(Supplementary Material) Domain Reduction Strategy for Non-Line-of-Sight Imaging

In this supplementary material, we first provide additional discussion and comparison with FFT-based methods in Section A. Then, we provide additional details of the proposed method in Section B, additional comparisons with more baseline methods in Section D, reconstruction time of our method in Section C, and additional evaluations and analysis in Section E.

A Comparison with FFT-based Methods

A.1 FFT-based Methods and Optimization

We would like to demonstrate that the goal of our work is not to achieve the fastest reconstruction speed. Instead, this work aims to identify efficiency bottlenecks inherent in previous optimization-based methods, particularly stemming from computations related to empty regions. Consequently, our domain reduction achieves substantial efficiency improvement (see Table 2 of the main paper), while inheriting the general applicability of the optimization framework.

While FFT-based methods benefit from computational efficiency based on the convolutional theorem or Stolt's method, several assumptions have to be made to achieve this, *e.g.* dense scanning points, planar relay walls, or ignoring surface normals. Some approximations could relax certain assumptions, but at the cost of sacrificing performance. In contrast, optimization-based methods are more generally applicable to various scenarios, with the potential for continuous modeling, joint optimization of noise parameters, arbitrary BRDF [8], and sparse samplings [5, 9]. Our domain reduction effectively addresses computational burdens of optimization frameworks, paving the way to unleash their full potential.

A.2 Analysis on Sparse Sampling

Reconstructing high-resolution volumes from undersampled sparse measurements is an ill-posed problem. Previous FFT-based methods require measurements to have the same resolution with the target volumes. One may apply interpolation techniques to upsample the measurements to the desired resolution, as commented by R3. However, this inevitably introduces approximation errors between upsampled and actual measurements. Therefore, solutions of all methods exhibit approximation errors, which become larger as scanning patterns become sparser. Optimization-based methods are capable of finding solutions with minimal errors through iterative minimization procedures. This is further demonstrated in the following paragraphs.



Fig. 7: Additional comparisons with FFT-based methods, which employs the bicubic interpolation as an additional technique to upsample the input transients to 128×128 scanning resolution.

Results on 32×32 with bicubic interpolation. For more precise comparisons, we deliver the results of FFT-based methods, using the additional bicubic interpolation technique. These results are obtained by first applying the bicubic interpolation to upsample the transients to the 128×128 spatial resolution, and then applying the FFT-based methods to the upsampled transients. As shown in Fig. 7, the additional bicubic interpolation improves the results of 3D convolution-based methods [6, 10], while the results of FK and Phasor field are less affected by the interpolation technique. Nevertheless, LCT and DLCT still produces blurry outputs, incorrect albedo values (see results of Bunny), and inaccurate fine details of the objects (*e.g.*, the ear of Bunny, the face of Serapis, the head of Dragon). Contrary to these methods, our method reconstructs hidden volumes directly from 32×32 transients, reconstructing high-quality volumes with fine details, while being robust to the effects of noise.

Results on 16×16 **sampling.** We also present the results of FFT-based methods on the 16×16 sparsely sampled measurements, which are upsampled to the lateral resolution of the target volumes by filling unsampled pixels with zero. As shown in Fig. 8 (a), approximation errors caused by the upsampling become evidently larger as scanning patterns become sparser, leading to unfavorable results of the FFT-based methods. On the other hand, our optimization-based framework successfully reconstructs clean shapes of the objects with many details in these challenging scenarios.



Fig. 8: (a) Results on 16×16 samplings, upsampled to 128×128 by filling unsampled pixels with zero. FFT-based methods produce unfavorable results with artifacts regardless of the upsampling techniques. (b) Results on 256×256 transients of ZNLOS Bunny.

Results on 256×256 measurements. Finally, we deliver the results on 256×256 measurements of ZNLOS [2] Bunny in Fig. 8 (b). As can be seen, all methods produce compelling results with sufficient scan points and sufficiently long scanning time.

B Method Details

B.1 Reconstruction Objective

Combining the noise parameters and the L1 regularization, the objective of our reconstruction pipeline can be described as

$$\mathcal{L}(\rho, \mathbf{n}) = ||(T+d) - \tau_{qt}||_2 + \alpha ||\rho||_1,$$
(8)

where T is the rendered transients, τ_{gt} is the ground truth measurements, and d is the noise parameter defined for each histogram. We set α to 0.8 for real-world scenes and 0.001 for synthetic measurements.

B.2 Additional Implementation Detail

In this section, we provide detailed explanations of our implementations for the future reproducibility. For albedo variables, we apply the ELU [1] activation function to suppress the negative albedo values. We set the orthogonal direction from the hidden objects to the relay wall as (0, 0, -1). With $\mathbf{n} = (n_x, n_y, n_z)$ representing the surface normal, points of the hidden objects which have positive n_z values (backfaces of the objects) are not visible from the measurements. To ensure negative n_z values, we apply the tanh to the variables and add (0, 0, -1)

4 F. Author et al.

Table 5: The ratio of active regions and the reconstruction time of all instances. The reconstruction time is measured using a single commercial RTX 3090 GPU. While the reconstruction time varies across the instances, our method typically takes about a minute to reconstruct 128×128 hidden volumes.

