RiFT - Radiance Field Tomography

Kevin Chew Figueroa, Zhipeng Dong, Greg Nero, Gordon Hageman, David J. Brady;
Camera Lab, Wyant College of Optical Sciences, The University of Arizona; Tucson, Arizona, USA

Abstract

Regression-based radiance field reconstruction strategies, such as neural radiance fields (NeRFs) and, physics-based, 3D Gaussian splatting (3DGS), have gained popularity in novel view synthesis and scene representation. These methods parameterize a high-dimensional function that represents a radiance field, from a low-dimensional camera input. However, these problems are ill-posed and struggle to represent high (spatial) frequency data; manifesting as reconstruction artifacts when estimating high frequency details such as small hairs, fibers, or reflective surfaces.

Here we show that classical spherical sampling around a target, often referred to as sampling a bounded scene, inhomogeneously samples the target's Fourier domain, resulting in spectral bias in the collected samples. We generalize the ill-posed problems of view-synthesis and scene representation as expressions of projection tomograpy and explore the upper-bound reconstruction limits of regression-based and integration-based strategies. We introduce a physics-based sampling strategy that we directly apply to 3DGS, and demonstrate high fidelity 3D anisotropic radiance field reconstructions with reconstruction PSNR scores as high as 44.04 dB and SSIM scores of 0.99, following the same metric analysis as defined in Mip-NeRF360.

Introduction

Radiance field reconstruction has advanced significantly through regression-based strategies, such as NeRFs [1, 2, 3, 4] and 3DGS [5]. While the interopolative nature of regression gives way to novel view synthesis, efficient processing, and estimation, radiance reconstruction pipelines have been shown to struggle on high frequency information [7]; at the time of this writing, Zip-NeRF and 3DGS are reported as current state-of-the-art strategies with their highest reconstruction PSNR scores being 28.54 and 27.21 respectively, on the Mip-NeRF360 benchmark dataset. In "Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains" Tancik *et al.*, utilizing Neural Tangent Kernel (NTK) theory [8], conducted rigorous analysis on the numerical behavior of multilayer perceptrons (MLPs) to understand and improve reconstruction of high frequency information. Utilizing spherical sampling around a target, it was determined regression on MLP architectures converges to the same mean-estimator as ridgeless kernel regression with the NTK; introducing a spectral bias on reconstructions and preferentially converging on eigenvectors with larger eigenvalues. From this analysis and the observed difficulty with reconstructing high frequency details, Tancik *et al.* qualitatively deduced that larger eigenvalues of low frequency information were present compared to the rest of the sampled domain, and introduced a tuneable Fourier features-based positional encoder that preprocesses the Fourier domain of the input measurements by tuneably amplifying the power spectrum of the high frequency components.
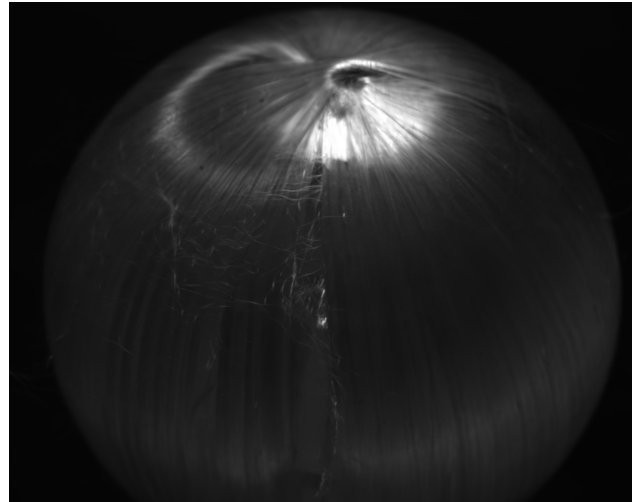

Figure 1. *RiFT generalizes the ill-posed problems of radiance field view-synthesis and scene representation as expressions of Fourier computed tomography; introducing a physics-based sampling strategy that is directly applicable to radiance field reconstruction methods and demonstrating high fidelity 3D anisotropic radiance field reconstructions with PSNR scores as high as 44.04 dB/ SSIM 0.99; Here, RiFT is implemented on 3D Gaussian Splatting to accurately reconstructing high frequency details including small hairs, fibers, and reflective surfaces on an ornament.*

With our work dubbed Radiant Field Tomography (RiFT), we apply Fourier analysis on computed tomography to quantitatively show how spherical sampling around a target results in a calculable spectral bias in the collected image measurements; specifically causing higher eigenvalues of the low frequency target information to be present in the measurement data compared to the rest of the domain. Our contribution here is to generalize the ill-posed problem of novel view synthesis and scene representation from the perspective of projection tomography. This leads to a more sophisticated forward model and physics-based sampling strategy that can be applied to already existing tomographic, neural, and 3DGS-based methods for orders of magnitude greater performance in achieving high-fidelity 3D anisotropic radiance field reconstructions.

