Supplemental Document: Self-Calibrating, Fully Differentiable NLOS Inverse Rendering

Kiseok Choi
KAIST
South Korea
kschoi@vclab.kaist.ac.kr

Inchul Kim
KAIST
South Korea
ickim@vclab.kaist.ac.kr

Dongyoung Choi
KAIST
South Korea
dychoi@vclab.kaist.ac.kr

Julio Marco
Universidad de Zaragoza - I3A
Spain
juliom@unizar.es

Diego Gutierrez
Universidad de Zaragoza - I3A
Spain
diegog@unizar.es

Min H. Kim
KAIST
South Korea
minhkim@vclab.kaist.ac.kr

ACM Reference Format:
Kiseok Choi, Inchul Kim, Dongyoung Choi, Julio Marco, Diego Gutierrez, and Min H. Kim. 2023. Supplemental Document: Self-Calibrating, Fully Differentiable NLOS Inverse Rendering. In *SIGGRAPH Asia 2023 Conference Papers (SA Conference Papers '23), December 12–15, 2023, Sydney, NSW, Australia.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3610548.3618140

This supplemental document provides additional information and results in support of the primary document. Refer to Table 1 for the notations and symbols used in this paper.

1 JOINT LASER-SENSOR CORRELATION MODEL

The joint laser-sensor model is derived as Equation 1 following the previous related work [Chen et al. 2020; Hernandez et al. 2017].

$$\begin{aligned} \boldsymbol{H}_R &= P_{PDE} \cdot (\Phi * (\Lambda * \boldsymbol{H}_r + L_a)) + L_{DCR} \\ &= (\Phi * \Lambda * \boldsymbol{H}_r + \Phi * L_a) + L_{DCR} \\ &= ((E_s * G_s) * G_l * \boldsymbol{H}_r + (E_s * G_s) * L_a) + L_{DCR} \\ &= ((E_s * (G_s * G_l) * \boldsymbol{H}_r + (E_s * G_s) * L_a) + L_{DCR} \\ &= ((E_s * G_{ls} * \boldsymbol{H}_r + (E_s * G_s) * L_a) + L_{DCR} \\ &= (\Psi(t; I_l, \kappa_s, \sigma_{ls}) * \boldsymbol{H}_r + L_a) + L_{DCR} \\ &= \Psi(t; I_l, \kappa_s, \sigma_{ls}) * \boldsymbol{H}_r + \eta_s \end{aligned} , \quad (1)$$

where P_{PDE} denotes the photon detection efficiency. L_a is the ambient light and L_{DCR} is the dark count rate. Φ is the sensor model function that can be expressed in the form of convolution between exponential function E_s and Gaussian function G_s. Λ is the laser function that has the shape of Gaussian G_l. Note that the convolution of two Gaussians G_s and G_l can be merged to a single Gaussian G_{ls}. The convolution of E_s and G_{ls} is then expressed as Ψ that has three parameters I_l, κ_s, and σ_{ls}. L_a and L_{DCR} can be summed to a single offset value η_s. Our joint laser-sensor correlation model finally has four parameters and these values are optimized in our self-calibrating pipeline.

SA Conference Papers '23, December 12–15, 2023, Sydney, NSW, Australia

ACM ISBN 979-8-4007-0315-7/23/12.
https://doi.org/10.1145/3610548.3618140

2 EXPERIMENTAL DETAILS

Table 2 summarizes the type of data (confocal or non-confocal), as well as the dimensions of the transient data, the dimensions of the reconstructed volume, the total reconstruction time, and the number of iterations before convergence; note that most of our scenes are significantly larger than previously reported results by transient optimization methods.

3 ADDITIONAL RESULTS

This section provides additional validations and results.

Manual parameter adjustment vs. our self-calibration. Figure 1 compares the estimated volumetric intensities of Bike and Resolution scenes by two different methods: the light cone transform (LCT) [O'Toole et al. 2018] and ours. To handle noise in the input dataset, we manually tweak the SNR parameter in the LCT method with a very wide range from 0.001 to 1.0. Our method yields clearer results than any of the results under the explored values for the SNR parameter of LCT, throughout all exposure levels.

Progressive optimization results. Figure 2 show detailed progress of the optimization in the Dragon and Erato scenes, displaying the evolution of the phasor-field kernel until the converged state. While the full optimization takes 100 iterations (1.28 hours), after only 50 iterations (39 minutes) the converged phasor-field kernel parameters already yield volumetric and geometric reconstructions very close to the converged result, while the remaining iterations refine more local details.

