D viewing direction (θ, ϕ). These are fed into the multi-layer perceptron (MLP) network to optimize its weights Θ and to map the 5D coordinates to the corresponding volume density σ and color emitted $c = (r, g, b)$. Equation 3.2 shows the described the fully connected layer model [**23**].

$$F_\Theta : (x, d) \rightarrow (c, \sigma) \tag{3.2}$$

The model is built with 8 fully-connected layers and one additional fully-connected layer, with 256 and 128 channels respectively. All layers used the ReLU activation function for each layer. The MLP first supplies the volume density σ and a feature vector with a dimension of 256, from the first 8 fully-connected layers with the 3D coordinates x. This feature vector is then used in the last layer along with the 2D viewing direction to construct the view-dependent RGB color outputs c. Figure 3.3 visualizes the network

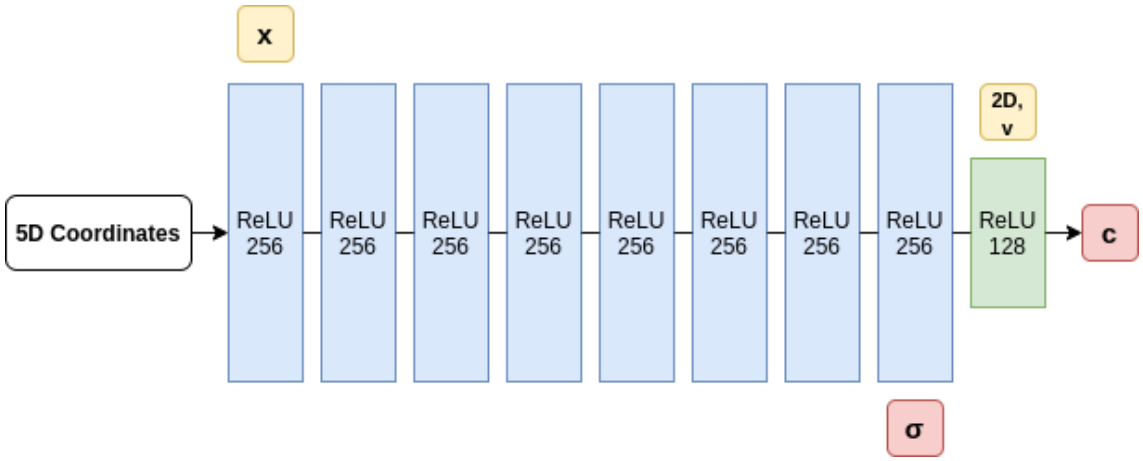

Figure 3.3. MLP F_Θ Network Architecture

architecture of the model [**23**]. After obtaining the volume density σ and the colors c via training the MLP model, classical volume rendering is done using the radiance field result. The differential probability of a ray terminating at an infinitesimal particle at point x is the volume density $\sigma(x)$. Within the near t_n and far t_f bounds, the expected color $C(r)$ of camera ray $r(t)$ can be explained with the differential equation 3.3.

$$C(r) = \int_{t_n}^{t_f} T(t)\sigma(r(t))c(r(t), d)\, dt, \text{where } T(t) = \exp\left(- \int_{t_n}^{t} \sigma(r(s))\, ds \right) \tag{3.3}$$

The function $T(t)$ denotes the accumulated transmittance, which is a probability that the ray travels from t_n to t. $C(r)$ renders the color of each ray. The estimation of this color $C(r)$ is required in order to render the view with continuous neural radiance fields [**23**].

This ordinary or "vanilla" NeRF procedure has a limitation when rendering the complex scene: it does not converge to a high resolution rendering. Thus, the NeRF method adds a new feature to overcome this limitation: positional encoding of coordinates, and

hierarchical volume sampling. As Rahaman et al. (2019) proposed, there is an empirical evidence of bias in deep neural networks; lower frequencies are learned first in deep networks [**26**]. Similarly, vanilla NeRF gave represented high-frequency (small, localized) color and geometry variations poorly. Thus, positional encoding is introduced here. This encoding was adopted from the popular Transformer architecture, where they provided the discrete positions of tokens in a sequence [**35**]. This converted input, the order of the sequence, is required in the model because the architecture contains neither recurrence nor convolution.

In NeRF, positional encoding is used to map continuous coordinates into a higher dimensional space. This allows MLP to approximate a higher frequency function more easily, and makes the model less biased [**23**]. Hierarchical volume sampling makes the rendering more efficient. Evaluating neural radiance field at N points along each camera is inefficient since free space and occluded regions will also be evaluated every time although they do not contribute significantly to the rendered output. In NeRF, "coarse" and "fine" networks are simultaneously optimized instead of using a single network. Each network samples N_c and N_f locations and evaluate the corresponding network at these locations. Finally, the fine network computes the final rendered color of the ray using all samples $N_c + N_f$. The loss function in equation 3.4 is simply the total squared error between the pixel colors of rendered and ground truth images.

$$L = \sum_{r \in R} \left[||\hat{C}_c(r) - C(r)||_2^2 + ||\hat{C}_f(r) - C(r)||_2^2 \right] \tag{3.4}$$

R denotes the set of rays in each batch and r denotes each ray of it. $C(r)$, $\hat{C}_f(r)$, $\hat{C}_c(r)$ refers to the ground truth, coarse and fine volume predicted in RGB colors, respectively.

Note that the loss of coarse $\hat{C}_c(r)$ should be minimized although the final rendering is done in the fine $\hat{C}_f(r)$. This will allocate samples in the fine network by utilizing the weight distributions from the coarse network. This process allocates more samples to the interested regions which contain more 'visible' content [**23**].

Although it requires the camera positions and information in addition to the image dataset, NeRF is able to produce renderings by representing 3D scenes as 5D neural radiance fields and sampling them to create images from any desired 3D viewpoint. Mildenhall et al. (2020) also claimed that NeRF outputs better renderings compared to the discretized voxel representations via training deep convolutional networks [**23**].

3.3. NeRF--: NeRF Without Known Camera Parameters

Another open-source software for representing scene as neural raidance field is called NeRF--, or NeRFMM. Unlike NeRF, NeRF-- proposed a novel view synthesis (NVS) without camera positions or intrinsic parameters. This state-of-art end-to-end framework is able to output a synthesized view only with the input RGB images, and without camera parameters. Instead, NeRF-- jointly optimizes the intrinsic and extrinsic camera parameters while training the NeRF model. Wang et al. (2021) proposed that NeRF-- results on par with the baseline trained with COLMAP, and also produces robust results in case where COLMAP fails [**37**].

