Fast and Uncertainty-Aware SVBRDF Recovery from Multi-View Capture using Frequency Domain Analysis – Appendix

ANONYMOUS AUTHOR(S)

CCS Concepts: • **Computing methodologies → Reflectance modeling**; **3D imaging**; **Appearance and texture representations**.

Additional Key Words and Phrases: Object acquisition, Appearance capture, Inverse rendering

ACM Reference Format:

Anonymous Author(s). 2024. Fast and Uncertainty-Aware SVBRDF Recovery from Multi-View Capture using Frequency Domain Analysis – Appendix. 1, 1 (June 2024), 3 pages. https://doi.org/XXXXXXX.XXXXXXX

A DERIVATION OF CONVOLUTION MODEL

The outgoing radiance at point p along direction ω_o, $L_o(p, \omega_o)$ is given by

$$B(p, \omega_o) = \int_{H^2(\mathbf{n})} f_r(p, \omega_o, \omega_i) L(p, \omega_i) \cos\theta_i d\omega_i, \tag{1}$$

where f is the BRDF and $L(p, \omega_i)$ is the incident radiance along direction ω_i. For the Torrance-Sparrow BRDF, f is defined as

$$f(p, \omega_o, \omega_i) = K_d + K_s \frac{D(\omega_m) F(\omega_o \cdot \omega_m) G(\omega_i, \omega_o)}{4 \cos\theta_i \cos\theta_o}, \tag{2}$$

where ω_m is the half-direction vector $\omega_m = (\omega_i + \omega_o)/||\omega_i + \omega_o||$; $D(\omega_m)$ is the normal distribution function; $F(\omega_o \cdot \omega_m)$ is the Fresnel term. Ramamoorthi and Hanrahan simplify this term to $F(\theta_o)$, as the angle θ_o is often close to the angle between ω_o and ω_m; $G(\omega_i, \omega_o)$ is the shadowing-masking term. Ramamoorthi and Hanrahan ignore G. We assume shadowing and masking are independent statistical events, so that $G(\omega_i, \omega_o) = G(\omega_i)G(\omega_o)$.

There are two important notes about the denominator in Equation 2:

(1) $1/(4\cos\theta_o)$ results from the half-direction transform: the distribution of microfacets with a normal ω_m is transformed to the distribution of outgoing directions ω_o that the incoming light ray ω_i reflects toward (see Pharr et al., Equation 9.27).
(2) $1/(\cos\theta_i)$ cancels out the cosine term applied to the incoming radiance (see Pharr et al., equation 9.30).

We now substitute Equation 2 into Equation 1 and split the equation into diffuse and specular

$$\begin{aligned} B(p, \omega_o) = {} & K_d \int_{H^2(\mathbf{n})} L(p, \omega_i) \cos\theta_i d\omega_i \\ & + K_s \int_{H^2(\mathbf{N})} \frac{D(\omega_m) F(\omega_o \cdot \omega_m) G(\omega_i, \omega_o)}{4 \cos\theta_i \cos\theta_o} L(p, \omega_i) \cos\theta_i d\omega_i \end{aligned} \tag{3}$$

This equation is simplified by Ramamoorthi and Hanrahan using the assumptions that F only depends on θ_o and shadowing-masking is ignored. We replace the integral of incoming radiance for diffuse with the symbol for irradiance E.

$$B(p, \omega_o) = K_d E(p) + K_s F(\theta_o) \int_{H^2(\mathbf{N})} \frac{D(\omega_m)}{4 \cos\theta_o} L(p, \omega_i) d\omega_i. \tag{4}$$

Ramamoorthi and Hanrahan rewrite the specular term as a convolution between a filter based on D, and L. Crucially, the domain of D in the Torrance-Sparrow model is the half-angle space. In Ramamoorthi and Hanrahan's derivation, the spherical harmonic representation for this filter, in the paper referred to as S is derived in incoming-direction space for normal exitance (Ramamoorthi and Hanrahan, Equation 27). This has two consequences:

(1) We do not have to account for a change of variables and $1/(4\cos\theta_o)$ can be removed.
(2) In reality, S depends on the outgoing direction that is observed and thus, the filter changes shape. This variation is ignored with the explanation that "the BRDF filter is essentially symmetric about the reflected direction for small viewing angles, as well as for low frequencies l. Hence, it can be shown by Taylor-series expansions and verified numerically, that the corrections to equation 20 [Equation 9 in our paper] are small under these conditions."