With this thinking, we assess the Mip-NeRF360 benchmark dataset and analyze the finite support of targets' frequency spectrum sampled from the available images. Next, we introduce a rigorous and generalized forward model for describing the full propagation for an anisotropic radiant field; from target of interest to a physically accurate camera model. Finally, utilizing our recommended sampling strategy, we explore the upper-bound resolution limits of regression-based and integration-based back-projection strategies; utilizing 3DGS as a physics-based approach

for radiance field reconstruction and filtered backprojection for non-biased sampling and estimation of object density. Under our sampling strategy we achieve a reconstruction score of 44.04 dB PSNR / 0.99 SSIM with 3DGS and accurately reconstruct small hairs, fibers, and reflective surfaces; see Fig. 1.

Theory

While "computed tomography" most often connotes x-ray imaging, Marks *et al.* [6] demonstrated that x-ray computed tomography algorithms can be applied to imaging in the visible spectrum with a camera, and does not necessarily require active illumination. While radiance estimation contrasts with object density estimation, generalizing this sampling theory to NeRF's conventional spherical/bounded sampling around a target, we can directly analyze the contribution each individual image serves in reconstructing the target 3D radiometric field in the Fourier domain.
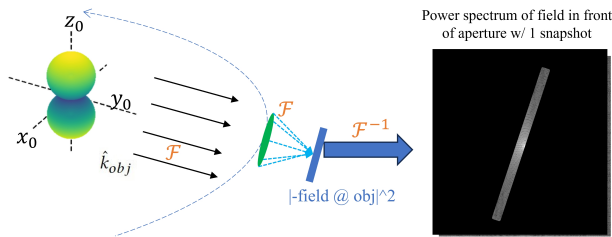

Figure 2. *The Fourier slice theorem generalized for camera-based computer tomography*

In Fig. 2, a single dipole serves as the target of interest. For more general targets, one can model every point of the target as a collection of independently radiating point sources. Considering a π solid angle radiating out from a point for each source point, we can model the propagation of their individual fields rays arriving at the camera's aperture. By performing an inverse Fourier transform on the image captured by the sensor, we can extract a slice of the power spectrum (auto correlation) of the radiance field that was originally in front of the lens's aperture. This process provides insight into the properties of the radiance field as manipulated by the lens.
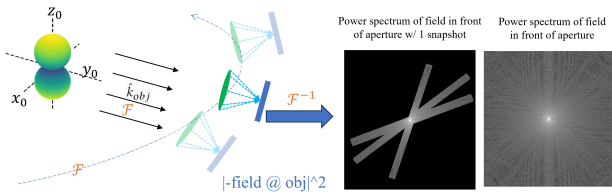

Figure 3. *Fourier Slice theorem generalized for multi-sampling computer tomography*

As we orbit the camera around the target, each image captured corresponds to a different slices of the target's power spectrum shown in Fig. 3. This process of sampling is described by the projection slice theorem, which states that an object f that has N dimensions, the Fourier transform of the projection in m dimension is equivalent to the slice in m dimension of the Fourier transform [10]. Intuitively as more slices encompass the target's

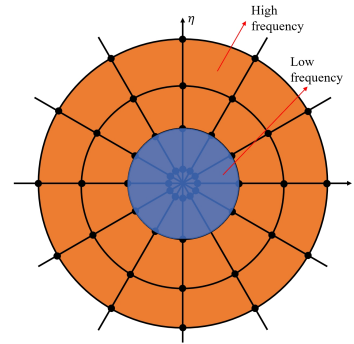Fourier spectrum, the more effectively we can reconstruct the target's radiance field.

Figure 4. *Conventional NeRF bounded scene sampling around a target leads to non-homogeneous sampling of the target Fourier field, specifically resulting in higher eigenvalues of the low frequency target information to be present in the measurement data compared to the rest of the domain.*

However, as shown in Fig. 4, NeRFs conventional bounded scene sampling results in a spectral bias in the measurement data, with a higher presence of eigenvalues in the low frequency target information compared to the high frequency information. As demonstrated in Fig. 5, x-ray computed tomography filtered backprojection, proper filtering on the collected measurement data is necessary. Combining Figs. 4 and 5, we show without filtering, spectrally biased measurement data will propagate its bias through the integration-based filtered backprojection, inducing spectrally biased reconstructions.
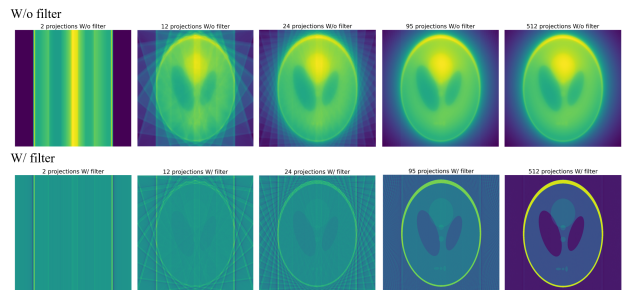