Scene	num. iter	active ratio	recon. time
Statue	1,000	0.9~%	$25 \mathrm{s}$
Dragon	1,000	2.6~%	$65 \ s$
Bunny	1,000	3.0~%	$54 \mathrm{s}$
Serapis	1,000	8.2~%	$130 \mathrm{~s}$
Bunny (non-confocal)	1,000	3.3~%	$91 \mathrm{s}$
NT (non-planar)	1,000	0.4~%	70 s

to obtain the surface normal vector. This process results in a range (-1, 1) for n_x and n_y , and a range (-2, 0) for n_z . Finally, we normalize the surface normal to make it as a unit vector.

Throughout all experiments, our method takes transients with 32×32 scanning points and reconstructs hidden volumes with a $128 \times 128 \times 333$ resolution, where the last is the resolution along z-axis. The standard deviation of the Gaussian kernel used in the soft domain reduction is set to 3. We empirically observe that slightly reducing the threshold of the domain reduction under the non-confocal setups yield better reconstruction quality. Therefore, we set the threshold to 5% for the confocal measurements and 3% for the non-confocal measurements. To report the results of fast Fourier transform (FFT)-based methods, we upsample spatial resolution of their outputs with bicubic interpolation to make 128×128 resolution. To measure the quantitative results of ZNLOS Bunny, we upsample the spatial resolution of results to 256×256 , matching that of the ground truth. By analyzing the last 10% transient histograms along t-axis of the real-world measurements, we empirically set b to 0.05 and λ to 0.06 for the noise regularization. These values are slightly reduced for reconstructing retroreflective targets $(b = 0.004, \lambda = 0.0012)$, which usually have higher maximum intensity values. To report the results of NeTF [8], we use the original source code provided by the authors, and train this model for 192 epochs with 2 stage training as in the original work. The training of NeTF takes more than 2 days in our environment. The optimization process of our method for revealing the albedo and surface normal of a single scene takes 1k iterations, which require about a minute using a single commercial RTX 3090 GPU.

C Reconstruction Time

We deliver the reconstruction time and the ratio of active regions for reconstructing all scenes in Table 5. Although the reconstruction time could vary according to characteristics of the scenes, *i.e.* remaining domain at each step, our method demonstrates its efficiency across all instances, typically taking about a minute to reconstructing 128×128 output volumes. Considering both reconstruction



Fig. 9: Additional comparisons including LCT, Phasor(FFT-based), FBP with a Laplacian filter (Lap.), FBP with a Laplacian-of-Gaussian filter (LoG), and Phasor field with wavelengths $\lambda = 2\Delta_p$ where Δ_p is the sampling distance.

time and scanning time required for the high-resolution outputs, our method can serve as an efficient and effective solution for reconstructing high-resolution volumes with 32×32 scanning points.

D Additional Comparison

To clearly demonstrate the effectiveness of our method, we provide comparisons with additional baseline methods. These include back-projection (BP) based methods that do not utilize Fourier transform and thus do not suffer from the lateral resolution issues, LCT [6], and the fast differentiable renderer [7].

D.1 Confocal Imaging Result

We report the results of LCT, Phasor with FFT, FBP with a Laplacian filter (Lap.) and a Laplacian of Gaussian (LoG) filter. We also deliver the results of Phasor with the wavelength $\lambda = 4\Delta_p$, while the results of Phasor in the main paper are with the wavelength $\lambda = 2\Delta_p$.

As reported in Fig. 9, our method clearly outperforms all other baseline methods, producing clearer results and successfully recovering fine details. The results of LCT are of low resolution, making it difficult to discern the details of the object. The results of Phasor with FFT fails to reconstruct specific parts, such as the bunny's ear. The results of FBP contain streak artifacts and noise, which often make the hidden objects difficult to be identified.



Fig. 10: (a) Reconstruction results of normal maps. We compare the results with DLCT [10] and the fast differentiable renderer proposed by Plack *et al.* [7] (denoted as Fast diff.). (b) Reconstruction results with non-planar relay walls. We additionally compare the results with Phasor field with a BP solver, using the wavelengths $\lambda = 2\Delta_p$ and $\lambda = 4\Delta_p$, where Δ_p is the sampling distance.

D.2 Surface Normal

We compare the reconstruction results of surface normals with the fast differentiable renderer [7]. The differentiable renderer, proposed by Plack *et al.*, reconstructs colors and surface geometry of the hidden objects through the differentiable rendering pipeline. As demonstrated in Fig. 10 (a), our method achieves compelling results in surface normal reconstruction, whereas other baselines, DLCT [10] and Plack *et al.* [7], only reconstruct coarse structures of the surfaces with artifacts, or lack several parts and details of the objects.

D.3 Non-Planar Relay Wall Result

We additionally provide comparisons of non-planar relay wall results, with Phasor field with a BP solver. We report the results of Phasor field with two different wavelengths, namely $\lambda = 2\Delta_p$ and $\lambda = 4\Delta_p$. As shown in Fig. 10 (b), our method delivers cleanest shapes of the "NT" instance, while other methods produce noisy results where the shapes of the instance are difficult to identify.