This means that we can rewrite Equation 4 with a convolution

$$B(p, \omega_o) = K_d E(p) + K_s F(\theta_o) \left[S * L\right]_{\omega_o}, \tag{5}$$

which equals Equations 21 and 22 in Ramamoorthi and Hanrahan.

B SAMPLING THEORY

The transformation from the directional domain to spherical harmonics begs the question: do we have the enough samples to accurately recover the coefficients of the outgoing radiance? We know from Equation 9 that the BRDF acts as a low-pass filter parameterized by α. We connect this knowledge with sampling theory to derive lower bounds on sampling counts.

The Nyquist-Shannon theorem provides a lower bound on the number of samples required to exactly recover a band-limited signal using a Fourier series. Similar theorems have been developed for spherical harmonics [Driscoll and Healy 1994; McEwen et al. 2011; McEwen and Wiaux 2011]. These state that, to recover a spherical signal with band-limit ℓ^*, the number of samples should be $O(\ell^{*2})$. The sampling rate and related band-limit have direct consequences for BRDF recovery. Assume that the incoming light has been sampled at a high enough rate to be accurately recovered, for example, from projected photographs or a gazing sphere. Then the outgoing light is the weakest link, as it is sampled by moving the camera along N positions around the object. Sampling theory tells us that we can only accurately recover outgoing radiance that is band-limited to $\ell^* < \sqrt{N}$ degrees. Signals with non-zero amplitude in higher degrees will suffer from aliasing.

Fortunately, the BRDF acts as a low-pass filter on the incoming radiance (Equation 9). That means the outgoing radiance can fall into two categories, based on the α parameter of the material (α = roughness^2): α is either too low or α is high enough to recover spherical harmonic coefficients. If α is too low, the low-pass filtering from the BRDF does not band-limit the signal enough to accurately

2024. XXXX-XXXX/2024/6-ART $15.00
https://doi.org/XXXXXXX.XXXXXXX

recover with the given sampling rate. The threshold for α can be determined based on Equation 9. Let t be an acceptable attenuation factor for degrees $\ell > \ell^*$. We solve Equation 9 for t to find the lower bound, α', for accurate recovery

$$\alpha' = \ell^{*-1}\sqrt{-\ln t}. \quad (6)$$

An acceptable threshold t can be determined empirically, by investigating the reconstruction error for a set of environment maps. To provide some intuition, for $N = 400$ samples and a threshold of $t = 0.5$, $\alpha' \approx 0.07$. Above this threshold, our method can recover α and K_s to an acceptable accuracy, provided that the incoming radiance has enough amplitude in the right degrees. This also extends to non-uniform samples, because the Nyquist-Shannon theorem holds for non-uniform samples [Marvasti 2012]. In other words: if a lower bound on α is known, it does not matter where the camera is placed, as long as the average distance to the closest sample is equal to $1/N$. It also means that one can determine the number of required views based on the lowest α that should be recovered: $N \sim \alpha^{-2}$.

It is important to understand what happens if $\alpha < \alpha'$. First, we would be uncertain where α lands between 0 and α', based on the power spectrum alone. For $0 < \alpha < \alpha'$, Equation 9 is close to 1 for all degrees below ℓ^*. Second, because this situation occurs for low α, the outgoing radiance should be similar to the incoming radiance, up to a scaling factor for absorption and transmission. It is unlikely that the spherical harmonics decomposition with significant aliasing will match a filtered version of the incoming radiance. Therefore, we can detect that $\alpha < \alpha'$. In this case, the MSE for any parameter combination ψ is relatively high. Once such a case is detected, we know that our spherical harmonic-based analysis provides no further insights on (un)certainty. There is still a chance for accurate BRDF recovery if $\alpha < \alpha'$. A sample might land on a fortunate spot in the outgoing radiance field. This is the case when there is high local variation in the incoming light around the sample locations, resulting in large changes in radiance for small changes in α. One could quantify this variation by comparing the difference between the sample location for $\alpha = 0$ and $\alpha = \alpha'$ and use this as a measure of certainty. In our work, we find that our certainty measure works well, even for $\alpha < \alpha'$ and thus, we do not add this measure.