REFERENCES

Wenzheng Chen, Fangyin Wei, Kiriakos N. Kutulakos, Szymon Rusinkiewicz, and Felix Heide. 2020. Learned Feature Embeddings for Non-Line-of-Sight Imaging and Recognition. *ACM Trans. Graph.* 39, 6 (2020).

Quercus Hernandez, Diego Gutierrez, and Adrian Jarabo. 2017. A Computational Model of a Single-Photon Avalanche Diode Sensor for Transient Imaging. *arXiv* preprint arXiv:1703.02635 (2017).

Matthew O'Toole, David B Lindell, and Gordon Wetzstein. 2018. Confocal non-line-of-sight imaging based on the light-cone transform. *Nature* 555, 7696 (2018), 338.

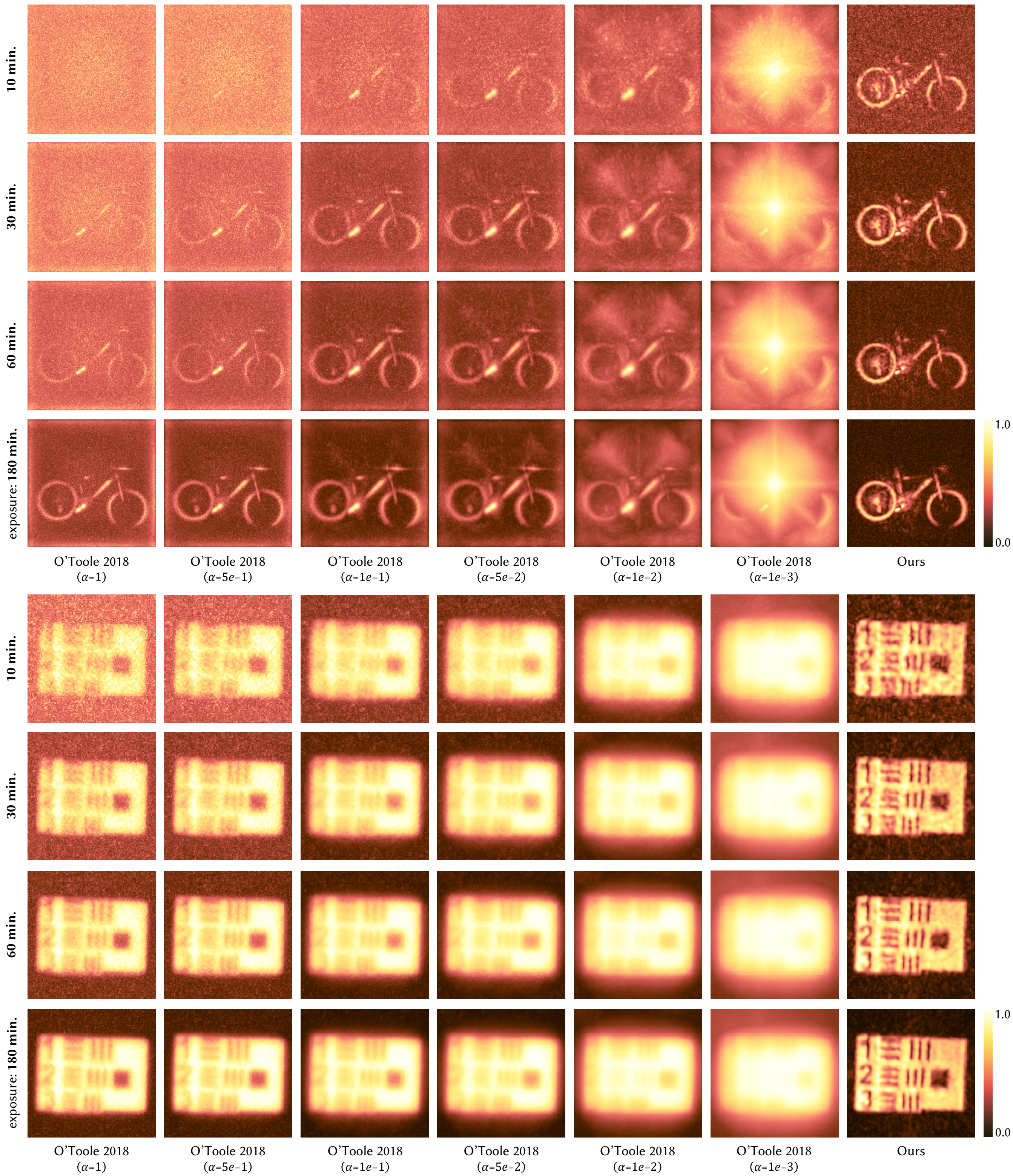

Figure 1: Comparisons of the estimated volumetric intensities of BIKE and RESOLUTION scenes by the light cone transform [O'Toole et al. 2018] and ours. To handle noise, we changed the SNR parameter α between 1 and 0.001 for different exposure times. In all exposure levels, our volume intensities outperform those of the LCT with manually selected parameters.