NeRF-- added the important step of a new joint optimization of camera parameters to the NeRF training step. Refer to Algorithm 1, which is taken from Wang et al. (2021), to see the detailed steps [**37**]. This joint optimization of model parameter Θ and camera intrinsic and extrinsic parameter of Π, with an input of image set I is denoted in

Algorithm 1 NeRF-- Pipeline [**37**]

Input: N Images, $I = \{I\}_{i=1}^{N}$
Output: NeRF Model F_Θ, camera parameters $\hat{\pi} = (\hat{f}_x, \hat{f}_y, \hat{\phi}_i, \hat{t}_i)$

```
import torch.nn as nn
f̂_x, f̂_y ← nn.Parameter(shape=(2,), ···)                ▷ estimate initial focal lengths
[φ̂_i] ← nn.Parameter(shape=(N, 3), ···)                 ▷ estimate initial rotation matrix
[t̂_i] ← nn.Parameter(shape=(N, 3), ···)                 ▷ estimate initial translation matrix
F_Θ ← NeRF                                               ▷ get NeRF network model
for i in range(N) do                                     ▷ training step
    for m in range(M) do                                 ▷ randomly selected pixel locations
        d̂_{i,m} = construct_ray(f̂_x, f̂_y, φ̂_i, t̂_i, ρ_{i,m})   ▷ get a ray from π̂ through the pixel ρ
        for h from h_n to h_f do                         ▷ within the ray range
            x_j ← sample_point(d̂_{i,m}, t̂_i, h)          ▷ joint optimization (Equation 3.5)
            c_h, σ_j ← F_Θ(x_h, d̂_{i,m})                 ▷ forward NeRF
        end for
        Î_{i,m} ← render_ray([c_h], [σ_h])               ▷ rendering views (NVS)
    end for
    L ← loss(Î_i, I_i)                                   ▷ loss from reconstructed pixel (Equation 3.5)
    L.backward()                                         ▷ backward loss
    optimizer.update(f̂_x, f̂_y, [φ̂_i], [t̂_i], Θ̂)          ▷ update the optimzer
end for
```

mathematical way in Equation 3.5.

$$\Theta^*, \Pi^* = \underset{\Theta^*, \Pi^*}{\operatorname{argmin}} L(\hat{I}, \hat{\Pi} | I) \tag{3.5}$$

There are four camera parameters handled in NeRF--: $\hat{f}_x, \hat{f}_y, \hat{\phi}_i, \hat{t}_i$. The focal lengths $\hat{f}_x$, and $\hat{f}_y$ can be directly optimized. Note that NeRF-- assumes that the camera principle points are as follows: $c_x \approx \frac{W}{2}$ and $c_y \approx \frac{H}{2}$, where W and H denote the width and height of the image. In case of extrinsic parameters, the rotation vector t_i and translation vector ϕ_i, for each image I_i, are also directly optimized. After initializing the intrinsic and extrinsic parameters, for each image $\hat{I}_i$, NeRF-- randomly selects M pixel locations $\{\rho_{i,m}\}_{m=1}^{M}$. These pixel locations are what NeRF-- wants to reconstruct from NeRF model

F_Θ. Rendering the color of each pixel $\rho_{i,m} = (u, v)$ involves the projection of a ray $\hat{r}_{i,m}(h)$, where h denotes the ray range from h_n to h_f, from the camera position through the pixel into the radiance field. Note that constructing the ray also requires the camera parameters $\hat{\pi}_i = (\hat{f}_x, \hat{f}_y, \hat{\phi}_i, \hat{t}_i)$. This can be resolved with Equation 3.6, after converting into the rotation matrix R from the representation of normalized rotation axis ω and a rotation angle α, as solved in Equation 3.7.

$$\hat{d}_{i,m} = \hat{R}_i \begin{pmatrix} \frac{(u-\frac{W}{2})}{\hat{f}_x} \\ \frac{-(v-\frac{H}{2})}{\hat{f}_y} \\ -1 \end{pmatrix} \tag{3.6}$$

$$R = I + \frac{sin(\alpha)}{\alpha}\phi^\wedge + \frac{1 - cos(\alpha)}{\alpha^2}(\phi^\wedge)^2, \tag{3.7}$$

where the skew operator $(\cdot)^\wedge$ in Equation 3.7 converts a vector ϕ into a skew matrix. Since the model may fall into local minima in case where the optimized camera parameters are sub-optimal, additional refinement step is introduced in NeRF--. Falling into local minima may result a blurry rendered output. This refinement step is done by dropping the trained NeRF model and re-initialize with random parameters, after the first training step. Note that camera parameters π are stay remained without re-initialization. After resetting the parameters, repeat the joint optimization step. This refinement results relatively sharpened rendering images [**37**].

Yet, there are some limitations in NeRF--. It struggles when rendering the scene with large texture-less regions or photometric inconsistent across frames -i.e. motion blur. Also, NeRF-- is limited to render the forward-facing scenes or short camera trajectories,

and struggles in rendering 360° scenes or large camera trajectories. Still, this NeRF-based pipeline results a NVS without known camera parameters, and estimates the parameters through joint optimization [**37**].

CHAPTER 4

Quantitative Analysis On Holography Interpolation

After introducing the limitations of preserving holography images as well as the ways of performing interpolations with images, this chapter introduces various methods applied in order to interpolate holography images including quantitative measures on the rendered result.

4.1. Experimental Setup

This section discusses the experimental setup. To perform this experiment, a photographed hologram image dataset was collected by illuminating the hologram with a fixed white LED light and photographing with a gantry-mounted camera that captured images from a 2D grid of positions in front of the hologram. To train the NeRF model with this dataset required some new hardware configurations that were difficult for us to achieve. I found that the workstation should be equipped with a GPU of at least 10GB RAM for this task. Lastly, the three open-source software packages required to perform the analysis must be carefully and correctly configured.

4.1.1. Obtaining Image Datasets

In this experiment, we named our three holography image datasets as: "train", "lion cubs," and "polaroid". Each dataset captured the visual appearance of an historic hologram made by Steven Benton in the late 1960s and 1970s, and was named to describe the

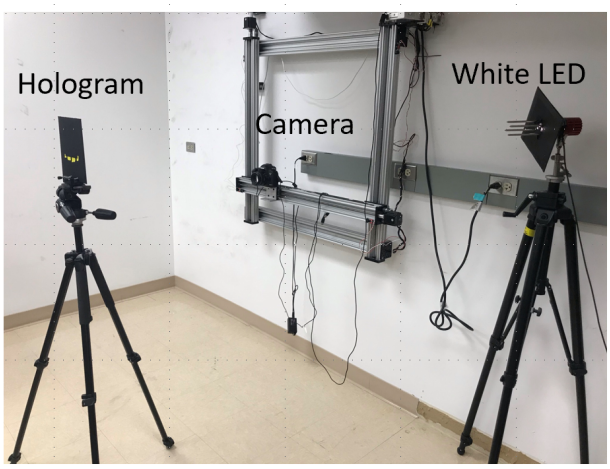

Figure 4.1. Image Dataset Capture Settings

3D scenes these holograms depicted. These image sets were taken using a DSLR camera with a 50mm lens from numerous viewpoints. In the case of train and lion cubs dataset, the camera moved in an x-y plane, with a vertical step size of 3.7cm and a horizontal step size of 2.3cm. The distance between the hologram plane and the DSLR camera lens is approximately 1.11m, and the camera lens and the holography panel are located in parallel. This setup is shown in Fig. 4.1. On the other hand, the polaroid dataset is not taken in a sequential order, but taken in a spiral view points with a different (tilted) camera angles. There are multiple fiducial markers, provided by Agisoft Metashape [1], attached around the holography panel. These landmarks have a role of features of the scene, to obtain the camera position later on. The list of datasets that are utilized in this experiment is addressed in Appendix B.

[1]Agisoft Metashape is a software that performs photogrammetric processing of image datasets and produces a 3D rendered output (`https://www.agisoft.com/`)

4.1.2. Hardware Specifications

The view synthesis step utilized two servers with the following configuration for training and rendering the image dataset. The first server has a configuration of Intel® Core™ i9-9940X CPU @ 3.30GHz with 128GB RAM, and three GPUs of NVIDIA® GeForce® RTX™ 2080 Ti. Another server has a configuration of Intel® Xeon® Processor (Cascade-lake) with 96GB RAM, and three GPUs of NVIDIA® Quadro RTX™ 8000. Although both servers lack of NVIDIA® NVLink® bridge and thus cannot distribute the task over multiple GPUs, the utilization of a single GPU suffices to train and render the novel view with the optimization of neural radiance fields. With these settings, the average run time of the training step takes around 15 to 20 hours, and rendering the photo-realistic view takes less than 30 minutes.