Title Suppressed Due to Excessive Length



Fig. 11: Results at several step during the optimization.



Fig. 12: Error map visualizations of the reconstructed depth maps on ZNLOS [2] Bunny.

E Additional Evaluation

We provide more evaluation results to clearly demonstrate the effectiveness of our method. First, we report the results at several steps during the optimization to illustrate that most details of the hidden objects can be reconstructed in the early stages of our optimization process. Second, we present error map visualizations of the reconstructed depth maps on ZNLOS [2] Bunny. Then we provide additional analysis and ablation study for a deeper understanding of our method.

E.1 Results at Different Steps

In Fig. 11, we report the results at several steps (100, 300, 500, 700, 900 steps) during the optimization process. As can be seen, our method already recovers almost all shapes of the objects at 500th iteration, and finer details are gradually revealed as the iteration progresses. While our method can reconstruct the satisfying outputs at the early stage, we continue the optimization process for high-fidelity results.

E.2 Depth Map Visualization

We provide visualizations of error maps of reconstructed depth maps in Fig. 12. While most of the baseline methods suffer from artifacts or missing details of the objects, our method delivers the accurately reconstructed depth maps while recovering most of the parts of the objects.



Fig. 13: Surface reconstruction results. We compare the reconstructed surfaces with DLCT [10].



Fig. 14: Ablation results on the continuous point sampling. We deliver the results with the continuous (denoted as cont.) and the fixed point (denoted as fixed) sampling.

E.3 Surface Reconstruction

We provide surface reconstruction results of our method and DLCT [10] in Fig. 13. Following [10], we obtain the reconstructed surfaces using the Poisson reconstruction method [3]. As evident, our method successfully reconstructs detailed surfaces using only 32×32 scanning points, whereas DLCT yields only coarse shapes of the objects, making it difficult to identify many details.

E.4 Additional Analysis

Ablation on continuous points sampling. We deliver the ablation results on the continuous points sampling in Fig. 14. Here, we compare our method, using the continuous grid-based random sampling, with the model using a fixed point sampling, which samples the center of each voxel. As can be seen, the fixed point sampling often produces inaccurate results at some part of the objects (see the circled part in Fig. 14), while the continuous sampling exhibits more accurate results in both synthetic and real-world datasets.

Results with various exposure time. To further validate the robustness of our method to noise, we present the results on the confocal real-world measurements with various exposure time. We use the 32×32 measurements with 28.1 s, 56.4 s, 168.8 s total exposure time, which correspond to 30, 60, 180 minute total exposure time of the original measurements. As shown in Fig. 15, our method

9



Fig. 15: Results on the confocal real-world measurements [4] with various exposure time. We use the measurements with 28.1 s, 56.4 s, 168.8 s total exposure time, which correspond to 30, 60, 180 minute total exposure time of the original measurements.

exhibits more clean and sharp outputs compare to DLCT [10], showing highquality results even with 28.1 s total exposure time. 10 F. Author et al.

References

- Clevert, D.A., Unterthiner, T., Hochreiter, S.: Fast and accurate deep network learning by exponential linear units (elus). arXiv preprint arXiv:1511.07289 (2015) 3
- Galindo, M., Marco, J., O'Toole, M., Wetzstein, G., Gutierrez, D., Jarabo, A.: A dataset for benchmarking time-resolved non-line-of-sight imaging (2019), https://graphics.unizar.es/nlos 3, 7
- Kazhdan, M., Bolitho, M., Hoppe, H.: Poisson surface reconstruction. In: Proceedings of the fourth Eurographics symposium on Geometry processing. vol. 7 (2006) 8
- Lindell, D.B., Wetzstein, G., O'Toole, M.: Wave-based non-line-of-sight imaging using fast fk migration. ACM Transactions on Graphics (TOG) 38(4), 1–13 (2019)
 9
- Liu, X., Wang, J., Xiao, L., Fu, X., Qiu, L., Shi, Z.: Few-shot non-line-of-sight imaging with signal-surface collaborative regularization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 13303–13312 (2023) 1
- O'Toole, M., Lindell, D.B., Wetzstein, G.: Confocal non-line-of-sight imaging based on the light-cone transform. Nature 555(7696), 338–341 (2018) 2, 5
- Plack, M., Callenberg, C., Schneider, M., Hullin, M.B.: Fast differentiable transient rendering for non-line-of-sight reconstruction. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 3067–3076 (2023) 5, 6
- Shen, S., Wang, Z., Liu, P., Pan, Z., Li, R., Gao, T., Li, S., Yu, J.: Non-line-ofsight imaging via neural transient fields. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 43(7), 2257–2268 (2021) 1, 4
- Ye, J.T., Huang, X., Li, Z.P., Xu, F.: Compressed sensing for active non-line-ofsight imaging. Optics Express 29(2), 1749–1763 (2021) 1
- Young, S.I., Lindell, D.B., Girod, B., Taubman, D., Wetzstein, G.: Non-line-of-sight surface reconstruction using the directional light-cone transform. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1407–1416 (2020) 2, 6, 8, 9