Figure 5. *Utilizing x-ray computed tomography, with integration-based backprojection, reconstructions of the target object density are achieved over 2, 12, 24, 95, and 512 projections from left to right. (Above) Reconstruction is implemented using integration-based backprojection without filtering. (Below) Reconstruction is implemented using integration-based backprojection with filtering.*

In Fig. 4, filtered backprojection measurement data is first preprocessed by utilizing a high-pass filter to suppress the low frequency eigenvalues; from the reconstruction it can be seen that by removing the inherent spectral bias over the entire measurement distribution, the spectral bias in the reconstructions are attenuated. Figure 6 demonstrates the numerical trend between reconstruction quality and number of projections. As we increase the number of projection angles, along with applying this filtering, the reconstruction's PSNR and SSIM increase; while the PSNR and SSIM for unfiltered does not show significant variation be-
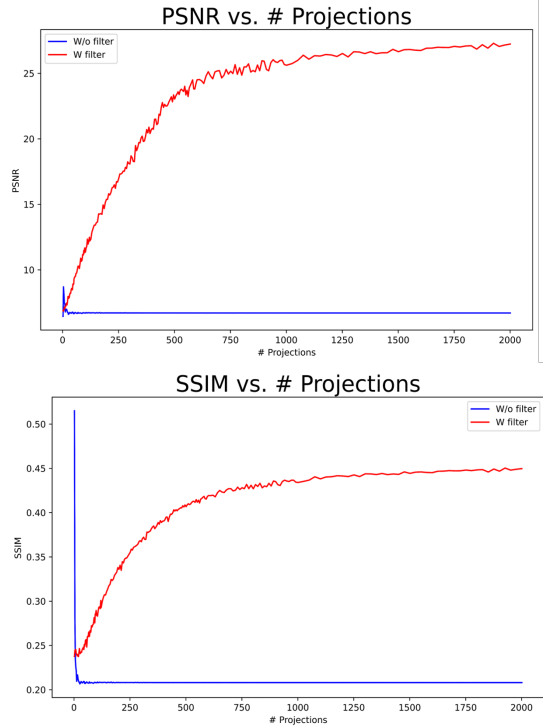
Figure 6. PSNR & SSIM vs. number of projections.

Forward Model

In this section we will introduce a rigorous and generalized forward model from the target anisotropic radiant field to a camera. A target scene can be represented as a collection of independently radiating point sources. The radiance field at any point in space can therefore be represented by the combination each point-radiator's contribution at that point. The target of interest generates a field which can then be propagated to the camera lens by following the procedure described in the next section.

Propagating from Target to the Camera Lens

$$\hat{k}_{obj} = \frac{1}{\sqrt{(x_a - x_0)^2 + (y_a - y_0)^2 + (z_a - z_0)^2}} \begin{bmatrix} x_a - x_0 \\ y_a - y_0 \\ z_a - z_0 \end{bmatrix} \quad (1)$$

where x_0, y_0, z_0 denote the objection location, and x_a, y_a, z_a denotes the location of the camera aperture. Let $f([x_0, y_0, z_0])$, represent the radiance field of the object we seek to recover. In object space, the frequency passing through the aperture of the lens $A(x_a, y_a, z_a)$ is

$$\xi = \frac{x_a + \frac{A}{2} - x_0}{\lambda (z_a - z_0)}; \ \eta = \frac{y_a + \frac{A}{2} - y_0}{\lambda (z_a - z_0)} \quad (2)$$

After transforming from the aperture to the image sensor, centered at r_i, we can express the image space frequencies as:

$$\xi' = \frac{||x_a - x_0| - x_i|}{\lambda f / \#mA}; \ \eta' = \frac{||y_a - y_0| - y_i|}{\lambda f / \#mA} \quad (3)$$

where m is the magnification of the imaging system. We can also simplify equation 3 to cancel A because $f/\#$ is a function of aperture size. Notice the image space frequency can be related back to the frequency passing through the aperture, via the magnification term m:

$$m\xi' = \xi; \ m\eta' = \eta \quad (4)$$

To obtain the image of the sensor, we integrate over wavelength and time to obtain the irradiance for each camera position. Figure 8 illustrates our entire model, which spans from the sun to the camera sensor, where k_{sun} stands for the propagation from the sun to the target. Since the sun is far from us, it can be considered as plane wave illumination with coherence length in the visible range from 40 to 60 micrometers [11]. As mentioned previously, the improved density of sampling enhances the resolution and accuracy of the reconstructed target radiance field.