Due to the lack of permission on those aforementioned servers, obtaining camera positions through COLMAP utilized the local system. The local desktop system has a configuration of AMD Ryzen™ 5 5600X 6-Core Processor, with 16GB × 2 DDR4 3600MHz RAM and a single NVIDIA® GeForce® GT 1030 GPU. Obtaining camera positions takes around 10 minutes, and converting these obtained positions through NeRF-formatted positions takes less than a minute.

4.1.3. Software Usage

Chapter 3 introduces three rendering methods: COLMAP, NeRF, and NeRF-- renderings are from all open-source software packages. Note that this also requires another software package, LLFF, to convert the COLMAP positions into the NeRF-formatted positions.

Table 4.1. Open-Source Rendering Software

Method	Open-Source Implementation
LLFF	`https://github.com/Fyusion/LLFF`
COLMAP	`https://github.com/colmap/colmap`
NeRF	`https://github.com/bmild/nerf`
NeRF--	`https://github.com/ActiveVisionLab/nerfmm`

The open-source URL of each method can be found in Table 4.1. Utilizing these software packages enables analysis of the interpolation of holography image data sets.

4.2. Quantitative Metrics

In this section, the quantitative metrics used to compare the performance of renderings will be introduced. The metrics include: MSE, PSNR, SSIM, and LPIPS. By comparing the result from multiple metrics here, we can verify whether the rendering with neural radiance field is able to interpolate hologram images. The calculation pipeline can be found on: `https://github.com/cjw531/neural-rendering`.

4.2.1. MSE

The mean squared error (MSE) simply calculates the differences between two images. MSE is computed as the average of the squared pixel intensity differences between a source image and a rendered image [**32**]. The mathematical notation is described in Equation 4.1.

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - K(i,j)]^2 \tag{4.1}$$

4.2.2. PSNR

The peak signal-to-noise ratio (PSNR) metric measure the quality of reconstruction, and is an extension of MSE. The Equation 4.2 shows that PSNR make a use of MSE, while it also incorporates the maximum value of the pixel -bits stored in the image, i.e. 8-bits image has a maximum pixel value of $2^8 - 1 = 256$.

$$PSNR = 10 \log_{10} \left(\frac{255^2}{MSE} \right) \tag{4.2}$$

The PSNR value approaches infinity, as the MSE goes to 0. This shows that the higher PSNR value provides the higher image quality, while a smaller value of the PSNR implies the given two images are different [**14**].

4.2.3. SSIM

The structure similarity index method (SSIM) evaluates the structural similarity based on a mathematical model of human perception. Equation 4.3 computes a normalized mean value of structural similarity between the two images.

$$SSIM = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \tag{4.3}$$

where μ_*, σ_*^2, σ_{xy} indicates the average, variance, and covariance of image. c variables are stabilizer offset for the weak denominator [**27**].

4.2.4. LPIPS

The learned perceptual image patch similarity (LPIPS) metric evaluates the distances between patches. LPIPS is one of the human perceptual metrics, which systematically evaluate deep features across different architectures. Higher values indicate that they are more varying, while the lower values indicate that two images are similar [**41**]. In this experiment, the open-source LPIPS library (ver. 0.1) is used to make comparison. This open-source library can be found in: `https://github.com/richzhang/PerceptualSimilarity`.

4.3. Results of Quantitative Measurements

In this section, the rendered results NeRF and NeRF-- will be discussed. The experiment utilized the three types of lion cubs datasets. Both visual and quantitative performance evaluations at a fixed central camera locations have done.

4.3.1. Input Dataset

To compare the neural radiance rendering performance in a numerical way, lion cubs dataset is used. In this experiment, the three different lion cubs datasets are provided to the NeRF model:

(1) `lioncubs-all`: dataset with all captured images (sample size: 60)

(2) `lioncubs-col`: dataset without center columns, a 8:2 split (sample size: 48)

(3) `lioncubs-row`: dataset without center rows, a 2:1 split (sample size: 40)

To make it easy to mention, we named the dataset as “all”, “column” (col), and “row”, respectively. The split of the dataset is also visualized in Figure 4.2. The numbering

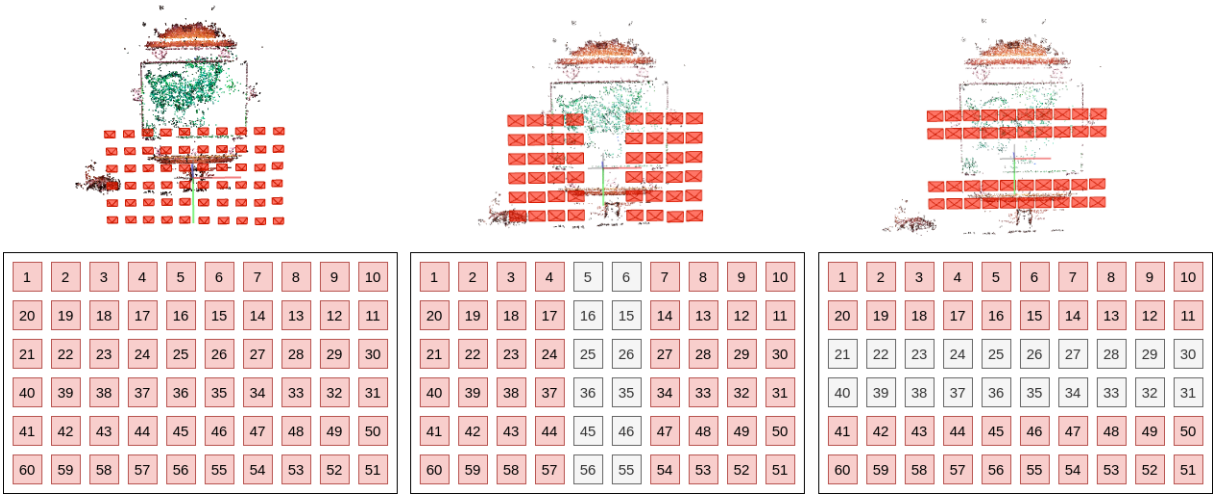

Figure 4.2. Lion Cubs Datasets (All, Col, Row) With COLMAP

of the images are the sequence of the images captured, as discussed in Section 4.1.1. The red-colored images belong to each dataset, all, col, row, respectively. The view point locations of images correspond to the camera position obtained through COLMAP. Interesting point to note is that the density of COLMAP point cloud renderings. As the number of input image decreases (figure from left-to-right) the more the rendering output becomes sparse.