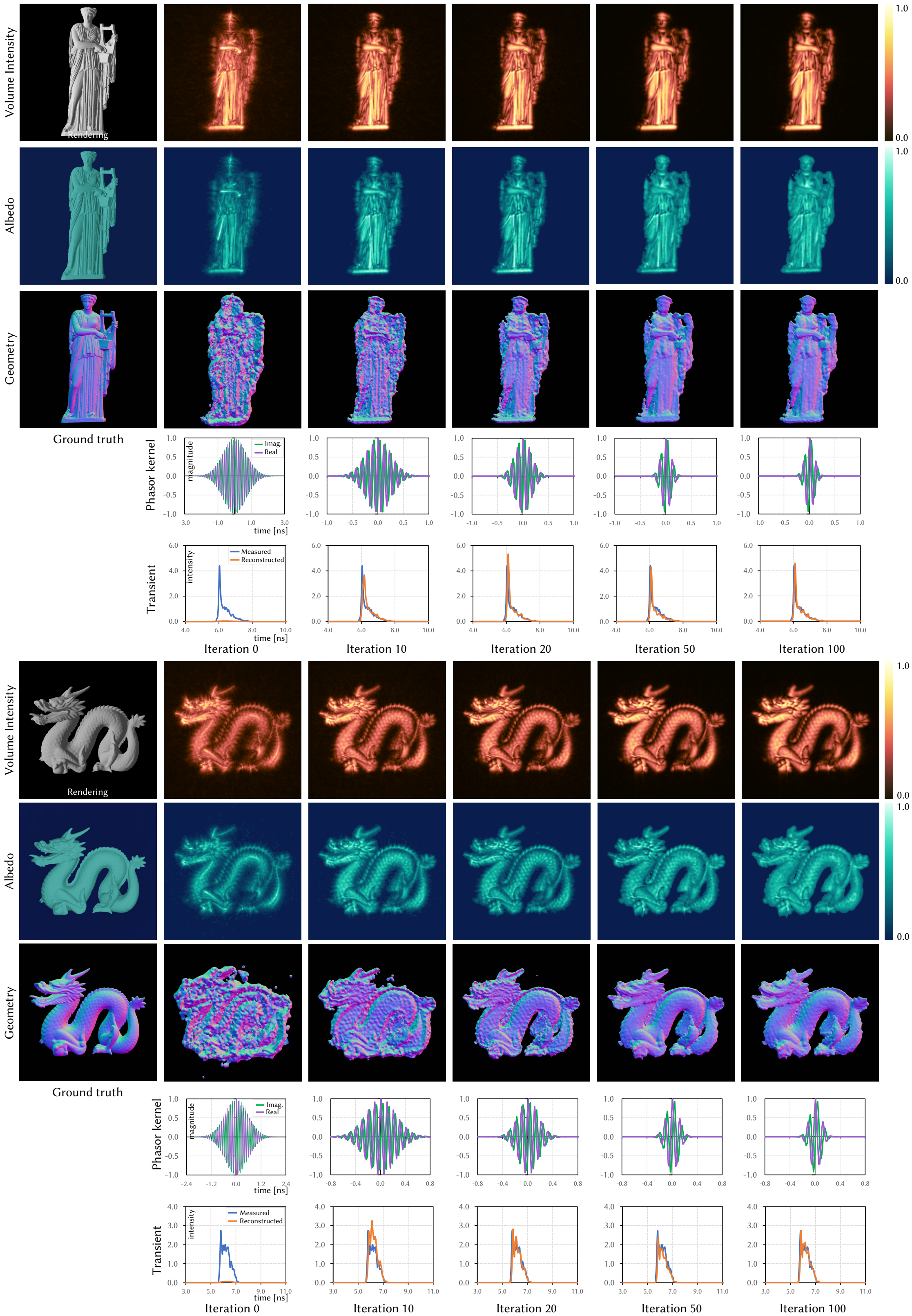

Figure 2: Progressive optimization of volumetric intensity, geometry, phasor kernel, and transient measurement samples of the ERATO and DRAGON scenes, showing how our reconstructions quickly converge after only 100 iterations.