In the following section we step through a rigorous camera reduction defining a direct physical mapping from a physically accurate camera model to a pinhole camera model, allowing for direct interpretability and interoperability between computational photographic, tomographic and pinhole model based reconstructions such as NeRFs and 3DGS.

Physics-based Camera Model

In our camera model, we begin with a basic configuration as depicted in Fig. 9 a), consisting of a simple lens group and a sensor. Using Gaussian reduction [12], we simplify this lens group

yond 20 projections. This is because the high contribution of the low frequency results in washing out the high frequency components of the reconstructions. While using the filtering strategy, on the other hand, the PSNR and SSIM grow significantly, until saturation, as the number of projection angles increase.

The benchmark data used to reconstruct the radiant fields in Mip-NeRF's datasets are taking slices in distinct directions of the power spectrum of the real world. As depicted in Fig. 7 a), each red line in the power spectrum can be considered as a picture from one direction. However, by slicing the power spectrum more densely or with greater extent, shown in Fig. 7 b), results in a greater finite support and improved reconstruction. To obtain more slicing information in the power spectrum, we can either try moving close towards the object (greater extent) or utilize a longer focal length imaging system. Denser slicing can be achieved by sampling more images in different perspectives. Ideally, denser and wider slices shown in Fig. 7 c) will result in the highest sampling of the Fourier domain, and thus the best reconstruction.
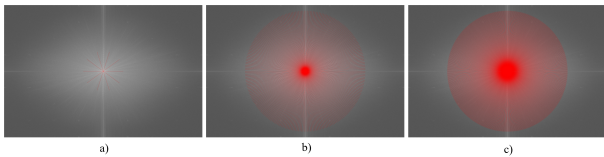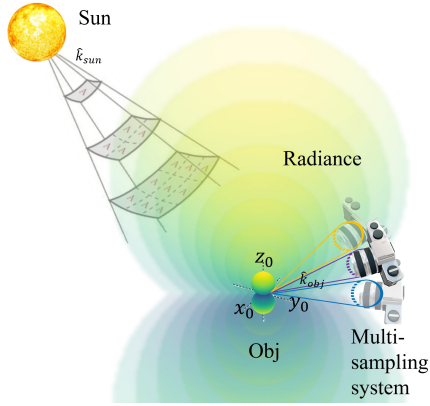

Figure 7. a) Power spectrum with 5 slices and far away from the object. b) Power spectrum with 250 slices and close to the object. c) Power spectrum with 570 slices and close to the object.

Figure 8. *Pipeline of model from the sun to high frequency-sampling strategy*

into two principal planes, as shown in Fig. 9 b). This process is then followed by a further simplification, reducing the system to a basic aperture stop in Fig. 9 c).

If we reduce the optical space inside the principle planes, we can obtain the pinhole camera model in Fig. 9 d). However, reducing to a pinhole camera model neglects the aperture size and depth of field, which is non ideal for real imaging conditions because the real camera system has a cutoff frequency and depth of focus band limiting the information from the target/object space to the image space. Furthermore, reducing the aperture size will decrease the highest frequency passing through the system but increase the depth of field, bringing more information along the optical axis direction to focus. The frequency in 3D space passed into the camera system is constrained by:

$$\frac{1}{\lambda^2} = \xi^2 + \eta^2 + \zeta^2 \tag{5}$$

where ξ, η, and ζ are frequencies in the Fourier transform of the real space. By decreasing the aperture size, ξ and η will decrease, but ζ will increase correspondingly; as a result, depth of field is increased. On the other hand, increasing the aperture size results in a narrower depth of field.

By swapping the aperture size, we create a natural filter for an imaging system because a smaller aperture size will pick up lower frequencies in the power spectrum, while larger apertures pass the higher frequencies as demonstrated in Fig. 4.

This step-by-step approach allows us to methodically simplify the camera model, making it easier to analyze and understand its fundamental optical properties.