4.3.2. Rendering Performance of NeRF

To compare the performance, the rendering at the central view point location is used. The rendered performance is compared against the ground truth captured hologram images in Figure 4.3. Using NeRF and NeRF--, three lion cubs dataset produced a rendering result as seen in Figure 4.4. To have the best rendering results, the model parameters, the number of training step (iterations), and resize factor is not modified when rendering this holography image. The parameters are set as default, given by each author. By

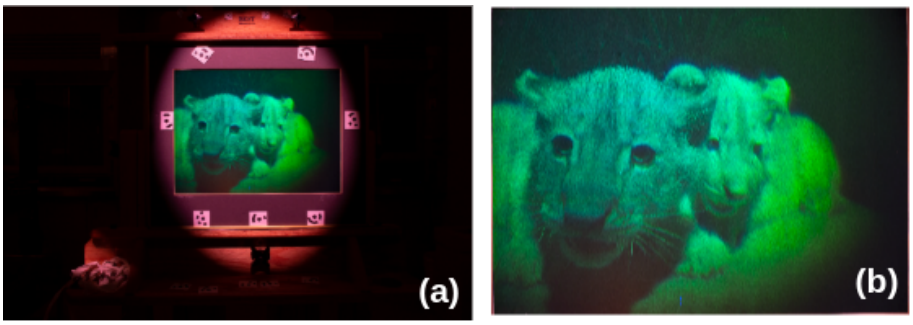

Figure 4.3. Ground Truth Images: (a) Original; (b) Cropped

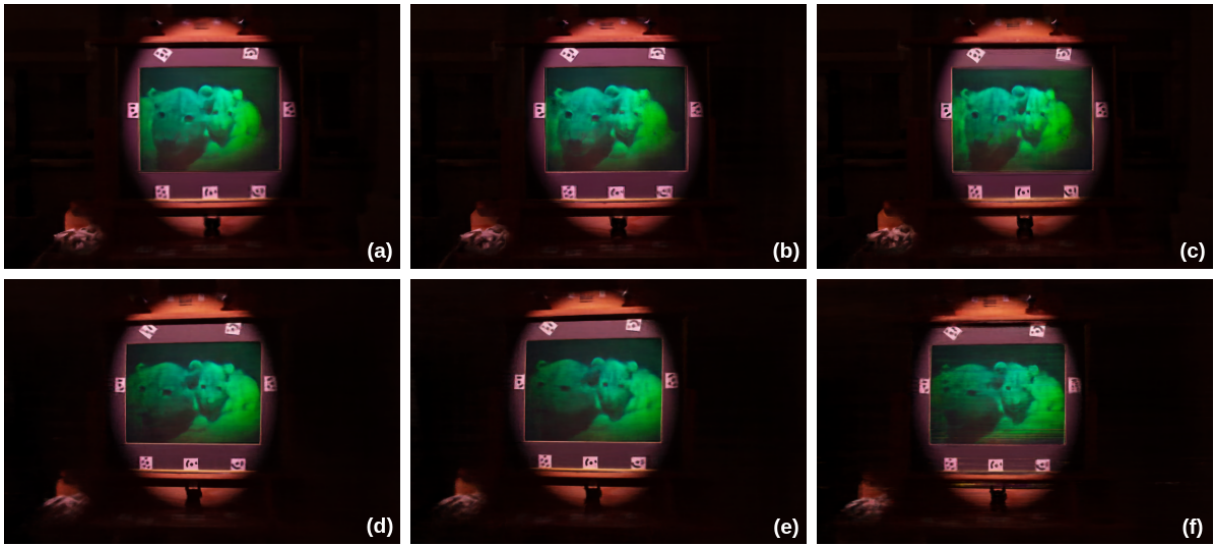

Figure 4.4. NeRF & NeRF-- Rendering Results With Different Datasets

NeRF rendering results at the center location with (a) all, (b) column, and (c) row dataset. NeRF-- rendering results at the center location with (d) all, (e) column, and (f) row dataset.

default, the resize ratio of NeRF and NeRF-- is 8 and 4, respectively. Thus, the resulted rendering image has a dimension of 720×480 and 1440×960, from the original input image dimension of 5760×3840. This resizing factor will be revisited later when performing the quantitative analysis.

As briefly discussed in Figure 4.2, NeRF and NeRF-- rendering result also demonstrate the similar trend. As the number of image samples decrease, the quality of rendering also

decreases. This is a nature of machine learning –the more the sample provided, the model will less likely be biased, and produce a prolific model after training. Surprisingly, both NeRF and NeRF-- resulted in a decently-interpolated central view point, even without having the information (the input data) of center column or row.

In the visual comparison aspect, NeRF outperforms NeRF--. This is due to the existence of camera parameters obtained via COLMAP in NeRF. With the camera parameters computed in mathematical way, COLMAP has more robust camera parameters than the trained parameters obtained through joint optimization in NeRF--. This can be evaluated by comparing the location of the holography panel in the ground truth versus the rendered result. This also affected the 'center' location of the view point. The rendering results of NeRF, at the fixed view point located in the center, is relatively more accurate than the one of NeRF--. NeRF-- initializes the camera parameter by assuming that the camera principle points are the midpoint of the pixel of the image, and this assumption may lead to the inaccurate optimization of camera parameters.

These NeRF and NeRF-- rendering results are best viewed as videos, so it is recommended to view the supplementary full-rendered results for visually convincing comparisons -i.e. continuous rendering results in various view points even without some images. The supplementary videos can be viewed here: `https://youtube.com/playlist?list=PLCVV8jHcNib2x8O69JZgkum6l1j63_K2V`.

The synthesized rendering results need to be evaluated in mathematical way as discussed in Section 4.2. Before the computation step, the preprocessing step is taken first. Since the resizing factors are differ by the rendering method, the ground truth image has resized into its $\frac{1}{8}$, and NeRF-- renderings have resized into its $\frac{1}{2}$, to match the image

Table 4.2. Quantitative Analysis On Rendering at Center View Point

Metrics	**All Dataset**		**Column Dataset**		**Row Dataset**	
	NeRF	NeRF--	NeRF	NeRF--	NeRF	NeRF--
MSE ↓	37.9767	44.1436	38.7402	45.9238	38.6376	47.3010
PSNR ↑	32.3356	31.6821	32.2492	31.5104	32.2607	31.3821
SSIM ↑	0.6550	0.6136	0.6443	0.6013	0.6452	0.5877
LPIPS ↓	0.2511	0.3345	0.2514	0.3449	0.2472	0.3575

dimension with the NeRF result. The quantitative rendering result of the rendered holography images of Figure 4.4 is stated in Table 4.2. Each of the 6 images are compared and calculated against the ground truth image, which is taken in the central viewpoint. The quantitative analysis show that NeRF outperforms NeRF--. For all dataset, MSE and PSNR values are significantly less than those of NeRF--'s. In case of PSNR and SSIM, NeRF's values are greater than those of NeRF--'s. By considering that MSE and LPIPS are better when the values are lower, while PSNR and SSIM are better with lower values, NeRF results comply to all of these principles. The same trend presence here as well, the more the dataset, the better the rendered result. Also, the aid of COLMAP camera positions allowed NeRF to render the scene closer to the ground truth.

Since the quantitative comparison done in Table 4.2 includes the uninteresting region (i.e. dark boundaries), the cropping of each rendered images are required for more accurate comparison. To get the interested area (i.e. holography panel), the four corner boundaries of ground truth and the rest of rendered images are required. Due to the usage of fiducial markers provided by Agisoft instead of ArUco marker, the detection of markers and cropping process cannot be fully automated but requires the manual verification step. Thus, the center circle in fiducial markers are detected, instead of the whole fiducial marker itself, and add an offset to obtain the four corners. In case of detecting markers