C STANFORD ORB OTHER RESULTS

We include the results table from StanfordORB for reference. These results were obtained under different acquisition condition and cannot be directly compared to our results.

D ABLATIONS

D.1 Spherical Harmonics fitting

Our method computes spherical harmonic coefficients using a least-squares fit and includes a regularizer. We would like to understand the effect of the maximum degree that is estimated, find the optimal weight for the regularizer, and see if the regularizer has the desired effect (improved accuracy). The results for the maximum degree are presented in Table 2. We find that more degrees help, but also that we obtain good results with a relatively low number of degrees (from 3 on). In Table 3, we find that our approach is not very sensitive to the specific setting of the regularizer, with an optimal value near $\lambda = 1 \times 10^{-4}$. When optimizing BRDF parameters, it is typical to weight samples based on their angle with the normal, θ. For example, samples at grazing angles are often associated with lower confidence and weighted less than samples near $\theta = 0$. We set up a general weighting function, $\max(0, 1 - (1 - \cos a\theta)^b)$ that ignores samples with $\theta > a\frac{\pi}{2}$ and weights the rest with a smooth falloff determined by b. We observe in Table 4 that $a = 1, b = 1$ gives the best results. We also observe that weighting is beneficial, compared to constant weight (top row).

Table 1. Benchmark Comparison for **Novel Scene Relighting** of Existing Methods from [Kuang et al. 2023]. † denotes models trained with the ground-truth 3D scans and pseudo materials optimized from light-box captures. The rest of results are obtained by optimizing jointly for illumination, geometry and material. *We report these numbers for reference, however they cannot be directly compared to our results.*

	PSNR-H↑	PSNR-L↑	SSIM↑	LPIPS↓
NVDiffRecMC [Hasselgren et al. 2022] †	25.08	32.28	0.974	0.027
NVDiffRec [Munkberg et al. 2022] †	24.93	32.42	0.975	0.027
PhySG [Zhang et al. 2021a]	21.81	28.11	0.960	0.055
NVDiffRec [Munkberg et al. 2022]	22.91	29.72	0.963	0.039
NeRD [Boss et al. 2021]	23.29	29.65	0.957	0.059
NeRFactor [Zhang et al. 2021b]	23.54	30.38	0.969	0.048
InvRender [Wu et al. 2023]	23.76	30.83	0.970	0.046
NVDiffRecMC [Hasselgren et al. 2022]	24.43	31.60	0.972	0.036
Neural-PBIR [Sun et al. 2023]	26.01	33.26	**0.979**	**0.023**

Table 2. Ablation max degree ℓ^*

ℓ^*	PSNR-H↑	PSNR-L↑	SSIM↑	LPIPS↓	Time
0	25.670	32.580	0.971	0.042	0.94s
1	25.531	32.536	0.971	0.042	1.08s
2	25.588	32.568	0.971	0.043	1.10s
3	25.881	32.983	0.972	0.041	1.13s
4	26.134	33.142	0.972	0.040	1.21s
5 (Ours)	26.182	33.215	0.972	0.040	1.42s
6	26.199	33.266	0.972	0.040	1.70s
7	26.147	33.227	0.972	0.040	2.05s
8	26.153	33.241	0.972	0.040	3.02s

Table 3. Ablation regularizer weight

λ	PSNR-H↑	PSNR-L↑	SSIM↑	LPIPS↓	Time
1×10^{-2}	26.498	33.683	0.976	0.033	**4.32s**
1×10^{-3}	26.524	33.703	0.976	0.033	4.33s
1×10^{-4}	26.582	33.762	0.976	0.033	4.33s
1×10^{-5}	26.484	33.667	0.975	0.033	4.33s
1×10^{-6}	**26.638**	**33.807**	**0.976**	**0.033**	4.33s
1×10^{-6} constant	26.611	33.788	0.976	**0.032**	4.36s
1×10^{-7}	26.585	33.750	0.975	0.033	4.33s