Experiments

Fourier computed tomography sampling strategies assist with amplifying eigenvalue contributions from a target's high frequency information. To explore the upper-bound resolution limits of regression-based and integration-based backprojection reconstruction strategies we conducted one set of regression-based experiments, using 3DGS, for reconstructing a target's radiance field and one set of integration-based experiments, using unfiltered and filtered backprojection, for reconstructing a target's object density.
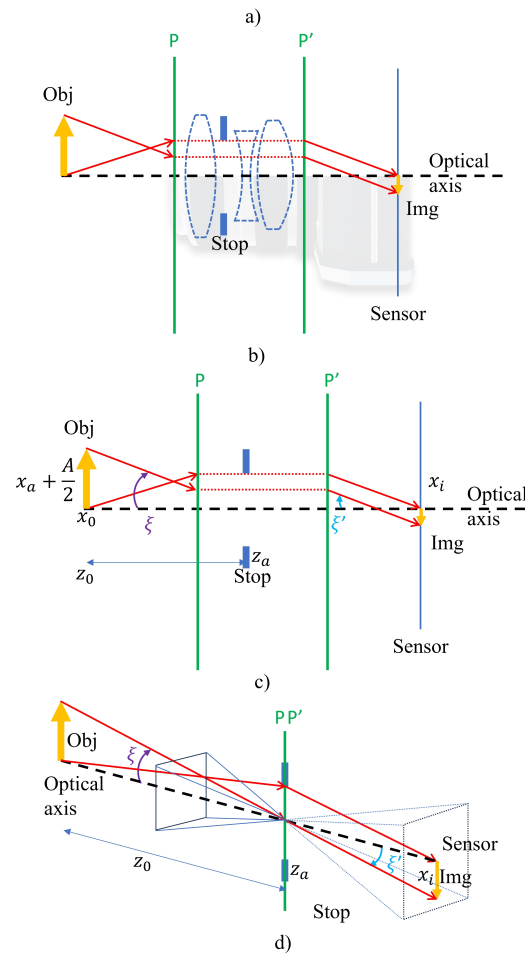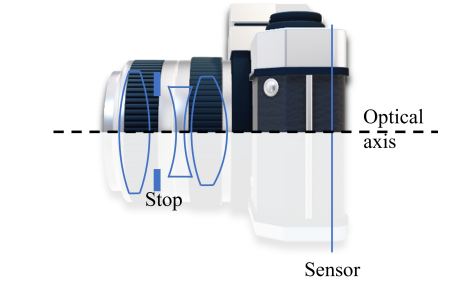

Figure 9. *a) An example of an camera with lens group, stop, and a sensor. b) Reduced lens group with an example object and image. c) Further reduce to a stop only. d) Reducing optical space between two principle plane we can obtain a pinhole camera visualized in 3D.*
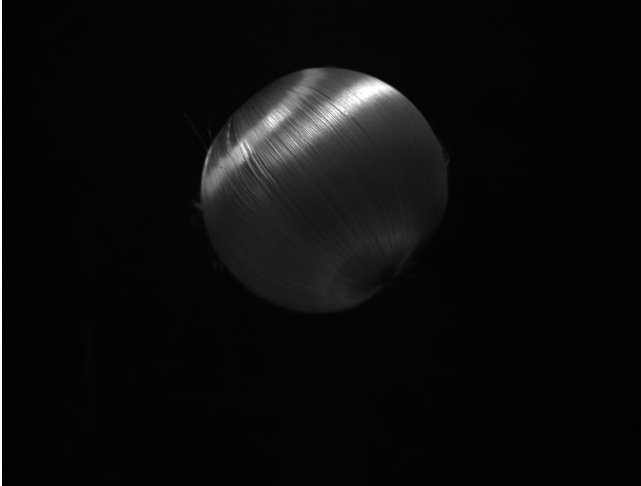
Figure 10. *An ornament is shown as an example of an imperfect lambertian object producing an aniostropic radiance field.*

For both experiments targets were sampled at 10, 64, 250 and 570 projections, respectively. For our regression-based experiments, we used an Allied Vision camera 1800U 1240m to sample a target aniostropic radiance field shown in Fig. 10. Note, as the ornament was the target subject for this experiment, sampled image projections were first preprocessed to remove an optical post holding the target object. Unlike the nominal 250 images provided in the Mip-NeRF360 benchmark dataset, we sampled 570 images at 4024×3036 pixel resolution over a full $360°$ rotation at a distance of approximately 27 centimeters from the target. Under our sampling condition, the collected samples provide a much higher frequency range than that of the Mip-NeRF360 benchmarks. For our integration-based experiments, we constructed a target object and simulated Magnetic Resonance Imaging (MRI) projections of the same sampling.

We evaluated the performance of 3DGS with reconstructing the radiance field by varying the number of input projected views. To evaluate its PSNR/SSIM reconstruction accuracy, reconstructed projections were compared with unseen views following the evaluation method described in Mip-NeRF360; utilizing a test/train split every 8th photo in the dataset was used for testing as an unseen view. For evaluating the performance of unfiltered and filtered backprojection we followed the same methodology of evaluating the performance over different number of input views, but instead calculated PSNR and SSIM reconstruction accuracy by directly comparing the whole reconstructed object density against the known ground truth.

Observations & Results

See Fig. 11 for a plot of our results. As discussed in Fig. 5, our integration-based backprojection results aligned with our initial hypothesis, which posits that more projections results in higher reconstruction accuracy up to a certain extent without proper filtering; as increased sampling without filtering leads to a bias with greater low frequency eigenvalues washing out the higher frequency information in reconstruction.