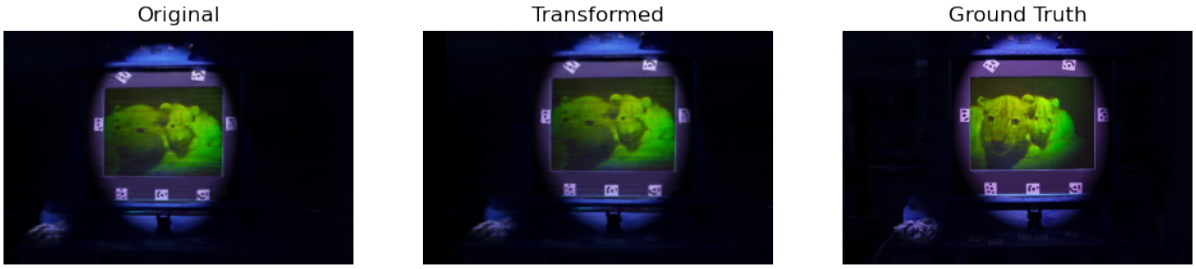

Figure 4.5. Homorgraphy Transformation

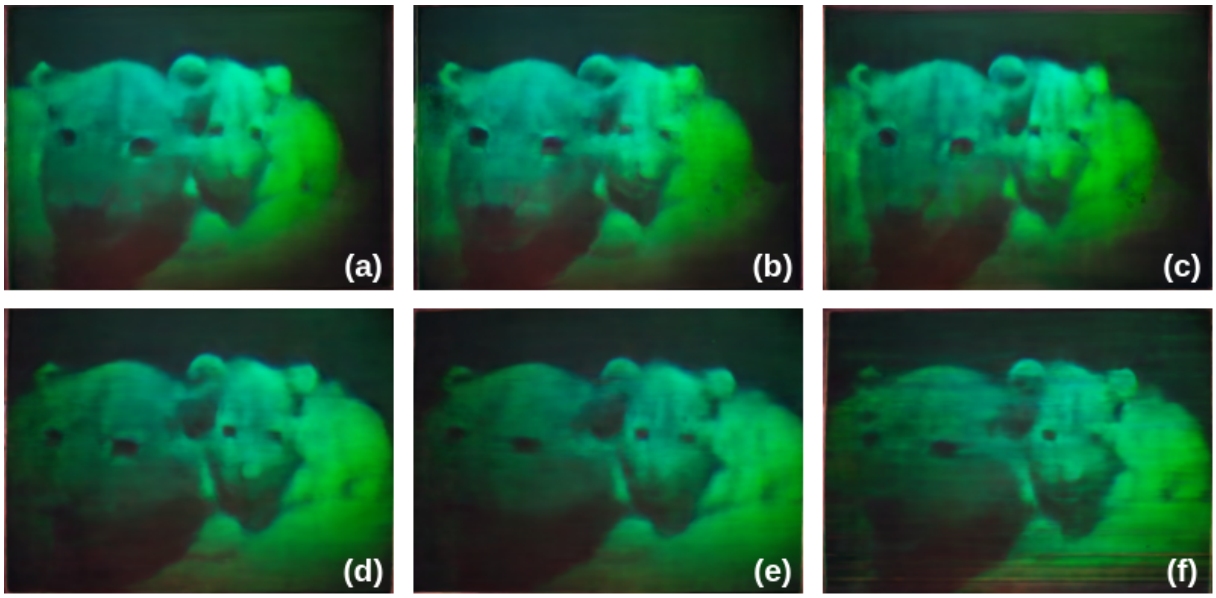

Figure 4.6. Cropped NeRF & NeRF-- Rendering With Different Dataset

Cropped NeRF rendering at the center location with (a) all, (b) column, and (c) row dataset. Cropped NeRF-- rendering at the center location with (d) all, (e) column, and (f) row dataset.

from rendered images, the manual validation of holography panel boundaries is required. With these corners, the homography has applied to align the rendered image based on the ground truth view. The example of the homography is demonstrated in Figure 4.5. In this example, the original holography panel in rendered view moved slightly up to match with the holography panel in ground truth image.

Table 4.3. Quantitative Analysis On Interest-Area at Center View Point

Metrics	**All Dataset**		**Column Dataset**		**Row Dataset**	
	NeRF	NeRF--	NeRF	NeRF--	NeRF	NeRF--
MSE ↓	54.1463	86.9383	54.0687	90.5415	63.4583	86.6946
PSNR ↑	30.7951	28.7386	30.8013	28.5623	30.1059	28.7509
SSIM ↑	0.6665	0.4795	0.6658	0.4608	0.6578	0.4627
LPIPS ↓	0.2323	0.4303	0.2426	0.4670	0.2346	0.4421

The homography-applied cropped rendered results can be found in Figure 4.6. These cropped scenes are further analyzed with quantitative metrics as done for original rendered scenes. The computation approach and methodology are the same, but does not involve the resizing preprocessing. The resizing factors are already addressed in homography process. The quantitative rendering results for the interested scenes are evaluated in Table 4.3. Compared to the analysis done in original scenes in 4.2, the SSIM values increase and LPIPS value decrease in NeRF for all dataset. This indicates that NeRF interplolates the scene well, in terms of human perceptual way –both SSIM and LPIPS metrics evaluate images in human-perceptual way. The other metrics which merely compares the pixel values, reported degraded values compared to the one from fully reconstructed scenes. NeRF-- reported worse scores in every dataset and in all metrics. The poor performance of NeRF-- can be correlated to the aforementioned limitations; it is limited to render the frontal scene only, since the model has to estimate and optimize the camera parameters from scratch. The lack of images due to the sub-sampled datasets might inaccurately optimize camera parameters with the discontinuous camera trajectory. This phenomena may yield another limitation, which is less-precisely rendered novel views.

Compared to other NeRF rendering experiments conducted by Mildenhall et al. (2020), the differences between measurements of holography reconstruction is minute [**23**]. Table

Table 4.4. NeRF Quantitative Measurement Results [**23**]

Metrics	**Diffuse Synthetic 360°**	**Realistic Synthetic 360°**	**Real Forward-Facing**
PSNR ↑	40.15	31.01	26.50
SSIM ↑	0.991	0.947	0.811
LPIPS ↓	0.023	0.081	0.250

4.4 conveys NeRF results driven from other types of datasets. Note that these values are adapted from Mildenhall et al. Although the captured holography images are lack of density in its surface, the evaluation result turned out well.

4.4. Limitations

There is one significant limitation I investigated in holography rendering with neural radiance field representation. Even considering the resizing factor of 8 and 4, the rendered result of both NeRF and NeRF-- failed to produce a high-dimension image. The assumption here is the existence of low-pass filter within the NeRF neural network model. To overcome this low-quality scene output, the experiment was done with adjusting the resizing factor into 1 and 2. However, due to the limitation of system memory (RAM), the operating system "kills" the process due to over-claiming system resources. This is happening because multiple high-quality images are getting loaded and allocated to the system memory at the same time.

Since the rendering experiment itself failed in case of train and polaroid dataset, quantitative measures cannot be applied to those. Train dataset firstly fed into COLMAP pipeline to acquire the camera position. Since rendering the interest area is critical, the holography panel area is cropped. This cropped train images are used to train the NeRF model, along with the camera positions obtained via raw and uncalibrated train images. The rendering failed, and it is illustrated in Figure 4.7, along with the COLMAP