Table 1: Main notations and symbols used in the paper.

Symbol	Description
$\bar{\mathbf{x}} = \mathbf{x}_0 ... \mathbf{x}_k$	Light path of $k+1$ vertices
$\mathbf{x}_l$	Light source point on the relay wall
$\mathbf{x}_g$	Surface point in the hidden scene
$\mathbf{x}_s$	Sensor point on the relay wall
$\mathbf{x}_v$	Voxel in a volumetric grid
$\mathbf{n}_g$	Surface normal in the hidden scene
G	Scene geometry parameters: points $\mathbf{x}_g$ and normals $\mathbf{n}_g$
$\mathbf{t} = t_0 ... t_k$	Time delays on $k+1$ vertices
d	Distance between the hidden surface and the relay wall
ψ	Space of all light paths
ψ_k	Space of light paths of $k+1$ vertices
$\mathcal{T}$	Space of temporal delays
c	Speed of light in vacuum
$\mathrm{tof}\,(\bar{\mathbf{x}})$	Total time of path $\bar{\mathbf{x}}$
$\mathcal{K}$	Time-resolved path contribution
$\boldsymbol{H}$	Transient measurements
$\boldsymbol{H}_{pf}$	Transient measurements filtered by a phasor kernel
$\boldsymbol{H}_r$	Rendered transient illumination
$\boldsymbol{H}_R$	Rendered transient after laser-sensor model applied
$D\,()$	Geometry estimation function
$\rho\,()$	Reflectance function at vertex
$V\,()$	Visibility function
$\mathfrak{T}\,()$	Path throughput with geometric attenuation/visibility
$R\,()$	Transient rendering function
I_{pf}	Volumetric intensity backprojected by Rayleigh-Sommerfeld integrals of phasor-field diffraction
Ω_{pf}	Illumination frequency of phasor field kernel
σ_{pf}	Illumination standard deviation of phasor field kernel
$\mathcal{P}()$	Filtering function with a phasor field kernel
I_l	Laser energy intensity
σ_l	Standard deviation of Gaussian laser pulse signal
κ_s	Sensor sensitivity decay rate
η_s	Sum of ambient light and sensor dark count rate
σ_{ls}	Standard deviation of Gaussian parameter for $\Psi\,()$
$\Lambda\,()$	Light source emission function
$\Phi\,()$	Sensor sensitivity function
$\Psi\,()$	Joint light-sensor correlation function
Θ_{pf}	Parameters of phasor field kernel: Ω_{pf}, σ_{pf}
Θ_{ls}	Parameters of laser and sensor models: σ_{ls}, I_l, κ_s, η_s
Θ_G	Parameters of per-voxel albedo ρ
Θ	Set of optimizing variables: $\Theta = \{\Theta_{\mathrm{pf}}, \Theta_{\mathrm{ls}}, \Theta_G\}$
$\mathcal{L}$	Loss function
$\lambda_{1...2}$	Loss-scale balance hyperparameters
Γ	Set of regularization terms

Table 2: Configurations of our input datasets, including converge time and the number of iterations needed.

	Scene	Confocal	Trans. measurement	Volume dimension	Time [hr. (#iter.)]
Synthetic	Bunny	Y	$256 \times 256 \times 1024$	$256 \times 256 \times 201$	1.93 (100)
	Dragon	Y	$256 \times 256 \times 1024$	$256 \times 256 \times 128$	1.28 (100)
	Erato	Y	$256 \times 256 \times 1024$	$256 \times 256 \times 128$	1.28 (100)
	Indonesian	Y	$256 \times 256 \times 1024$	$256 \times 256 \times 128$	1.93 (150)
Real	34	Y	$64 \times 64 \times 500$	$64 \times 64 \times 105$	1.05 (300)
	Bike	Y	$256 \times 256 \times 512$	$256 \times 256 \times 64$	1.73 (170)
	Resolution	Y	$256 \times 256 \times 512$	$256 \times 256 \times 26$	1.33 (300)
	SU	Y	$64 \times 64 \times 2048$	$64 \times 64 \times 584$	6.10 (200)
	44i	N	$130 \times 180 \times 4096$	$180 \times 180 \times 417$	3.76 (150)
	NLOS	N	$130 \times 180 \times 4096$	$180 \times 180 \times 417$	4.38 (180)