For our regression-based backprojection results, interesting trends were observed. Although 3DGS does not have a built-
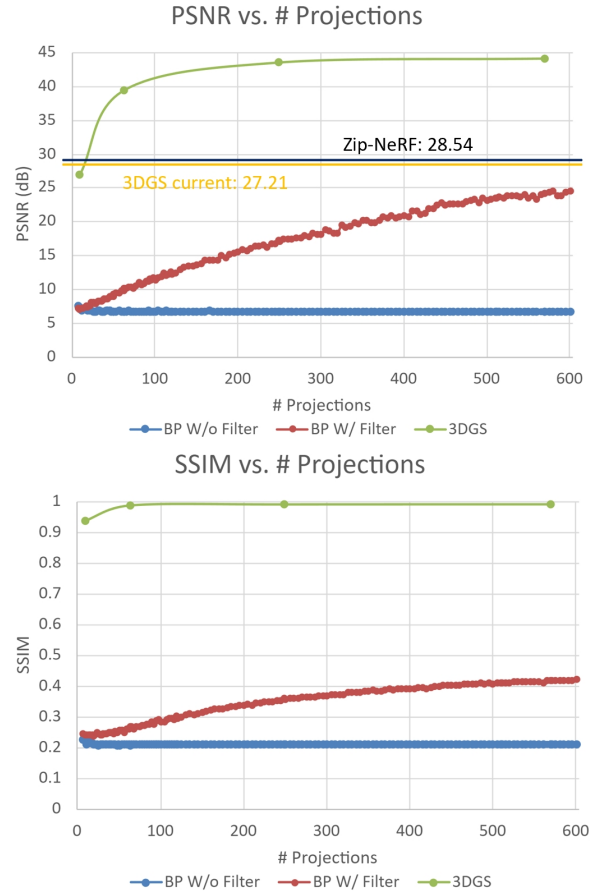

Figure 11. *PSNR (above) with PSNR of current 3DGS and Zip-NeRF & SSIM (below) vs number of input projections.*
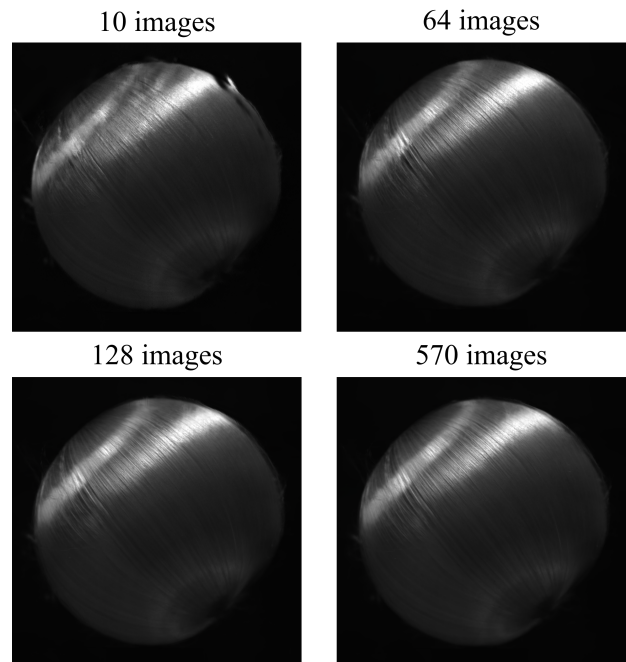

Figure 12. *Reconstructed ornament results with varying number of input projections. Fig. 10 shows the nearest ground truth.*

in preprocessing step to homogenize the spectral sampling, like that of [7] or filtered backprojection, its reconstruction accuracy at 570 images achieved a high fidelity reconstruction of 44.04dB PSNR / 0.99 SSIM and also followed the trend of integration-based filtered backprojection; growing in reconstruction accuracy until saturation. In Fig. 12, we demonstrate the reconstructed radiance field results with over the various number of input projections into 3DGS. With 250 and 570 images provided to 3DGS, small hairs, fibers, and reflective surfaces were reconstructed with extremely high fidelity.

Discussion & Future Work

The similar PSNR/SSIM regression and integration trends observed in Fig. 11 suggest the existence of an optimal compromise for various applications, balancing acceptable quality against the requisite number of sampling points and camera sampling. From the physics-based camera model presented, optical filtering techniques can be applied to directly homogenize measurement samples at capture time rather than as a post-processing step, thereby potentially further increasing the PSNR/SSIM reconstruction score than what is reported here.