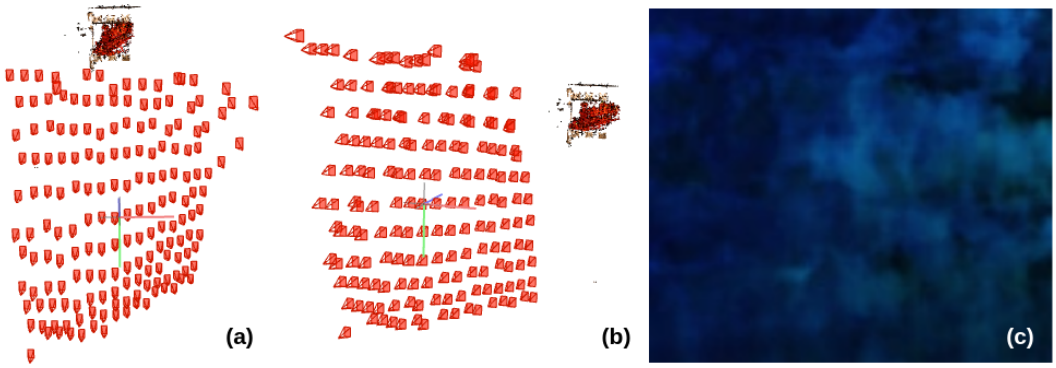

Figure 4.7. COLMAP from Raw Train Dataset & NeRF With Cropped Train Dataset

(a) COLMAP Front; (b) COLMAP 45°; (c) Train Cropped NeRF Rendering

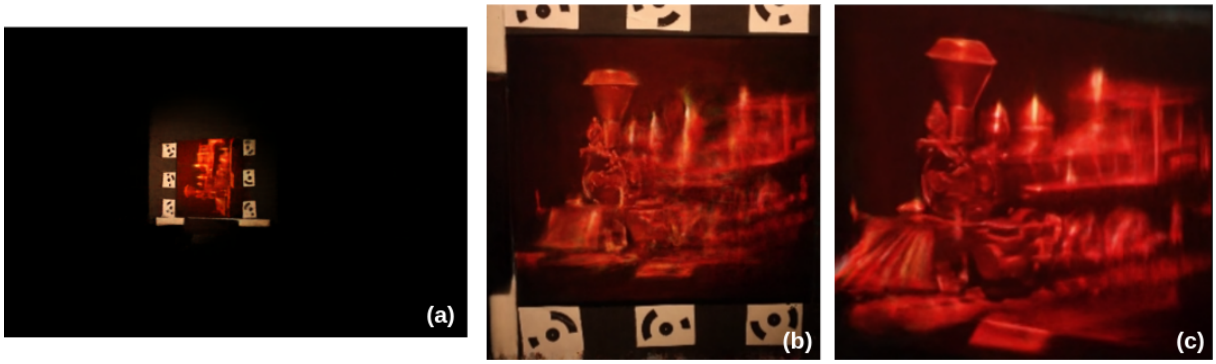

Figure 4.8. NeRF-- With Raw, Calibrated, and Cropped Train Dataset

(a) Raw Dataset; (b) Calibrated Dataset; (c) Calibrated & Cropped Dataset

output from the raw images. Although the camera lens and the holography panels are located in a parallel plane, the COLMAP result shows that the camera views are not aligned. This might also be a reason for the failure. Since NeRF failed, NeRF-- pipeline was executed in multiple train dataset. The three dataset that are used here is: raw dataset, calibrated dataset which contains holography panel with fiducial markers, and calibrated dataset only with the innner-holography panel region. The view synthesis is portrayed in Figure 4.8. Among these view synthesis, the only successful rendering was

produced by the raw train dataset. Although this view synthesis with the raw data was able to depict the appropriate depth reconstruction, the interest area is only a small portion out of the whole image. Once the inner-holography area gets cropped for the quantitative analysis, the quality of image became deteriorated, and result in a blurry image. This raw synthesis is thus not appropriate for the quantitative evaluations. The view synthesis with calibrated and cropped dataset are problematic. As seen in the rendering, the scene gets warped and the original shape of train gets ruined. The warped region is the back side of train and the bottom-front part, underneath train, for (b) and (c) in Figure 4.8, respectively. These failed rendering results are best viewed as videos, so it is recommended to view the supplementary full-rendered video here: `https://youtube.com/playlist?list=PLCVV8jHcNib3maFLJ1xNK-xh1M_MonHtX`.

Although polaroid dataset produced a decent spiral-COLMAP output, both NeRF and NeRF-- failed to render the novel views. From the original dataset, the central region is cropped uniformly all over the images. The intention behind was maximizing the interest region. Afterwards, the experiment firstly ran with the dataset will all images. However, it failed to produce scenes, so some outer camera trajectories are removed, based on the COLMAP result. The reasoning behind was to avoid any potential aperture occlusion. The example of aperture occlusion is depicted in Figure 4.9. With the sub-sampled dataset, both NeRF and NeRF-- pipeline were executed. The rendered views are illustrated in Figure 4.10. For both NeRF and NeRF--, the rendering failed and produced a black output. The best inference can be made here is the input images are too dark; less-varying pixel values may yield failure in training the NeRF model, since it is difficult for the neural network to capture the features.

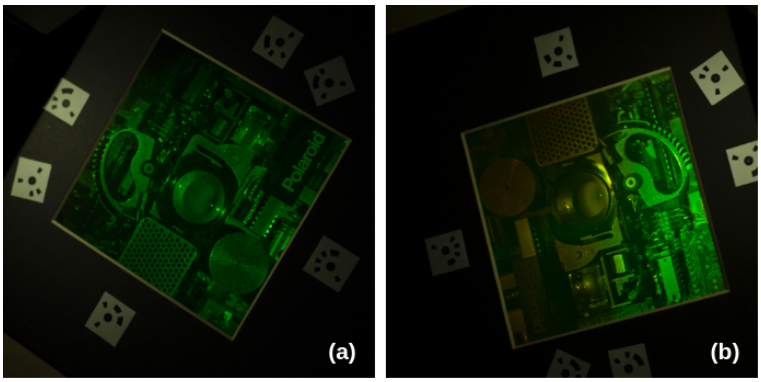

Figure 4.9. Aperture Occlusion In Polaroid Dataset
(a) Normal Image; (b) Aperture Occlusion

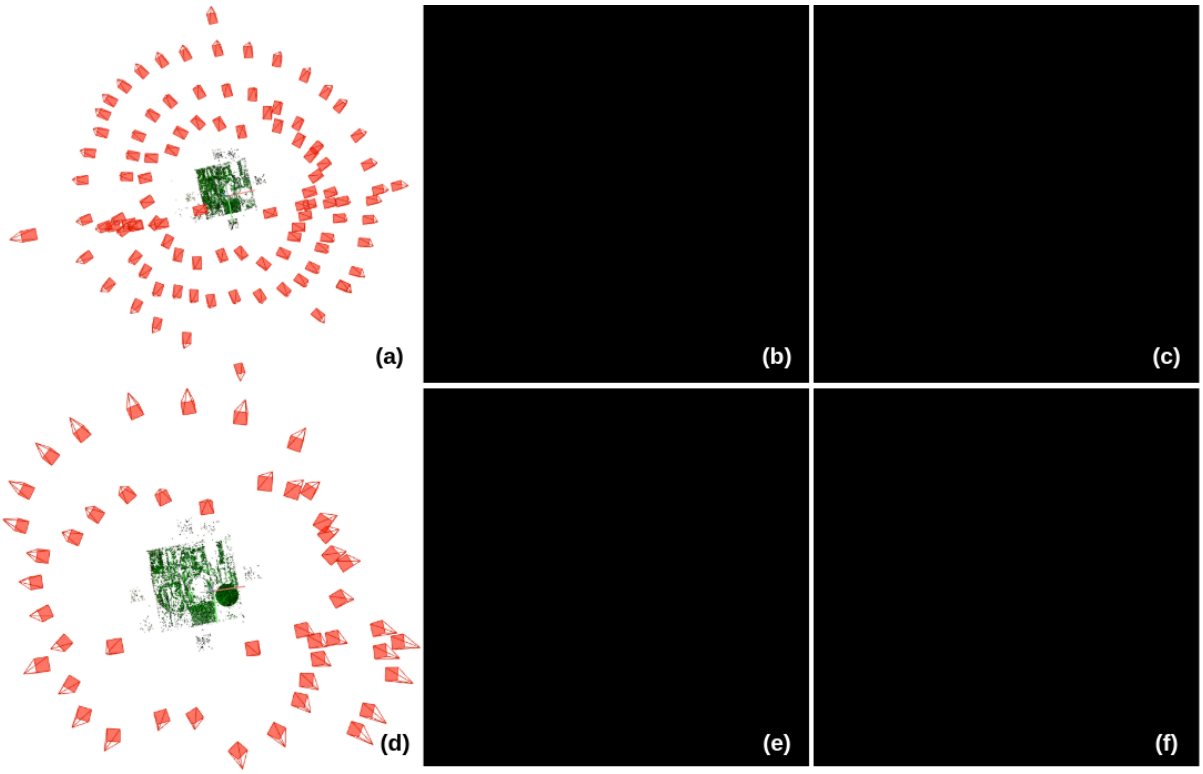

Figure 4.10. NeRF & NeRF-- Rendering With Polaroid Datasets

All polaroid dataset produces (a) COLMAP, (b) NeRF, and (c) NeRF--.
Inner polaroid dataset produces (d) COLMAP, (e) NeRF, and (f) NeRF--.