Table 4. Ablation sample weighting

$\max(0, 1 - (1 - \cos a\theta)^b)$	PSNR-H↑	PSNR-L↑	SSIM↑	LPIPS↓	Time
No weighting	26.457	33.607	0.975	0.034	4.33s
$a = 0.8, b = 1$	26.445	33.705	0.976	**0.032**	4.33s
$a = 0.9, b = 1$	26.529	33.725	0.976	0.032	4.33s
$a = 1, b = 1$ (Ours)	**26.638**	**33.807**	**0.976**	**0.033**	4.33s
$a = 1, b = 2$	26.618	33.800	0.976	0.032	4.33s
$a = 1, b = 3$	26.588	33.778	0.976	0.033	4.33s
$a = 1, b = 4$	26.558	33.750	0.975	0.033	4.33s
$a = 1, b = 5$	26.530	33.719	0.975	0.033	4.33s

REFERENCES

Mark Boss, Raphael Braun, Varun Jampani, Jonathan T Barron, Ce Liu, and Hendrik Lensch. 2021. Nerd: Neural reflectance decomposition from image collections. In *Proceedings of the IEEE/CVF International Conference on Computer Vision.* 12684–12694.

J.R. Driscoll and D.M. Healy. 1994. Computing Fourier Transforms and Convolutions on the 2-Sphere. *Advances in Applied Mathematics* 15, 2 (1994), 202–250. https://doi.org/10.1006/aama.1994.1008

Jon Hasselgren, Nikolai Hofmann, and Jacob Munkberg. 2022. Shape, light, and material decomposition from images using Monte Carlo rendering and denoising. *Advances in Neural Information Processing Systems* 35 (2022), 22856–22869.

Zhengfei Kuang, Yunzhi Zhang, Hong-Xing Yu, Samir Agarwala, Shangzhe Wu, and Jiajun Wu. 2023. Stanford-ORB: A Real-World 3D Object Inverse Rendering Benchmark. arXiv:2310.16044 [cs.CV]

Farokh Marvasti. 2012. *Nonuniform sampling: theory and practice.* Springer Science & Business Media.

Jason D. McEwen, Gilles Puy, Jean-Philippe Thiran, Pierre Vandergheynst, Dimitri Van De Ville, and Yves Wiaux. 2011. Sampling theorems and compressive sensing on the sphere. In *Wavelets and Sparsity XIV*, Manos Papadakis, Dimitri Van De Ville, and Vivek K. Goyal (Eds.), Vol. 8138. International Society for Optics and Photonics, SPIE, 81381F. https://doi.org/10.1117/12.893481

Jason D McEwen and Yves Wiaux. 2011. A novel sampling theorem on the sphere. *IEEE Transactions on Signal Processing* 59, 12 (2011), 5876–5887.

Jacob Munkberg, Jon Hasselgren, Tianchang Shen, Jun Gao, Wenzheng Chen, Alex Evans, Thomas Müller, and Sanja Fidler. 2022. Extracting triangular 3d models, materials, and lighting from images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 8280–8290.

Matt Pharr, Wenzel Jakob, and Greg Humphreys. 2023. *Physically based rendering: From theory to implementation.* MIT Press.

Ravi Ramamoorthi and Pat Hanrahan. 2001. A signal-processing framework for inverse rendering. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques (SIGGRAPH '01).* Association for Computing Machinery, New York, NY, USA, 117–128. https://doi.org/10.1145/383259.383271

Cheng Sun, Guangyan Cai, Zhengqin Li, Kai Yan, Cheng Zhang, Carl Marshall, Jia-Bin Huang, Shuang Zhao, and Zhao Dong. 2023. Neural-PBIR Reconstruction of Shape, Material, and Illumination. In *Proceedings of the IEEE/CVF International Conference on Computer Vision.* 18046–18056.

Haoqian Wu, Zhipeng Hu, Lincheng Li, Yongqiang Zhang, Changjie Fan, and Xin Yu. 2023. NeFII: Inverse Rendering for Reflectance Decomposition with Near-Field Indirect Illumination. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 4295–4304.

Kai Zhang, Fujun Luan, Qianqian Wang, Kavita Bala, and Noah Snavely. 2021a. Physg: Inverse rendering with spherical gaussians for physics-based material editing and relighting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 5453–5462.

Xiuming Zhang, Pratul P Srinivasan, Boyang Deng, Paul Debevec, William T Freeman, and Jonathan T Barron. 2021b. Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. *ACM Transactions on Graphics (ToG)* 40, 6 (2021), 1–18.