For future work, we are interested in exploring multiple areas, including extending analysis on NeRFs. We also plan to utilize heterogeneous array cameras with varying focal lengths and the physics-based camera model presented, to serve as a direct analog methodology for managing aliasing artifacts from multi-scale sampling. Additionally, we are also exploring various alternative bases for the 3DGS algorithm, focusing particularly on the Laguerre-Gaussian and Hermite-Gaussian bases due to their unique properties and interconvertibility [14]. Laguerre-Gaussian functions, LG(p,l), characterized by their concentric ring intensity profiles, are crucial in laser optics and quantum mechanics, especially in describing the orbital angular momentum of photons. Hermite-Gaussian functions, HG(m,n), on the other hand, display a rectangular symmetry with a pattern resembling a chessboard, making them vital in applications requiring precise beam shape control or on-axis sampling, such as in remote sensing, optical tweezers, and quantum computing. Figure 13 and Fig. 14 illustrate examples of the Laguerre-Gaussian and Hermite-Gaussian bases, respectively. The Gaussian splatting algorithm treats each point source on the object as a Gaussian function. Since Laguerre and Hermite-Gaussian functions are also types of Gaussian functions, they may offer a more effective representation and differentiability compared to directly using the first three rows of spherical harmonics. See Fig. 15. This indicates their potential superiority in accurately modeling complex structures in 3D reconstruction. Lastly, we plan to explore a combination of NeRFs and 3DGS strategies, and implement Generative AI methods for direct transformation on the estimated anisotropic radiance field, to further expand the limit of resolution.

References

[1] J. T. Barron, B. Mildenhall, D. Verbin, P. P. Srinivasan, and P. Hedman, "Zip-NeRF: Anti-Aliased Grid-Based Neural Radiance Fields," *ICCV*, 2023.

[2] J. T. Barron, B. Mildenhall, D. Verbin, P. P. Srinivasan, and P. Hedman, "Mip-NeRF 360: Unbounded anti-aliased neural radiance fields," *arXiv preprint arXiv:2103.13415*, 2021.

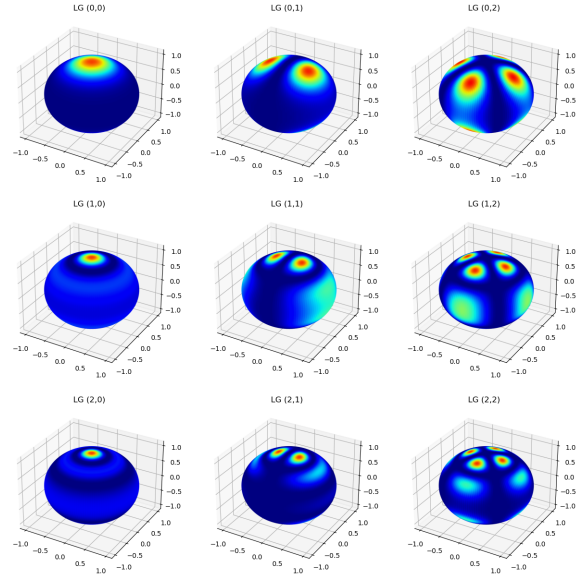[3] T. Müller, A. Evans, C. Schied, and A. Keller, "Instant neural graph-

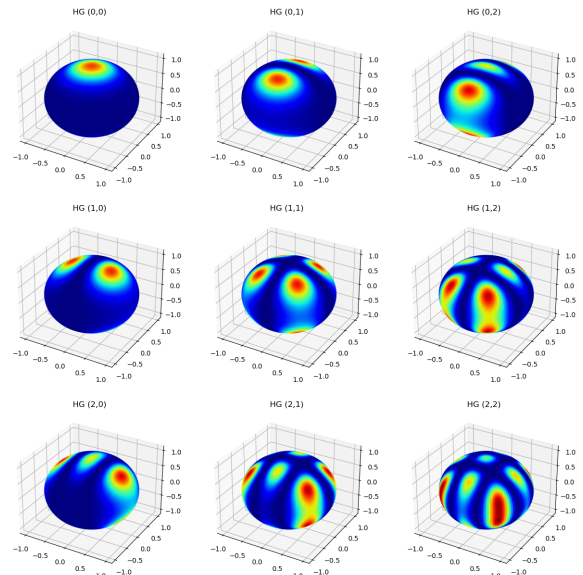**Figure 13.** Laguerre-Gaussian basis for mode and order 0 to 2.

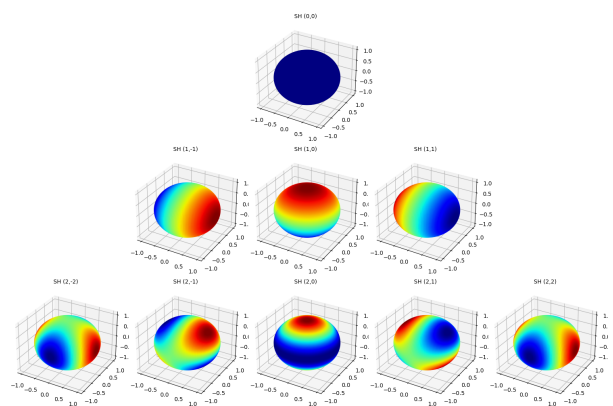**Figure 14.** Hermite-Gaussian basis for mode and order 0 to 2.