4.5. Future Work

As discussed, train dataset has failed to reconstruct the COLMAP precisely. The parallel relationship between the camera lens and view points are not reflected. Also, the cropped images' synthesized views are warped. Further research is needed to decompose the reasons of the failure behind. In case of polaroid dataset, it failed to render views although acquiring the camera parameters via COLMAP was done successfully. Thus, another further investigations are needed.

In order to analyze the interesting region in a better way and with many more rendered results, having a universally-used marker system is required. This will allow automatic marker detection with a Python library, yet only require a simple offset to compensate for the point where the markers are detected. After this, we can easily get the boundary of holography panel to crop to this area-of-interest.

The NeRF network needs to be further investigated to avoid producing low-quality reconstructed scenes. The two recommendations can be made here. First, run the network with a higher configuration RAM settings, so that the resizing factor can be minimized as much as possible. Another recommendation is re-structuring the NeRF neural network to get rid of the low-pass filter.

Yet, it is still challenging to render only the inner area of holography panel. Further research is needed to enhance the quality of rendering as well as minimize the free space. Minimizing this boundary area may also lead to a reduction in computation time and enhance the efficiency in training step.

CHAPTER 5

Conclusion

This paper analyzes the performance of two rendering approaches, with and without COLMAP: NeRF and NeRF--. The multi-mode quantitative analysis performed supports the claim that the neural radiance field representation of view reconstruction is able to accurately interpolate the photographed hologram images. The difference between the captured hologram images and real-capture images is negligible, by considering that the surface of the captured hologram images are not as continuous as real-capture images.

The contribution of this thesis is twofold. First, a holography preservation method of sparse image capture can be substantially improved by applying neural radiance fields for novel view synthesis. Second, I confirmed the quality and validity of this method by multi-dimensional assessment of the performance in mathematical approach with four different metrics. I believe that this work makes progress towards a trustworthy, long-term hologram preservation pipeline.

References

[1] Ballester, M., Schiffers, F., Wang, Z., Hasani, H., Fiske, L., Shedligeri, P., Tumblin, J., Willomitzer, F., Katsaggelos, A. K., and Cossairt, O. Fast simulations in computer-generated holograms for binary data storage. In *Computational Optical Sensing and Imaging* (2021), Optical Society of America, pp. CTh4A–7.

[2] Beder, C., and Steffen, R. Determining an initial image pair for fixing the scale of a 3d reconstruction from an image sequence. In *Joint Pattern Recognition Symposium* (2006), Springer, pp. 657–666.

[3] Benton, S. A. Survey of holographic stereograms. In *Processing and Display of Three-Dimensional Data* (1983), vol. 367, International Society for Optics and Photonics, pp. 15–19.

[4] Benton, S. A. Holography. *SPIE's International Technical Group Newsletter* (2004).

[5] Britannica. Holography, Apr 2019.

[6] Chen, J.-S., Chu, D., and Smithwick, Q. Y. Rapid hologram generation utilizing layer-based approach and graphic rendering for realistic three-dimensional image reconstruction by angular tiling. *Journal of Electronic Imaging 23*, 2 (2014), 023016.

[7] Crandall, D., Owens, A., Snavely, N., and Huttenlocher, D. Discrete-continuous optimization for large-scale structure from motion. In *CVPR 2011* (2011), IEEE, pp. 3001–3008.

[8] Elmorshidy, A. Holographic projection technology: the world is changing. *arXiv preprint arXiv:1006.0846* (2010).

[9] Fischler, M. A., and Bolles, R. C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM 24*, 6 (1981), 381–395.

[10] Gherardi, R., Farenzena, M., and Fusiello, A. Improving the efficiency of hierarchical structure-and-motion. In *2010 IEEE computer society conference on computer vision and pattern recognition* (2010), IEEE, pp. 1594–1600.

[11] Hariharan, P. *Basics of holography*. Cambridge university press, 2002.

[12] Hartley, R. I., and Sturm, P. Triangulation. *Computer vision and image understanding 68*, 2 (1997), 146–157.

[13] Holzbach, M. Evaluation of holographic technology in tactical mission planning and execution. Tech. rep., ZEBRA IMAGING INC AUSTIN TX, 2008.

[14] Hore, A., and Ziou, D. Image quality metrics: Psnr vs. ssim. In *2010 20th international conference on pattern recognition* (2010), IEEE, pp. 2366–2369.

[15] Huebschman, M. L., Munjuluri, B., and Garner, H. R. Dynamic holographic 3-d image projection. *Optics express 11*, 5 (2003), 437–445.

[16] James, R. 3d holographic projection: The future of advertising. *Retrieved November 10* (2009), 2009.

[17] Johnston, S. F. A cultural history of the hologram. *Leonardo 41*, 3 (2008), 223–229.

[18] Kang, L., Wu, L., and Yang, Y.-H. Robust multi-view l2 triangulation via optimal inlier selection and 3d structure refinement. *Pattern Recognition 47*, 9 (2014), 2974–2992.

[19] Lepetit, V., Moreno-Noguer, F., and Fua, P. Epnp: An accurate o (n) solution to the pnp problem. *International journal of computer vision 81*, 2 (2009), 155.

[20] Lucente, M., and Galyean, T. A. Rendering interactive holographic images. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques* (1995), pp. 387–394.

[21] Matsuda, N., Wheelwright, B., Hegland, J., and Lanman, D. Vr social copresence with light field displays. *ACM Transactions on Graphics (TOG) 40*, 6 (2021), 1–13.

[22] Matsushima, K., Nishi, H., and Nakahara, S. Simple wave-field rendering for photorealistic reconstruction in polygon-based high-definition computer holography. *Journal of Electronic Imaging 21*, 2 (2012), 023002.

[23] Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., and Ng, R. Nerf: Representing scenes as neural radiance fields for view synthesis. In *European conference on computer vision* (2020), Springer, pp. 405–421.

[24] Nishi, H., Matsushima, K., and Nakahara, S. Rendering of specular surfaces in polygon-based computer-generated holograms. *Applied optics 50*, 34 (2011), H245–H252.

[25] Peixeiro, J., Brites, C., Ascenso, J., and Pereira, F. Digital holography: Benchmarking coding standards and representation formats. In *2016 IEEE International Conference on Multimedia and Expo (ICME)* (2016), IEEE, pp. 1–6.

[26] Rahaman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F., Bengio, Y., and Courville, A. On the spectral bias of neural networks. In *International Conference on Machine Learning* (2019), PMLR, pp. 5301–5310.

[27] Sara, U., Akter, M., and Uddin, M. S. Image quality assessment through fsim, ssim, mse and psnr—a comparative study. *Journal of Computer and Communications 7*, 3 (2019), 8–18.

[28] Schonberger, J. L., and Frahm, J.-M. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 4104–4113.

[29] Slama, C. C. Manual of photogrammetry. Tech. rep., America Society of Photogrammetry,, 1980.

[30] Snavely, K. N. *Scene reconstruction and visualization from internet photo collections*. University of Washington, 2008.

[31] Snavely, N., Seitz, S. M., and Szeliski, R. Photo tourism: exploring photo collections in 3d. *ACM Transactions on Graphics (TOG) 25*, 3 (2006), 835–846.

[32] Tan, H. L., Li, Z., Tan, Y. H., Rahardja, S., and Yeo, C. A perceptually relevant mse-based image quality metric. *IEEE Transactions on Image Processing 22*, 11 (2013), 4447–4459.

[33] Tres Media Group Company. Holographic medical moa's, 2008.

[34] Triggs, B., McLauchlan, P. F., Hartley, R. I., and Fitzgibbon, A. W. Bundle adjustment—a modern synthesis. In *International workshop on vision algorithms* (1999), Springer, pp. 298–372.

[35] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. In *Advances in neural information processing systems* (2017), pp. 5998–6008.

[36] Walton, M., Hao, P., Vermeulen, M., Willomitzer, F., and Cossairt, O. Characterizing the immaterial. noninvasive imaging and analysis of stephen benton's hologram engine no. 9. *arXiv preprint arXiv:2110.06080* (2021).

[37] Wang, Z., Wu, S., Xie, W., Chen, M., and Prisacariu, V. A. Nerf–: Neural radiance fields without known camera parameters. *arXiv preprint arXiv:2102.07064* (2021).

[38] Willomitzer, F., Rangarajan, P. V., Li, F., Balaji, M. M., Christensen, M. P., and Cossairt, O. Fast non-line-of-sight imaging with high-resolution and wide field of view using synthetic wavelength holography. *Nature Communications 12*, 1 (2021), 1–11.

[39] Ye, Y., Shan, J., Hao, S., Bruzzone, L., and Qin, Y. A local phase based invariant feature for remote sensing image matching. *ISPRS Journal of Photogrammetry and Remote Sensing 142* (2018), 205–221.

[40] Zhang, H., Cao, L., and Jin, G. Computer-generated hologram with occlusion effect using layer-based processing. *Applied optics 56*, 13 (2017), F138–F143.

[41] Zhang, R., Isola, P., Efros, A. A., Shechtman, E., and Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018), pp. 586–595.

[42] Zheng, Y., Kuang, Y., Sugimoto, S., Astrom, K., and Okutomi, M. Revisiting the pnp problem: A fast, general and optimal solution. In *Proceedings of the IEEE International Conference on Computer Vision* (2013), pp. 2344–2351.

APPENDIX A

Holography Renderings With Different Methods

In this section, the result of holography renderings from different methods will be covered, to visually compare the rendered against the rendered view with neural radiance field. Note that the rendered views are adapted from the original corresponding authors.

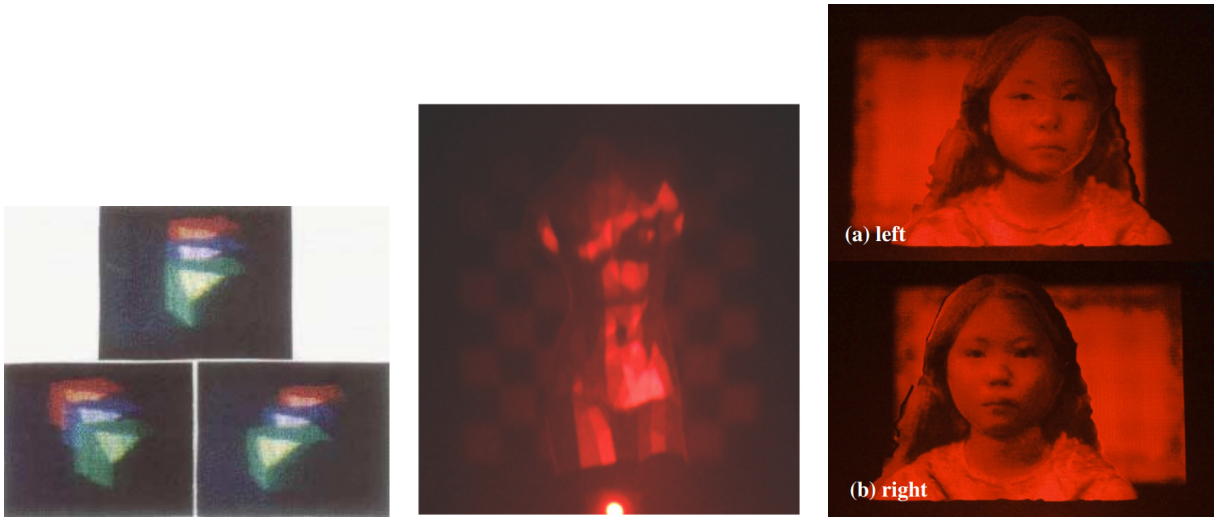

Figure A.1. Holography Rendering Result by Lucente et al. [**20**], Nishi et al. [**24**], and Matsushima et al. [**22**]

(1) Image with three cut cubes located at different depths, photographed from different view locations. Top: center. Left: left. Right: right.
(2) Photographs of optical reconstructions of The Metal Venus I.
(3) Optical reconstruction of the polygon-based high-definition computer hologram which rendered by texture mapping and Gouraud shading.

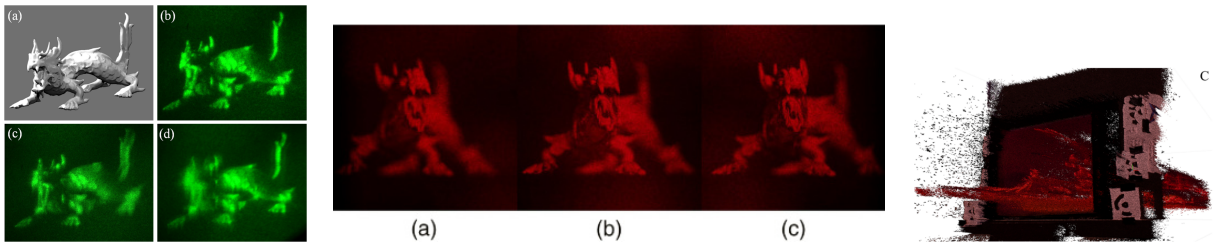

Figure A.2. Holography Rendering Result by Chen et al. [**6**], Zhang et al. [**40**], and Walton et al. [**36**]

(1) Reconstructed holography images. (a) target image as rendered by computer graphics; (b) holographic reconstruction with a single plane; (c) and (d) holographic reconstruction with depth information, around the head and the tail, respectively. (b)-(d) are projected on a diffusive screen.
(2) Holography optical reconstruction results with a following viewing directions: (a) left, (b) center, and (c) right.
(3) The holography of MIT Engine no. 9 copy rendered with Agisoft Metashape Professional Edition.

APPENDIX B

Holography Image Datatsets

In this section, the list of captured data that are used in the experiment is discussed. There are three big categories: lion cubs, train, and polaroid. This name indicates the different holography images. Each of these datasets are further sub-sampled to discover meaningful rendering results.

Table B.1. List of Holography Image Dataset

Name	**Dimension**	**Sample**	**Description**
Lion Cubs All	5760×3840	60	All captured raw data
Lion Cubs Column	5760×3840	48	All captured data without central columns
Lion Cubs Row	5760×3840	40	All captured data without central rows
Train Raw	3840×5760	176	All captured data
Train Calibrated	1064×1136	176	Raw data with calibration & registration
Train Cropped	943×861	176	Cropped data after calibration
Polaroid All	2500×2500	100	All captured raw data
Polaroid Inner	2500×2500	47	All captured data without occluded ones