Figure 15. Spherical harmonic basis for mode from 0 to 2.

ics primitives with a multiresolution hash encoding," *ACM Trans. Graph.*, vol. 41, no. 4, Jul. 2022, doi: 10.1145/3528223.3530127.

[4] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "NeRF: Representing scenes as neural radiance fields for view synthesis," *arXiv preprint arXiv:2003.08934*, 2020.

[5] B. Kerbl, G. Kopanas, T. Leimkuehler, and G. Drettakis, "3D Gaussian Splatting for Real-Time Radiance Field Rendering," *ACM Trans. Graph.*, vol. 42, no. 4, Jul. 2023, doi: 10.1145/3592433.

[6] D. L. Marks, R. Stack, A. J. Johnson, D. J. Brady, and D. C. Munson, "Cone-beam tomography with a digital camera," *Applied Optics*, vol. 40, no. 11, pp. 1795–1805, United States, 2001, DOI: 10.1364/AO.40.001795.

[7] M. Tancik, P. P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. T. Barron, and R. Ng, "Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains," *arXiv preprint arXiv:2006.10739*, 2020.

[8] A. Jacot, F. Gabriel, and C. Hongler, "Neural Tangent Kernel: Convergence and Generalization in Neural Networks," *arXiv preprint arXiv:1806.07572*, 2020.

[9] J. W. Goodman, *Introduction to Fourier optics*, 4th ed., includes bibliographical references (pages 517-533) and index., 2017, ISBN: 9781319119164.

[10] R. Ng, "Fourier slice photography," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 735–744, Jul. 2005, DOI: 10.1145/1073204.1073256.

[11] T. D. Milster, "Evaluation of the Coherence," Slide 4-24-13, College of Optical Sciences, The University of Arizona, 2022. Online.

[12] J. E. Greivenkamp, *Field Guide to Geometrical Optics*, vol. v.FG01, SPIE, Bellingham, 1st ed., 2004, ISBN: 9780819452948.

[13] D. J. Brady, "Optical imaging and spectroscopy," Wiley-Blackwell, Hoboken, NJ, Mar. 2009.

[14] D. Shen, T. He, X. Yu, and D. Zhao, "Mode Conversion and Transfer of Orbital Angular Momentum Between Hermite-Gaussian and Laguerre-Gaussian Beams," *IEEE Photonics Journal*, vol. 14, no. 1, pp. 1-6, 2022, DOI: 10.1109/JPHOT.2022.3140359.

Author Biography

Kevin Chew Figueroa earned a BS in computer & electrical engineering from California Polytechnic University Pomona, focusing on circuit analysis, signal processing, and FPGA design. He received a MS from the University of Southern California's Computer Science (Data Science) program with a focus on Deep Learning, Computer Graphics, and Discrete Differential Geometry. Currently, he is completing his Ph.D. under the mentoring of Professor Brady, at the Wyant College of Optical Sciences, University of Arizona with a focus on computational imaging, photography, and information theory as it pertains to optics and electronic systems. He is a Fellow of NASA, ARCS, GEM, and was one of two finalists in the international Meta (Facebook) Photonics & Optics 2022 Ph.D. Research Fellowship Competition.

Zhipeng Dong received his B.Sc. degree in optical sciences engineering and applied mathematics from the University of Arizona (UA), Tucson, the United States, in 2021. He is currently working toward a Ph.D. degree with the Wyant College of Optical Sciences, the University of Arizona, Tucson, the United States. His research interests include super-resolution, radiance field reconstruction, and imaging processing.

Gregory Nero received the B.S. degree in imaging science from the Rochester Institute of Technology, Rochester, NY, USA, in 2020. They are currently working toward the Ph.D. degree in optical sciences with the James C. Wyant College of Optical Sciences, University of Arizona, Tucson, AZ, USA. Their research interests include designing SLM-based optical engineering solutions for display and communication systems as well as developing computational imaging techniques for super-resolution.

Gordon Charles Hageman received a BS in Astronomy and Physics from Northern Arizona University in 2022 and is currently pursuing a Masters in Optical Sciences at The University of Arizona. Under the supervision of Dr. David Brady, he works in the Camera Lab, focusing on the use of telescope arrays for remote sensing. His work includes synthetic tracking, unresolved object characterization, and synthetic aperture applications.

David J. Brady earned a B.A. in physics and math from Macalester College (1984) and M.S.(1986) and Ph.D.(1990) from Caltech in applied physics. He was on the faculty of the University of Illionis and Duke University prior to joining the Wyant College of Optical Sciences at the University of Arizona in 2021. His work focuses on computational imaging. He is a Fellow of Optica, SPIE and IEEE.